Tratamiento de datos y azar

Emiliano Ávila Arreola
Silvia Ochoa Hernández

SEP

conalep
Michoacán

Semestre 4

conalep
Michoacán

PRESENTA:

Tratamiento de datos y azar

Autores

Emiliano Ávila Arreola
Silvia Ochoa Hernández

Coordinadores:
Lic. José Azahir Gutiérrez Hernández
Ing. Eduardo Ochoa Hernández
Lic. Filho Enrique Borjas García

Título original de la obra:

Registro: **CONALEP-LIB-PRO-01A**

Programa: Profesor escritor, objetos de aprendizaje. Desarrollo de la competencia de la producción de información literaria y lectura.

Ávila Arreola, Emiliano; *et al.* (2014) **Tratamiento de datos y azar.** México: CONALEP/CIE

xii, 213 p.; carta

Registro: **CONALEP-LIB-PRO-01A** Documentos en línea

Editores:
Ing. Eduardo Ochoa Hernández
Lic. Filho Enrique Borjas García

Editorial: CONALEP Michoacán
Col. Chapultepec norte, Gral. Nicolás Bravo No. 144, Morelia, Michoacán.
http://www.cie.umich.mx/conalepweb2013/

Registro: **CONALEP-LIB-PRO-01A**
ISBN: En trámite
Impreso en______________________
Impreso en México –Printed in Mexico

DIRECTORIO

Lic. Fausto Vallejo Figueroa
Gobernador Constitucional del Estado de Michoacán

Lic. J. Jesús Sierra Arias
Secretario de Educación

Mtro. Álvaro Estrada Maldonado
Subsecretario de Educación Media Superior y Superior

Ing. Fernando Castillo Ávila
Director de Educación Media Superior

M.A. Candita Victoria Gil Jiménez
Directora General del Sistema CONALEP

M.C. Víctor Manuel Lagunas Ramírez
Titular de la Oficina de Servicios Federales en Apoyo a la Educación en Michoacán

Dr. Salvador Jara Guerrero
Rector de la Universidad Michoacana de San Nicolás de Hidalgo

Lic. José Arturo Villaseñor Gómez
Director General del CONALEP Michoacán

Lic. José Azahir Gutiérrez Hernández
Director Académico

L.E. Rogelio René Hernández Téllez
Director de Planeación, Programación y Presupuesto

Lic. Faradeh Velasco Rauda
Directora de Promoción y Vinculación

Ing. Mónica Leticia Zamudio Godínez
Directora de Informática

Lic. Víctor Manuel Gómez Delgado
Director de Servicios Administrativos

Ing. Genaro González Sánchez
Secretario General del SUTACONALEPMICH

Tec. Juan Pineda Calderón
Secretario General del SUTCONALEP

Prefacio

Estimado estudiante:

Las palabras que sombrean estas páginas, no son simple ciencia dentro del diálogo como depósitos de datos e información, ni son cuestión de vocabulario o listado de definiciones, son la experiencia generosa de la comunidad CONALEP Michoacán, esa realidad oculta pero necesaria que respaldó las tareas de investigación y composición literaria del discurso que integra este libro. Nos referimos a los profesores, administrativos y sindicatos que hoy convergen en el umbral de la existencia para apoyar a un grupo de profesores escritores que han creado en el sereno libre, arquitecturas de conocimientos como un viaje de aprendizaje que exigirá del estudiante, lo mejor de sí mismo ante la presencia luminosa del texto, ese que pretende enseñarle a caminar con la frente en alto.

Las ideas asociadas en este libro, equivalen a la imaginación lograda en el acto de escribir desde otros textos, al decodificarlas el estudiante, se le exige más vocabulario para enriquecer su habla y hacer ver a sus ojos más allá de la estrechez de la información que inunda a la sociedad moderna. El libro no presenta la superficie de la existencia como cruda observación, procura que su dificultad incite a perforar la realidad hasta reflexiones que renueven los modos inciertos de dar significado al mundo. La ciencia, la literatura y la tecnología no las percibimos como mundos incomunicables, los valores son explícitos caminos que las vinculan entorno al currículo del técnico bachiller. Tienen estos textos organización de premisas, técnicas, justificaciones, normas, criterios y como Usted se dará cuenta, también mostrará nuestros límites para seguir haciendo puentes entre las incesantes creaciones de nuevas fronteras de la investigación científica y técnica. Se pretende que estos libros sean contenido y no un libro de prácticas escolares, sean la herramienta de complementación para enriquecer los discursos de la enseñanza-aprendizaje.

Los profesores de CONALEP enfrentan a diario las carencias visibles de medios tecnológicos, materiales y documentales, sería fácil usar las palabras para señalar hasta el cansancio nuestras apremias, pero se ha decidido mejor producir libros como testimonios vivos y luminosos que renueven el rol social de la academia colegiada sensible a la condición social, susceptible de ir perfeccionándose con la acumulación de esta experiencia literaria, para servir de mejor manera al enriquecimiento de las competencias necesarias para realizar el sueño de éxito de tantos jóvenes Michoacanos.

Lic. José Arturo Villaseñor Gómez
Director General del CONALEP Michoacán

Mensaje a la comunidad académica

Se vive una época difícil para los libros, no son momentos de lectura tampoco, sin embargo, esto no debe ser un motivo para no producir literatura para educar mentes creativas, nuestra generación considera importante que sea la literatura la que emocione a realizar los grandes sueños de nuestros jóvenes. Si el libro escrito por mexicanos para mexicanos se perdiera, la conciencia de nuestra sociedad quedaría en orfandad, congelada de sentido y seriamos habitantes de un mundo siempre detrás de mascaras en rituales de simulación. Este argumento presente en *El laberinto de la soledad*, Octavio Paz nos dice al hombre moderno, seamos universales sin dejar de ser diferentes. El efecto de producir literatura en una sociedad, es como dice Steven Pinker, es "sacar los ángeles que llevamos dentro", es lo mejor de nuestro ser como oferta de conocimiento y valores al servicio de nuestra sociedad. El efecto Malinche que prefiere libros de traducción, ocasiona la perdida de identidad y crear la cultura de producir experiencias de conocimiento como herencia educativa para nuestros estudiantes. Los Aztecas en su derrota se sintieron abandonados por sus dioses, en el presente si cancelásemos que la voz de nuestro profesores que hablaran desde libros a las nuevas generaciones, es algo equivalente, al negar que las ideas en su crecimiento son una respuesta sociolingüística económica y democrática de identidad de una nación. Michoacán es un proyecto histórico de libertad, sin ignorar la realidad del malestar de la cultura actual, CONALEP desarrolla un programa académico para impulsar su capacidad y compromiso social para generar las ideas curriculares para enriquecer la sensibilidad y la imaginación científica, técnica y humanista de su comunidad.

Producir literatura curricular en CONALEP Michoacán, es asumir su personalidad moral, como dueña de una conciencia movida para una educación con la pasión de razones y expresiones culturales con la competencia para transformar positivamente la realidad. Aun cuando esta realidad adversa, de actitudes de autoridades renuentes a transformar la realidad educativa y de presiones económicas, hoy entregamos esta literatura escrita con profesores de CONALEP, máxime reconocimiento, porque trabajaron de manera altruista durante extenuantes y largas jornadas de trabajo; la comunidad CONALEP y la sociedad Michoacana toda, reciban esta obra como testimonio de la grandeza de este histórico Michoacán.

Lic. José Azahir Gutiérrez Hernández
Director Académico

Enfoque por competencias

Este texto de apoyo es una introducción al tratamiento de datos numéricos y al cálculo de probabilidades, que sirve como antecedente y base para el cálculo de variables. Tales tratamientos de datos incluyen muestreo aleatorio, distribuciones de frecuencias, representación gráfica, medidas de posición central y no central, medidas de dispersión y forma, así como técnicas de contar y cálculo de probabilidades.

Características del libro de apoyo para el estudiante:

Cada capítulo cuenta con una introducción que incluye contexto histórico y aplicaciones, definiciones, ejemplos de ejercicios resueltos, problemario, autoevaluación y soluciones, y conclusiones, que sin ser finales, más bien son una invitación al análisis de otras posibilidades y aplicaciones de este tema.

En su versión digital, las referencias son accesibles siguiendo la liga en la red.

En toda obra literaria se afirma una realidad independiente de la lengua y del estilo: la escritura considerada como la relación que establece el escritor con la sociedad, el lenguaje literario transformado por su destino social. Esta tercera dimensión de la forma tiene una historia que sigue paso a paso el desgarramiento de la conciencia: de la escritura transparente de los clásicos a la cada vez más perturbadora del siglo XIX, para llegar a la escritura neutra de nuestros días.

Roland Barthes, *El grado cero de la escritura*

Palabra escrita bajo luz

En un mundo cada día con más canales de comunicación, la palabra escrita camina por los muros que denuncian el drama catastrófico sobre el medio ambiente y sobre el control de la vida humana; el combustible de esta desesperanza produce apatía profunda por tener contacto con el mundo de la literatura, esta distorsión moral parece reflejarse entre los que no quieren sentir responsabilidad ni pensar, dejando a otros su indiferencia al ser prisioneros de ligeras razones y tirria justificada en la empresa de sobrevivir.

Especular en un mundo sin libros es exponer al mundo a la ausencia de pensamiento, creatividad y esperanza. Los libros, dedicados a ser arrebatados por el lector, están expuestos a ser poseídos por las bibliotecas vacías y que con el tiempo opacan sus páginas y empolvan la cubierta de lo que alguna vez fue un objeto de inspiración. Resulta difícil transcribir este instante de un peligroso espacio donde ya muy pocas palabras sobreviven dentro de la reflexión y las pocas sobrevivientes han abandonado la unión del sentido de vivir y el sentido del pensar científico. Escribir pasa de ser un placer repentino a ser una necesidad inminente, es el puente entre lo conocido y lo inexplorado. Es un reto de hoy en día inmiscuirse en lo que una vez fue lo cercano y dejar de lado la novedad tecnológica para poder a través de las barreras que nos ciegan abrir fronteras literarias. Es un proyecto que conspira a favor de la libertad creativa, de la felicidad lúcida cargada de libros embajadores de nuevas realidades.

Entre un mar de razones dentro del libro escolar en crisis, se percibe la ausencia de esa narrativa del cuerpo del texto, misma que alimenta al lector de una experiencia de conocimiento, su ausencia, es más un mal glosario, de un mal armado viaje literario científico o de ficción. En esos viajes de libros en crisis, nos cansamos de mirar espacios vacíos de talento, emociones y sensibilidad para responder a un entorno adverso; son muchas veces un triunfalismo de autoevaluación y una falsa puerta de una real competencia para actuar en la realidad. Uno no solo vive, escucha la voz interior de un libro, uno es fundado en el manejo del lenguaje que explica, crea, aplica o expande los límites del horizonte de nuestro imaginario actuante en lo real. No vivimos leyendo texto, sino leyendo el paisaje de una realidad, el libro toma la voz del progreso en una siempre reconstrucción lingüística del sujeto que explica, transforma y comunica desde los desafíos de su generación.

La información cruda que tanto rellena los libros grises, oscuros y papel pintado; requiere ser dotada de conceptos que permitan alimentar al sujeto que toma decisiones, que explora con paso lento, que mira por dentro del lenguaje y aplica la información que cobra sentido en la siempre expansión de las ideas.

Escribir un libro es siempre reconstruir un discurso, sus lectores en este discurso son el puente a un texto profundo que demanda esfuerzo en la reconstrucción de los procesos de razonamiento y el entretejido del discurso que involucra información de fondo, esas fuentes que justifican su análisis y poseen significado privilegiado para la comprensión de una realidad.

El lector puede hacer uso del libro con su propia experiencia y con su autoayuda, al precisar términos y conceptos para prolongar su horizonte de interpretación, el libro se hace cargo de la memoria de un plan de estudios, es un discurso de diferentes capas de argumentos, tras este texto se anuncia un orden de experiencia propuesto para su aprendizaje. El libro está conformado para jóvenes con memoria sin dolor para nuevas palabras, aborda el olvido como una deficiencia de interactividad entre el discurso argumentativo y los referentes conceptuales. Esto es el reto en la producción de los libros CONALEP. La propuesta es una reconstrucción de una semántica más profunda, como el principal reto del estudiante técnico bachiller del siglo XXI.

Libro,…
todos te miran,
nosotros te vemos bajo la piel.

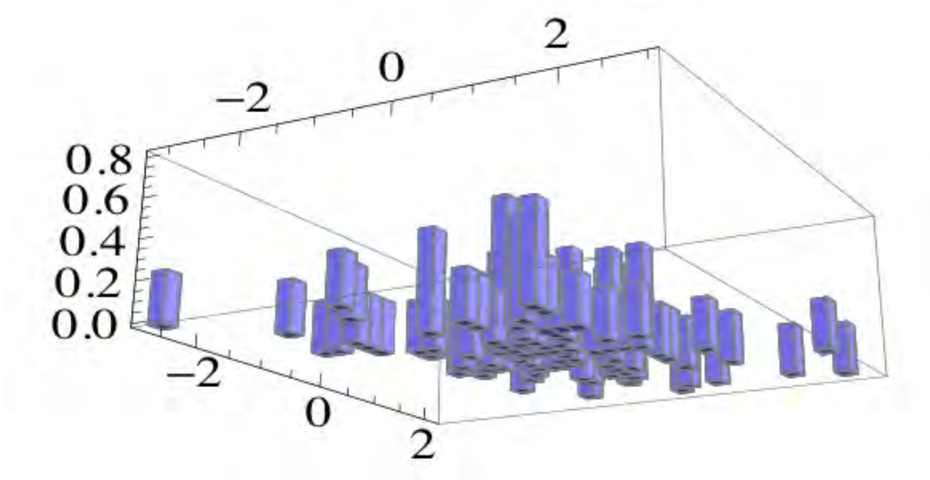
SUMARIO

Capítulo 3

Capítulo 4

Capítulo 1:

TRATAMIENTO DE DATOS Y AZAR



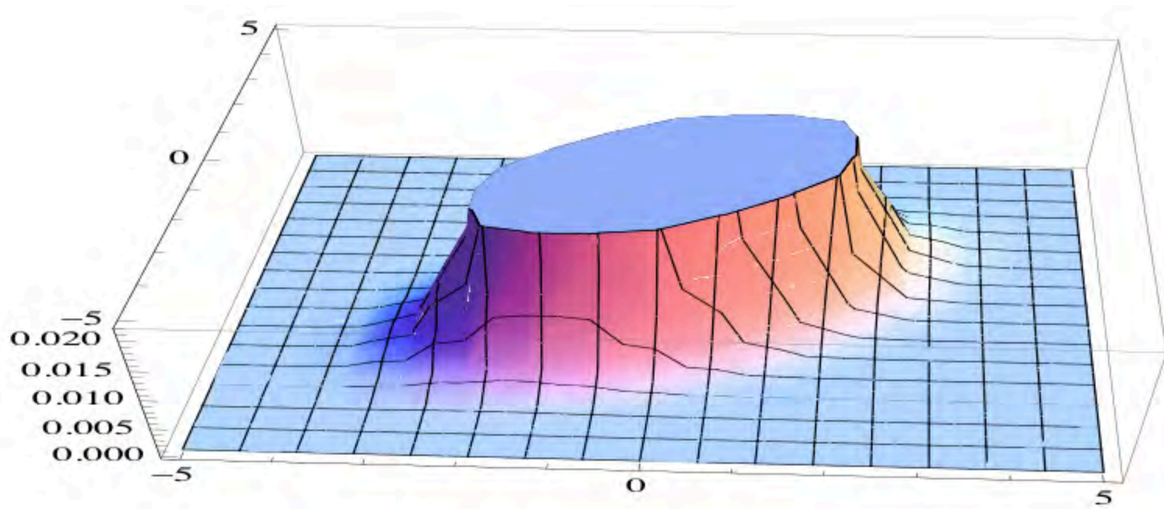
Distribución variables random. Gráfica hecha con mathematica 9

1. Introducción a la estadística descriptiva

La estadística, puede ser considerada el "arte" o métodos de recopilar información, interpretarla, representarla y hasta tratar de predecir el comportamiento de los datos para poder tomar decisiones sin necesidad de tener toda la información completa[1].

La estadística[2] se divide en dos grandes ramas: estadística descriptiva[3] y estadística inferencial, la estadística descriptiva consta de la metodología para organizar mediante tablas, gráficas y de forma numérica los datos obtenidos, y la estadística inferencial a partir de resultados obtenidos del comportamiento de una muestra puede predecir el comportamiento de la población.

La palabra estadística tiene varias acepciones, en un sentido se refiere a los resultados obtenidos por algún censo o recuento, seguramente en alguna ocasión escuchaste o viste los resultados de tu equipo favorito, cuántos juegos han ganado, cuántos han perdido, cuántas medallas de oro hemos ganado en los juegos olímpicos, las estadísticas de desempleo en el estado, etc. En este texto nos referiremos a los métodos de recopilación e interpretación de datos numéricos, tratando de hacer conclusiones del comportamiento de las variables.



Histograma en 3D. Gráfica hecha con mathematica 9

1.1. Variables cualitativas y cuantitativas

Clasificación de las variables:

Variables cualitativas[4]: como el nombre lo indica, muestran cualidades a las que no se les puede asignar valores numéricos, por ejemplo, cuando hacemos la distinción del idioma hablado por un grupo de personas, su lugar de nacimiento, sexo al que pertenecen, comida que les gusta, color de piel, género de música que prefieren escuchar, etc.

Variables cuantitativas[5]: son otro tipo de variables a las que se puede asignar un valor numérico, por ejemplo la estatura, el peso, la cantidad vendida de un producto, los gastos semanales de un restaurante, la calificación obtenida en una asignatura. Estas variables cuantitativas a su vez se clasifican en discretas y continuas.

Las variables discretas toman valores numéricos enteros, por ejemplo el número de hijos en una familia podría ser 1, 2 o 3, pero no podríamos decir que tiene 4.567 hijos.

Las variables continuas toman cualquier valor numérico real, por ejemplo, la velocidad que alcanza una motocicleta al pasar por un punto de revisión podría ser 40.3km/h o 50km/h.

Las variables también se pueden clasificar de acuerdo al número de características que recogen de un objeto o individuo:

Variables unidimensionales[6], estas muestran una sola característica de lo que estemos analizando, por ejemplo, la edad de un grupo de personas que asisten a una conferencia.

Variables bidimensionales[3], estas recaban información sobre dos características de los elementos que se están analizando, por ejemplo, en un grupo de personas se puede recabar su C.I. y su edad.

Variables pluridimensionales[3], son las que recogen información sobre más de dos características de una población, por ejemplo, al extraer una muestra de un metal de una zona *x* podremos obtener su densidad, capacidad calorífica, dureza, etc.
Lo anterior se puede esquematizar como se muestra en el siguiente esquema:

Para resolver: en equipos recorten de revistas o periódicos resultados de estadísticas y de gráficas , péguenlas en su cuaderno y clasifiquen el tipo de variables, al finalizar el capítulo serán capaces de identificar el nombre que reciben las estadísticas y gráficas que encontraron.

Datos curiosos de estadísticas:
Somos el país de América con el mayor número de patrimonios de la humanidad (UNESCO) y el cuarto en el mundo, solo por debajo de España, Italia y Francia.

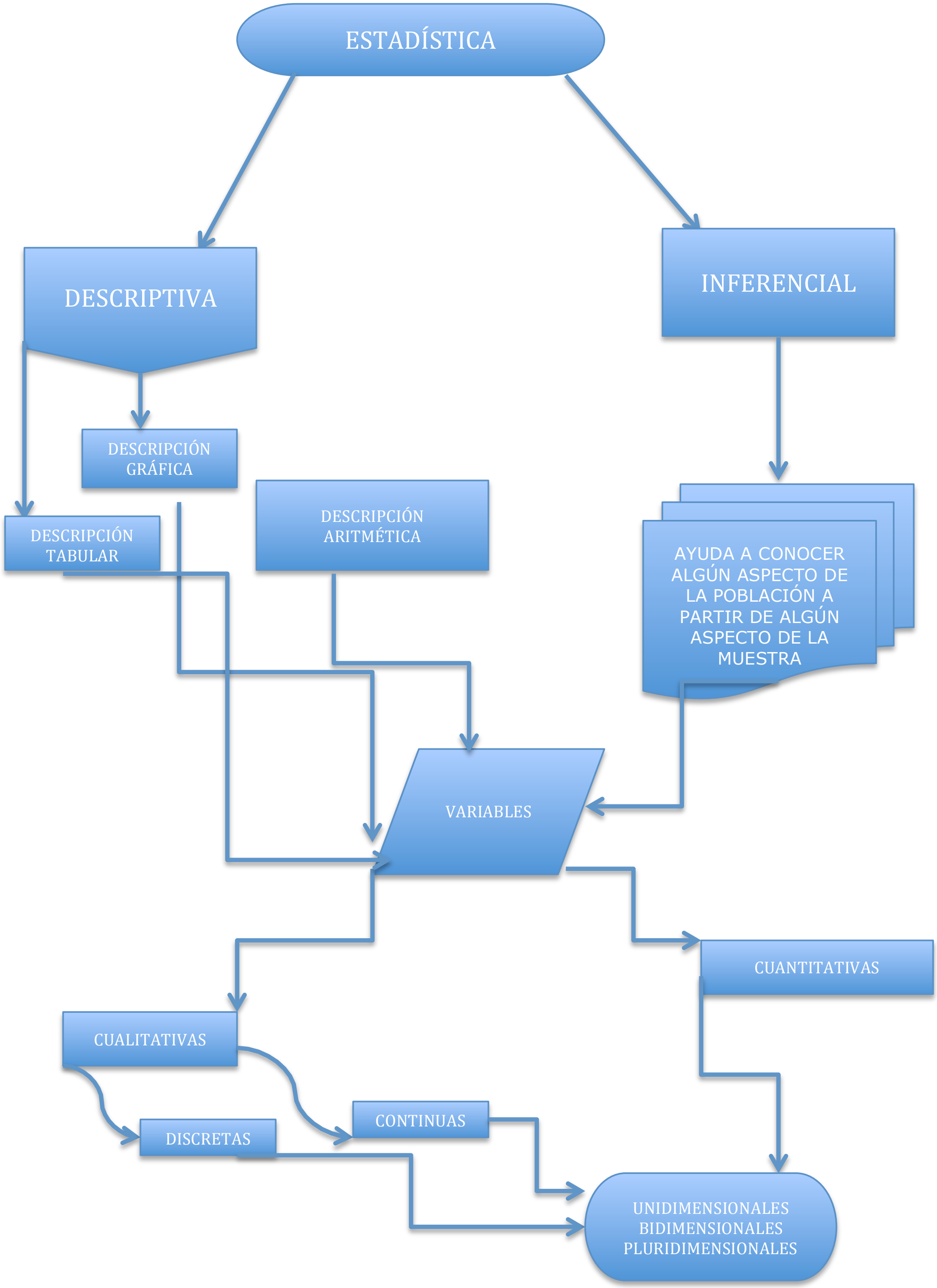


Diagrama que muestra la clasificación de la estadística y las variables que intervienen en ella.

1.2. Conceptos básicos

1.2.1. Población

Al estudiar el comportamiento de las variables es importante que definamos el concepto de población[7], como "la totalidad de las observaciones posibles", pero también como sucede con la palabra estadística, tiene varios significados, puede venir a nuestra mente cuando se menciona la palabra población, las personas que habitan en una zona determinada, por ejemplo, de acuerdo con el INEGI (2011), Michoacán tiene una población de 4 351 037[8] habitantes.

En estadística al conjunto de todas las observaciones posibles del evento que se esté considerando le llamamos población, nótese que no necesariamente una población está representada por personas, pueden ser objetos, animales o cualquier cosa de la situación que se esté analizando, por ejemplo si queremos verificar el control de calidad en la producción de clavos de acero, la población puede ser los clavos producidos por la máquina A durante el turno de 8:00am a 11:00am, otro ejemplo puede ser si queremos evaluar el coeficiente intelectual de los niños de la escuela primaria "Melchor Ocampo" de la ciudad de Morelia, todos los niños de la escuela constituyen la población.

En las poblaciones podemos considerar varias características que pueden ser medidas, a dichas características les llamamos parámetros[9].

En algunos casos es muy difícil considerar a todos los elementos que forman una población, por ejemplo si deseamos conocer aproximadamente cuántas personas cuentan con televisión analógica en nuestro país, para poder producir los convertidores de señal digital y dar el abasto suficiente, los costos y el tiempo para hacer el cálculo puede no ser conveniente, ya que se tiene que pagar a las personas que hacen las encuestas, encuentren o no a los habitantes de cada casa, el tiempo de desplazamiento a comunidades alejadas implica gastos de transporte y tiempo de los encuestadores, por lo que podemos estudiar una parte de toda la población que esté representada y elegida aleatoriamente para hacer un cálculo aproximado.

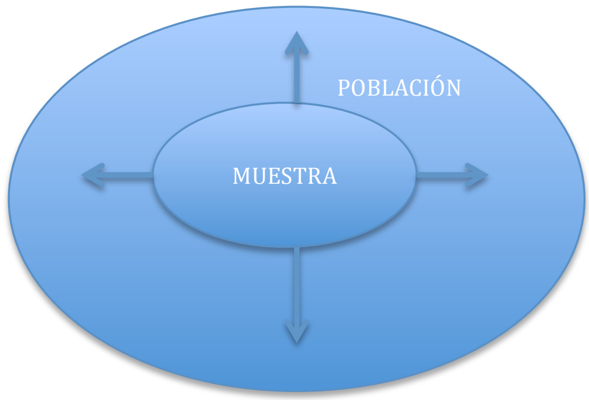
Imagine ahora que quiere conocer el tiempo promedio de duración de las bisagras producida en la fábrica "Aceros de Michoacán", si lo hacemos con toda la población no es conveniente ya que provocaríamos el desgaste completo de cada bisagra, esto implicaría destruir la población, un tiempo excesivo en la verificación y baja en la ganancia de producción, por lo que es conveniente tomar una parte de dicha población para ser analizada.

Cuando no conviene por tiempo, costos, etc., tomar la totalidad de las observaciones, es decir la población, se puede tomar una parte de ella que la represente, a esa parte de la población se le llama muestra[10].

Una población puede ser los habitantes de la ciudad de Morelia Michoacán

1.2.2. Muestra

La muestra nos ayuda a estimar qué resultados podríamos tener si consideramos a la población completa.



El simple hecho de tomar una parte de la población no garantiza que la muestra represente a la población, debemos elegirla con mucho cuidado para que sea representativa y que cada elemento tenga la misma probabilidad de ser elegido.

Posiblemente alguna vez tu o alguien de tu familia ha sido encuestado por teléfono fijo, aquí se está eligiendo una población con una característica en particular, que tengan línea telefónica en casa, puede convenir dependiendo del estudio que estén haciendo los encuestadores, tal vez esto les garantice un nivel socioeconómico, o un área con mejores servicios, pero puede no arrojar resultados que representen a una población por que están dejando fuera a los que no cuentan con dicho servicio.

Supongamos que queremos conocer en una escuela que tiene 2,500 alumnos, cuántos libros leen los estudiantes en promedio por año, y no tenemos tiempo ni el personal suficiente para realizar la encuesta, si los compañeros de tu grupo son los encargado de realizarla, ellos podrían decidir, tal vez inconscientemente, entrevistar a sus amigos o conocidos y esto podría no brindar información representativa de la población estudiantil.

Así podemos concluir que una muestra[6] es un subconjunto de la población, observe que la muestra siempre es menor que la población, de otra forma sería la población misma.

Uno de los procedimientos que nos puede ayudar a elegir una muestra aleatoria es seleccionar la muestra con una tabla de números aleatorios.

Muestra de una máscara del diablo usada para las representaciones de la pastorela del 2 de febrero en Tócuaro Michoacán, exhibida en el Museo de la ciudad de Pátzcuaro Michoacán.

1.2.3. Muestreo aleatorio y uso de tablas de números aleatorios

Los números aleatorios[11] son números generados al azar, la siguiente tabla muestra números entre 0 y 10,000, los cuales se generaron en Excel, si deseas hacer tu propia tabla, el procedimiento es el siguiente:

En f_x poner la fórmula =ALEATORIO.ENTRE(0,10000), elegir las celdas que se requieren.

2575	933	5066	7451	9525	1536	9608
2336	594	8198	6454	1347	9027	3729
3461	2619	8511	8582	2794	9326	8184
5181	7136	5620	6492	3568	3869	7222
6407	3241	2789	6628	596	676	6042
952	3551	5380	2261	2831	1628	737
9285	556	3518	9173	6987	1958	1440
4471	5882	5811	820	6602	1511	5208
2575	3166	756	6565	5138	807	1718
1993	8058	1442	355	8636	8585	9460
3821	4338	6813	3395	4169	2126	1206
946	9427	5558	4804	1601	8352	5421
8731	4971	372	3182	6077	5814	9698
8512	9654	2838	7400	1809	3424	7950
6778	8352	289	4399	437	2586	5543
4278	8302	9698	5666	8809	4924	3980
633	9248	4422	3938	5599	2958	4542
8234	8163	6784	4681	3850	5523	8004
8219	6878	9903	6532	885	9834	7304
6228	8149	964	3263	6503	6185	9862

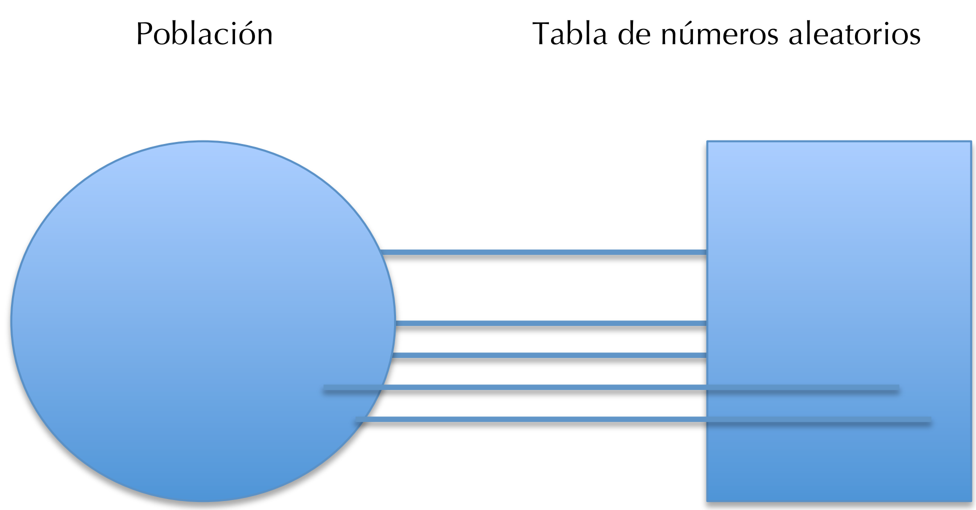
¿Cómo podemos elegir una muestra aleatoria con una población dada?

Suponga que se tiene una población de 250 lámparas producidas por el turno A, para llevar el control de calidad de dicho producto el verificador desea tomar una muestra del 5% (12 lámparas), a cada lámpara se le asigna un número de producción y lote, así que se pueden numerar del 1 al 150 y tomar 12 usando una tabla de números aleatorios, para hacerlo observemos que nuestra población de 250 lámparas es un número de 3 dígitos, por lo que en nuestra tabla de números aleatorios buscaremos los números de 3 dígitos que sean menores o iguales a nuestra población, en este caso menores o iguales al número 250, que son de acuerdo a la tabla anterior los números sombreados.

2575	933	5066	7451	9525	1536	9608
2336	594	8198	6454	1347	9027	3729
3461	2619	8511	8582	2794	9326	8184
5181	7136	5620	6492	3568	3869	7222
6407	3241	2789	6628	596	676	6042
952	3551	5380	2261	2831	1628	737
9285	556	3518	9173	6987	1958	1440
4471	5882	5811	820	6602	1511	5208
2575	3166	756	6565	5138	807	1718
1993	8058	1442	355	8636	8585	9460
3821	4338	6813	3395	4169	2126	1206
946	9427	5558	4804	1601	8352	5421
8731	4971	372	3182	6077	5814	9698
8512	9654	2838	7400	1809	3424	7950
6778	8352	289	4399	437	2586	5543
4278	8302	9698	5666	8809	4924	3980
633	9248	4422	3938	5599	2958	4542
8234	8163	6784	4681	3850	5523	8004
8219	6878	9903	6532	885	9834	7304
6228	8149	964	3263	6503	6185	9862

El primer número de la columna tomando 3 dígitos es 257, como es un número mayor a la población (250) lo descartamos y seguimos con 233, como es menor que 250 lo tomamos para nuestra muestra, a continuación se indican los 12 números que representan a la muestra aleatoria:

1. 233
2. 95
3. 199
4. 94
5. 63
6. 93
7. 59
8. 55
9. 75
10. 144
11. 37
12. 28



Para resolver: trabaja en equipos y sugiere otro método para elegir una muestra aleatoria de una población, coméntalo con el resto del grupo.

1.2.4. Usos y abusos de la estadística

El uso de los métodos para recopilar información puede estar viciado, por ejemplo al tomar muestras que no son aleatorias, se tiene pérdidas en la calidad de la información, también en la forma de interpretarla o representarla. El mejor ejemplo de observar esto es en las votaciones simuladas, ya que con las respuestas voluntarias se pretende predecir cómo quedarán las votaciones finales, el periodo pasado de elecciones para presidente de la república las casas encuestadoras[12] se equivocaron 8 puntos respecto a los resultados obtenidos, y estas encuestas sí pueden comprobarse.

En diciembre de 2011, el candidato[13] por la alianza Compromiso por México, tuvo una respuesta muy pobre cuando contestó cuáles libros le marcaron como persona y como político, su respuesta "equivocada" para los medios y muy criticada en las redes sociales hizo parecer que afectaría su preferencia, y los resultados mostraron lo contrario. Aquí los medios juegan un papel muy importante de manipulación y es importante tomar en cuenta el mayor número de variables.

Como estos hay muchos más ejemplos y queda al lector la tarea de investigar, observar con cuidado cuando se tengan resultados para formar un criterio sólido.

Datos curiosos de estadísticas:
El número de matrimonios es el doble que el de divorcios; por lo tanto, uno de cada dos matrimonios acaba en divorcio.

1.3. Distribuciones de frecuencias

Con lo anterior se ha mostrado algunas formas de obtener datos, pero cuando se tienen datos en bruto (o sin agrupar) estos no son significativos, por lo que habría que organizarlos para poder manejar adecuadamente la información obtenida.

Considérese una muestra de los pesos de niños recién nacidos de madres solteras en el municipio de Morelia. Supóngase que 20 niños se eligen aleatoriamente y sus pesos en kilogramos son:

3.2	2.4	3.0	2.8	2.8	3.3	1.9	2.7	2.5	3.4
2.7	3.4	3.5	2.9	2.4	2.7	1.5	2.6	2.8	3.0

Los datos anteriores (información ficticia) no tienen sentido si no los tenemos organizados de manera adecuada. Estos datos así mostrados se llaman datos sin agrupar o datos en bruto, una forma muy sencilla de organizarlos es mediante un arreglo de los datos por orden de magnitud, esto es ascendente o descendente, una vez hecho esto podemos ver que la diferencia entre el menor peso y el mayor es de 2 kg, también observamos que el bebé con menor peso es de 1.5 kg y el de mayor peso 3.5 kg, esta es una forma muy elemental de organización de los datos, y sería muy complicado si el número de observaciones fuera mucho mayor. Cuando sucede que tenemos grandes cantidades de datos lo más aconsejable es presentarlos en una distribución de frecuencias; como veremos a continuación los datos organizados en distribuciones de frecuencias muestran el patrón de distribución de manera relevante.
Un conjunto de datos se puede organizar de varias maneras, aquí mostraremos algunas, pero la forma que se elija para organizarlos dependerá de la naturaleza de los datos, qué aspecto se quiere precisar o la cantidad de datos que se tenga.

1.3.1. Distribución de frecuencias simple

Una forma de organizar los datos es mediante una distribución de frecuencias simple.

Al número de veces que aparece un dato en el conjunto inicial se le llama frecuencia y la denotaremos con f.

Una distribución de frecuencias simple[14], nos indica con qué frecuencia aparecen los datos desde el menor hasta el mayor.

Vamos a organizar una distribución de frecuencias simple con el siguiente conjunto de datos sin organizar:

Consideremos las calificaciones obtenidas por los estudiantes de primer grado de secundaria en la asignatura de matemáticas:

ALMANZA NÚÑEZ LISSET	10
CÁZAREZ DURÁN DIEGO	7
DELGADO VERDUZCO MORGAN	5
ESPINOSA SÁNCHEZ SAMARA	7
ESPINOSA VÁZQUEZ HILARY VANESSA	7
FERNÁNDEZ NÚÑEZ LOURDES MARIANA	7
GOVEA VEGA JOSUÉ	8
GUZMÁN FERNÁNDEZ EVA ALESSANDRA	9
JARQUIN CENDEJAS LEONARDO	9
LABORDE MASERA DIEGO SALVADOR	8
MÁRQUEZ VÁZQUEZ AMANDA	7
MARTÍNEZ RODRÍGUEZ REGINA	8
MOLINA GARCÍA ANDRÉS	9
PÉREZ BLANCO RODRIGO	10
PINTOR ANDALUZ DIEGO DE JESÚS	7
RINCÓN VAQUERO SANTIAGO	7
SOTO ROVIRA XIUH U	5
TERÁN CHÁVEZ PAULA	8
TORRES NEGRETE NATALIA	8
VILLALOBOS ALMANZA JULIO	10

CALIFICACIONES									
10	7	5	7	7	7	8	9	9	8
7	8	9	10	7	7	5	8	8	10

Iniciamos localizando el dato menor y el dato mayor. El menor es 5 y el mayor es 10. En seguida enlistamos los números cuyos extremos son 5 y 10 ordenados de mayor a menor (puede ser al contrario pero por conveniencia para próximas formas de organizar usaremos orden descendente).

Calificación

10 dato mayor
9
8
7
6 debemos incluir los valores intermedios no importa que no aparezcan en los datos obtenidos
5 dato menor

El dato 10 aparece 3 veces, se escribe a su derecha en la lista anterior la marca (/) por cada vez que aparece, el dato 6 no aparece, no ponemos ninguna marca, con el resto de números hacemos lo mismo que con el 10.

Calificación

10 ///
9 ///
8 ~~////~~
7 ~~////~~ //
6
5 //

Las marcas que aparecen a la derecha de las calificaciones indican la frecuencia con que aparecen en el conjunto, así la siguiente tabla muestra finalmente la distribución de frecuencia simple al cambiar las marcas por el número que representan:

DISTRIBUCIÓN DE FRECUENCIAS SIMPLE

CALIFICACIÓN	FRECUENCIA
10	3
9	3
8	5
7	7
6	0
5	2
	N= 20

N=número total de datos

La distribución de frecuencias simple así construida, muestra entre otras cosas, que la mayoría de los alumnos obtuvo calificaciones aprobatorias, y que la mayoría de los alumnos obtuvo 7.

El ejemplo anterior mostró de manera muy sencilla la construcción de una distribución de frecuencias simple con un número muy pequeño de datos, de la misma manera se procede para grandes cantidades de datos, pretendemos que se observe el procedimiento y no la "talacha" que es elaborarla.

Una herramienta muy valiosa para cantidades grandes de datos es utilizar Excel:

La función FRECUENCIA en Excel

La función FRECUENCIA en Excel nos permitirá obtener fácilmente una **tabla de frecuencias** con tan solo especificar dos argumentos. El primer de ellos será la matriz de que contiene los datos y el segundo argumentos el rango de celdas que contiene las categorías.

En el siguiente ejemplo, en la columna A, tengo una lista de 200 números entre 1 y 10. En la columna C podrás observar que he definido 5 rangos: 2, 4, 6, 8, 10.

D7 f_x =SUMA(D2:D6)

	A	B	C	D	E
1	Números		Categorías	Frecuencia	
2	10		2		
3	8		4		
4	4		6		
5	4		8		
6	3		10		
7	7		Total:	0	
8	2				
9	1				
10	10				
11	10				
12	8				
13	3				
14	2				
15	2				

Toca al lector investigar en un tutorial la forma de hacerlo[15]

1.3.2. Distribución de frecuencias con clases o intervalos

Cuando el número de datos es muy grande conviene organizarlos en clases o intervalos, consideremos el siguiente conjunto de datos obtenidos al medir a 30 estudiantes de un grupo, los resultados se indican en metros:

ESTUDIANTES	ESTATURA
Estudiante 1	1.24
Estudiante 2	1.28
Estudiante 3	1.27
Estudiante 4	1.21
Estudiante 5	1.22
Estudiante 6	1.29
Estudiante 7	1.3
Estudiante 8	1.24
Estudiante 9	1.27
Estudiante 10	1.29
Estudiante 11	1.23
Estudiante 12	1.26
Estudiante 13	1.3
Estudiante 14	1.21
Estudiante 15	1.28
Estudiante 16	1.3
Estudiante 17	1.22
Estudiante 18	1.25
Estudiante 19	1.2
Estudiante 20	1.24
Estudiante 21	1.21
Estudiante 22	1.29
Estudiante 23	1.26
Estudiante 24	1.22
Estudiante 25	1.28
Estudiante 26	1.27
Estudiante 27	1.26
Estudiante 28	1.23
Estudiante 29	1.22
Estudiante 30	1.21

De nuevo los datos así obtenidos están sin agrupar, comencemos por agruparlos en frecuencia simple con el método indicado anteriormente, en clases de ancho tres construimos una distribución en clases o intervalos:

ESTATURA		ESTATURA(m)	frecuencia simple (f)
1.3		1.3	3
1.3		1.29	3
1.3		1.28	3
1.29		1.27	3
1.29		1.26	3
1.29		1.25	1
1.28		1.24	3
1.28		1.23	2
1.28		1.22	4
1.27		1.21	4
1.27		1.2	5
1.27			N= 34
1.26			
1.26			
1.26	INTERVALOS O CLASES	FRECUENCIAS	f
1.25	1.28-1.30		9
1.24	1.25-1.27		7
1.24	1.22-1.24		9
1.24	1.19-1.21		9
1.23			N=34
1.23			
1.22			
1.22			
1.22			
1.22			
1.21			
1.21			
1.21			
1.21			
1.2			
30			
30			
1.2			

Para construir la tabla en clases o intervalos[16] se contaron cuántos datos pertenecen a cada uno de los intervalos, así en la clase 1.28-1.30 el número de datos que pertenecen a esa clase son los datos que son iguales a 1.28, 1.29 y 1.30, que son 9 datos, este conteo se hace de la tabla de distribución de frecuencias simple cuando se tiene a la mano los datos originales.

Examinemos las características de la tabla en clases:

1) Cada clase está definida por dos números, por ejemplo en la primera clase en orden descendente 1.28-1.30, los números 1.28 y 1.30 reciben el nombre de límites aparentes[17] inferior y superior respectivamente, la distancia entre los límites de cada clase es siempre la misma, observe que la última clase para que tenga el mismo ancho de clase se le agrega el número descendente que sigue a 1.20, que es 1.19 así todas las clases tienen el mismo ancho.

2) El ancho o longitud de la clase[18] es la cantidad de datos que tiene la clase desde el límite aparente inferior, el o los valores intermedios y el límite aparente superior, por ejemplo en la última clase en orden descendente va desde 1.21, 1.20, hasta 1.19, son tres datos, ese es el ancho o longitud de la clase.

3) Observe que existe una proporción entre el número de clases y el ancho de clase, si aumentan las clases disminuye el ancho, y si aumentamos el ancho el número de clases disminuye.

Podemos calcular de manera aproximada el número de clases o el ancho que debemos dar mediante la fórmula siguiente, donde aparece un nuevo concepto que es el Rango, este se define como la diferencia entre el dato mayor y el dato menor.

Número de clases$\cong \frac{Rango}{longitud\ del\ intervalo}$

Si despejamos obtenemos la relación:

Longitud del intervalo$\cong \frac{Rango}{Número\ de\ clases}$

En nuestro ejemplo Rango=1.30-1.19 =0.11

Veamos el siguiente ejemplo:

CLASE O INTERVALO	FRECUENCIA
71-73	10
68-70	18
65-67	20
62-64	30
59-61	15
56-58	12
	N= 105

Los datos muestran que las clases tienen ancho o longitud 3, el Rango 17, así si utilizamos la relación anterior

$$\text{Número de clases} \cong \frac{Rango}{Número\ de\ clases} = \frac{17}{3} \cong 5.6 \cong 6$$

El número 6 es el número aproximado de clases, nótese que exactamente quedaron 6 como se indica en la tabla anterior.

Si ahora hacemos el cálculo del ancho que necesitamos para obtener 4 clases:

$$\text{Longitud del intervalo} \cong \frac{Rango}{Número\ de\ clases} \cong \frac{17}{4} \cong 4.25 \cong 4$$

Aproximadamente deberemos hacer las clases de longitud 4, esto es muy útil, ya que aunque es aproximado nos da una idea antes de hacer las tablas tanto del ancho como del número de clases que nos quedarán.

1.3.3. Distribución de frecuencias acumuladas

La siguiente forma de organizar los datos es muy útil para identificar cuántos datos son menores o iguales a un determinado límite superior, para lo cual introduciremos un nuevo concepto, la frecuencia acumulada[19] fa, que se define como el número de datos que son menores o iguales que un determinado límite superior.

Por ejemplo con la siguiente tabla construiremos una distribución de frecuencias acumuladas:

DISTRIBUCIÓN DE FRECUENCIAS EN CLASES

CLASE O INTERVALO	FRECUENCIA
71-73	10
68-70	18
65-67	20
62-64	30
59-61	15
56-58	12
	N= 105

DISTRIBUCIÓN DE FRECUENCIAS ACUMULADAS

CLASES	FRECUENCIA	FRECUENCIA ACUMULADA
71-73	10	105
68-70	18	95
65-67	20	77
62-64	30	57
59-61	15	27
56-58	12	12
	N=105	

Las frecuencias acumuladas se obtuvieron sumando los datos menores o iguales que 73, que es el límite superior de la primera clase, como todos los datos son menores o iguales que 73, su frecuencia acumulada es 105 o bien fa = 10+18+20+30+15+12=105, de la misma manera por ejemplo, para la clase 59-61 su fa=15+12=27, lo que significa que existen 27 datos menores o iguales al límite superior 61.

Si leemos la tabla anterior podemos concluir que por ejemplo hay 77 datos que son menores o iguales a 67.

Entonces una distribución de frecuencias acumuladas muestra las frecuencias acumuladas de cada uno de los límites superiores de cada clase.

También podemos calcular de dichas tablas qué porcentaje del total de los datos pertenece a una determinada clase, por ejemplo:

¿Qué porcentaje del total de los datos pertenece a la clase 59-61?

Primero debemos responder cuántos datos del total (105) tiene dicha clase, la respuesta es 15, después podemos usando una regla de tres simple hacer el cálculo, donde nuestro 100% es el total de los datos y queremos saber 15 datos qué porcentaje representan respecto al total de los datos:

$$100\% = 105\ datos$$
$$x\ \% \ = \ 15\ datos$$

$$x = \frac{15\ datos \times 100\%}{105\ datos} = \frac{1500\%}{105} = 14.28\%$$

1.3.4. Distribución porcentual acumulativa

Si a la tabla anterior de distribución acumulada le agregamos una columna donde indiquemos el porcentaje de la frecuencia acumulada, a esta forma de organizar los datos se le conoce como una distribución porcentual acumulativa[7].

CLASES	FRECUENCIA	FRECUENCIA ACUMULADA	PORCENTAJE ACUMULADO
71-73	10	105	100%
68-70	18	95	90.47%
65-67	20	77	73.33%
62-64	30	57	54.28%
59-61	15	27	25.71%
56-58	12	12	11.42%
	N=105		

Con ayuda de esta tabla podemos concluir y responder preguntas, como

¿Qué porcentaje del total de los datos es menor o igual al dato 64? La respuesta es 54.28%.

¿Qué porcentaje del total de los datos pertenece a la clase 68-70?

La clase 68-70 tiene 18 elementos (frecuencia simple), hacemos una regla de 3 donde nuestro 100% siempre es el total de los datos y tendremos que en dicha clase hay un 17.14% del total de los datos.

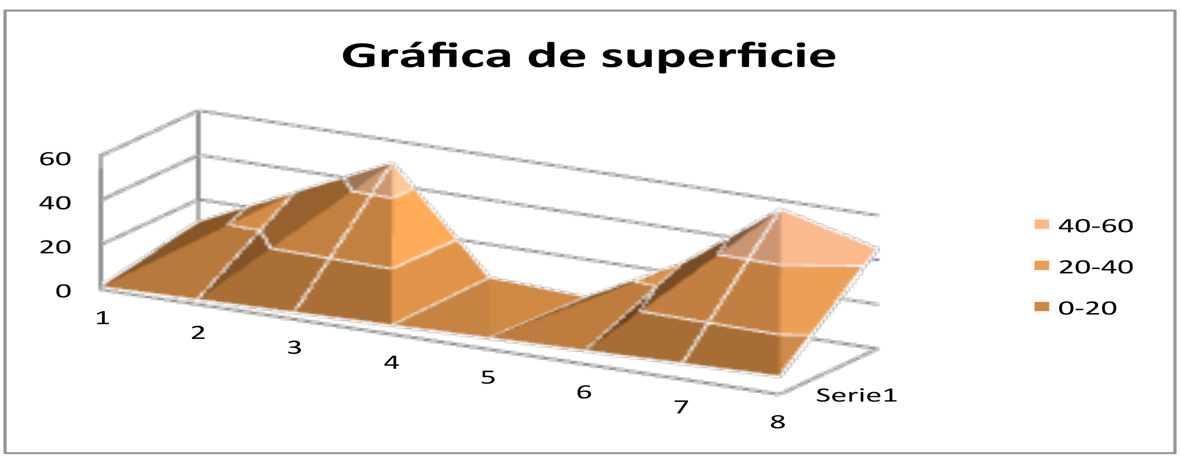
Para resolver: pide a tu maestro que te comparta la información de las calificaciones obtenidas por tu grupo, y organízala mediante distribuciones de frecuencias.

1.4. Representación gráfica

Paralelamente a la organización de datos mediante tablas, podemos representar los datos mediante una descripción gráfica, que puede ser útil para observar los resultados de manera resumida más sencilla [20], para interpolar o extrapolar resultados, las gráficas son muy utilizadas en educación, resultados de encuestas, resultados de efectividad de algún producto, en ventas, resultados de análisis de suelo, monitoreo de procesos, y muchos más que sería difícil encontrar dónde no se utilizan.

Dicho sea de paso también tiene desventajas la representación gráfica de datos, como cuando en el eje vertical que suele representar las frecuencias de los datos, las escalas son aproximadas a los verdaderos resultados obtenidos, aún mejor es colocar tanto la gráfica como la tabla que muestre los datos.

Existen muchos tipos de representaciones gráficas, no obstante, se pueden generar distintas dependiendo de la creatividad del diseñador, nosotros por nuestra parte mostraremos las más utilizadas en dos dimensiones y algunas de ellas en tres.



Otro tipo de gráficas es la gráfica de superficie. Gráfica hecha con Excel.

1.4.1. Gráfica de sectores circulares

Esta gráfica también es conocida con el nombre de gráfica de pie o pastel[21], por cierto, se usa mayormente para representar datos cualitativos, sin embargo los datos cuantitativos también pueden ser representados así.
Se le pueden poner colores, hacerla en dos o tres dimensiones[22] .

Para construirla en primer lugar debemos convertir las frecuencias de los datos en porcentaje. En segundo lugar mediante una regla de tres simple donde 360º = N, es decir, el número total de datos, convertir a grados sexagesimales los porcentajes de las frecuencias. Por último en un círculo trazamos ángulos centrales proporcionales al tamaño de las frecuencias o los porcentajes de ellas[9], lo ilustraremos con el siguiente ejemplo simulado:

Los alumnos de CONALEP de un grupo de 30 alumnos, votaron por sus compañeros Santiago Torres, Juan Pérez y Alicia Ochoa para elegir el jefe del grupo, los resultados son los siguientes:

CANDIDATOS	VOTOS (f)
SANTIAGO	15
JUAN	8
ALICIA	7
	N=30

En este ejemplo tenemos datos cuantitativos, es indistinto colocar primero a Santiago, Juan o Alicia, por lo que la gráfica de sectores circulares es adecuada para representar los resultados de la tabla anterior.

Como se mencionó anteriormente el primer paso es convertir las frecuencias a porcentaje mediante la regla de tres simple, donde el 100% = 30 alumnos, la pregunta a responder es: ¿qué porcentaje del total de los datos representan 15, 8 y 7 alumnos?

100% = 30 alumnos

x% = 15 alumnos

$$x = \frac{15\ alumnos \times 100\%}{30\ alumnos} = 50\%$$

De la misma manera se procede para hacer el cálculo de las frecuencias faltantes y tendremos:

CANDIDATOS	VOTOS f	$\%f$
SANTIAGO	15	50%
JUAN	8	26.6%
ALICIA	7	23.3%
	N=30	$99.9\% \cong 100\%$

El segundo paso es convertir a grados sexagesimales la frecuencia simple o el porcentaje de la frecuencia simple, aquí lo realizaremos con el porcentaje de la frecuencia de la siguiente manera:

$$360^{o}=100\%$$

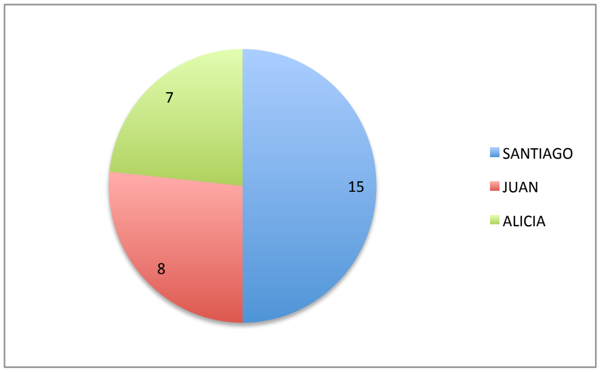
$$x^{o}=50\%$$

$$x=\frac{50\%\times360^{0}}{100\%}=180^{0}$$

hacemos lo mismo con los datos faltantes y tendremos:

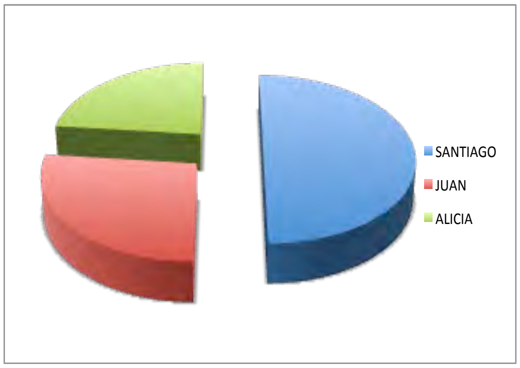
CANDIDATOS	VOTOS (f)	$\%f$	GRADOS
SANTIAGO	15	50%	180º
JUAN	8	26.6%≅ 27%	95º45′
ALICIA	7	23.3%≅ 23%	83º52′
	N=30	99.9% ≅ 100%	359º37′≅360º

Por último en un círculo hacemos el trazo con ayuda de un transportador y compás o por medio de algún programa para graficar, dicha gráfica se muestra a continuación:



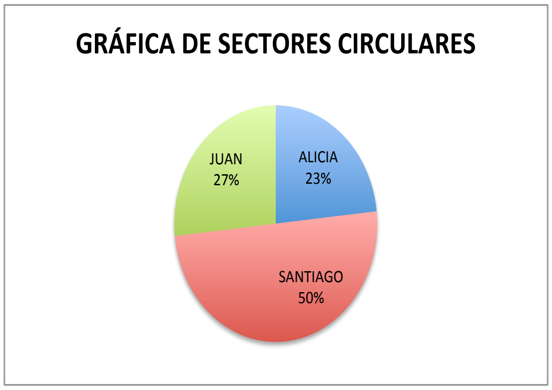
Gráfica de sectores circulares, se indica candidatos y número de votos que obtuvieron

Al leer la gráfica anterior no obstante que la tabla no esté a la vista, es claro que la mitad del grupo votó por Santiago y que es el ganador, ya que el área marcada en el círculo es la mitad y representa entonces al 50% de los votantes, los sectores de Juan y Alicia son casi iguales por lo que podemos concluir que la preferencia por ellos fue muy cercana, inclusive igual si la gráfica no indica los votos, como ocurre en la siguiente gráfica en 3D seccionada, que a simple vista Juan y Alicia parecen haber obtenido igual número de votos.



Gráfica 3D de sectores seccionados

La creatividad y finalidad de representar los resultados de forma gráfica, hará que quien presente la gráfica añada datos explícitos de los resultados obtenidos además de poner a la vista la tabla.



Gráfica de sectores circulares, se muestra los candidatos y porcentaje que votó por cada uno de ellos

CANDIDATOS	VOTOS (f)
SANTIAGO	15
JUAN	8
ALICIA	7
	N=30

Para resolver: utiliza los datos de tus calificaciones del periodo para representarlas en una gráfica de sectores circulares.

Curiosidades estadísticas:

Los babilonios utilizaban pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y los artículos vendidos, hace más de 3000 años

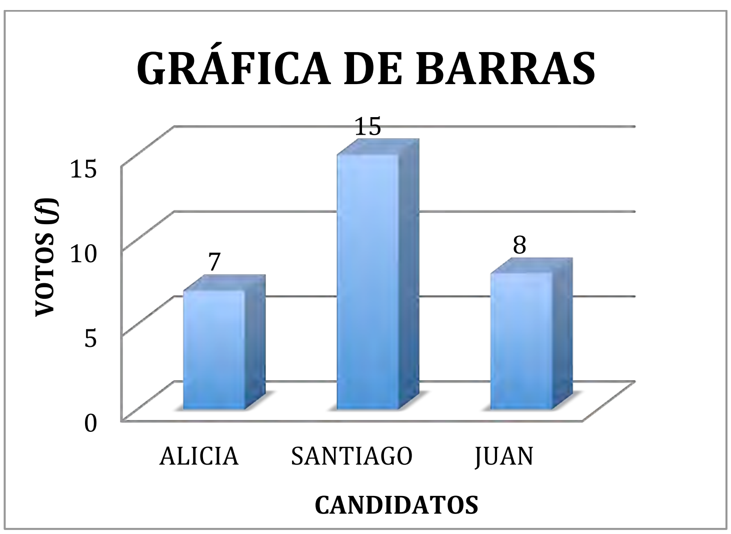
1.4.2. Gráfica de barras

Este tipo de gráficas es utilizado también para representar datos cualitativos, mediante rectángulos cuya base indica la categoría y la altura es proporcional a la frecuencia[23].

Para ejemplificar esta gráfica utilizaremos los datos:

CANDIDATOS	VOTOS (f)
SANTIAGO	15
JUAN	8
ALICIA	7
	N=30

En resumen, cada candidato representa la base de cada rectángulo que se grafica en el eje horizontal, con una altura proporcional a la frecuencia, la cual deberá graficarse en el eje vertical y comenzar en cero. Nótese que cada barra tiene el mismo ancho y el eje vertical representa las frecuencias, los candidatos pueden estar colocados en cualquier orden como se indica en la siguiente gráfica:



1.4.3. Histograma

Se atribuye al historiador y matemático, fundador de la bioestadística Karl Pearson, el haber acuñado el término histograma[24], tipo de gráfica usada para representar datos cuantitativos, formada por rectángulos de base igual a los límites verdaderos y con una altura proporcional a la frecuencia de la clase[9].

Anteriormente cuando organizamos datos en clases o intervalos definimos los límites aparentes de una clase[12], pero dichos datos pudieron haber sido resultado del redondeo, por ejemplo si un alumno obtiene 6.9 suele escribirse 7 en su boleta de calificaciones.

Para representar datos en un histograma es necesario convertir los límites aparentes a verdaderos, esto se hace restando 0.5 al límite aparente inferior y sumando 0.5 al límite aparente superior[9], logrando así que las clases no tengan espacio entre ellas, las clases estarán representadas en el eje horizontal , y las alturas de cada rectángulo serán proporcionales a las frecuencias de las clases en el eje vertical.

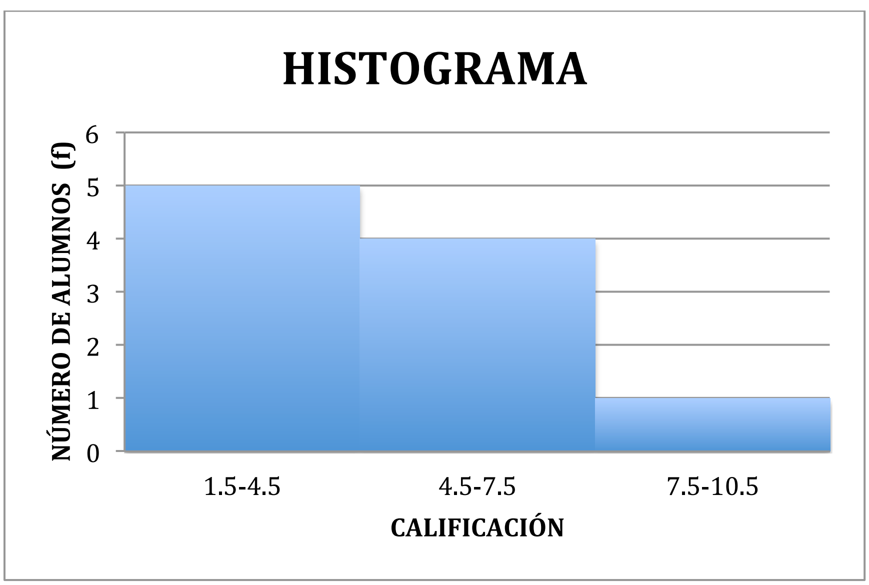
Consideremos la siguiente tabla de distribución en clases, que representa las calificaciones obtenidas por un grupo de 10 alumnos en una determinada asignatura:

CALIFICACIÓN	f
8-10	5
5-7	4
2-4	1
	N=10

Primeramente convertimos los límites aparentes a reales y tendremos:

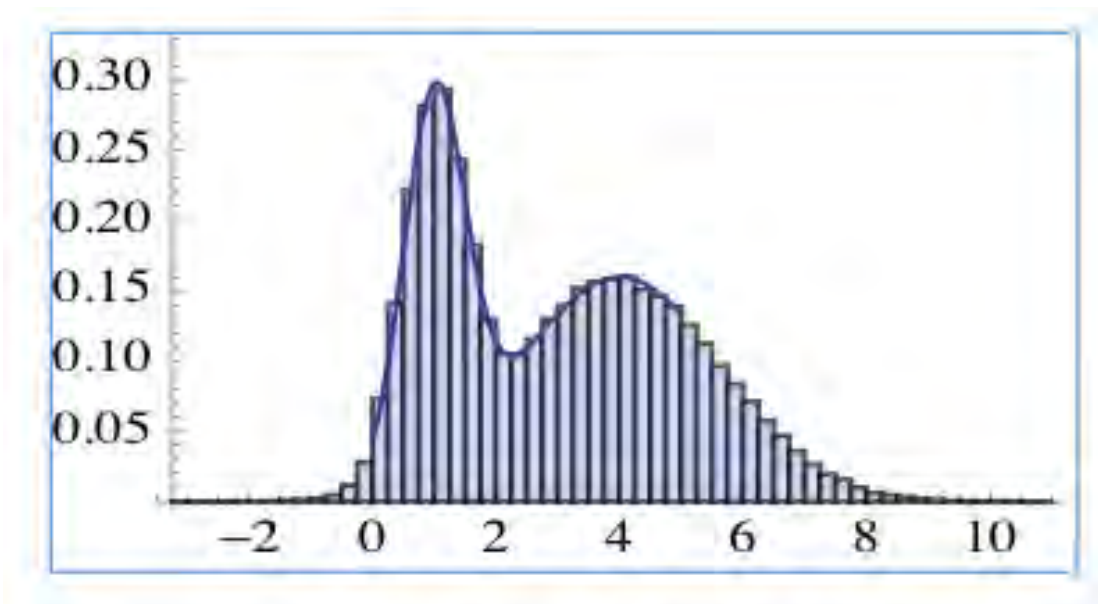
CALIFICACIÓN	f
7.5-10.5	5
4.5-7.5	4
1.5-4.5	1
	N=10

Observe que las frecuencias de clase no cambian, y que ahora cada clase termina donde comienza la que sigue en orden ascendente, por lo que la gráfica es muy parecida a la de barras, sin embargo no hay espacio entre las barras, los histogramas suelen usarse en datos cuantitativos, el histograma de la tabla anterior es el que se muestra enseguida:



Observe que aquí los intervalos no pueden cambiar de orden ya que existe un orden creciente sobre el eje horizontal, pero que al igual que la gráfica de barras cada intervalo deberá tener el mismo ancho, también note que la frecuencia comienza en cero y está representada en el eje vertical.

De la gráfica podemos hacer varias aseveraciones sin tener presente la tabla de resultados, como que la mayoría de los alumnos (5) obtuvieron una calificación entre 7.5 y 10.5, o que un solo alumno obtuvo calificación menor a 4.5 pero mayor que 1.5, etc.



Histograma hecho con mathematica 9

1.4.4. Polígono de frecuencias

Cuando tenemos datos organizados mediante tablas en clases o intervalos, podemos suponer que el punto medio o marca de clase puede representar a todos los datos que pertenecen a dicha clase, esto es subjetivo sobre todo cuando existen pocos datos, aunque parece no tener muchas alteraciones a la hora de interpretar los datos cuando el conjunto de datos es muy grande[9,18].

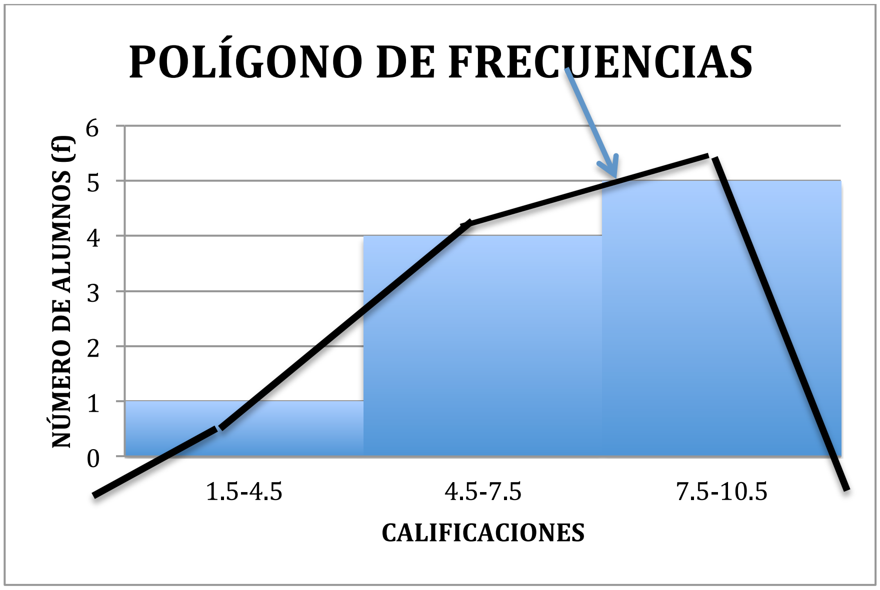
Si tenemos un intervalo con los límites aparentes o con los límites reales, el punto medio o marca de clase se calcula sumando los límites y dividiendo el resultado de la suma entre dos[25], veamos un ejemplo utilizando la siguiente tabla de datos:

CALIFICACIÓN	f
8-10	5
5-7	4
2-4	1
	N=10

En la primera clase en orden descendente 8-10 la marca de clase o punto medio es: $\frac{8+10}{2}$ =9, así podemos suponer que los datos 8, 9 y 10 son iguales o representados por el valor 9, note que si trabajamos con los límites reales tendremos: $\frac{7.5+10.5}{2} = 9$ por lo que podemos concluir que no importa con qué tipo de límites estemos trabajando la marca de clase o punto medio es la misma. La siguiente tabla muestra las marcas de clase y la frecuencia necesaria para graficar un polígono de frecuencias, donde se localizan los puntos en el plano XY formados por el valor de la marca de clase y su respectiva frecuencia, unidos los puntos obtenidos mediante segmentos. El polígono comienza y termina en el eje horizontal:

CALIFICACIÓN CLASES	MARCA DE CLASE O PUNTO MEDIO	f
8-10	9	5
5-7	6	4
2-4	3	1
		N=10

El polígono de frecuencias que a continuación se muestra fue construido sobre el histograma del mismo conjunto de datos, lo que facilita su construcción:



Tanto los histogramas como los polígonos de frecuencias se utilizan para representar las distribuciones de frecuencia en clases, donde en el eje horizontal se representan las clases y la frecuencia simple, o el porcentaje de la frecuencia en el eje vertical[18].

Para resolver: investiga qué otros tipos de gráficas se pueden elaborar, elige una y usa tu creatividad para mejorarla, puede ser agregando imágenes, dibujos, color, etc.

Datos curiosos de estadísticas:

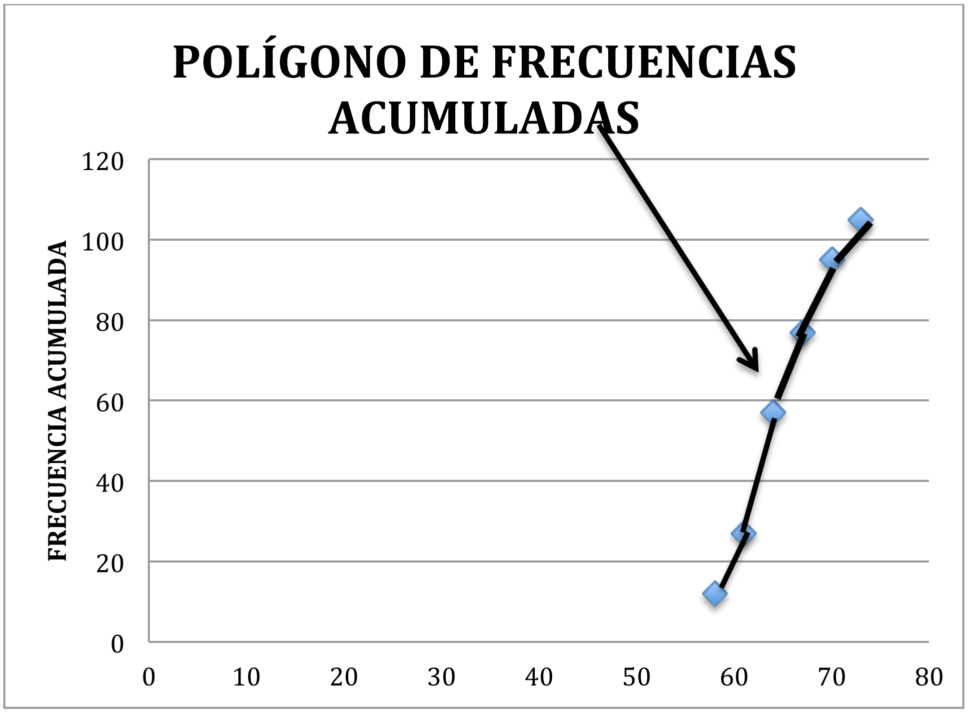
El 60% de los aproximadamente 900 millones de analfabetos en el mundo son mujeres, según datos de la UNESCO.

1.4.5. Polígono de frecuencias acumuladas (ojivas)

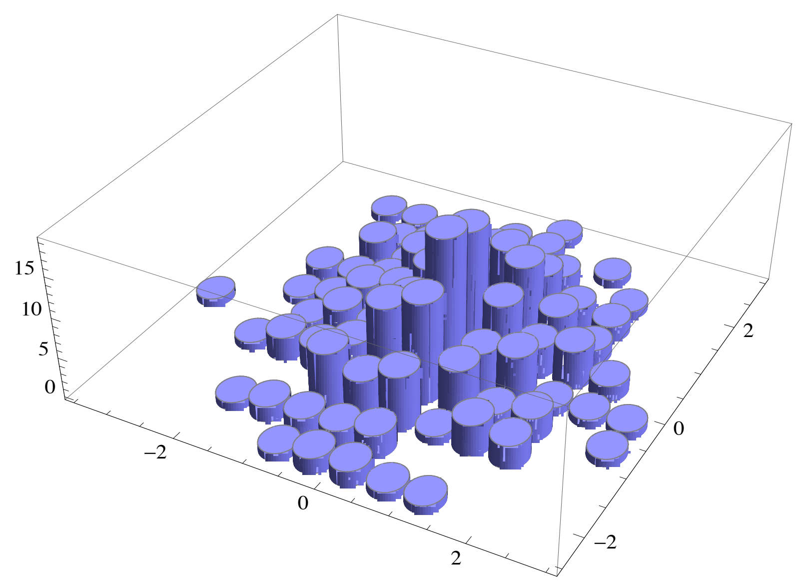
Este tipo de gráficas es la representación de una distribución de frecuencias acumuladas, para construirla primero se localizan los límites superiores de cada clase sobre el eje horizontal, después, se localizan en el plano los puntos formados por los límites superiores de la clase con su respectiva frecuencia acumulada que deberá estar representada en el eje vertical, por último unimos mediante segmentos de recta los puntos y la gráfica así obtenida recibe el nombre de polígono de frecuencias acumuladas, cuando la variable estadística es continua la gráfica recibe el nombre de ojiva [9,15].

Consideremos la tabla siguiente:

CLASES	FRECUENCIA	FRECUENCIA ACUMULADA
71-73	10	105
68-70	18	95
65-67	20	77
62-64	30	57
59-61	15	27
56-58	12	12
	N=105	



Note que la frecuencia inicia en cero, y que el punto final (73) corresponde al 100% del total de los datos, al leer la gráfica se puede identificar cuántos datos son menores o iguales a un determinado límite superior[7,9,18].



Histograma en 3D. Gráfica hecha con mathematica 9

1.4.6. Gráfica de tallo y hojas

Esta forma de esquematizar los datos es muy parecida al histograma y es utilizada para representar datos cuantitativos cuando tenemos un gran número de datos[26]; más bien parece una forma de organizar y no una gráfica, la explicaremos utilizando un ejemplo:

En una sala de cine donde presentan la película X, llevan el conteo de cuántos clientes asisten durante 20 días que duró la proyección y los resultados son los siguientes:

100	65	100	70	80
80	100	90	72	83
50	62	60	60	81
57	63	61	65	77

Para organizar en tallo y hojas identificamos el menor y el mayor de los datos, en este caso 50 y 100 respectivamente, como el número menor tiene dos dígitos tomamos el primero para formar el tallo en este caso 5, como van los datos de 10 en 10 ponemos las decenas que siguen y como 100 tiene 3 dígitos tomamos los primeros dos que son 10, y la derecha de la línea ‖ los dígitos que forman el número, observemos cuántos comienzan con 5, son el 50 y 57, por lo tanto colocamos 0, 7 y hacemos lo mismo con los números restantes obteniendo lo siguiente:

5	0,7
6	5,2,0,0,3,1,5
7	0,2,7
8	0,0,3,1
9	0
10	0,0,0

Ordenando las hojas en orden creciente tenemos:

5	0,7	se lee 50 y 57 personas que asistieron
6	0,0,1,2,3,5,5	se lee 60,60,61,62,63,65,65
7	0,2,7	se lee 70,72,77
8	0,0,1,3	se lee 80,80,81,83
9	0	se lee 90
10	0 ,0,0	se lee 100,100,100

Esta forma de organizar los datos es conocida como gráfica de tallo y hojas.

1.5. Problemario

1. Obtener una muestra aleatoria del 3% para una población de 500. Utiliza la siguiente tabla de números aleatorios por columna y de izquierda a derecha.

2575	933	5066	7451	9525	1536	9608
2336	594	8198	6454	1347	9027	3729
3461	2619	8511	8582	2794	9326	8184
5181	7136	5620	6492	3568	3869	7222
6407	3241	2789	6628	596	676	6042
952	3551	5380	2261	2831	1628	737
9285	556	3518	9173	6987	1958	1440
4471	5882	5811	820	6602	1511	5208
2575	3166	756	6565	5138	807	1718
1993	8058	1442	355	8636	8585	9460
3821	4338	6813	3395	4169	2126	1206
946	9427	5558	4804	1601	8352	5421
8731	4971	372	3182	6077	5814	9698
8512	9654	2838	7400	1809	3424	7950
6778	8352	289	4399	437	2586	5543
4278	8302	9698	5666	8809	4924	3980
633	9248	4422	3938	5599	2958	4542
8234	8163	6784	4681	3850	5523	8004
8219	6878	9903	6532	885	9834	7304
6228	8149	964	3263	6503	6185	9862

2. 20 alumnos de un grupo obtuvieron las siguientes calificaciones en la clase de representación simbólica y angular del entorno, construir la distribución de frecuencias simple.

9	10	10	6	5
6	8	8	7	6
5	10	9	7	9
8	6	7	6	7

3. ¿Qué porcentaje del total de los alumnos obtuvo 7 de calificación respecto a los datos del ejercicio anterior?

4. La siguiente tabla muestra los resultados de una encuesta realizada, construye una distribución en clase de ancho 3:

EDAD (años)	*f*
65	3
64	2
63	4
62	3
61	1
60	5
59	6
58	0
57	3
56	3
55	10
54	5
53	7
52	6
51	12
50	0
49	12
48	1
	N=83

5. Con la tabla siguiente construye una distribución de frecuencias acumuladas:

PESO (kg)	*f*
80	10
79	4
78	3
77	6
76	23
75	8
	N= 54

6. Construye una distribución porcentual acumulativa, con los datos de la siguiente tabla:

x	f
12	5
11	3
10	2
9	7
8	8
7	9
	N=34

7. Construir una gráfica de sectores circulares utilizando los datos de la siguiente tabla:

CANDIDATOS	Votos (f)
Sonia Torres	15
Rogelio Ochoa	15
Pedro González	5
	N=35

8. Con los datos del ejercicio anterior construir una gráfica de barras.

9. Construir un histograma y un polígono de frecuencias en la misma gráfica , utilizando la siguiente tabla de datos:

CLASES	*f*
63-65	9
60-62	9
57-59	9
54-56	18
51-53	25
48-50	13
	N=83

10. Construir un polígono de frecuencias acumuladas para la siguiente tabla de datos:

DATOS	*f*
12	26
11	21
10	16
9	12
8	11
7	4
	N=90

11. Construir un polígono de porcentajes acumulados, utilizando la siguiente tabla:

DATOS	f
10	3
9	3
8	3
7	4
6	5
5	2
	N=20

1.6. Autoevaluación

1. Construir una distribución de frecuencias simple con las 20 calificaciones obtenidas en un grupo al aplicar un examen de historia universal:

7	8	6	5	7	5
9	5	8	9	6	8
6	7	6	7	4	7
7	8	10	6	9	4
8	9	6	8	5	9

2. Si utilizamos los datos anteriores, aproximadamente ¿cuántas clases quedarán si construimos una distribución en clases de ancho 2?

3. Construir una distribución en clases de ancho 2 con los datos del ejercicio 1, tomar como primera clase la 9-10, hacerlo en orden descendente.

4. Construir una distribución de frecuencias acumuladas para los datos obtenidos en el ejercicio 3.

5. Construir una gráfica de sectores circulares, con los datos obtenidos de la preferencia de un color:

Color	f
Azul	14
Rojo	12
Verde	10
Amarillo	7
	N=43

6. Los resultados obtenidos para identificar las tribus urbanas en la ciudad de Morelia son los siguientes, construir una gráfica de barras:

TRIBUS URBANAS	f
Emos	600
Metaleros	450
Rastafaris	40
Frikiss	234
Punks	8
Skatos	150
Góticos	347
	N=1,829

7. ¿Qué tipos de gráficas deben utilizarse de preferencia para representar datos cualitativos?

8. ¿Cómo se llaman los números que definen una clase o intervalo?

9. ¿Qué nombre recibe el número que representa a una clase o intervalo?

10. Indica cuándo es mejor tomar toda la población y cuándo una muestra.

1.7. Soluciones del problemario

1. El 3% de 500 son 15 y son los elementos que tengan el número: 257,233,246,095,447,257,199,382,094,427,063,093,059,261,324.

2.

x	f
10	3
9	3
8	3
7	4
6	5
5	2
	N= 20

3. Cuatro alumnos obtuvieron la calificación 7 igual al 20%

4.

CLASES	f
63-65	9
60-62	9
57-59	9
54-56	18
51-53	25
48-50	13
	N=83

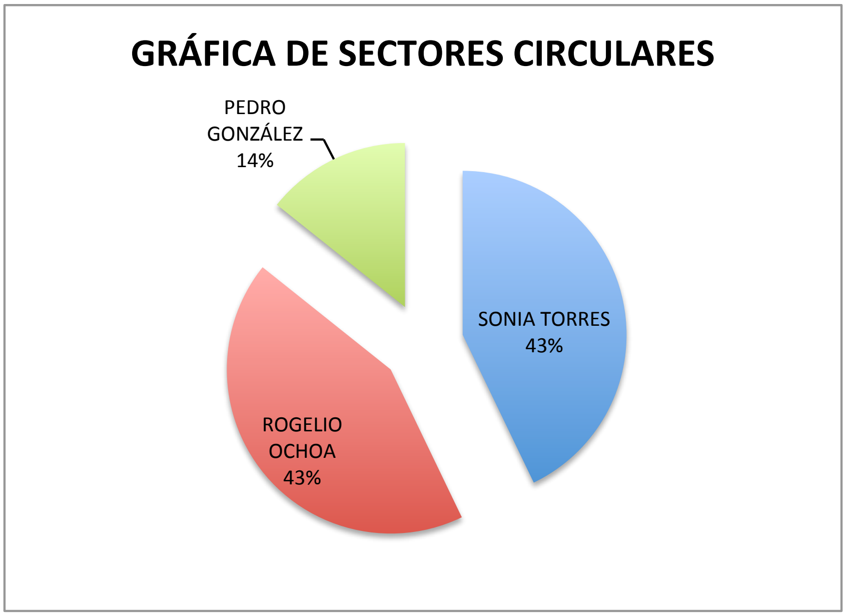
5. Distribución de frecuencias acumuladas

PESO (kg)	f	fa
80	10	54
79	4	44
78	3	40
77	6	37
76	23	31
75	8	8
	N= 54	

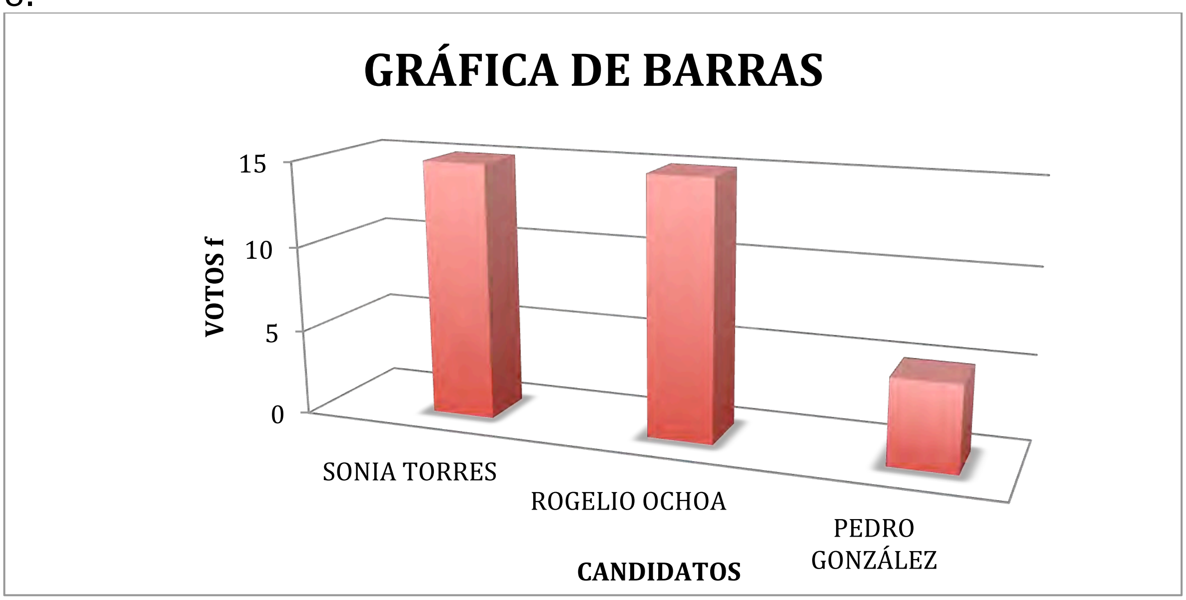
6. Distribución porcentual acumulativa

x	f	fa	$\%fa$
12	5	34	100%
11	3	29	85.29%
10	2	26	76.47%
9	7	24	70.58%
8	8	17	50%
7	9	9	26.47%
	N=34		

7.

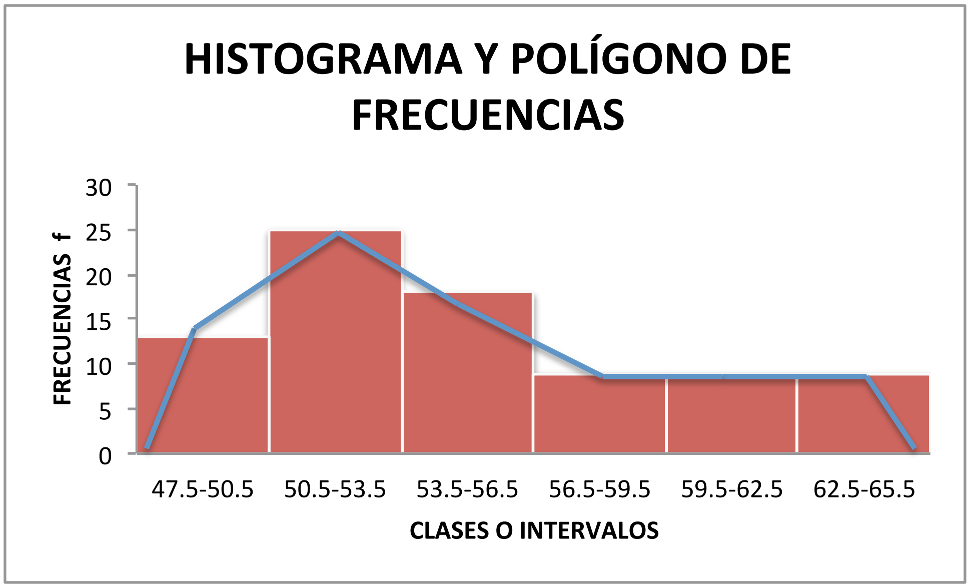

GRÁFICA DE SECTORES CIRCULARES
PEDRO GONZÁLEZ 14%
SONIA TORRES 43%
ROGELIO OCHOA 43%

8.

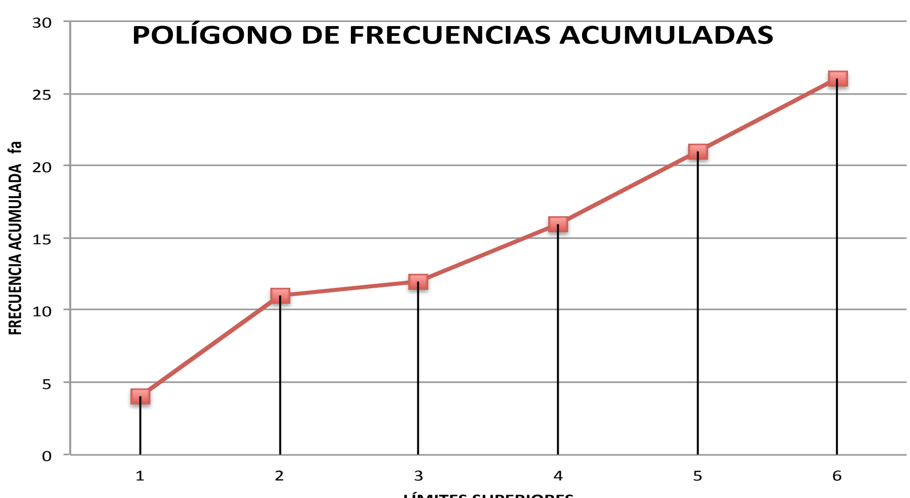

GRÁFICA DE BARRAS
15
10
5
0
VOTOS f
SONIA TORRES
ROGELIO OCHOA
PEDRO GONZÁLEZ
CANDIDATOS

9. Convertimos los límites aparentes a verdaderos y ordenamos de forma descendente.

CLASES	f
47.5-50.5	13
50.5-53.5	25
53.5-56.5	18
56.5-59.5	9
59.5-62.5	9
62.5-65.5	9
	N=83



10. Polígono de frecuencias acumuladas



11.Distribución porcentual acumulativa

DATOS	f	fa	$\%fa$
10	3	20	100%
9	3	17	85%
8	3	14	70%
7	4	11	55%
6	5	7	35%
5	2	2	10%
	N=20		

1.8. Soluciones de autoevaluación

1. Distribución de frecuencias simple:

CALIFICACIÓN	f
10	1

9	5
8	6
7	6
6	6
5	4
4	2
	N=30

2. Número de clases$\cong \frac{Rango}{longitud\ del\ intervalo} \cong \frac{10-4}{2} \cong \frac{6}{2} \cong 3$

3. Distribución en clases de ancho 2:

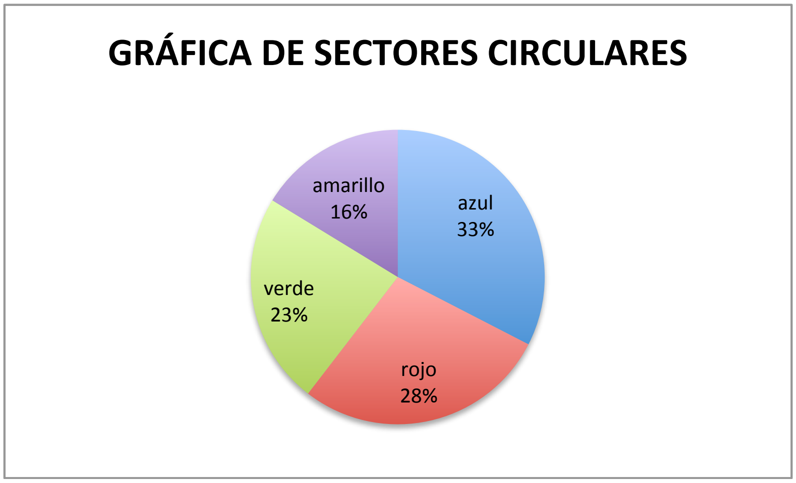
CLASES	f
9-10	6

7-8	12
5-6	10
3-4	2
	N=30

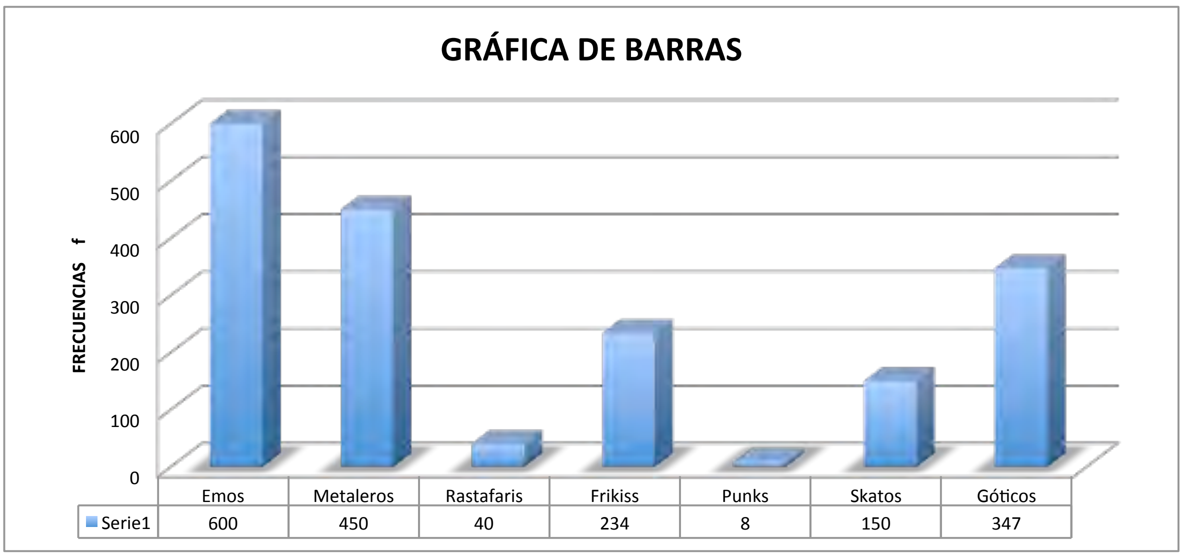
4. Distribución de frecuencias acumuladas:

CLASES	f	fa
9-10	6	30
7-8	12	24
5-6	10	12
3-4	2	2
	N=30	

5. Gráfica de sectores circulares:



6. Gráfica de Barras:



7. Gráfica de sectores circulares y de barras.

8. Límites aparentes o verdaderos, inferior y superior.

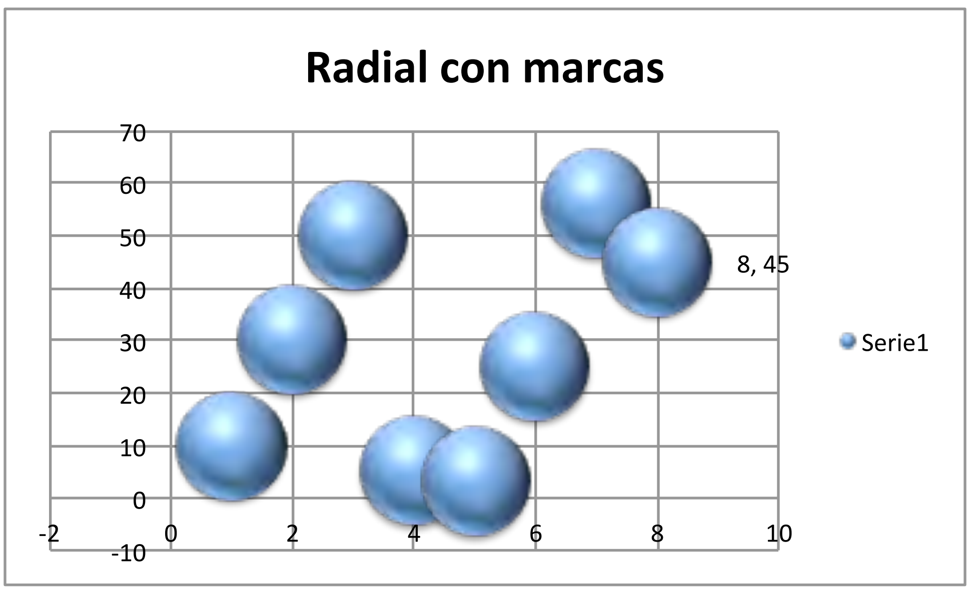
9. Marca de clase o punto medio del intervalo.

10. Cuando una población es muy pequeña es conveniente tomar toda la población, y cuando es muy grande la población conviene tomar una muestra aleatoria.

1.9. Conclusión

Existen otras formas de organizar los datos, también más tipos de gráficas como la de caja y brazos, aquí la creatividad es el recurso que hace que sean mejores o llamativas , y puedan tener buenos resultados, por ejemplo en publicidad.

Te invitamos a profundizar tanto en la metodología como en la interpretación cuidadosa de resultados.



REFERENCIAS

[1] Abreud José L, Barot Michael, et al. (2010) Enciclopedia de conocimientos fundamentales. Volumen 5 página 108. México: UNAM-Siglo veintiuno Editores

[2] SaundersnMark, Brown Brandon(2008)Dealing with Statistics: What You Need to Know. USA: McGraw-Hill Education.

[3] Jackson Sherri L. (2012) Research Methods and Statistics: A Critical Thinking Approach. USA: Wadsworth.

[4] Johnson Robert, Kuby Patricia(2004)Estadística elemental: lo esencial. México: THOMSON

[5] Guardia Joan, Freixa Montserrat et al.(2008)Análisis de datos en psicología. Madrid: DELTA PUBLICACIONES.

[6] Alea Mª Victoria, Guillén Montserrat et al. (2001) Estadística descriptiva: aplicaciones prácticas. España: Editions de la Universidad de Barcelona.

[7] Quintana Carlos(1996)Elementos e inferencia estadística.Costa Rica: Editorial de la Universidad de Costa Rica.

[8] INEGI,(2010)cuentame.inegi.gob.mx. Recuperado el 4 de junio de 2013, de http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/censos/poblacion/2010/panora_socio/mich/Panorama_Mich.pdf

[9] Naresh K, Malhotra et al. (2004) Investigación de Mercados. México: Pearson Educación.

[10] Sheldon M: Ross(2007) Introducción a la Estadística. España: Editorial Reverté, S.A.

[11] Grasso Livio (2006) Encuestas: elementos para su diseño y análisis. Argentina: Encuentro Grupo Editor

[12] http://www.colloqui.org/colloqui/2012/8/16/analisis-de-las-estimaciones-de-las-casas-encuestadoras-vs-l.html. Consultado el 23 de julio de 2013.

[13]Parametría SA de CVhttp://www.parametria.com.mx/DetalleEstudio.php?E=4331. Consultado el 23 de julio de 2013.

[14] Portilla Enrique (1984) Estadística, primer curso. México: Nueva Editorial INTERAMERICANA S.A. de C.V.

[15] *exceltotal.com*. Consultado el12 de junio de 2013, http://exceltotal.com/distribucion-de-frecuencias-en-excel/

[16] Walpole Ronald E. (1999)Probabilidad y estadística para ingeniería. México: Prince-hall Hispanoamericana S.A.

[17] Hernández Óscar (2004)Estadísitca elemental para Ciencias Sociales. Costa Rica: Editorial de la Universidad de Costa Rica.

[18] Ulloa Victor, Quijada Verónica, et al. (2006) Estadísitca Aplicada a la Comunicación. México: UNAM Facultad de Estudios Superiores de Acatlán.

[19] Vladimirovna Olga (2005)Fundamentos de Probabilidad y Estadísitica. México: Universidad Autónoma del Estado de México.

[20] Vargas Antonio (1995) Estadística Descriptiva e inferencial. España: Universidad de Castilla-La Mancha.

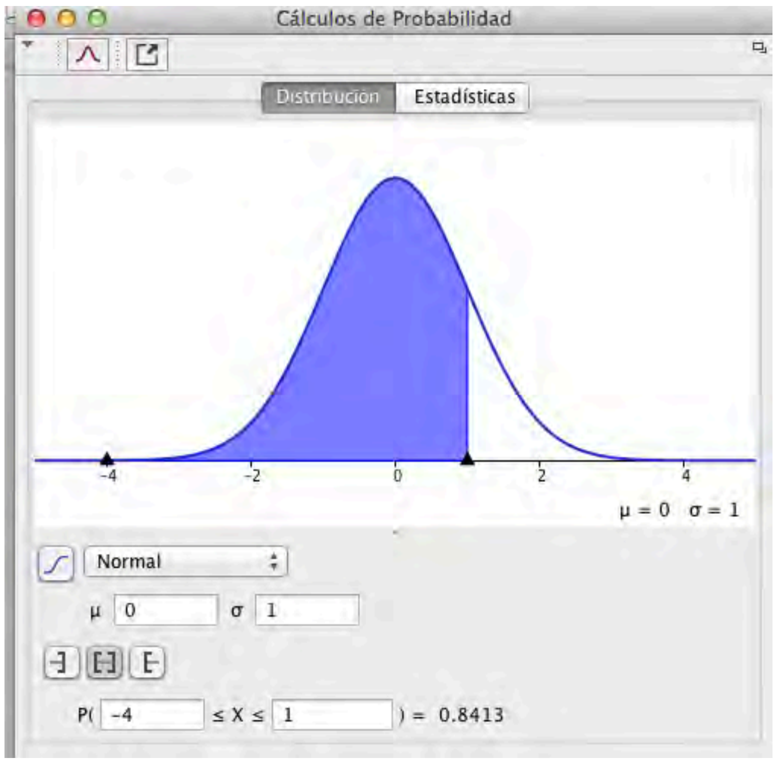
[21] Lincoln L. Chao (1997)Introducción a la Estadística. México: CECSA.

[22] Freund John, Simon Gary (1994) Estadística Elemental. México: PRENTICE HALL HISPANOAMERICANA, S.A.

23 Berenson Mark, Levine David(1996)Estadística básica en Administración, conceptos y aplicaciones. México: Prentice Hall Hispanoamericana, S.A.
24 Universidad Virtual de Colombia sede Bogotá. Consultado el 7 de julio de 2013, de: http://www.virtual.unal.edu.co/cursos/ciencias/2001065/html/un1/cont_114_14.html
25 Álvarez Rafael (2007)Estadística Aplicada a las Ciencias de la Salud. España: Ediciones Díaz de Santos.
26 Monegal Mariona (1999) Introducción al SPSS: Manipulación de datos y estadística descriptiva. España: Editions de la Universitat de Barcelona

Capítulo 2:

Medidas estadísticas



2. Introducción

En el capítulo anterior organizamos datos numéricos, los representamos mediante tablas y gráficamente, en este capítulo realizaremos cálculos de algunas medidas numéricas que nos brindan información sobre las características de las series de datos, estas son las medidas de tendencia central, medidas de posición no central, medidas de dispersión y medidas de forma.

Estas medidas se utilizan tanto para analizar las series de datos, como para comparar entre series distintas de datos.

Las utilizan los psicólogos[1] para interpretar sus pruebas psicológicas, analistas de negocios para evaluar el funcionamiento de sus empresas y promedio de ventas, así como los ingresos y gastos. Los investigadores en general de distintas áreas, muestran sus informes de sus observaciones (de laboratorio o sociales), por medio de estas medidas.

Con este tipo de medidas podemos describir con un solo número dónde está ubicado el valor central de los datos, qué dato balancea a la serie, cuán alejados respecto de los valores centrales se encuentran el resto de los datos numéricos. El cálculo de estas medidas no es difícil, interpretar los resultados y tomar más de una medida es lo importante para hacer una descripción adecuada y completa de los datos.

Este tipo de medidas requiere de cálculo de operaciones con sumatorias cuyo signo es la letra griega sigma Σ, se lee la suma de, así la suma de $x_1 + x_2 + x_3$ se puede representar con la simbología $\sum_1^3 x_i$ o si queremos indicar que la suma es de todos los valores de x, simplemente se denota $\sum x$.

2.1. Medidas de posición central

2.1.1. Media aritmética

Dado un conjunto de datos $x_1, x_2, x_3, \ldots, x_N$ *no agrupados*, la media aritmética, promedio o simplemente media, se define como :

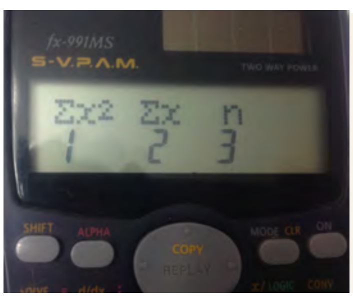
$$\bar{x} = \frac{\sum x}{N}$$

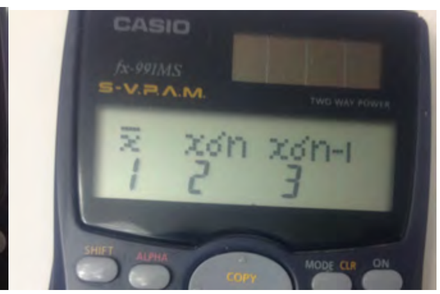
$\bar{x} = se\ lee\ x\ testada$

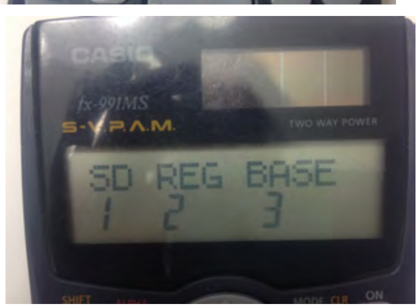
donde:

$N = número\ total\ de\ datos$

Las calculadoras científicas son de gran ayuda en los cálculos laboriosos, lee tu manual e identifica la manera de introducir los datos (modo) y la forma de obtener la media aritmética







Ejemplo:

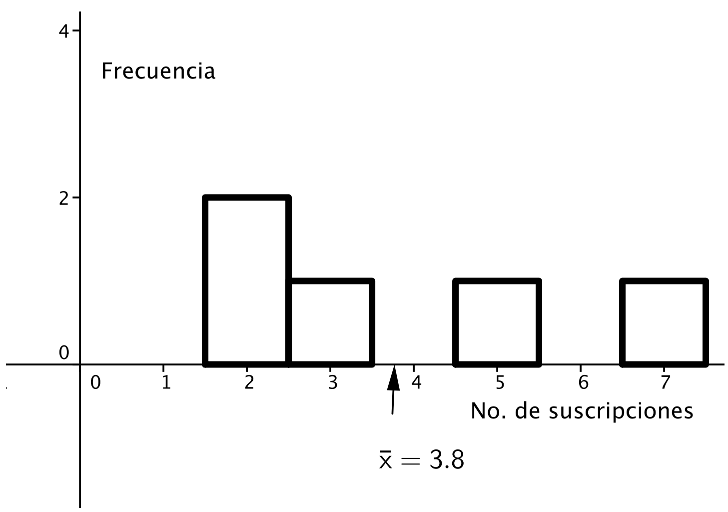
Obtener la media aritmética de la siguiente muestra de 5 observaciones, que representan el número de suscripciones a un servicio de telefonía en 5 días.

$$x_1 = 2,\ x_2 = 2,\ x_3 = 7,\ x_4 = 5,\ x_5 = 3$$

De acuerdo con la definición debemos sumar los datos y a dicha suma dividirla entre el número total de datos, esto es:

$$\bar{x} = \frac{\sum x}{N} = \frac{2+2+7+5+3}{5} = \frac{19}{5} = 3.8$$

esto significa que en promedio se suscriben aproximadamente 4 personas diarias.



La media aritmética para datos agrupados en una distribución de frecuencias simple se calcula mediante la expresión:

$$\bar{x} = \frac{\sum fx}{N}$$

donde f es la frecuencia o número de veces que aparece un dato x en el conjunto de datos, esto es:

Datos	f
x_1	f_1
x_2	f_2
$\vdots$	$\vdots$
x_N	f_N

Ejemplo:

Las calificaciones de un grupo de 30 alumnos son las siguientes, calcular la media

x	f
10	8
9	4
8	6
7	9
6	2
5	1
	N=30

La fórmula para calcular la media aritmética

$$\bar{x} = \frac{\sum fx}{N}$$

nos indica que cada dato se multiplica por su respectiva frecuencia, se obtiene la suma de dichos productos y el resultado se divide entre el número total de datos.

Por lo que en la tabla original es conveniente agregar una columna que indique el producto fx, como se muestra a continuación:

x	f	fx
10	8	80
9	4	36
8	6	48
7	9	63
6	2	12
5	1	5
	N=30	$\Sigma fx = 244$

$$\bar{x} = \frac{\sum fx}{N} = \frac{244}{30} = 8.13$$

Cuando los datos están agrupados en clases o intervalos la media aritmética se calcula con la fórmula:

$$\bar{x} = \frac{\sum fx}{N}$$

donde x representa la marca de clase o el punto medio de la clase.

Ejemplo:

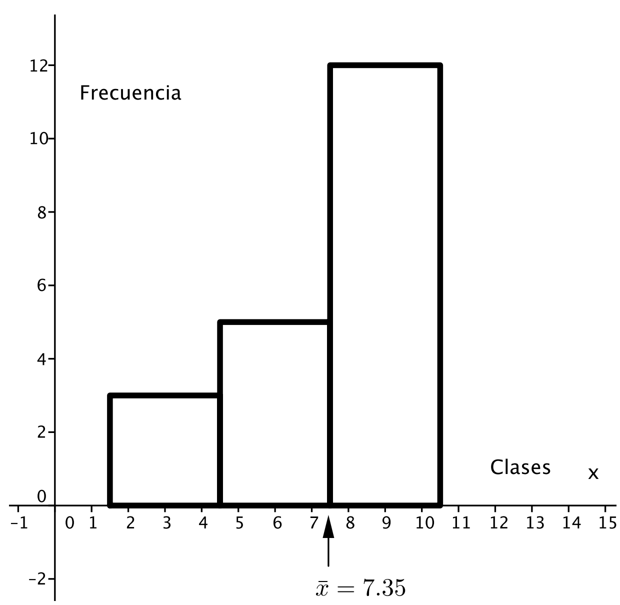
Calcular la media del siguiente conjunto de datos

Clase	f
8-10	12
5-7	5
2-4	3
	N=20

En la tabla anterior agregamos una columna para indicar x la marca de clase o punto medio, y otra para el producto de fx.

Marca de clase x	Clase	f	fx
9	8-10	12	108
6	5-7	5	30
3	2-4	3	9
		N=20	$\Sigma fx = 147$

$$\bar{x} = \frac{\sum fx}{N} = \frac{147}{20} = 7.35$$



2.1.2. Media geométrica

La media geométrica ($M.G.\ o\ X_G$) es utilizada con datos relativos, como porcentajes. Mide la razón de cambio de una variable en el tiempo y se utiliza[2] para promediar razones, tasas de variación e índices económicos, la media geométrica es menor o igual a la media aritmética, y se calcula por medio de la fórmula:

$$M.G. = \sqrt[n]{x_1 x_2 x_3 \ldots x_N} \text{ para datos } sin\ agrupar$$

Esta fórmula nos indica que la media geométrica es la raíz n-ésima del producto de los N datos (positivos).

Por ejemplo:

Una cuenta bancaria ha registrado los siguientes valores durante 5 años, una vez que se le ha agregado el interés correspondiente a la ganancia anual:

2009=\$101.5, 2010=\$135.8, 2011=\$150, 2012= \$151.05 y 2013= \$152

Calcular la media geométrica M.G.

$$M.G. = \sqrt[5]{(101.5)(135.8)(150)(151.05)(152)}$$

$$M.G. = \sqrt[5]{4.7470 \times 10^{10}} = \$136.54$$

Si calculamos la media aritmética para comparar dichos resultados tendremos:

$$\bar{x} = \frac{\sum x}{N} = \frac{101.5 + 135.8 + 150 + 151.05 + 152}{5} = \$138.07$$

Como se puede observar se encuentran muy cercanos los dos valores pero la media geométrica es menor que la aritmética.

$$M.G. \leq \bar{X}$$

La media geométrica para *datos agrupados* en clases se define como:

$$M.G. = \sqrt[N]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \dots x_N^{f_N}}$$

Por ejemplo:

Calculemos la media geométrica del siguiente conjunto de datos:

x	f
10	8
9	4
8	6
7	9
6	2
5	1
	N=30

$$M.G. = \sqrt[N]{x_1^{f_1} x_2^{f_2} x_3^{f_3} \dots x_N^{f_N}} = \sqrt[30]{(10)^8(9)^4(8)^6(7)^9(6)^2(5)^1}$$

$$M.G. = \sqrt[30]{1.2492 \times 10^{27}} = 8.0024$$

que al comparar con $\bar{x} = 8.13$ se verifica que se cumple que $M.G. \leq \bar{X}$

Cuando los datos están *agrupados en clases* o intervalos, recuerda que el punto medio o marca de clase representa a la clase, la media geométrica[3] está dada por la expresión:

$$\log M.G. = \frac{\sum f_i \log x_i}{N}$$

Donde:

$$f_{i=} Frecuencia\ absoluta\ de\ la\ clase$$

$$x_i = Marca\ de\ clase$$

$$N = Número\ de\ datos$$

Ejemplo:

Calcular la media geométrica para el siguiente conjunto de datos

Clase	f	x	$f\log x$
8-10	12	9	11.4509
5-7	5	6	3.8907
2-4	3	3	1.4313
	$N=20$		$\sum f\ \log x=16.7729$

$$\log M.G.=\frac{\sum f_i \log x_i}{N}=\frac{16.7729}{20}=0.8386$$

$$M.G.=\ 10^{0.8386}=6.8960$$

Si lo comparamos con $\bar{x}=7.35$ se verifica que $M.G.\leq \bar{X}$.

2.1.3. Moda

La moda Mo es otra medida de tendencia central que se define como el valor que mayor frecuencia tiene en el conjunto de datos.

Es un término muy utilizado cotidianamente en cuanto a moda de vestir se refiere, así al observar a un grupo de personas en un lugar se puede decir que por ejemplo el pantalón de mezclilla está de moda porque es el que más se repite, en los zapatos podemos decir que el más vendido o de moda es el del número 24 cm.

Ejemplo:

Una taquería vende tacos de cabeza , un empleado lleva la relación de las ventas en el horario de 8 a 9 y obtuvo los siguientes resultados respecto al número de tacos que pide cada cliente en dicha hora, calcular la moda.

Registros: 1,3,5,3,4,5,2,3,2,3,2,4,5,7,1,2,2,3,3,4,4,5,3.

La moda $Mo = 3$ ya que es el dato que más se repite (7 veces) o mayor frecuencia tiene.

Ejemplo:

Calcular la moda Mo para el siguiente conjunto de datos:

x	f
10	8
9	4
8	6
7	9
6	2
5	1
	N=30

Claramente podemos observar que la moda está en 7, ya que es el dato que mayor frecuencia tiene.

Algunos conjuntos de datos pueden tener más de una moda, por ejemplo si se tienen dos datos que tienen la misma frecuencia y tienen el valor más grande, se dice que los dos están de moda o que el conjunto de datos es bimodal, pero debemos tener mucho cuidado cuando tenemos datos con frecuencias muy grandes pero muy cercanas, ya que es correcto afirmar que ambos datos están de moda.

Cuando los datos se encuentran agrupados en clases o intervalos la moda está en el punto medio o marca de clase que mayor frecuencia tiene.

Ejemplo:
Calcular la moda del siguiente conjunto de datos:

Clase	f
8-10	12
5-7	5
2-4	3
	N=20

La moda $Mo = 9$ ya que 9 es el punto medio o marca de clase (8-10) que mayor frecuencia tiene ($f = 12$).

2.1.4. Mediana (Mdn)

La mediana es el valor central de un conjunto de datos, este ocupa un lugar tal que el 50% de los datos son menores que él y el otro 50% son mayores. En muchas ocasiones es preferible el cálculo de esta medida respecto de la media.

Para localizar la mediana es importante notar que los datos deben estar ordenados, no importa si el orden es ascendente o descendente.

Ejemplo:

Tenemos la siguiente muestra de 5 observaciones, que representan el número de suscripciones a un servicio de telefonía en 5 días (d).

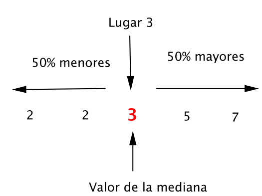
$$d_1 = 2,\ d_2 = 2,\ d_3 = 7, d_4 = 5,\ d_5 = 3$$

Ordenando los datos:

$$2, 2, 3, 5, 7$$

Claramente se observa que el valor central es 3 por lo tanto $Mdn = 3$

Nótese que en el ejemplo se tiene un número impar de datos (5) , una vez ordenado el conjunto de datos, se puede localizar el dato que ocupa el valor central, mediante la fórmula $\frac{N+1}{2}$, esto es $lugar\ que\ ocupa\ la\ Mdn\ \frac{N+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3$, no se confunda el lugar que ocupa con el valor de la mediana, ya que en este ejemplo coinciden.



Consideremos ahora un conjunto de datos con un número par de observaciones.

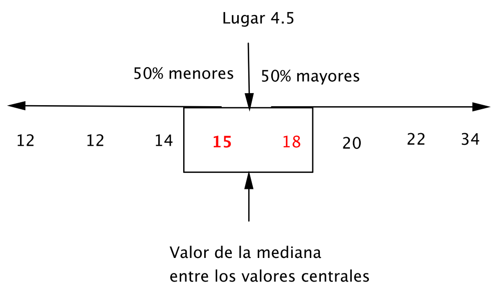
Dado el conjunto de datos siguiente calcular su mediana.

$$12,\ 14,\ 12,\ 15,\ 18,\ 20,\ 22,\ 34$$

Recuerde ordenar antes de localizar que valor ocupa la mediana

$$12, 12, 14, 15, 18, 20, 22, 34$$

El lugar que ocupa la mediana es el $\frac{N+1}{2} = \frac{8+1}{2} = \frac{9}{2} = 4.5$ por lo tanto el valor que ocupa el lugar 4.5 es el promedio de los dos valores centrales.



El valor que ocupe el lugar 4.5 es el promedio de los dos valores centrales, por lo tanto Mdn es igual a $\bar{x} = \frac{15+18}{2} = 16.5$

Cuando los datos están agrupados en frecuencia simple es muy útil calcular la frecuencia acumulada para localizar el dato que ocupa el lugar de la mediana: $\frac{N+1}{2}$

Ejemplo:

Calcular la mediana del siguiente conjunto de datos:

x	f
10	8
9	4
8	6
7	9
6	2
5	1
	N=30

A la tabla anterior agregamos la columna para la frecuencia acumulada

x	f	fa
10	8	30
9	4	22
8	6	18
7	9	12
6	2	3
5	1	1
	N=30	

El lugar $\frac{N+1}{2} = \frac{30+1}{2} = 15.5$, este lugar está ocupado por el valor 8, $\therefore Mdn = 8$.

Cuando los datos se encuentran agrupados en clases la mediana se calcula mediante la siguiente fórmula[4]:

$$Mdn = L + \frac{\frac{N}{2} - fa}{f} i$$

Donde:

$L = \text{límite real inferior de la clase donde se localiza la } Mdn$

$N = \text{número total de datos}$

$fa = \text{frecuencia acumulada de la clase inmediata inferior a la clase de la } Mdn$

$f = \text{frecuencia de la clase}$

$i = \text{el ancho de la clase o intervalo}$

Ejemplo:

Calcular la Mdn del siguiente conjunto de datos:

Clase	f
8-10	12
5-7	5
2-4	3
	N=20

Agregamos una columna donde se indique la frecuencia acumulada

Clase	f	fa
8-10	12	20
5-7	5	8
2-4	3	3
	N=20	

Comenzamos por localizar la clase que ocupa la mediana, esto es $\frac{N}{2} = 10$

L= 7.5

$fa = 8$

$f = 12$

$i = 3$

Sustituyendo: $Mdn = L + \frac{\frac{N}{2} - fa}{f}\ i$

$$Mdn = 7.5 + \frac{10 - 8}{12}\ (3)$$

$$Mdn = 7.5 + 0.5 = 8$$

2.2. Medidas de posición no central

Otro tipo de medidas que nos permiten conocer otras características de las distribuciones de los datos, son las medidas de posición no central, estos indicadores dividen las muestra en tramos iguales.

2.2.1. Cuartiles

Los cuartiles son 3 valores que distribuyen a la serie de datos en cuatro tramos iguales, por lo tanto cada tramo concentra un 25% de los datos.

Para localizar los valores que llamaremos Q_1, Q_2, Q_3 es conveniente calcular el porcentaje de la frecuencia acumulada y necesario ordenar los datos no importa si el orden es ascendente o descendente.

Ejemplo: calcular los cuartiles de la siguiente serie de datos que hacen referencia a las veces que un grupo de familias va a consulta médica en un año (x).

x	f	fa	$\%f$	$\%fa$
15	6	79	7.59%	100%
14	2	73	2.53%	92.40%
13	2	71	2.53%	89.87%
12	15	69	18.98%	87.34%
11	10	54	12.65%	68.35%
10	3	44	3.79%	55.69%
9	12	41	15.18%	51.89%
8	6	29	7.59%	36.70%
7	13	23	16.45%	29.11%
6	10	10	12.65%	12.65%
	N=79		99.94% ≈ 100%	

Los cuartiles son los valores que ocupan las posiciones 25%, 50% (mediana) y 75% y son:$Q_1 = 7$, $Q_2 = 9$, $Q_3 = 12$

Cuando los datos estén agrupados en clases la x es la marca de clase.

2.2.2. Deciles

Los deciles son 3 valores que distribuyen a la serie de datos en diez tramos iguales, por lo tanto cada tramo concentra un 10% de los datos.

Para localizar los valores que llamaremos $Q_1, Q_2, Q_3, Q_4, Q_5, Q_6, Q_7, Q_8, Q_9$

2.2.3. Percentiles

Los percentiles son 99 valores que distribuyen a la serie de datos en cien tramos iguales, por lo tanto cada tramo concentra un 1% de los datos.

Para localizar los valores que llamaremos $Q_1, Q_2, Q_3, \ldots, Q_{99}$

2.3. Medidas de dispersión

Al hacer análisis de datos se debe tomar en cuenta además de las medidas de tendencia central, ciertas medidas que indican si los datos están concentrados o dispersos, en este capítulo abordaremos cuatro medidas de dispersión. Las más utilizadas son el rango, también conocido como recorrido o amplitud, la desviación media, la varianza y el coeficiente de Pearson.

2.3.1. Rango

Esta medida se calcula determinando la diferencia entre el dato mayor y el dato menor del conjunto de datos.

Por ejemplo si en una empresa se desea conocer la variabilidad de los salarios de los empleados, esta medida es la adecuada, cuidando que los datos no estén concentrados algunos y otros muy separados del resto, ya que la interpretación de dicha medida no representaría la variabilidad de los salarios.

En el próximo ejemplo se muestra la observación anterior de manera numérica.

Ejemplo:

Los salarios de una compañía son los siguientes:

Salario	f
$39,000	10
$30,000	6
$24,000	6
$20,000	9
$15,000	12
$12,000	13
	N=47

El rango R= $39,000-$12,000= $27,000

Además de haber hecho el cálculo del rango debemos observar las frecuencias, en este caso los salarios bajos y altos tienen valores cercanos, así podemos decir que la diferencia entre el que más gana y entre el que menos gana es $27,000.

Si queremos calcular el rango de la siguiente tabla donde el gerente de la compañía es el que tiene el mayor salario y el resto de los empleados son los que ganan los salarios menores, es importante observar y dejar fuera del cálculo el dato que está "descosido" del resto de los salarios.

Salario	f
\$39,000	1
\$20,000	6
\$19,000	6
\$15,000	9
\$13,000	12
\$12,000	13
	N=56

En este caso es conveniente tomar como dato mayor \$20,000 y dato menor \$12,000, así rango R=\$20,000- \$12,000= \$8,000 y no R= \$39,000-\$12,000=\$27,000 .

2.3.2. Desviación media

Otra medida que se usa para medir la variabilidad es la desviación media, primero veamos qué es la desviación de un dato respecto de la media.
Dado un dato x con una media $\bar{x}$, decimos que la desviación de x respecto a la media $\bar{x}$ es $d = x - \bar{x}$, se puede decir que es la distancia que hay del dato a la media.

La desviación media de un conjunto de datos $x_1, x_2, x_3, \dots, x_N$ se calcula mediante la expresión:

$$D.M. = \frac{\sum|x-\bar{x}|}{N} \quad para\ datos\ \text{sin}\ agrupar$$

o bien

$$D.M. = \frac{\sum|d|}{N}$$

Esto es, el promedio de las desviaciones o de las distancias entre los datos y la media.

Ejemplo:

Las calificaciones de un alumno en el primer periodo son: 10,10,9,7,8,6,5,7,10, calcular la desviación media.

La expresión que nos permite calcular la D.M. nos indica que hay que calcular la media, a cada dato restar la media (desviación), hacer la suma de los valores absolutos de las desviaciones y dividir entre el número de datos.

x	f	$\lvert x-\bar{x}\rvert$
10	3	$\lvert 10-8\rvert = 2$
9	1	1
8	1	0
7	2	1
6	1	2
5	1	3
	N=9	$\sum\lvert x-\bar{x}\rvert = 9$

$$\bar{x} = \frac{\sum x}{N} = \frac{10+10+10+9+8+7+7+6+5}{9} = 8$$

$$D.M. = \frac{\sum|x-\bar{x}|}{N} = \frac{9}{9} = 1$$

Cuando los datos se encuentran agrupados en frecuencia simple, la desviación media se calcula mediante la expresión:

$$D.M. = \frac{\sum f|x-\bar{x}|}{N}$$

Ejemplo:

En un grupo de 20 alumnos se obtuvieron las calificaciones siguientes:

x	f	fx	$\|x-\bar{x}\|$	$f\|x-\bar{x}\|$
10	3	30	$\|10-7.55\| = 2.45$	$(3)(2.45) = 7.35$
9	5	45	1.45	7.25
8	2	16	0.45	0.9
7	3	21	0.55	1.65
6	4	24	1.55	6.2
5	3	15	2.55	7.65
	N=20	$\sum fx = 151$		$\sum f\|x-\bar{x}\| = 31$

$$\bar{x} = \frac{\sum fx}{N} = \frac{151}{20} = 7.55$$

$$\therefore\ D.M. = \frac{\sum f|x-\bar{x}|}{N} = \frac{31}{20} = 1.55$$

Cuando los datos se encuentran organizados mediante *clases* utilizamos la expresión:

$$D.M. = \frac{\sum f|x-\bar{x}|}{N}$$

donde $x = punto\ medio\ o\ marca\ de\ clase$

Ejemplo:

Con el siguiente conjunto de datos calcular la desviación media

Clase	f	x	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
13-15	3	14	42	$\lvert 14-9.2\rvert$ $=4.8$	$(3)(4.8)=14.4$
10-12	5	11	55	1.8	9
7-9	2	8	16	1.2	2.4
4-6	5	5	25	4.2	21
	N=15		$\sum fx=138$		$\sum f\lvert x-\bar{x}\rvert=46.8$

$$\bar{x}=\frac{\sum fx}{N}=\frac{138}{15}=9.2$$

$$D.M.=\frac{\sum f|x-\bar{x}|}{N}=\frac{46.8}{15}=3.12$$

2.3.3. Desviación estándar

Otra medida de dispersión es la desviación estándar s, dado un conjunto de datos $x_1, x_2, x_3, \dots, x_N$ la desviación estándar se calcula mediante la expresión:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{N}} \; para\ datos\ no\ agrupados$$

o bien

$$s = \sqrt{\frac{\sum d^2}{N}}$$

Ejemplo:

Dado el siguiente conjunto de datos calcular su desviación estándar

3,6,9,12,15,17 y 19

Es conveniente organizarlos en una tabla para simplificar los cálculos

x	$(x-\bar{x})$	$(x-\bar{x})^2$
3	$(3-11.5) = -8.5$	$(-8.5)^2 = 72.25$
6	-5.5	30.25
9	-2.5	6.25
12	0.5	0.25
15	3.5	12.25
17	5.5	30.25
19	7.5	56.25
		$\sum(x-\bar{x})^2 = 207.75$

$$\bar{x} = \frac{\sum x}{N} = \frac{3 + 6 + 9 + 12 + 15 + 17 + 19}{7} = \frac{81}{7} = 11.5$$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{N}} = \sqrt{\frac{207.75}{7}} = \sqrt{29.67} = 5.44$$

Cuando los datos están agrupados en *frecuencia simple* se calcula la desviación estándar mediante la expresión:

$$s = \sqrt{\frac{\sum f(x - \bar{x})^2}{N}}$$

Ejemplo:

Calcular la desviación estándar del siguiente conjunto de datos:

x	f
10	3
9	1
8	1
7	2
6	1
5	1
	N=9

Es conveniente agregar a la tabla donde los datos están organizados en frecuencia simple, las columnas que indiquen por pasos los cálculos correspondientes, como se muestra a continuación:

x	f	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
10	3	30	$(10-8)=2$	$(2)^2=4$	$(3)(4)=12$
9	1	9	1	1	1
8	1	8	0	0	0
7	2	14	-1	1	2
6	1	6	-2	4	4
5	1	5	-3	9	9
	N=9	$\sum fx = 72$			$\sum f(x-\bar{x})^2 = 28$

$$\bar{x} = \frac{\sum fx}{N} = \frac{72}{9} = 8 \qquad s = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}} = \sqrt{\frac{28}{9}} = 1.76$$

Si los datos se encuentran organizados en clases utilizamos la expresión anterior tomando en cuenta que x representa la marca de clase o punto medio.

Ejemplo:

Clase	f
13-15	3
10-12	5
7-9	2
4-6	5
	N=15

Agregamos las columnas necesarias para hacer dicho cálculo

Clase	f	x	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
13-15	3	14	42	$(14-9.2)$ $=4.8$	$(4.8)^2$ $=23.04$	$(3)(23.04)$ $=69.12$
10-12	5	11	55	1.8	3.24	16.2
7-9	2	8	16	-1.2	1.44	2.88
4-6	5	5	25	-4.2	17.64	88.2
	N=15		$\sum fx$ $=138$			$\sum f(x-\bar{x})^2=176.4$

$$\bar{x}=\frac{\sum fx}{N}=\frac{138}{15}=9.2 \qquad s=\sqrt{\frac{\sum f(x-\bar{x})^2}{N}}=\sqrt{\frac{176.4}{15}}=3.42$$

2.3.4. Varianza

La varianza mide la distancia entre los valores de la serie de datos y la media.

La varianza es el cuadrado de la desviación estándar s^2 así para datos no agrupados tendremos la expresión:

$$s^2=\frac{\sum(x-\bar{x})^2}{N}\ \textit{para datos no agrupados}$$

Así tomando los ejemplos anteriores veremos que se realizan los mismos cálculos.

Ejemplo:

Dado el siguiente conjunto de datos calcular su desviación estándar
3,6,9,12,15,17 y 19

x	$(x-\bar{x})$	$(x-\bar{x})^2$
3	$(3-11.5)=-8.5$	$(-8.5)^2=72.25$
6	-5.5	30.25
9	-2.5	6.25
12	0.5	0.25
15	3.5	12.25
17	5.5	30.25
19	7.5	56.25
		$\sum(x-\bar{x})^2=207.75$

$$\bar{x}=\frac{\sum x}{N}=\frac{3+6+9+12+15+17+19}{7}=\frac{81}{7}=11.5$$

$$s^2=\frac{\sum(x-\bar{x})^2}{N}=\frac{207.75}{7}=29.67$$

La varianza se calcula para datos agrupados en frecuencia simple, mediante la expresión:

$$s^2=\frac{\sum f(x-\bar{x})^2}{N}$$

Utilizando uno de los ejemplos anteriores mostraremos el cálculo de la varianza para datos agrupados en frecuencia simple.

Ejemplo:

Calcular la varianza del siguiente conjunto de datos:

x	f
10	3
9	1
8	1
7	2
6	1
5	1
	N=9

Es conveniente agregar a la tabla donde los datos están organizados en frecuencia simple, las columnas que indiquen por pasos los cálculos correspondientes, como se muestra a continuación:

x	f	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
10	3	30	$(10-8)=2$	$(2)^2=4$	$(3)(4)=12$
9	1	9	1	1	1
8	1	8	0	0	0
7	2	14	-1	1	2
6	1	6	-2	4	4
5	1	5	-3	9	9
	N=9	$\sum fx=72$			$\sum f(x-\bar{x})^2=28$

$\bar{x} = \frac{\sum fx}{N} = \frac{72}{9} = 8$ $s^2 = \frac{\sum f(x-\bar{x})^2}{N} = \frac{28}{9} = 3.11$

Si el conjunto de datos está organizado en clases debemos recordar que x representa la marca de clase.

Ejemplo:

Clase	f
13-15	3
10-12	5
7-9	2
4-6	5
	N=15

Agregamos las columnas necesarias para hacer dicho cálculo

Clase	f	x	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
13-15	3	14	42	$(14-9.2)$ $= 4.8$	$(4.8)^2$ $= 23.04$	$(3)(23.04)$ $= 69.12$
10-12	5	11	55	1.8	3.24	16.2
7-9	2	8	16	-1.2	1.44	2.88
4-6	5	5	25	-4.2	17.64	88.2
	N=15		$\Sigma fx = 138$			$\Sigma f(x-\bar{x})^2 = 176.4$

$\bar{x} = \frac{\sum fx}{N} = \frac{138}{15} = 9.2$ $s^2 = \frac{\sum f(x-\bar{x})^2}{N} = \frac{176.4}{15} = 11.76$

2.3.5. Coeficiente de variación de Pearson

El coeficiente de variación de Pearson C_V se calcula dividiendo la desviación estándar entre la media, esto lo podemos expresar mediante la expresión:

$$C_V = \frac{s}{\bar{x}}$$

Como ya se hicieron los cálculos de la desviación estándar y de la media mostraremos un ejemplo apoyándonos en ejercicios anteriores.

Ejemplo:

Dado el siguiente conjunto de datos calcular el coeficiente de Pearson

3, 6, 9, 12, 15, 17 y 19

Es conveniente organizarlos en una tabla para simplificar los cálculos

x	$(x - \bar{x})$	$(x - \bar{x})^2$
3	$(3 - 11.5) = -8.5$	$(-8.5)^2 = 72.25$
6	-5.5	30.25
9	-2.5	6.25
12	0.5	0.25
15	3.5	12.25
17	5.5	30.25
19	7.5	56.25
		$\Sigma(x - \bar{x})^2 = 207.75$

$$\bar{x} = \frac{\sum x}{N} = \frac{3 + 6 + 9 + 12 + 15 + 17 + 19}{7} = \frac{81}{7} = 11.5$$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{N}} = \sqrt{\frac{207.75}{7}} = \sqrt{29.67} = 5.44$$

$$C_V = \frac{s}{\bar{x}} = \frac{5.44}{11.5} = 0.47$$

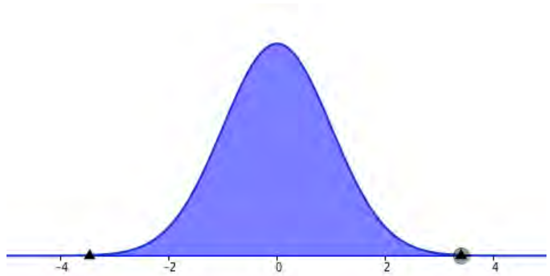
La importancia del coeficiente de variación de Pearson es que por ser un porcentaje nos permite comparar qué nivel de dispersión hay entre dos muestras.

2.4. Medidas de forma

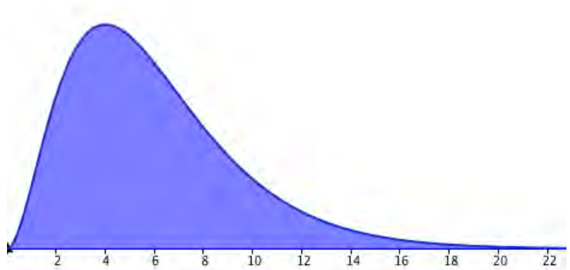
Otro tipo de medidas muy útiles son las medidas de forma, que como su nombre lo dice describen qué forma tiene la curva que representa a las series de datos.

2.4.1. Asimetría: índice de Fisher

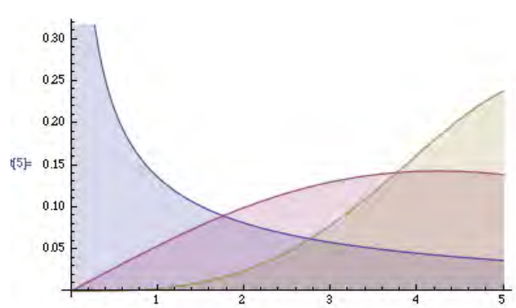
Este índice nos indica si la curva tiene forma simétrica, respecto de un punto de simetría (media aritmética), los segmentos de curva que se localizan a la izquierda y a la derecha son semejantes.



Curva simétrica



Curva asimétrica negativa



Curvas asimétrica a la izquierda y asimétricas a la derecha

Para medir el nivel de simetría utilizaremos el coeficiente de Fisher, que se define mediante la siguiente expresión:

$$g_1 = \frac{\frac{1}{N}\sum f(x-\bar{x})^3}{\left(\frac{1}{N}\sum f(x-\bar{x})^2\right)^{3/2}}$$

El índice se interpreta como sigue:

$g_1 > 0$ Significa que hay una asimetría positiva, es decir, existe mayor concentración de datos a la derecha de la media.

$g_1 = 0$ Significa que los datos están distribuidos simétricamente, por lo tanto existe la misma concentración de datos a la izquierda y a la derecha de la media.

$g_1 < 0$ Significa que hay una asimetría negativa, esto es que la mayoría de los datos se encuentran cargados a la izquierda de la media.

Ejemplo:

Calcular el coeficiente de asimetría de Fisher del siguiente conjunto de datos

x	f
10	3
9	1
8	1
7	2
6	1
5	1
	N=9

Agregando las columnas necesarias para realizar los cálculos tenemos:

x	f	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$f(x-\bar{x})^3$	$f(x-\bar{x})^2$
10	3	30	$(10-8)=2$	$(2)^2=4$	$(2)^3=8$	$(3)(8)=24$	12
9	1	9	1	1	1	1	1
8	1	8	0	0	0	0	0
7	2	14	-1	1	-1	-2	2
6	1	6	-2	4	-8	-8	4
5	1	5	-3	9	-27	-27	9
	N=9	Σfx $=72$				$\Sigma f(x-\bar{x})^3$ $=-12$	$\Sigma f(x-\bar{x})^2$ $=28$

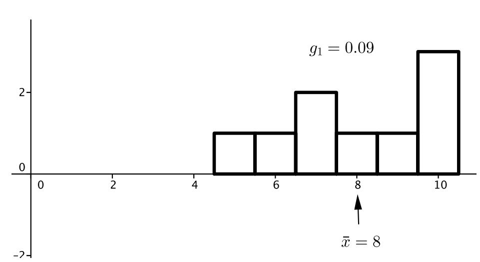
$$\bar{x}=\frac{\sum fx}{N}=\frac{72}{9}=8$$

sustituyendo en:

$$g_1=\frac{\frac{1}{N}\sum f(x-\bar{x})^3}{\left(\frac{1}{N}\sum f(x-\bar{x})^2\right)^{3/2}}$$

$$g_1=\frac{\frac{1}{9}(-12)}{\left(\frac{1}{9}(28)\right)^{3/2}}=\frac{-1.33}{5.48}=-0.24$$

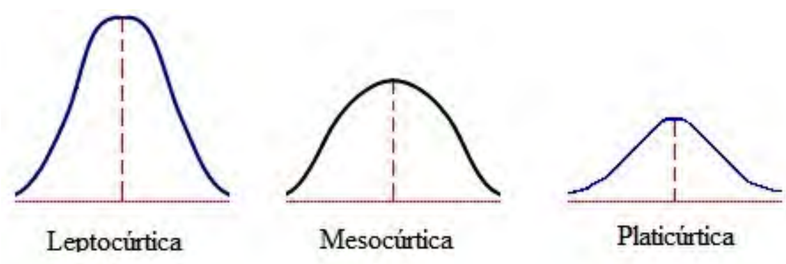
El valor negativo menor que cero, nos indica que la simetría de los datos está cargada ligeramente a la izquierda de la media.



2.4.2. Grado de concentración de Curtosis

El coeficiente de concentración de Curtosis g_2 ayuda en el análisis del grado de concentración de una serie de datos, respecto de un valor central.

Dependiendo del grado de Curtosis podemos tener tres tipos de distribuciones:



La distribución mesocúrtica presenta un grado de concentración medio alrededor de los valores centrales y su valor es $g_2 = 0$.
La distribución es leptocúrtica cuando presenta una alta concentración alrededor de los valores centrales y su valor es $g_2 > 0$.
La distribución es platicúrtica cuando presenta un grado reducido de concentración alrededor de los valores centrales y su valor es $g_2 < 0$.

El coeficiente de Curtosis está definido mediante la expresión:

$$g_2 = \frac{\frac{1}{N}\sum f(x-\bar{x})^4}{\left(\frac{1}{N}\sum f(x-\bar{x})^2\right)^2} - 3$$

Ejemplo:

Calcular el coeficiente de Curtosis del siguiente conjunto de datos:

x	f
10	3
9	1
8	1
7	2
6	1
5	1
	N=9

Agregando las columnas necesarias para realizar los cálculos tenemos:

x	f	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$(x-\bar{x})^4$	$f(x-\bar{x})^4$	$f(x-\bar{x})^2$
10	3	30	$(10-8)$ $=2$	$(2)^2=4$	$(2)^4=16$	$(3)(16)=48$	$(3)(4)$ $=12$
9	1	9	1	1	1	1	1
8	1	8	0	0	0	0	0
7	2	14	-1	1	1	2	2
6	1	6	-2	4	16	16	4
5	1	5	-3	9	81	81	9
	N=9	$\sum fx=72$				$\sum f(x-\bar{x})^4=148$	$\sum f(x-\bar{x})^2$ $=28$

$$\bar{x}=\frac{\sum fx}{N}=\frac{72}{9}=8$$

sustituyendo en:

$$g_2 = \frac{\frac{1}{N}\sum f(x-\bar{x})^4}{\left(\frac{1}{N}\sum f(x-\bar{x})^2\right)^2} - 3$$

$$g_2 = \frac{\frac{1}{9}(148)}{\left(\frac{1}{9}(28)\right)^2} - 3$$

$$g_2 = \frac{16.44}{9.67} - 3 = 1.7 - 3 = -1.29$$

Por lo tanto se trata de una distribución platicúrtica, esto es tiene concentración reducida alrededor de los valores centrales de la distribución.

2.5. Problemario

1. Calcular la media aritmética, la moda y la mediana del siguiente conjunto de datos:

2,2,3,5,7,2,4,3,7,5,6,8

2. Calcular la media, moda y mediana del siguiente conjunto de datos:

x	f
10	3
9	8
8	2
7	2
6	5
5	5
	N=25

3. Calcular la media aritmética, moda y mediana del siguiente conjunto de datos:

Clase	f
9-10	6
7-8	3
5-6	5
3-4	8
	N=22

4. Calcular la media geométrica $M.G.$ del siguiente conjunto de datos:

2,2,3,5,7,2,4,3,7,5,6,8

5. Calcular la media geométrica $M.G.$ del siguiente conjunto de datos:

x	f
10	3
9	8
8	2
7	2
6	5
5	5
	N=25

6. Calcular la media geométrica $M.G.$ del siguiente conjunto de datos:

Clase	f
9-10	6
7-8	3
5-6	5
3-4	8
	N=22

7. Calcular los cuartiles del siguiente conjunto de datos:

x	f
130	3
129	3
128	4
127	3
126	3
125	2
124	1
123	2
122	4
121	4
120	1
	N=30

8. Calcular el rango del siguiente conjunto de datos:

2,2,3,5,7,2,4,3,7,5,6,8

9. Calcular el rango del siguiente conjunto de datos:

x	f
10	3
9	8
8	2
7	2
6	5
5	5
	N=25

10. Calcular el rango del siguiente conjunto de datos:

Clase	f
9-10	6
7-8	3
5-6	5
3-4	8
	N=22

11. Calcular la desviación media del siguiente conjunto de datos:

x	f
15	3
14	2
13	3
12	5
11	1
10	1
	N=15

12. Calcular la desviación media del siguiente conjunto de datos:

x	f
15	3
14	2
13	3
12	5
11	1
10	1
	N=15

13. Calcular la desviación media del siguiente conjunto de datos:

Clase	f
9-10	6
7-8	3
5-6	5
3-4	8
	N=22

14. Calcular la varianza, la desviación estándar y el coeficiente de variación de Pearson C_V del siguiente conjunto de datos:

x	f
20	3
19	2
18	1
17	2
16	1
15	1
	N=10

2.6. Autoevaluación

1. Calcula la media aritmética de:

x	f
5	2
4	5
3	1
	N=8

2. Calcular la media geométrica de: 3,4,5,6,7,9

3. Calcular moda del siguiente conjunto de datos:

x	f
5	2
4	5
3	1
	N=8

4. Calcula el coeficiente de asimetría de Fisher del siguiente conjunto de datos:

x	f
12	3
10	2
8	1
	N=6

5. Calcular el coeficiente de concentración de Curtosis

x	f
5	3
4	3
3	2
	N=8

2.7. Soluciones del problemario

1. $\bar{x} = 4.5$, $Mo = 2$, $Mdn = 5$ lugar 6.5

x	f	fa
8	1	12
7	2	11
6	1	9
5	2	8
4	1	6
3	2	5
2	3	3
	N=12	

2.

x	f	fx	fa
10	3	30	25
9	8	72	22
8	2	16	14
7	2	14	12
6	5	30	10
5	5	25	5
	N=25	$\sum fx = 187$	

$$\bar{x} = \frac{\sum fx}{N} = \frac{187}{25} = 7.48$$

$$Mo = 9$$

$$Lugar\ de\ Mdn\ \frac{N+1}{2} = \frac{25+1}{2} = \frac{26}{2} = 13\ \therefore Mdn = 8$$

3.

Clase	f	x	fx	fa
9-10	6	9.5	57	22
7-8	3	7.5	22.5	16
5-6	5	5.5	27.5	13
3-4	8	3.5	28	8
	N=22		$\sum fx = 135$	

$$\bar{x} = \frac{\sum fx}{N} = \frac{135}{22} = 6.13$$

$$Mo = 3.5$$

$$Lugar\ de\ Mdn\ \frac{N+1}{2} = \frac{22+1}{2} = \frac{23}{2} = 11.5\ \therefore Mdn = 5.5$$

4.

2,2,3,5,7,2,4,3,7,5,6,8

$$M.G. = \sqrt[12]{(2)(2)(3)(5)(7)(2)(4)(3)(7)(5)(6)(8)}$$

$$M.G. = \sqrt[12]{16{,}934{,}400}$$

$$M.G. = 4.004$$

5.

$$M.G. = \sqrt[25]{10^3 9^8 8^2 7^2 6^5 5^5}$$

$$M.G. = \sqrt[25]{3.280366 \times 10^{21}}$$

$$M.G. = 7.25$$

6.

Clase	f	x	$\log x$	$f\ logx$
9-10	6	9.5	0.9777	5.8663
7-8	3	7.5	0.8750	2.6251
5-6	5	5.5	0.7403	3.7018
3-4	8	3.5	0.5440	4.3525
	N=22			$\sum f\ logx = 16.5457$

$$\log M.G. = \frac{\sum f\ logx}{N}$$

$$\log M.G. = \frac{16.5457}{22} = 0.7520$$

$$M.G. = 5.6493$$

7. $Q_1 = 122, Q_2 = 126, Q_3 = 128$

x	f	fa	$\%fa$
130	3	30	100%
129	3	27	90%
128	4	24	80%
127	3	20	66.66%
126	3	17	56.66%
125	2	14	46.66%
124	1	12	40%
123	2	11	36.66%
122	4	9	30%
121	4	5	16.66%
120	1	1	33.33%
	N=30		

8. R=8-2=6

9. R=10-5=5

10. R= 10-3=7

11.

x	f	fx	$\lvert x-\bar{x}\rvert$
15	3	45	$\lvert 15-12.86\rvert = 2.14$
14	2	28	1.14
13	3	39	0.14
12	5	60	0.86
11	1	11	1.86
10	1	10	2.86
	N=15	$\Sigma fx = 193$	$\Sigma\lvert x-\bar{x}\rvert = 9$

$$\bar{x} = \frac{\sum fx}{N} = \frac{193}{15} = 12.86$$

$$D.M. = \frac{\sum\lvert x-\bar{x}\rvert}{N} = \frac{9}{15} = 0.6$$

12.

x	f	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
15	3	45	$\lvert 15-12.86\rvert$ $= 2.14$	$(3)(2.14) = 6.42$
14	2	28	1.14	2.28
13	3	39	0.14	0.42
12	5	60	0.86	4.3
11	1	11	1.86	1.86
10	1	10	2.86	2.86
	N=15	$\sum fx = 193$	$\sum\lvert x-\bar{x}\rvert = 9$	$\sum f\lvert x-\bar{x}\rvert = 18.14$

$$\bar{x} = \frac{\sum fx}{N} = \frac{193}{15} = 12.86 \qquad D.M. = \frac{\sum f\lvert x-\bar{x}\rvert}{N} = \frac{18.14}{15} = 1.20$$

13.

Clase	f	x	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
9-10	6	9.5	57	3.37	20.22
7-8	3	7.5	22.5	1.37	4.11
5-6	5	5.5	27.5	0.63	3.15
3-4	8	3.5	28	2.63	21.04
	N=22		$\Sigma fx = 135$		$\Sigma f\lvert x-\bar{x}\rvert = 21.04$

$$\bar{x} = \frac{\sum fx}{N} = \frac{135}{22} = 6.13$$

$$D.M. = \frac{\sum f\lvert x-\bar{x}\rvert}{N} = \frac{21.04}{22} = 0.95$$

14.

x	f	fx	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
20	3	60	$(20-18.1)^2$ $=(1.9)^2 = 3.61$	$(3)(3.61)$ $= 10.83$
19	2	38	0.81	1.62
18	1	18	0.01	0.01
17	2	34	1.21	2.42
16	1	16	4.41	4.41
15	1	15	9.61	9.61
	N=10	$\sum fx = 181$		$\sum f(x-\bar{x})^2$ $= 28.9$

$$\bar{x} = \frac{\sum fx}{N} = \frac{181}{10} = 18.1$$

$$s^2 = \frac{\sum f(x - \bar{x})^2}{N} = \frac{28.9}{10} = 2.89$$

$$s = \sqrt{\frac{\sum f(x - \bar{x})^2}{N}} = \sqrt{\frac{28.9}{10}} = 1.7$$

$$C_V = \frac{s}{\bar{x}} = \frac{1.7}{18.1} = 0.093$$

2.8. Soluciones de autoevaluación

1. La media aritmética de:

x	f	fx
5	2	10
4	5	20
3	1	3
	N=8	$\Sigma fx = 33$

$$\bar{x} = \frac{\sum fx}{N} = \frac{33}{8} = 4.12$$

2. La media geométrica de: 3,4,5,6,7,9

$$M.G. = \sqrt[6]{(3)(4)(5)(6)(7)(9)} = \sqrt[6]{22{,}680} = 5.32$$

3. La moda del siguiente conjunto de datos es:

x	f
5	2
4	5
3	1
	N=8

$Mo = 4$

4. El coeficiente de asimetría de Fisher del siguiente conjunto de datos

x	f	fx	$(x-\bar{x})$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$	$(x-\bar{x})^3$	$f(x-\bar{x})^3$
12	3	36	1.34	1.7956	5.3868	2.40	7.21
10	2	20	-0.66	0.4356	0.8712	-0.2874	-0.5749
8	1	8	-2.66	7.0756	7.0756	-18.82	-18.82
	N=6	$\Sigma fx = 64$			$f(x-\bar{x})^2 = 13.33$		$\Sigma f(x-\bar{x})^3 = -12.18$

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{64}{6} = 10.66$$

$$g_1 = \frac{\frac{1}{N}\Sigma f(x-\bar{x})^3}{\left(\frac{1}{N}\Sigma f(x-\bar{x})^2\right)^{3/2}} = \frac{\frac{1}{6}(-12.58)}{\left(\frac{1}{6}(13.33)\right)^{3/2}} = \frac{-2.0966}{3.3127} = -0.6329$$

5. Calcular el coeficiente de concentración de Curtosis

x	f	fx	$(x-\bar{x})$	$f(x-\bar{x})^2$	$(x-\bar{x})^4$	$f(x-\bar{x})^4$
5	3	15	0.9	2.43	0.6561	1.9683
4	3	12	-0.1	0.03	0.0001	0.0003
3	2	6	-1.1	2.42	1.4641	2.9282
	N=8	$\Sigma fx = 33$		$\Sigma f(x-\bar{x})^2 = 4.88$		$\Sigma f(x-\bar{x})^4 = 4.8968$

$$\bar{x} = \frac{\Sigma fx}{N} = \frac{33}{8} = 4.1$$

$$g_2 = \frac{\frac{1}{N}\Sigma f(x-\bar{x})^4}{\left(\frac{1}{N}\Sigma f(x-\bar{x})^2\right)^{2}} - 3 = \frac{\frac{1}{8}(4.8968)}{\left(\frac{1}{8}(4.88)\right)^{2}} - 3 = \frac{0.6121}{0.3721} - 3 = 1.644 - 3 = -1.35$$

2.9. Conclusión

En este capítulo se mostró una introducción al cálculo de algunas medidas de tendencia central, de dispersión y de concentración, cuantas más herramientas tengas para realizar un análisis de datos, mejores serán tus conclusiones acerca de ellos, te invitamos a profundizar en el tema y a buscar aplicar los conocimientos adquiridos en tu vida cotidiana. Observa cuando leas o escuches noticias, los indicadores que usan. Si te gusta algún deporte encontrarás que estos parámetros ayudan a predecir el comportamiento de los equipos deportivos, así como su rendimiento.

Referencias

[1] Múria Josep & Gil Roberto (1998) Preparación, tabulación y análisis de encuestas para directivos. España: Editorial ESIC
[2] John Freund & Simon Gary (1994) Estadística Elemental. México: Prentice Hall Hispanoamericana, S. A.
[3] Hernández Emil (2006) Manual de Estadística. Colombia: Universidad Cooperativa de Colombia (Educe)
[4] Portilla Enrique (1980) Estadística primer curso. México: Interamericana

Capítulo 3:

Técnicas de contar

3. Técnicas de contar

En el próximo capítulo nos encontraremos con problemas que requieren de contar los elementos que pertenecen a un conjunto o evento, para ello es útil contar con una herramienta que facilite los cálculos, desafortunadamente no existe una técnica universal para dichos conteos, aquí te brindaremos una herramienta muy útil y que se vuelve necesaria en dichos cálculos. Una técnica es la regla fundamental del conteo[1] o principio de multiplicación[2].

Principio de la multiplicación

Si puede realizarse un evento de n_1 maneras diferentes y luego otro puede realizarse de n_2 maneras diferentes, y así sucesivamente, entonces el número de maneras que los eventos pueden realizarse (uno inmediatamente después del otro en el orden indicado) es de $n_1 n_2 \ldots$

Por ejemplo en el Distrito Federal las placas de automóvil tienen 3 dígitos seguidos de tres letras, si suponemos que se pueden usar las 27 letras del abecedario y cualesquiera de los 10 dígitos del 0 al 9. ¿Cuántas placas diferentes son posibles?

27	27	27	10	10	10

=19,683,000 placas

Haciendo una similitud con colocar en casilleros el número de formas en que sucede cada evento, los casilleros arriba mostrados indican que la primera letra puede ser escogida o colocada en ese lugar de 27 formas diferentes, como no hay una condición para la segunda ni tercera letra estas pueden también ser colocadas o escogidas de 27 maneras diferentes, el primer número puede ser escogido o colocado de 10 formas diferentes, al igual que el segundo y el tercero, por lo tanto el número total de placas diferentes que son posibles es el producto de las formas en que suceden cada uno de los eventos en el orden indicado, esto es de:

$$27 \times 27 \times 27 \times 10 \times 10 \times 10 = 19{,}683{,}00 \ \text{placas}$$

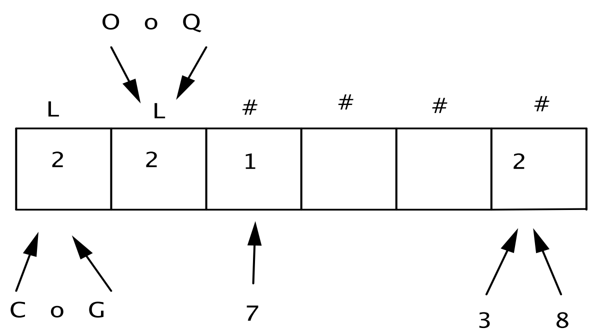
Si en el problema anterior modificamos condiciones veremos que el resultado es distinto, considere que se quiere conocer cuántas placas de tres letras diferentes seguidas de tres dígitos distintos donde el primero no puede ser el cero se pueden hacer.

27	26	25	9	9	8	=13,226,976 placas

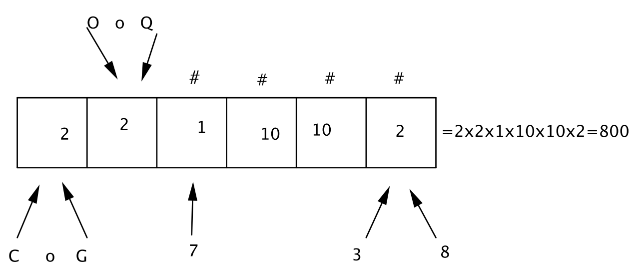
La primera letra puede ser cualesquiera de las 27 del abecedario, por esa razón llenamos el casillero con el número 27, el segundo casillero puede ser ocupado por cualquiera de las 26 letras restantes ya que no se admiten repeticiones y la tercera letra (ya se colocaron 2) puede ser escogida de 25 formas distintas. El primer dígito no puede ser el 0, por lo tanto puede ser cualquiera de los 9 dígitos que quedan, una vez colocado ahí un dígito no podrá ser escogido o colocado en las otras casillas por la condición de ser diferentes, el segundo casillero puede ser ocupado por cualquiera de los nueve números que quedan ya que ahí sí puede ir el cero, y el último de 8 formas diferentes, por lo tanto hay:

$$27 \times 26 \times 25 \times 9 \times 9 \times 8 = 13{,}226{,}976 \, placas$$

Diego y Santiago vieron a dos hombres alejarse en automóvil, frente a una joyería, justo antes de que sonara una alarma contra robos. Cuando fueron interrogados por la policía, los dos jóvenes dieron la siguiente información acerca de la placa (que constaba de dos letras seguidas de cuatro dígitos). Diego estaba seguro de que la segunda letra de la placa era una O o una Q, y que el último dígito era un 3 o un 8. Santiago dijo que la primera letra de la placa era una C o una G y que el primer dígito era definitivamente un 7. ¿Cuántas placas diferentes tendrá que verificar la policía? Considerar 27 letras del abecedario



La primera letra puede ser la C o la G por lo que hay 2 formas de elegir, la segunda letra puede ser O o Q, por lo que tiene 2 formas de ser colocado, como estaba seguro que el primer dígito es un 7 solo hay una forma de que ahí vaya un número, el último número creen que puede ser un 3 o un 8, así que 2 formas de vaya ahí un número, ahora debemos considerar todos los números que tendrá que buscar la policía con y sin repeticiones por lo que deberán ser las siguientes:



Para resolver:

¿Cuántos números telefónicos diferentes de 7 dígitos se pueden formar, si el primer número debe ser 2 o 3, y el último no puede ser 0?

Calcularemos de cuántas formas podemos ordenar o acomodar 3 autos en 3 cajones de estacionamiento, suponga que podemos distinguirlos por color, por ejemplo rojo, blanco y amarillo.

3	2	1

El primer cajón de estacionamiento puede ser ocupado por cualquiera de los tres autos, por poner un ejemplo puede ser colocado el amarillo, quedarían dos autos el rojo y el blanco, así que hay dos formas en que pueden colocar en ese cajón cualquiera de los autos, suponga que se colocó el blanco, por lo tanto el último cajón solo puede ser ocupado por un auto por el rojo. Por el principio de la multiplicación hay $3\times2\times1 = 6$ formas de colocar los autos, enseguida se muestran dichas opciones:

amarillo, blanco, rojo
amarillo, rojo, blanco
blanco, rojo, amarillo
blanco, amarillo, rojo
rojo, amarillo, blanco
rojo, blanco, amarillo.

En este tipo de ejemplo aún es fácil cuantificar y describir los resultados, pero en otros no es posible por los números tan grandes.

Al realizar conteos pueden surgir patrones como el siguiente. En una fila con 5 sillas se desea acomodar a 5 personas. ¿Cuántas formas hay de hacerlo?

5	4	3	2	1	=120

La primera silla puede ser ocupada de 5 formas distintas por cualquiera de las 5 personas, una vez sentada una de ellas la segunda silla solo podrá ser ocupada por cualquiera de las 4 personas que quedan sin sentarse, la tercera silla puede ser ocupada de 3 formas distintas, la segunda de 2 y la última de 1 sola forma por la persona que quede de pie, por lo tanto hay $5\times4\times3\times\times\times2\times1 = 150$ formas distintas en que se pueden sentar.

Note que nos queda el producto de los enteros consecutivos desde 1 hasta 5, estos productos de números sucesivos desde la unidad hasta un número finito determinado se le conoce como el *factorial* de un número.

Factorial

Dado un número $n \in \mathbb{N}$, diremos que $n!$ es el factorial de n, cuando:

$$n! = n(n-1)(n-2)\ldots(3)(2)(1)$$

también podemos decir que el factorial de un número es el producto de los enteros positivos desde 1 hasta n. El símbolo $n!$ se lee *n factorial*.

Axiomáticamente[3] se definen: $0! = 1$ y $1! = 1$

Calcular :

a) 6!

Como el orden de los factores no altera el producto podemos desarrollarlo de varias formas:

$$6! = 6\times5\times4\times3\times2\times1 = 720$$

$$6! = 1\times2\times3\times4\times5\times6 = 720$$

b) $7! = 7\times6\times5\times4\times3\times2\times1\times= 5{,}040$

c) $0!\,1!\,2! = (1)(1)(2) = 2$

d) $\frac{6!}{4!} = \frac{6\times5\times4\times3\times2\times1}{4\times3\times2\times1} = \frac{720}{24} = 30$

e) $\frac{8!}{3!(8-3)!} = \frac{8\times7\times6\times5\times4\times3\times2\times1}{3\times2\times1\times5\times4\times3\times2\times1} = 56$

Para practicar:

Sin calculadora realiza las operaciones siguientes

a) $4!3!=$

b) $\frac{7!5!}{6!4!} =$

c) $0!\,3!\,1! =$

d) $5!\,2! =$

e) $\frac{8!}{6!(8-6)!} =$

f) $\frac{1!}{0!} =$

g) $1!\,2!\,3! =$

3.1. Permutaciones sin repeticiones

Una permutación de un conjunto o de objetos distintos es un ordenamiento, permutar es sinónimo de cambiar de lugar o de posición.

En el modelo del ADN de James Watson y Francis Crick (1953) en 3D, donde la doble hélice del esqueleto interno del ADN permite combinar entre sí los 4 aminoácidos; 2 purinas (adenina y guanina) y 2 pirimidinas (citosina y timina), de forma específica:

A	T

T	A

G	C

C	G

permitió encontrar primero, cada aminoácido se puede acomodar de dos formas distintas por ejemplo A-T Y T-A llamadas bases, a su vez cada base tiene 4! = 24 posibles permutaciones[4], estos análisis permiten calcular las permutaciones de objetos, en ocasiones es posible representarlos; en nuestro curso no trataremos de representarlos sino de calcular el número de formas en que pueden ocurrir los eventos.

Teorema: El número de permutaciones distintas de r objetos tomados de un conjunto de n elementos es :

$$P(n,r) = \frac{n!}{(n-r)!}$$

en caso de que r=n

$$P(n,n) = n!$$

En algunas calculadoras la tecla

$$\boxed{{}_nP_r}$$

hace el cálculo directo del número de permutaciones de r objetos de un conjunto de n elementos.

Ejemplificaremos lo anterior con el siguiente problema: Ana, Beatriz, Carla, Delia y Eugenia participan en una carrera de 100 metros, solamente se premia los 3 primeros lugares, ¿cuántas formas hay de que pudieran recibir premio?

Usando la fórmula de permutaciones donde n=5 y r=3 ,

$$P(n,r) = \frac{n!}{(n-r)!}$$

$$P(5,3) = \frac{5!}{(5-3)!}$$

$$P(5,3) = \frac{5!}{2!}$$

$$P(5,3) = 60$$

Si usamos el principio de multiplicación

5	4	3

El primer lugar puede ser para cualquiera de las 5, una vez ocupado el primer lugar, el segundo lugar puede ser ocupado por cualquiera de las 4 que faltan y por último el tercer lugar puede ser ocupado por cualquiera de las 3 que falta, por lo tanto hay $5 \times 4 \times 3 = 60$ formas en que ocupen un lugar en la premiación.

En el siguiente ejemplo no se permiten repeticiones, dados los números 1,2,3,4 y 5

a) ¿Cuántos números de 3 cifras se pueden hacer?

5	4	3

Tenemos 5 números 1,2,3,4 y 5, el primer número puede comenzar con cualquiera de ellos, supongamos que el primer número que se colocó es el 4, nos quedan el 1,2,3 y 5, por lo tanto el segundo número puede ser cualquiera de los números anteriores por lo que hay 4 formas de ser ocupada la segunda posición, suponga que se colocó el número 1, nos quedan los números 2,3 y5, así que hay tres formas en que pueda ser colocado el último número, aplicando el principio de la multiplicación tenemos $5 \times 4 \times 3 = 60$ número distintos de 3 dígitos.

Ejemplos de algunos de los 60 números que podemos hacer son los siguientes: 543,421,312,245,123,etc.

b) ¿Cuántos de esos números son menores de 500?

Para que el número de 3 dígitos sea menor que 500 debe comenzar con cualesquiera de los números 1,2,3 o 4 , una vez colocado uno de ellos el segundo puede ser colocado de 4 formas ya que ahora el 5 puede ir en esa posición, así que el tercer número puede ser colocado de 3 formas distintas.

4	4	3	= 48 números menores de 500

c) ¿Cuántos números de 3 dígitos son pares?

La condición para que sean pares es que deben terminar en número par, por lo que hay dos formas en que puede terminar, el primer número puede ser cualquiera de los 4 que quedan una vez colocado uno al final, así hay 4 formas de comenzar, una vez colocado ahí un número, el segundo dígito puede ser cualquiera de los 3 que quedan, usando el principio de la multiplicación hay:

4	3	2	= 24 números pares

d) ¿Cuántos son múltiplos del número 5?

Los números múltiplos de 5 terminan en 0 o en 5, así que solo existe una forma de que termine, una vez colocado el número el primero puede ser colocado de 4 formas, y una vez colocado el primero el segundo tendrá 3 formas de ser colocado, por lo tanto hay:

4	3	1	=12 múltiplos de 5

En el siguiente ejemplo tenemos a cinco personas: A=Ana, B=Beatriz, C=Carla, D=Delia y E= Eugenia se van a subir a la montaña rusa que tiene 5 sillas , de cuántas maneras pueden sentarse

a) ¿En total?

La primera silla puede ser ocupada por cualquiera de las 3 personas, una vez colocada una, el segundo sitio puede ser ocupado por cualquiera de las 4 que quedan, y así sucesivamente, quedando:

5	4	3	2	1	=5!=120

b) ¿Si Ana no puede colocarse en ninguno de los extremos de las sillas de la montaña rusa?

El primer y último lugar no puede ser ocupado por A, así que pueden ser ocupados por cualquiera de las 4 que quedan, una vez ocupado el último lugar el primero puede ser ocupado por cualquiera de las 3, note que ya se ocuparon el lugar 1 y el 5, dos personas ya están ocupando esos lugares, quedan 3 personas por sentarse ya que A sí puede ocupar los lugares que siguen, por lo que hay 3, 3 y 2 formas de que se sienten en los 3 espacios que restan, por lo tanto hay:

3	3	2	1	4	= 72

c) ¿Si Carla tiene que ir en la primera silla?

El primer sitio solo puede ser ocupado de 1 forma, que es sentando ahí a C, y los lugares que siguen pueden ser ocupados de 4, 3, 2 y 1 formas, por lo tanto hay:

$$1\times4\times3\times2\times1=24$$

d) ¿Si Ana y Carla deben ir siempre juntas?

Cuando tenemos elementos que deben ir juntos podemos considerarlos como uno solo para facilitar los cálculos, así que $(A|C)$, B,D y E son 4 los elementos a considerar, por lo tanto hay $4\times3\times2\times1\times2!=48$ al final se multiplica por 2! ya que el elemento $(A|C),(C|A)$ puede ir de dos formas primero A y después C, o viceversa.

Es importante mencionar que el orden en que son colocados los resultados es importante, ya que un mal razonamiento puede llevar al resultado correcto por la casualidad que resulta de multiplicar los factores.

¿De cuántas maneras se pueden acomodar 6 botellas de vino (A,B,C,D,E,F) en una hilera de seis casilleros, si dos de ellas (A y B) no pueden ir juntas y otras dos (C y D) no pueden ir separadas?

Los elementos que tenemos son A,B,C,D,E,F y las condiciones son que A y B estén separados y C y D juntos.

Comencemos por calcular todas las formas en que pueden acomodarse en total con C y D juntos, para lo cual los consideramos un solo elemento y así garantizamos que siempre estarán juntos, de manera que tendremos los nuevos elementos: A, B, $(C|D)$,E,F.

Hay: $5\times4\times3\times2\times1 = 5! = 120\times2! = 240$ donde $2! = (C|D),(D|C)$

En los 240 están incluidos A y D juntos y queremos que vayan separados, así que al total quitaremos cuando A y D estén juntos, de manera que nuestros elementos serán ahora 4: $(A|B),(C|D),E,F$ y la forma de acomodarlos es:

$$4!\times2!\times2! = 96$$

4!= son las formas en que los cuatro elementos se pueden acomodar,

2!= la simetría de C-D, D-C,

2!= la simetría de D-C,C-D.

∴ 240 -96 = 144.

En este tipo de ejemplos es más fácil hacer el cálculo total y quitar los elementos no deseados.

Calculemos ahora cuántas palabras de seis letras se pueden formar con las letras del nombre JULIAN?

a) Que lleven dos consonantes juntas (no 3) , pero separadas la U y la I.

Vocales=U,I,A

Consonantes=J,L,N

C 3	C 2	V 3	3	2	1	= 108
V 3	C 3	C 2	V 2	2	1	=72
2	V 3	C 3	C 2	V 2	1	=72
2	1	V 3	C 3	C 2	V 2	=72
3	2	1	V 3	C 3	C 2	=108

Total = 432

En las 432 van las dos consonantes juntas y garantizamos que no vayan 3 al poner a los lados vocales, ahora debemos quitar cuando UI van juntas, por ejemplo $(J|L)$, $(U|I)$, A, N Tenemos cuatro elementos

$$\therefore \quad 4!\times 3!\times 2! = 144$$

4!= las permutaciones de los cuatro elementos

3!= aquí son las consonantes juntas JL,JN, LN

2! = por las simetrías

Finalmente hay 432-144=288

b) Que terminen con consonante y que la U vaya después de la A.

A 1	4	3	2	1	CONSONANTE 3	$=4!\times 3=72$
No vocal 3	A 1	3	2	1	CONSONANTE 3	$=3!\times 9$ $=54$
No vocal 3	No vocal 2	A 1	2	1	CONSONANTE 3	$=3!\times 6$ $=36$
No vocal	No vocal	No vocal	A 1	V 1	CONSONANTE 3	$3!\times 3=18$
						Total=180

El siguiente ejemplo es especial porque de los elementos totales algunos quedan fuera de las ordenaciones y se deben hacer todas las consideraciones posibles.

¿Cuántas palabras se pueden formar con las letras COMER?

a) Que lleven juntas las letra ME por una parte y CO por la otra, todo en la misma palabra.

COMER:

Vocales: O,E

Consonantes: C,M,R

juntas ME y CO los tomaremos como uno solo para garantizar que estén juntos ∴ se reducen a 3 elementos : $(M|E),(C|O)$,R

en la palabra es condición que estén ME y CO así que R quedará fuera y tendremos:

2	1	2! ME EM	2! CO OC	=8 palabras

b) Que lleven dos vocales juntas.

Para garantizar que estén juntas las vocales las consideramos un solo elemento y tendremos: C,M,$(O|E)$,R.

Como las palabras formadas deben tener a las dos vocales juntas, tres de ellas quedarán fuera, L,M y R además debemos considerar la simetría de OE-EO y tendremos:

3	2	1	2 EO- OE	3 Fuera: L,M,R	=36

c) Que terminen con consonante y que la O vaya después de la letra R.

R 1	 2	 1	Terminan con C o M 2	Letras que quedan fuera 2	 =8
 1	R 1	O 1	 1	 1	 =4
				TOTAL	=12

El ejemplo siguiente es muy diferente a los anteriores porque consideraremos que los elementos están en círculo, hecho que puede confundir por no haber un inicio y un final en la posición de los elementos, lo cual tiene la salvedad cuando fijamos un elemento y a partir de él consideramos el orden del resto, veamos un ejemplo.

¿De cuántas maneras se pueden acomodar los 6 aderezos que se muestran en la imagen siguiente? Para identificarlos les nombraremos con las letra: A,B,C,D,E,F.

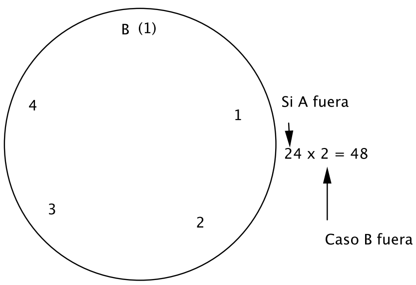
Si solamente podemos colocar 5 de ellos en el círculo y

a) El aderezo A y B no deben ir juntos.

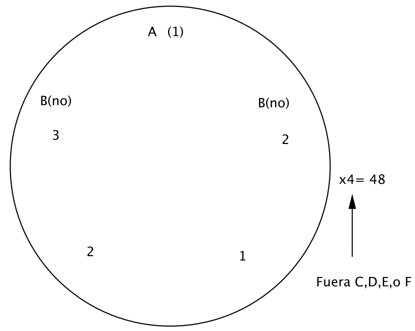
Si A y B no pueden ir juntos consideraremos dos casos:

Caso I: si A queda fuera será el mismo caso que cuando B quede, por lo que hay dos formas.

Si colocamos por ejemplo el aderezo B (1 forma) quedan 4 elementos por acomodar ya que estamos considerando que A queda fuera.



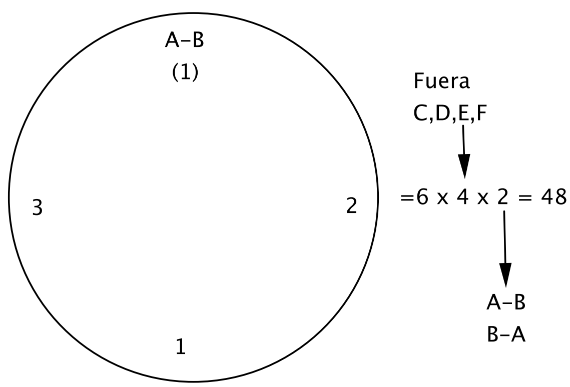
Caso II: si A y B dentro



Total= caso I + caso II = 96

b) Si los aderezos A y B deben ir juntos.

si los aderezos deben ir juntos hacemos $(A|B)$, un solo elemento por lo tanto C,D,E o F son los aderezos que pueden quedar fuera.



∴ hay 48 formas.

En la ciudad de Morelia, los números telefónicos deben comenzar con 2,3,4,5, o 6 y deben tener seis dígitos a condición de que el segundo dígito no sea igual al primero, ni tampoco que terminen en 000. ¿Cuántos números telefónicos se pueden hacer con esas condiciones?

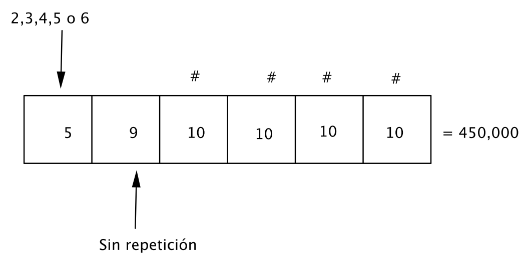
Tenemos los dígitos 0,1,2,3,4,5,6,7,8,9 para construir los números, observa las condiciones:

1) Comenzar con 2,3,4,5,o 6.

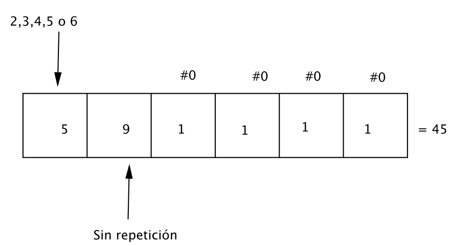
2) El segundo dígito debe ser diferente del primero, es decir no hay repeticiones.

3) No deben terminar en 000.

Calcularemos primeramente cuando los números telefónicos contienen todas las terminaciones incluida la terminación 000.



Calculemos ahora los que sí tienen la terminación 000 para restarla del anterior:



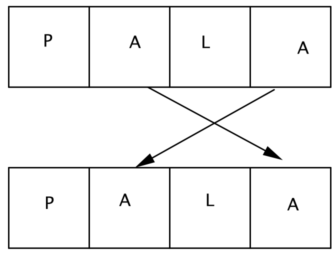
450,000-45= 449,955.

3.2. Permutaciones con repeticiones

Teorema: El número de permutaciones de n objetos de los cuales n_1 son iguales, n_2 son iguales, …, n_r son iguales, es

$$\frac{n!}{n_1!\,n_2!\ldots n_r!}$$

En las permutaciones que tienen elementos repetidos al intercambiar elementos se pueden obtener los mismos resultados, por ejemplo :



Cuántas formas diferentes , cada una de 8 pelotas colocadas de forma vertical, se pueden hacer con 2 pelotas verdes, 2 blancas , 2 rojas, 1 azul y 1 amarilla todas son iguales y no se pueden diferenciar.

Aplicando el teorema para permutaciones con repeticiones tenemos:

$$\frac{8!}{2!\,2!\,2!\,1!\,1!} = \frac{8\times7\times6\times5\times4\times3\times2\times1}{2\times1\times2\times1\times1\times1} = 10{,}080$$

¿Cuántas palabras distintas se pueden hacer con las letras de la palabra PALA?

Tenemos:

n=4 P,A,L,A

n_1=2 A,A $\frac{4\times3\times2\times1}{2\times1} = 12$ o $\frac{n!}{n_1} = \frac{4!}{2!} = 12$

¿Cuántas palabras se pueden formar con las letras de la palabra CABAL?

a) En total.

C=1

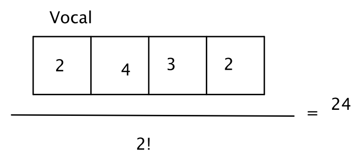
A= 2

B=1 $\frac{5!}{2!} = 60$

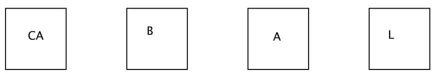
L=1

n=5

b) Que comiencen con vocal

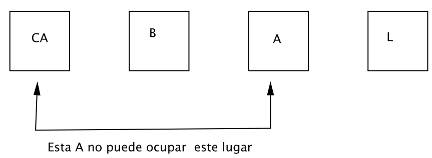


c) Que lleven la sílaba CA



Vocales= A, A

Consonantes= C,B,L



El hacer un solo elemento CA garantizamos que estarán juntos, y tendremos 4 elementos por organizar no habiendo repeticiones con esa condición.

$$4\times3\times2\times1 = 24$$

Cuántas permutaciones distintas se pueden formar con las letras de las palabras siguientes:

a) gato

$4! = 24$

b) campanas

c=1

a=3

m=1

p=1

n=1

s=1

n=8

$$\frac{8!}{3!} = \frac{8\times7\times6\times5\times4\times3\times2\times1}{3\times2\times1} = 6{,}720$$

c) matemáticas

m=2

a=3

t=2

e=1

i=1

c=1

s=1

n=11

$$\frac{11!}{2!\,3!\,2!} = 1{,}663{,}200$$

Para practicar:

a) ¿De cuántas formas se pueden acomodar en una fila de 3 sillas 3 personas A, B y C?

b) ¿De cuántas formas distintas de pueden acomodar horizontalmente 3 libros de matemáticas y 2 de física?

c) En un casillero de 5 huecos se desean colocar señales con banderas , si tenemos 3 rojas y 2 negras, ¿cuántas señales diferentes se pueden formar?

d) 3 hombres y tres mujeres se quieren sentar en una fila de 6 sillas, ¿cuántas formas tienen de hacerlo si deben ir alternados?

e) ¿Cuántas palabras de 6 letras se pueden formar con las letras de SILVIA?

f) ¿Cuántas formas hay de acomodar a 4 niños en una fila con 3 sillas?

g) ¿Cuántas formas hay de acomodar en un círculo a 4 niños?

h) ¿Cuántas formas hay de acomodar en un círculo con 4 sillas a 5 niños?

3.3. Combinaciones

Cuando seleccionamos elementos de algún conjunto sin importar el orden se trata de combinaciones que a diferencia de las permutaciones si importaba el orden. Para realizar dichos cálculos usaremos el siguiente teorema, donde denotaremos las combinaciones de n objetos tomados r por $C(n,r)$ o ${}_nC_r$, se lee las combinaciones de n objetos tomados de r en r. En este tema es importante identificar palabras claves como seleccionar, escoger.

Teorema: $C(n,r) = {}_nC_r = \frac{P(n,r)}{r!} = \frac{n!}{r!(n-r)}!$

Por ejemplo calculemos el número de combinaciones que podemos hacer con los colores amarillo (A), blanco(B), crema(C), Dorado(D), formando las combinaciones de 3 en 3.

COMBINACIONES	PERMUTACIONES
ABC	ABC, ACB, BCA, BAC,CAB,CBA
ABD	ABD,ADB,BAD,BDA,DAB,DBA
ACD	ACD,ADC,CDA,CAD,DAC,DCA
BCD	BCD,BDC,CDB,CBD,DBC,DCB

Como se muestra en el ejemplo no son lo mismo las combinaciones que las permutaciones, las primeras son una selección y las segundas tienen que ver con orden.

Un estudiante realiza un examen de geografía, en las instrucciones se indica que debe contestar:

a) 9 de los 10 reactivos ,¿cuántas maneras tiene de contestar?

$$ {}_{10}C_{9}=\frac{10!}{9!(10-9)!}=10 $$

b) 8 de los 10 reactivos, ¿cuántas maneras tiene de contestar?

$$ {}_{10}C_{8}=\frac{10!}{8!(10-8)!}=45 $$

c) 3 de las primeras 5 y 4 de las últimas 5, ¿cuántas maneras tiene de contestar?

${}_{5}C_{3}\times{}_{5}C_{4}=10\times5=50$

d) 8 de los 10 reactivos, si tiene que responder 4 de los primeros 5, ¿cuántas maneras de escoger tiene?

1

2

3 $\boldsymbol{{}_{5}C_{4}=5}$

4

5

_ _ _ _ _ _ _ _ _ _ _ _

6

7

8 ${}_{5}C_{4}=5$

9

10

Por el principio del conteo hay 5 formas de escoger los primeros 5 reactivos y 5 de escoger los segundos 5 ∴ $5\times5=25$.

e) 7 de los 10 reactivos, de los cuales es obligatorio que conteste por lo menos 3 de los primeros 5 reactivos, ¿cuántas formas tiene de escoger?

Como pide al menos 3 de los primeros 5 reactivos debemos considerar los casos en que escoja 3, 4 y 5 de ser así le quedarán por escoger de los últimos 5 reactivos 4, 3 y 2 respectivamente, al final sumaremos todos los casos.

${}_5C_3 \times {}_5C_4 + {}_5C_4 \times {}_5C_3 + {}_5C_5 \times {}_5C_2 = 50 + 50 + 10 = 110$

Consideremos un grupo de 9 alumnos se quiere saber cuántos equipos de 6 alumnos se pueden formar.

El problema de combinaciones ya que vamos a escoger de los 9 alumnos solamente 6, así que n=9 y r=6

$${}_9C_6 = \frac{9!}{6!(9-6)!} = 84$$

Un grupo de 20 alumnos desean formar grupos de apoyo en una contingencia ocurrida en la ciudad de Morelia, se forman al mismo tiempo 3 subgrupos, uno de 9 alumnos apoyará a la Cruz Ámbar, otro de 5 apoyará a la Cruz Roja y el último de 4 alumnos que apoyará a los Ángeles Verdes, investigue cuántas formas hay de que se distribuyan.

Es necesario calcular el número de formas en que sucede cada evento y después aplicar el principio de la multiplicación para calcular el número total de formas.

De los 20 alumnos 9 apoyarán a la Cruz Ámbar ∴ hay ${}_{20}C_9 = 167{,}960$

quedan 11 alumnos de los cuales apoyará a la Cruz Roja ∴ hay ${}_{11}C_5 = 462$

quedan 6 alumnos 4 de ellos apoyará, a los Ángeles Verdes ∴ hay ${}_6C_4 = 15$

$$167{,}960 \times 462 \times 15 = 1'962{,}800 \text{ número total de formas}$$

A continuación realizaremos cálculos con el juego de la baraja inglesa, para lo cual describiremos sus jugadas, se llaman *palos* o *figuras* a las picas, tréboles, corazones y rombos (columnas) y *valores* a las horizontales (renglones).

La baraja consta de 54 cartas como se muestra a continuación:

Picas/espadas ♠	Tréboles ♣	Corazones ♥	Diamantes/rombos ♦
A	A	A	A
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
J	J	J	J
Q	Q	Q	Q
K	K	K	K

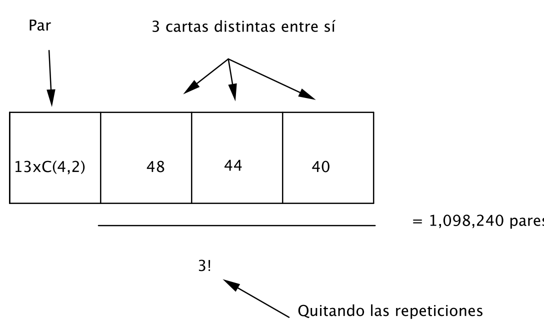
Explicaremos las jugadas al mismo tiempo que vayamos haciendo los cálculos de las forman en que pueden ocurrir o "caer" cuando nos dan las cartas o una "mano".

En el juego de la baraja consideraremos que nos entregan 5 cartas de la baraja.

a) Calcular el número de maneras en que puede salir un par.

Un par es cuando se tienen dos cartas del mismo valor y las otras tres diferentes al par y distintas entre ellas para que no se forme otro par.

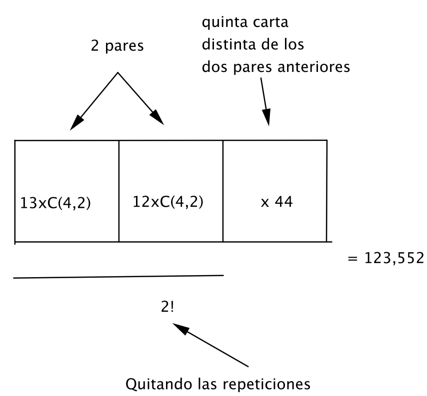
Suponga que tenemos dos cartas del valor 5 esto sucede con ${}_4C_2 = 6$ aquí tenemos un par de cincos, pero un par en general puede suceder con cualesquiera de las 13 figuras así que hay $13\,{}_4C_2$, ahora debemos garantizar que las tres cartas que faltan sean diferentes al par y distintas entre sí, en nuestro ejemplo quitamos los 5 que restan y que darán entonces $12\times 4 = 48$ cartas así que la tercera podrá ser escogida de 48 formas diferentes, una vez que se ponga esa carta hacemos lo anterior para evitar que se repitan que es quitar las que son pertenecientes a esa figura y tendremos 44 cartas que son las formas de elegir la siguiente carta, seguimos el proceso y la quinta carta tendrá 40 formas de ser escogida, por lo tanto hay:

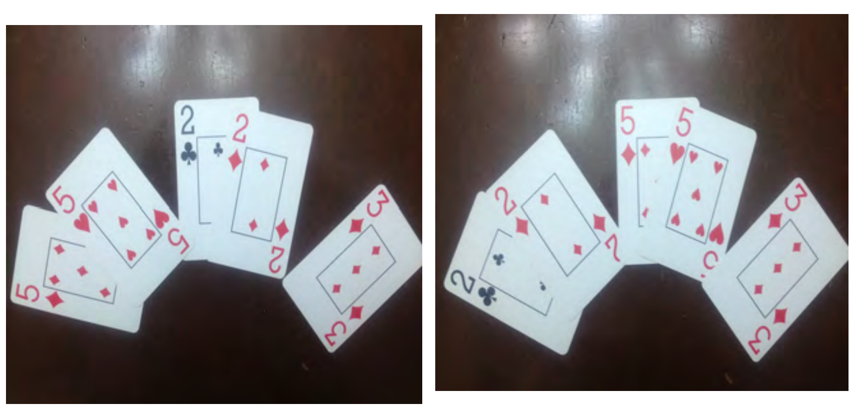


b) Calcular el número de maneras en que pueden salir dos pares.

Dos pares significa dos cartas del mismo valor, otras dos del mismo valor pero de diferente respecto del primer par, y una quinta carta distinta a todas las anteriores.

El primer par puede salir de $13 \times {}_4C_2 = 78$ formas
el segundo par de $12 \times {}_4C_2 = 72$
hasta aquí deberemos dividir entre 2! para quitar las repeticiones ya que será lo mismo, por ejemplo par de doses 2-2 y par de cincos 5-5, que par de cincos 5-5 y par de doses 2-2, la quinta carta tiene 44 formas de ser escogida una vez que se quitaron las cartas que evitan repeticiones, esto es 52-4= 48, 48-4=44.
por lo tanto hay:





Note que es la misma mano no importa el orden

c) Calcular el número de maneras en que puede salir una tercia.

Para tener una tercia deberemos tener tres cartas del mismo valor y dos diferentes entre sí y distintas a las de la tercia.

Las tres primeras cartas se pueden seleccionar de 13_4C_3=52
la cuarta debe escogerse entre las 48 que quedan ya que se deben quitar las del mismo valor para no formar 2 pares, y la quinta carta de 44 maneras ya que también se quitan las del mismo valor de la carta anterior para no formar un par con ella. En las imágenes anteriores se muestra la misma mano pero debemos quitar las posibles repeticiones con las dos últimas cartas, por lo que dividimos entre 2!

$$13_4C_2\times\frac{48\times x44}{2!} = 54{,}912$$


Ejemplo de la misma mano tercia de ases y dos cartas distintas

d) Calcular el número de maneras en que puede salir una corrida o escalera.

Una corrida o escalera se forma cuando tenemos las 5 cartas con valores consecutivos pero que no pertenecen todas a la misma figura, de hecho cuando pertenecen a la misma figura se llama flor imperial.

Por ejemplo de dos corridas 5,6,7,8 ,9 y A,2,3,4,5, ver la siguiente imagen:

Observemos las 10 corridas que pueden hacerse:

1= A,2,3,4,5

2= 2,3,4,5,6

3= 3,4,5,6,7

4= 4,5,6,7,8

5= 5,6,7,8,9

6= 6,7,8,9,10

7= 7,8,9,10,J

8= 8,9,10,J,Q

9= 9,10,J,Q,K

10=10,J,Q,K,A

Por ejemplo la primera corrida A,2,3,4,5, el as debe seleccionarse 1 entre los cuatro existentes, lo mismo sucede con el 2 hay que seleccionar 1 entre los cuatro existentes, y así sucesivamente hasta llegar al 5. Lo mismo pasa con las demás corridas.

En los cálculos anteriores están incluidas 4 *flores imperiales* que habrá que restar al total:

$10\ ({}_4C_1 \times {}_4C_1 \times {}_4C_1 \times {}_4C_1 \times {}_4C_1\ -4) = 10{,}200$

e) Calcular el número de maneras en que puede salir una flor.

Una flor se forma cuando las 5 cartas son de la misma figura, pero no son consecutivas, ya que al serlo formarían una corrida que de hecho tendremos que quitarlas las 10 que se forman después de haber calculado cuántas flores se pueden hacer.

Como cada figura tiene 13 cartas y hay 4 figuras tenemos:

${}_{13}C_5$ -10= 1,277 por cada figura por lo que debemos multiplicar por las 4 que hay.

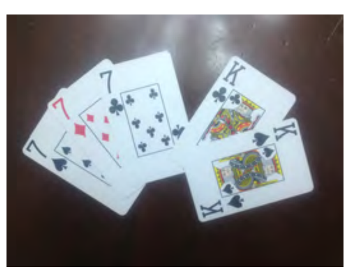
$4({}_{13}C_5 - 10) = 5{,}108$

Ejemplo de una flor de tréboles

f) Calcular el número de maneras en que puede salir full.

Un *full* se forma cuando nos sale una tercia y un par, así que una tercia se puede formar de 13 valores y de ${}_4C_3$ formas y un par de 12 valores y de ${}_4C_2$, así tenemos:

$$13\,{}_4C_3 \times 12\,{}_4C_2 = 3{,}744$$



Ejemplo de tercia de sietes y par de reyes

g) Calcular el número de maneras en que puede salir póker.

Se dice que tenemos *póker* cuando nos tocan 4 cartas del mismo valor y por consecuencia, ya que solo hay 4 en cada valor, otra carta diferente.

Ejemplo de póker de reinas

Esto puede ocurrir con los 13 valores por las combinaciones de los 4 valores tomados los 4 a la vez, por 48 formas de elegir la otra carta, una vez que se han quitado las 4 iguales.

$$13\,{}_4C_4 \times 48 = 624$$

h) Calcular el número de maneras en que puede salir flor imperial.

Una *flor imperial* es cuando tenemos en la jugada las cinco cartas de la misma figura y además de forma consecutiva.

En el inciso d ya se mostraron las 10 corridas que pueden salir en cada figura y como hay 4 figuras tendremos $4 \times 10 = 40$

i) Calcular el número de maneras en que no sale ningún juego de los anteriores.

Algunos llaman *pancha* a la jugada que no es ninguna de las anteriores, y esta se forma tomando 1 carta de cada valor diferente quitando las repeticiones que suceden y tendremos:

$$\frac{52\times48\times44\times40\times36}{5!} = 1'317{,}888$$

Al haber hecho lo anterior estamos considerando los juegos verticales y hay que quitar los que son flores imperiales, flores y las corridas.

Por lo tanto el número de formas en que puede suceder que no nos salga ningún juego son:

$$1'317{,}888 - 15{,}348 = 1'302{,}540$$

Estas son todas las formas que pueden ocurrir al darnos una mano de 5 cartas y lo podemos comprobar sumando todos los juegos y haciendo el cálculo de todas las combinaciones que pueden formarse al tomar 5 cartas de una baraja de 52 que son $_{52}C_5$=2'598,960

Tienes el reto de probar que coinciden los cálculos .

Para resolver:

a) ¿Cuántas formas hay de escoger a un comité formado por un presidente, 3 vocales y 2 tesoreros, entre un grupo de 10 personas?

b) En el juego de dominó un jugador toma 7 fichas, ¿de cuántas formas puede escoger su jugada?

c) De un grupo con 5 hombres y 4 mujeres, ¿cuántos equipos de 3 personas se pueden formar?

3.4. Problemario

Permutaciones con y sin repeticiones

1. ¿De cuántas formas se pueden acomodar 5 personas

a) En una fila con 5 sillas?

b) En una mesa redonda?

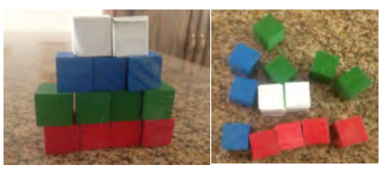
2. 

a) ¿De cuántas formas pueden acomodarse 3 naranjas y 3 manzanas en una fila?

b) ¿De cuántas maneras si las naranjas deben estar juntas y las manzanas también?

c) ¿De cuántas maneras si las manzanas deben ir siempre juntas?

3. Un diseñador tiene que hacer señalizaciones con banderas colocadas de forma vertical, ¿Cuántas señales puede hacer si tiene 4 verdes, 3 blancas y 2 rojas?

4. 

a) ¿De cuántas formas pueden acomodarse 2 cubos blancos, 4 verdes , 4 rojos y 3 azules en fila, de modo que los del mismo color queden siempre juntos?

b) ¿De cuántas formas se pueden acomodar los cubos anteriores en un círculo si los cubos del mismo color deben ir juntos?

Combinaciones

5. De una baraja inglesa se te dan 6 cartas, ¿cuántas formas hay de que te salgan dos tercias?

6. Si de una baraja inglesa se te reparten 6 cartas, ¿de cuántas maneras te pueden salir tres pares?

7. Si de una baraja inglesa te dan 4 cartas, ¿de cuántas formas te pueden salir dos pares?

3.5. Autoevaluación

1.

a)¿Cuántas palabras de seis letras se pueden hacer con las letras del nombre MIGUEL?

b) ¿Que comiencen con vocal y terminen con consonante?

c) ¿ Que empiecen con 2 vocales (no tres) y terminen con consonante.

2. Se desean acomodar 3 cubos de color verde, 4 de color blanco y 2 de color rojo, de tal manera que estén en fila y siempre queden juntos los del mismo color. ¿De cuántas formas se pueden acomodar, si

a) El cubo V_2 verde no debe quedar junto al cubo B_3 blanco;

b) Los cubos B_2 y B_3 blancos no se deben separar en ningún sentido B_2B_3-B_3B_2.

c) El cubo R_1 rojo debe quedar exactamente en medio?

3. ¿Cuántos equipos de futbol (11 elementos) se pueden formar con un grupo de 20 personas?

4. ¿Cuántas placas de dos letras seguidas de 3 dígitos se pueden hacer si el primer dígito y el último no pueden ser cero?

5. Hallar el número de permutaciones que se pueden formar con las palabras:

a) perro

b) gato

c) pez

d) espárrago (considere á=a)

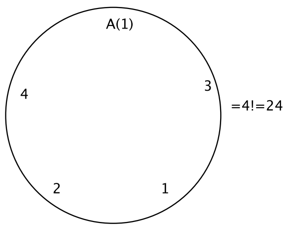
6. ¿Cuántas palabras de cinco letras se pueden formar con las letras de la palabra PATADA?

3.6. Soluciones del problemario

1.

a) 5!=120

b)



2.

a) 6!=720

b) Tenemos dos formas en las que se pueden acomodar, primero las naranjas y después las manzanas, primero las manzanas y después las naranjas. En cada uno de los casos las naranjas tienen 3!=6 formas de acomodarse y las manzanas 3!=6 formas de colocarse, así en total hay $2 \times 3! \times 3! = 72$

c) Tenemos 4 formas de que esto suceda

Así que hay $4 \times 3! \times 3! = 144$

3. Tenemos permutaciones con elementos repetidos donde:

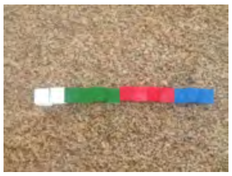
$n=9$

$n_1=4$

$n_2=3$

$n_3=2$

$$\frac{n!}{n_1!\,n_2!\,n_3!}=\frac{9!}{4!\,3!\,2!}=1{,}260$$

4. 

a) Los cuatro colores se pueden acomodar de 4!, en el caso de los blancos hay 2! formas de acomodarse, los verdes 4! formas, los rojos 4! y los azules 3!, así que en total hay:

$$4!\,2!\,4!\,4!\,3!=165{,}888$$

b) Hay 3! formas de acomodarse por colores en un círculo, y cada color tiene 2!, 4!,4! y 3! formas de acomodarse, así que en total hay:

$$3!\times 2!\times 4!\times 4!\times 3!=41{,}472$$

5. $\frac{13(4C3\times 124C3)}{2!}=2{,}496$

6. $\frac{13\times 4C2\times 12\times 4C2\times 11\times 4C2}{3!}=370{,}656$

7. $\frac{13\times 4C2\times 12\times 4C2}{2!}=5{,}616$

3.7. Soluciones de autoevaluación

1.

a) 720

b) 216

c) 24

2.

a) $3V=V_1V_2V_3$, $4B=B_1B_2B_3B_4$ y $2R=R_1R_2$

Hay 3!=6 formas de acomodarlos por colores

VBR	BRV	RVB
VRB	BVR	RBV

Total 3!×3!×4!×2! = 1728

Quitando cuando V_2 queda junto con B_3

Hay cuatro casos:

V	B	R
R	V	B
B	V	R
R	B	V

Terminan o comienzan con V_2

1	2	1	=2

Terminan o comienzan con B_3

1	3	2	1	=6

4(casos)×2(terminan o comienzan con V_2)×6(terminan o comienzan con B_3)×

2(rojas)= 96

total – juntas= 1728-96= 1632

b) Los cubos B_2B_3 no deben separarse, los hacemos un solo elemento, tendremos los tres colores de 3! formas por 2 de los simétricos de B_2B_3 y B_3B_2

$$3\times2\times1\times2 = 12$$

Todas las formas de acomodarlos por colores son 3!=6, los cubos de color verde de 3!, los blancos 12 (cálculo anterior), y 2!=2 los rojos.

$$3!\times3!\times12\times2! = 864$$

c) El cubo R_1debe quedar en medio

Tenemos dos casos: VRB y BRV

$V_1V_2V_3R_2R_1B_1B_2B_3B_4$ y $B_1B_2B_3B_4R_1R_2V_1V_2V_3$

$$2(3!\times1\times1\times4!) = 288$$

3. ${}_{20}C_{11} = \frac{20!}{11!(20-11)!} = 167{,}960$

4.

L	L	NO CERO		NO CERO	
27	26	9	8	8	=404,352

5.

a) perro, p=1, e=1, r=2, o=1, n=5

$$\frac{5!}{2!} = 60$$

b) gato

$$4\times3\times2\times1 = 4! = 24$$

c) pez

$$3\times2\times1 = 3! = 6$$

d) espárrago, e=1, s=1,p=1, a=2, r=2, g=1, o=1, n=9

$$\frac{9!}{2!\,2!} = 90{,}720$$

6. C=2, A=3, B=1

$$\frac{5\times4\times3\times2\times1}{3!\times2!}\times6 = 60$$

EL 6 es por las letras que pueden quedar fuera.

3.8. Conclusión

En este capítulo encontraste algunos métodos de contar haciéndolo de forma indirecta, es decir, sin construir los resultados; otra forma de obtener el número de formas en que se pueden acomodar objetos es con diagramas de árbol que veremos en el próximo capítulo.

La aplicación de estas técnicas de conteo es muy amplia en programación, genética y el cálculo de eventos en probabilidad.

Busca a tu alrededor y encontrarás que puedes poner en práctica estos métodos con objetos que tienes a tu lado, observa los diseños en construcción cómo muestran patrones y calcula el número de formas en que pueden hacerse de forma diferente; cambia variables y encontrarás que se modifican enormemente las formas en que se pueden acomodar.

Referencias

[1] Lipschutz Seymour (1971).Probabilidad.México: McGRAW-HILL

[2] Larson Harold(1992). Introducción a la teoría de probabilidades e inferencia estadística. México: Editorial Limusa,S.A. de C.V.

[3] JohnsonBaugh Richard(1999). Matemáticas discretas. México: Prentice Hall.

[4] Sanmartín José(1992).Los nuevos redentores: reflexiones sobre la ingeniería genética, la sociobiología y el mundo feliz que nos prometen. España: Anthropodos.

Capítulo 4:

Probabilidad

4. Probabilidad

Algunos de los precursores en la teoría de la probabilidad reconocidos en la historia son: Luca Pacioli[1] (1487) quien trata de predecir resultados en los juegos de azar aunque sus resultados no fueron acertados, Girolamo Cardano (1565) que publicó el *Libro de los juegos de azar*, Galileo Galilei, resolvió problemas sobre dados, publicando sus resultados en el libro: *Sobre la puntuación en tiradas de dados*, pero su contribución más importante fue la invención de la teoría de la medida de errores.

Blaise Pascal, Pierre Fermat y Christian Huygens sentaron las bases para la teoría de la probabilidad, siendo Bernoulli, De Moivre y Bayes quienes formulan los teoremas básicos de la probabilidad, el teorema de la suma, de la multiplicación y de la probabilidad condicionada respectivamente.

El uso del cálculo de probabilidades es amplio, algunas áreas de aplicación son: medicina[2], Derecho jurídico[3] e ingeniería[4].
No solo es una teoría la probabilidad sino que ofrece aplicaciones a problemas reales concretos de origen aleatorio con técnicas probabilísticas que pueden ser modeladas.
Este capítulo Incluye la teoría axiomática de la probabilidad, cálculo de probabilidades de eventos equiprobables, dependientes, independientes, variables aleatorias y algunas distribuciones de probabilidad.

En un experimento se pueden tener desde un único resultado llamado suceso elemental[5] o muchos resultados, en consecuencia, cada uno de esos resultados forman parte de un conjunto que contiene a todos los posibles resultados del experimento, llamado espacio muestral[6] se le denota con Ω[1] , en consecuencia, cada evento o suceso siempre es un subconjunto de Ω . Como los resultados son conjuntos[7] podemos utilizar las operaciones del álgebra de conjuntos o álgebra de

Boole para establecer algunas operaciones que permitirán calcular probabilidades de ciertos eventos, tales como la unión, intersección, resta.

George Cantor[8] y otros crearon la teoría de conjuntos[9], la cual aplicaremos como base en la teoría de la probabilidad, es importante hacer notar que en teoría de conjuntos Ω es el universo.

La probabilidad como teoría axiomática se basa en los siguientes tres axiomas de Kolmogorov:

1. Si A es un elemento de la colección Ω existe un número P(A)≥0 denominado probabilidad del suceso S.
2. P(Ω)=1.
3. Dada una sucesión numerable de sucesos A_1, $A_2,\ldots,A_n,\ldots$, disyuntos dos a dos, $A_i \cap A_j = \emptyset$ se verifica que

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

El primer axioma hace referencia a que el valor de la probabilidad de todo evento se encuentra entre los valores 0 y 1, es decir $0 \leq P(A) \leq 1$, por lo que no existen probabilidades negativas, siendo los valores 0 la probabilidad del evento imposible P(imposible)=0, y 1 la probabilidad del evento seguro[10] P(Ω)=1, esto se puede observar en el segundo axioma. Por último el tercer axioma indica que si los eventos no pueden suceder simultáneamente, la probabilidad de que sucedan es la suma de las probabilidades de que suceda cada uno.

Partiendo de estos axiomas se deducen teoremas que dan estructura del cálculo de probabilidades.

Teorema 1. La probabilidad del evento imposible es cero: $P(\emptyset) = 0$.

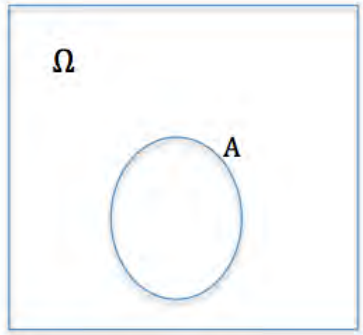


Diagrama de Venn

Demostración:

Sea A un conjunto o suceso; el conjunto vacío $\emptyset$ y el conjunto A son conjuntos disyuntos , así que $A \cup \emptyset = A$

1. $A \cup \emptyset = A$

aplicando la función probabilidad a ambos miembros:

2. $P(A \cup \emptyset) = P(A)$

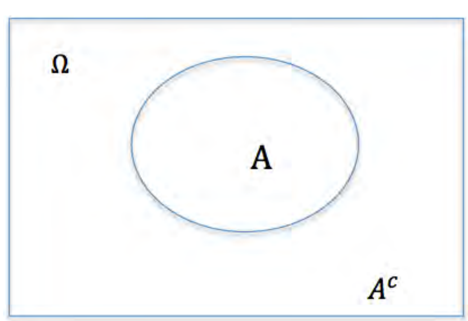
por axioma 3 $P(A \cup \emptyset) = P(A) + P(\emptyset)$

Sustituyendo lo anterior en 2 tenemos:

3. $P(A) + P(\emptyset) = P(A)$ restando de ambos miembros P(A)

4. $P(\emptyset)=0$.

Teorema 2: Si A^c es el complemento de un evento A, entonces $P(A^c)=1-P(A)$.



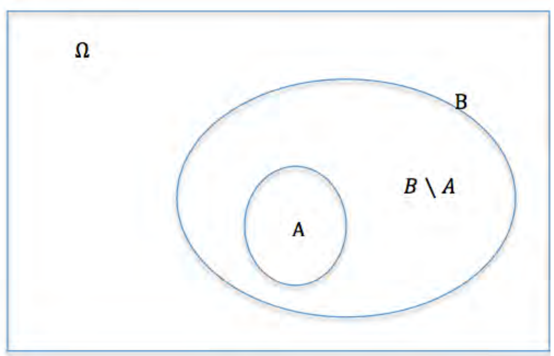
Demostración:

1. $\Omega = \mathrm{A} \cup A^c$ donde A y A^c son eventos mutuamente excluyentes, es decir no tienen elementos en común o no pueden suceder al mismo tiempo.
2. $P(\Omega) = \mathrm{P}(\mathrm{A} \cup A^c)$ aplicando la función probabilidad a los dos miembros

por axiomas 2 y 3 tenemos:

3. $1 = \mathrm{P(A)} + P(A^c)$ despejando
4. P(A^c)=1-P(A).

TEOREMA 3: Si un suceso A está contenido en otro B ($A \subset B$), se verifica que $P(A) \leq P(B)$.

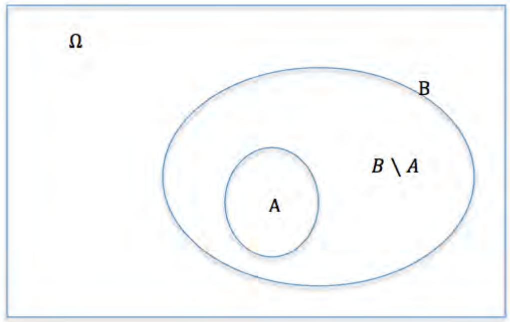


Demostración:

1. $B = A \cup A \setminus B$
2. $P(B) = P(A \cup A \setminus B)$ aplicando la función probabilidad
3. $P(B) = P(A) + P(A \setminus B)$ con lo que se prueba puesto que

$$P(A \setminus B) \geq 0$$

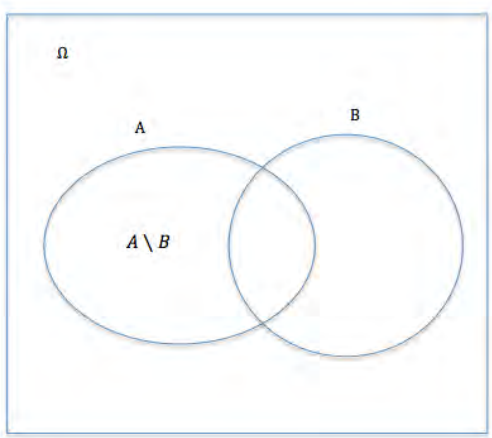
Teorema 4: Si A y B son dos eventos, entonces $P(A \setminus B) = P(A) - P(A \cap B)$



Demostración:

1. A= $(A \setminus B) \cup (A \cap B)$ donde los eventos $A \setminus B$ y $A \cap B$ son eventos mutuamente excluyentes
2. P(A)= $P((A \setminus B) \cup (A \cap B))$ aplicando la función probabilidad
3. P(A)= $P(A \setminus B) + P(A \cap B)$ por axioma 3
4. P$(A \setminus B) = P(A) - P(A \cap B)$ despejando

Teorema 6. La probabilidad de la unión de dos sucesos cualesquiera A y B, $P(A \cup B)$, es igual a $P(A) + P(B) - P(A \cap B)$.



Demostración:

1. $A \cup B = (A \setminus B\,) \cup B$ donde $A \setminus B$ y B son disyuntos

2. $\mathrm{P}(A \cup B) = P((A \setminus B\,) \cup B)$ aplicando la función probabilidad

3. $\mathrm{P}(A \cup B) = P(A \setminus B\,) + P(\,B)$ por axioma 3

4. $\mathrm{P}(A \cup B) = P(A) - P(A \cap B\,) + P(\,B)$

Por teorema $P(A \setminus B) = P(A) - P(A \cap B)$

5. $\mathrm{P}(A \cup B) = P(A) + P(B) - P(A \cap B\,)$

Espacio muestral

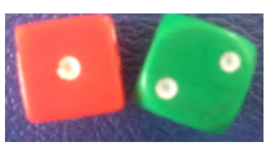
El conjunto de todos los posibles resultados de un experimento se llama *espacio muestral*, se representa con Ω, y aunque no es posible predecir en un experimento que resultado sucederá, si podemos saber cuáles son los posibles resultados, experimentos de este tipo se llaman *aleatorios*.

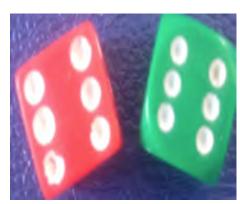
Un *experimento* puede ser cualquier proceso que nos conduce a resultados bien determinados.

Ejemplos de espacios muestrales de experimentos aleatorios:

a) Si el experimento es lanzar un dado normal una vez y observar el número de puntos que aparecen en su cara superior.
El espacio muestral consta de 6 posibles resultados, $\Omega = \{1,2,3,4,5,6\}$

b) Experimento: lanzar 2 dados normales y observar el número de caras superiores que caen.





El espacio muestral consta de 36 eventos o puntos del espacio muestral

$$\Omega = \begin{Bmatrix} 1,1 & 2,1 & 3,1 & 4,1 & 5,1 & 6,1 \\ 1,2 & 2,2 & 3,2 & 4,2 & 5,2 & 6,2 \\ 1,3 & 2,3 & 3,3 & 4,3 & 5,3 & 6,3 \\ 1,4 & 2,4 & 3,4 & 4,4 & 5,4 & 6,4 \\ 1,5 & 2,5 & 3,5 & 4,5 & 5,5 & 6,5 \\ 1,6 & 2,6 & 3,6 & 4,6 & 5,6 & 6,6 \end{Bmatrix}$$

c) Lanzar una moneda normal 2 veces o lanzar dos monedas una vez (este es un ejemplo de un experimento equivalente).

Si llamamos S= sello y C= cara el espacio muestral está formado por 4 posibles resultados $\Omega = \begin{Bmatrix} SS, SC \\ CS, CC \end{Bmatrix}$

Es importante mencionar que CS y SC no es el mismo resultado, CS significa que la primera moneda cayó cara y la segunda sello, o que en el primer lanzamiento salió cara y en el segundo sello.

Definición clásica de probabilidad

La regla clásica de Laplace: cociente entre casos favorables y casos probables, así la probabilidad de un evento A se calcula[11]:

$$P(A) = \frac{número\ de\ formas\ en\ que\ puede\ ocurrir\ el\ evento\ A}{número\ total\ de\ formas\ en\ que\ puede\ ocurrir\ el\ experimento}$$

4.1. Eventos equiprobables

Cuando todos los eventos elementales o los posibles resultados de un experimento tienen igual probabilidad de suceder, se dice que son equiprobables[12].

Por ejemplo al lanzar un dado normal, el espacio muestral es $\Omega = \{1,2,3,4,5,6\}$ la probabilidad de un evento, que denotaremos P(A) de cada punto o evento elemental es:

$P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$, $P(3) = \frac{1}{6}$, $P(4) = \frac{1}{6}$, $P(5) = \frac{1}{6}$, $P(6) = \frac{1}{6}$

Note que la probabilidad de cada evento es ≥ 0 y la suma de las probabilidades de todos los eventos posibles es $\frac{6}{6} = 1$.

Si al lanzar un dado normal queremos calcular la probabilidad de que salga un número par, ya vimos que $\Omega = \{1,2,3,4,5,6\}$, si llamamos A al evento salir número par:

A={2,4,6}

$$P(A) = \frac{3}{6} = \frac{1}{2} = 0.5 = 50\%$$

Si del mismo experimento de lanzar un dado normal queremos calcular la probabilidad de que salga un número primo, llamaremos B al evento salir número primo, y contaremos de cuántas formas distintas puede ocurrir, del total de posibles resultados.

B={2,3,5}

$$P(B) = \frac{3}{6} = \frac{1}{2} = 0.5 = 50\%$$

Utilizando las técnicas de contar del capítulo anterior calculemos probabilidades de que suceda alguna da las jugadas con una baraja inglesa.

Se reparten 5 cartas de una baraja inglesa (52 cartas), calcular la probabilidad de que salga el juego:

Picas/espadas ♠	Tréboles ♣	Corazones ♥	Diamantes/rombos ♦
A	A	A	A
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
J	J	J	J
Q	Q	Q	Q
K	K	K	K

a) Un par

A= salir un par

$$P(A) = \frac{\textit{número de maneras en que se puede sacar un par}}{\textit{número de maneras en que es posible sacar dos cartas}}$$

$$= \frac{13 \times C(4,2) \times \frac{48x44x40}{3!}}{C(52,5)} = \frac{1{,}098{,}240}{2{,}598{,}960} = 0.422569 \cong 42.26\%$$

b) Salgan dos pares

B= salir dos pares

$$P(B)=\frac{\frac{13\times C(4,2)\times 12\times C(4,2)\times 44}{2!}}{C(52,5)}=\frac{123{,}552}{2{,}598{,}960}=\ 0.047539015\cong 4.75\%$$

Se hace una inspección de calidad en una fábrica de celulares, para probar en los aparatos la recepción del wi-fi, se escogen tres celulares entre 15 que hay en una caja de los cuales se conoce que 5 son defectuosos. Calcular la probabilidad de que:

a) Ningún celular sea defectuoso.

El número total de formas en que se pueden escoger los 3 celulares de entre los 15 que hay en la caja es ${}_{15}C_3=455$

Llamaremos A al evento escoger 3 celulares no defectuosos, esto nos deja 10 celulares buenos de los cuales seleccionaremos 3.

A puede suceder de ${}_{10}C_3= 120$ ∴

$$P(A)=\frac{120}{455}=\frac{24}{91}=0.2637\cong 26.37\%$$

b) Por lo menos uno sea defectuoso

Tenemos dos formas de resolverlo, una es calculando la probabilidad de que salgan, uno, dos y tres celulares defectuosos,

$$P(1\ defectuoso)=\frac{C(5,1)\times C(10,2)}{C(15,3)}=\frac{225}{455}=\frac{45}{91}$$

$$P(2\ defectuosos)=\frac{C(5,2)\times C(10,1)}{C(15,3)}=\frac{100}{455}=\frac{20}{91}$$

$$P(3\ defectuosos)=\frac{C(5,3)}{C(15,3)}=\frac{10}{455}=\frac{2}{91}$$

$$P(por\ lo\ menos\ uno)=P(1)+P(2)+P(3)=\frac{45}{91}+\frac{20}{91}+\frac{2}{91}=\frac{67}{91}$$

Otra forma es:

Al pedirnos por lo menos uno, su complemento es ninguno, así aplicando el teorema

Teorema 2: Si A^c es el complemento de un evento A, entonces P(A^c)=1-P(A).

y habiendo calculado en a) la probabilidad de que ninguno sea defectuoso, tenemos:

$$P(al\ menos\ uno\ sea\ defectuoso) = 1 - P(ninguno\ sea\ defectuoso)$$

$$P(al\ menos\ uno\ sea\ defectuoso) = 1 - \frac{24}{91} = \frac{67}{91}$$

Consideremos el experimento de lanzar una moneda y un dado al mismo tiempo.

a) Construir el espacio muestral

$$\Omega = \{S1, S2, S3, S4, S5, S6, C1, C2, C3, C4, C5, C6\}$$

Indicar que elementos pertenecen a los eventos:

b) Sea A el evento salir cara y número par, calcular su probabilidad.

A={C2,C4,C6}

$$P(A) = \frac{3}{12} = \frac{1}{4} = 0.25 = 25\%$$

c) Sea B el evento aparece un número par, calcular su probabilidad.

B={S2,S4,S6,C2,C4,C6}

$$P(B) = \frac{6}{12} = \frac{1}{2} = 0.5 = 50\%$$

d) Sea C el evento en que sale sello y un número impar, calcular su probabilidad.

C={S1,S3,S5}

$$P(C) = \frac{3}{12} = \frac{1}{4} = 0.25 = 25\%$$

e) Calcular la probabilidad de que suceda A o C.

A o C $=A \cup C = \{C2, C4, C6, S1, S3, S5\}$

$$P(A \cup B) = P(A) + P(B) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = 0.5 = 50\%$$

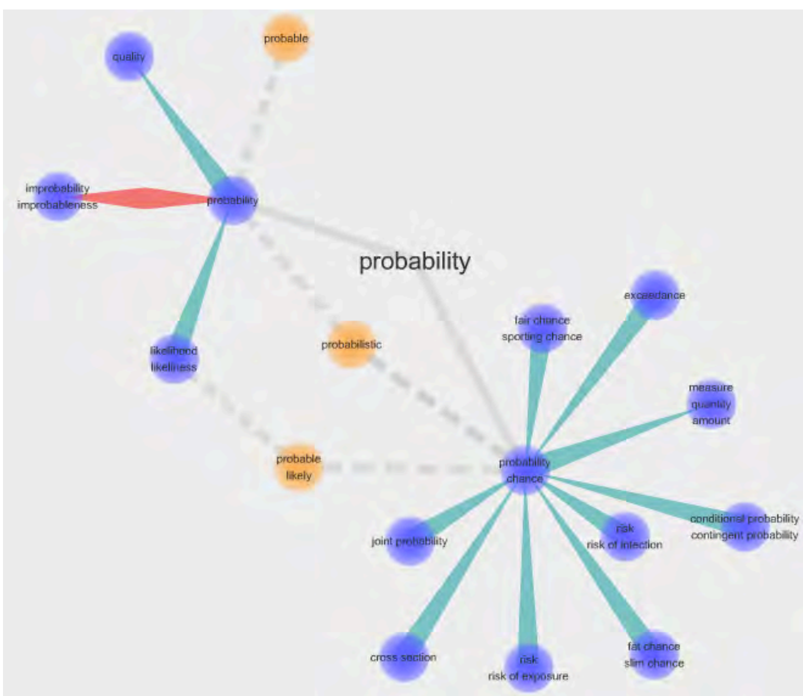
$$P(A \cup B) = \frac{6}{12} = \frac{1}{2} = 0.5 = 50\%$$

Note que A y C son exclusivos o disyuntos, esto es $A \cap C = 0$

f) Calcular la probabilidad de que suceda $A \cap B$.

$A \cap B = \{C2, C4, C6\}$

$$P(A \cap B) = \frac{3}{12} = \frac{1}{4} = 0.25 = 25\%$$



4.2. Probabilidad condicional e independencia

En algunos experimentos la probabilidad de que ocurra algún evento se ve alterada por la ocurrencia de un evento previo, este tipo de eventos se llaman *condicionales* o *dependientes*. La probabilidad de que ocurra un evento A una vez que haya sucedido B se denotará con: $P(A/B)$, se lee; la probabilidad de que ocurra A dado que ya ocurrió B y se define como:

$$P(A/B) = \frac{P(A \cap B)}{P(\mathrm{B})} = \frac{\frac{número\ de\ elementos\ de\ A \cap B}{número\ de\ elementos\ de\ \Omega}}{\frac{número\ de\ elementos\ de\ \mathrm{B}}{número\ de\ elementos\ de\ \Omega}} = = \frac{número\ de\ elementos\ de\ A \cap B}{número\ de\ elementos\ de\ \mathrm{B}}$$

Esto es, la probabilidad de que ocurra el evento A, una vez que haya sucedido B, es igual al número de formas en que pueden ocurrir A y B, dividido entre el número de formas que sucede el evento B o bien, la probabilidad de que ocurra A dado que ya sucedió B, es igual a la probabilidad de que ocurran A y B, entre la probabilidad de que ocurra el evento B.

Por ejemplo, si nuestro experimento es lanzar un par de dados normales, calcularemos la probabilidad de que si la suma de sus caras superiores es 7, uno de los dados sea un 5.

En nuestro ejemplo el evento B={las caras superiores suman 7 puntos} está formado por los eventos:

$$B = \{1{,}6, 2{,}5, 3{,}4, 4{,}3, 5{,}2, 6{,}1\}$$

y nuestro evento $A = \{uno\ de\ los\ dados\ por\ lo\ menos\ es\ el\ 5\}$

$$A = \{1{,}5,\ 2{,}5,\ 3{,}5,\ 4{,}5,\ 5{,}5,\ 6{,}5,\ 5{,}1,\ 5{,}2,\ 5{,}3,\ 5{,}4,\ 5{,}6\}$$

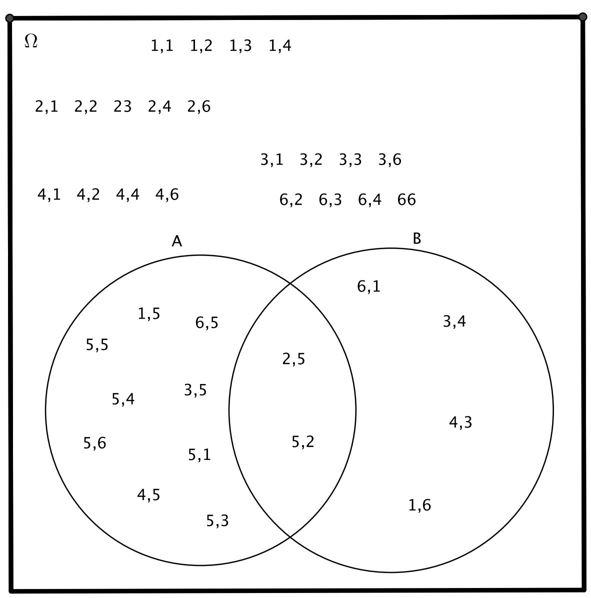
el espacio muestral Ω consta de 36 posibles resultados que se muestran a continuación:

$$\Omega = \begin{Bmatrix} 1,1 & 2,1 & 3,1 & 4,1 & 5,1 & 6,1 \\ 1,2 & 2,2 & 3,2 & 4,2 & 5,2 & 6,2 \\ 1,3 & 2,3 & 3,3 & 4,3 & 5,3 & 6,3 \\ 1,4 & 2,4 & 3,4 & 4,4 & 5,4 & 6,4 \\ 1,5 & 2,5 & 3,5 & 4,5 & 5,5 & 6,5 \\ 1,6 & 2,6 & 3,6 & 4,6 & 5,6 & 6,6 \end{Bmatrix}$$

El evento $A \cap B$ sucede cuando sucede A y sucede B, así $A \cap B = \{\, sale\ un\ 5\ y\ la\ suma\ es\ 7\}$, $\therefore\ A \cap B = \{2,5\ y\ 5,2\}$ tiene dos elementos ya que tenemos dos resultados que suman 7 sus puntos ,y uno de ellos es el 5.

Por lo tanto la probabilidad de que salga un 5, dado que la suma de las caras es 7, está dada por:

$$P(A/B) = \frac{P(A \cap B)}{\mathrm{P(B)}} = \frac{\frac{2}{36}}{\frac{6}{36}} = \frac{2}{6} = \frac{1}{3}$$



El siguiente ejemplo, nos muestra dos eventos dependientes:

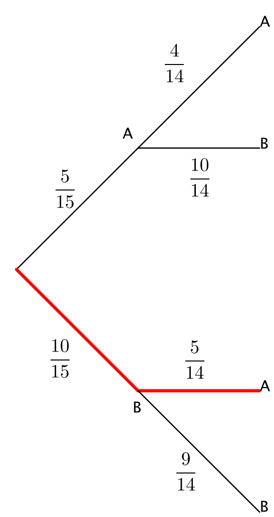
En un laboratorio están colocadas 15 muestras de cultivos, de las cuales 10 son de hepatitis B y 5 de hepatitis A, se eligen al azar dos muestras una por una. ¿Cuál es la probabilidad de que:

a) La primera muestra elegida sea de hepatitis B?

La probabilidad de elegir una muestra por primera vez y que esta sea de hepatitis B es:

$$P(B) = \frac{10}{15} = \frac{2}{3}$$

b) La segunda sea de hepatitis A, dado que la primera elegida fue de hepatitis B?



La probabilidad de que la primera muestra sea de hepatitis B es $P(B) = \frac{10}{15}$, y la probabilidad de que la segunda muestra elegida sea de hepatitis A se ve condicionada a que la primera haya sido de B, así su probabilidad es $P(A) = \frac{5}{14}$ ya que quedan 5 muestras de hepatitis A de 14 muestras dado que sabemos que la primera fue de hepatitis B. De manera que la probabilidad de que la segunda sea de hepatitis A dado que la primera fue de hepatitis B, es el producto de las dos probabilidades:

$$P = \left(\frac{10}{15}\right)\left(\frac{5}{14}\right) = \frac{5}{21}$$

Si en la fórmula de probabilidad condicional:

$$P(A/B) = \frac{P(A \cap B)}{P(\mathrm{B})}$$

despejamos $P(A \cap B)$, tenemos:

$$P(A \cap B) = P(\mathrm{B})P(A/B)$$

Esto se lee: la probabilidad de que sucedan A y B es igual, a la probabilidad de que suceda B, por la probabilidad de que suceda A dado que ya sucedió B.

Es importante hacer notar que $A \cap B = B \cap A$.

La fórmula obtenida

$$P(A \cap B) = P(\mathrm{B})P(A/B)$$

se conoce como *teorema de la multiplicación* para probabilidad condicional.

En el caso particular de que los eventos A y B sean excluyentes o disyuntos, esto es $A \cap B = 0$, la fórmula anterior se convierte a:

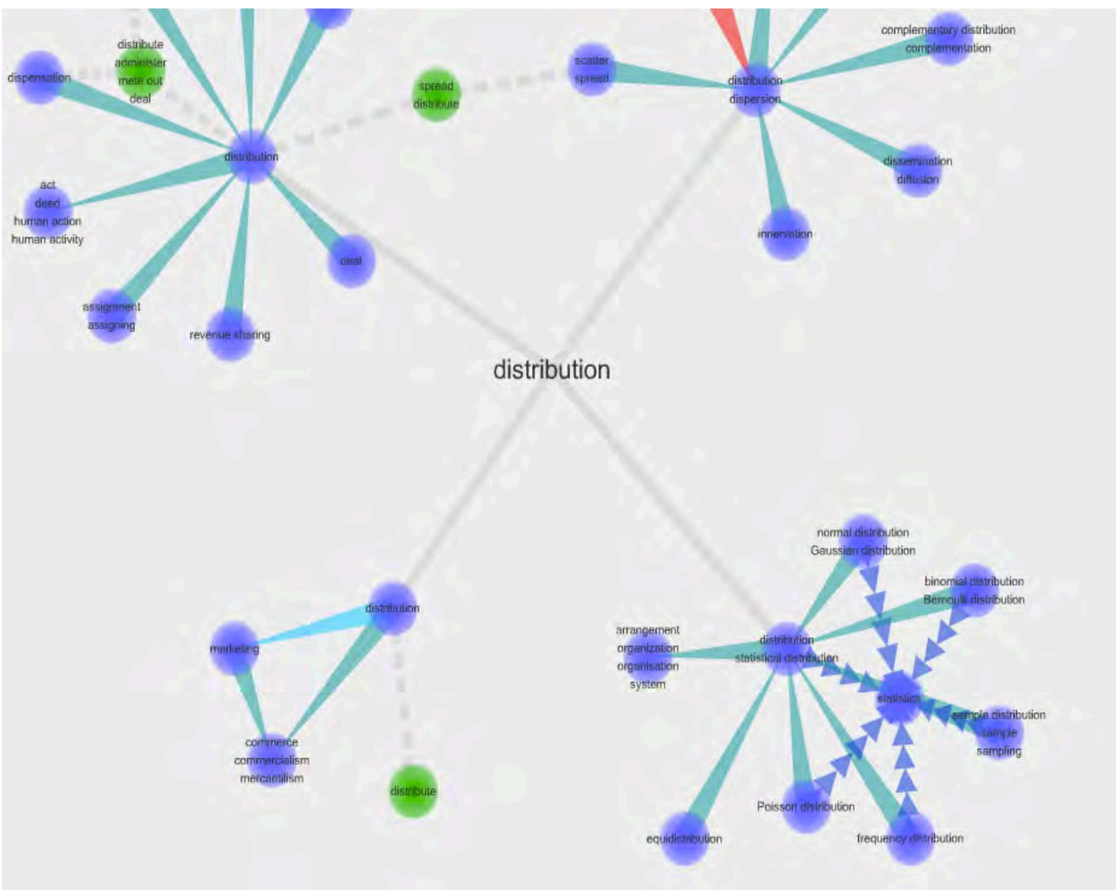
$$P(A \cap B) = P(A)P(B)$$

que es la probabilidad de que dos eventos *independientes* ocurran, y se lee: la probabilidad de que ocurra A y B es igual a la probabilidad de que ocurra A, por la probabilidad de que ocurra B.

Por ejemplo si tenemos una urna con 120 boletas de votos, de los cuales 40 están anulados por haber sido llenados incorrectamente y se toman al azar tres boletas de la urna una tras otra. Hallar la probabilidad de que las tres boletas sean de las llenadas correctamente.

La probabilidad de que el primer voto sea correcto es $P = \frac{80}{120}$ puesto que hay 80 votos correctos de un total de 120 boletas. Si la primera boleta es correcta, entonces la probabilidad de que la segunda boleta extraída sea correcta es $P = \frac{79}{119}$. Si las dos primeras boletas son correctas, la probabilidad de que la tercera extraída sea correcta es $P = \frac{78}{118}$ así por el teorema de la multiplicación.

$$P = \frac{80}{120} \times \frac{79}{119} \times \frac{78}{118} = \frac{2054}{7021} = 0.2925 \cong 29.25\%$$



4.3. Procesos estocásticos finitos y diagramas de árbol

Se llaman *procesos estocásticos finitos*[13] a un conjunto de experimentos finito, donde cada uno de ellos tiene también un número finito de resultados, los cuales son contables y calculables.

Para hacer los cálculos de probabilidades es conveniente hacer una descripción gráfica de los eventos y sus procesos, para lo cual utilizaremos los *diagramas de árbol*.

En una fábrica donde se procesan microchips para celulares, verifican el proceso de producción, analizando en los tres turnos de los empleados, cuántos microchips no cumplen con el estándar de calidad. El día que se hizo la verificación se obtuvieron los siguientes resultados:

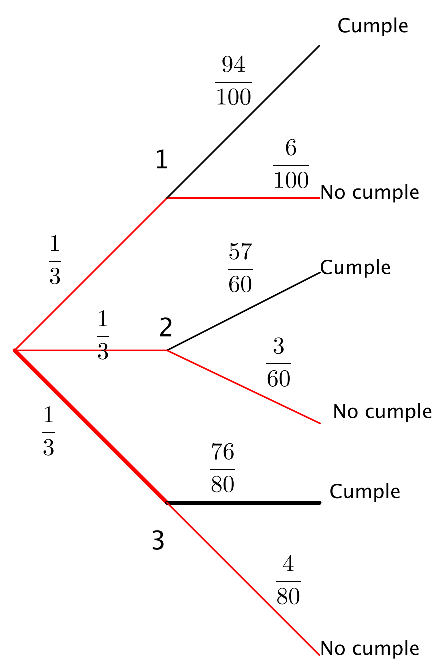
Turno 1: produjo 100 chips de los cuales 6 no cumplen el estándar de calidad.
Turno 2: produjo 60 chips de los cuales 3 no cumplen con el estándar de calidad.
Turno 3: produjo 80 chips de los cuales 4 no cumplen con el estándar de calidad.

El encargado de control de calidad escoge al azar un turno para hacer su inspección y toma al azar un microchip de la producción, ¿cuál es la probabilidad de que el microchip no cumpla con el estándar de calidad?

En este problema tenemos dos experimentos
1) Escoger el turno;
2) Escoger un microchip que cumpla o no cumpla con el estándar de calidad.

Para describir el proceso haremos un diagrama de árbol. El primer experimento es escoger al azar un turno, por lo que se deben abrir tres ramas ya que hay tres turnos, después por cada rama abrimos dos ramas que indican el segundo experimento; tomar un microchip al azar, ya que podemos obtener un microchip que cumpla y otro que no cumpla con los estándares de calidad. En cada rama colocamos la probabilidad que tiene cada evento. Por el teorema de la multiplicación la probabilidad de cada trayectoria que nos lleve a que suceda el evento es el producto de las probabilidades de cada rama de la trayectoria. Cada trayectoria que nos lleve al evento que queremos calcular su probabilidad es exclusiva, por lo que la probabilidad es la suma de todas las trayectorias que cumplen con los eventos pedidos.

El siguiente diagrama de árbol muestra el proceso de los experimentos, sus probabilidades y están marcadas las trayectorias que llevan a dichos eventos.



$$P = \frac{1}{3}\times\frac{6}{100} + \frac{1}{3}\times\frac{3}{60} + \frac{1}{3}\times\frac{4}{80} = \frac{53}{900} = 0.0588 \cong 6\%$$

4.4. Variables aleatorias[14]

Sean (Ω, P) un espacio probabilístico asociado a un experimento aleatorio

$$\mathrm{X}: \Omega \rightarrow \mathbb{R}$$

$$a_1 \rightarrow X(a_1)$$

esto es, a cada elemento de Ω se le asocia un número real.

Definición. Una variable aleatoria[15] es una función numérica definida en el espacio de muestreo.

Las variables aleatorias son base para modelar experimentos aleatorios, se les asocia un valor numérico, resultado del experimento. Las variables aleatorias pueden ser de dos tipos, discretas o continuas, unidimensionales, n-dimensionales (ver el capítulo 1). Las variables aleatorias tienen asociadas dos funciones: la función de distribución de probabilidad y la función de densidad. En este capítulo trabajaremos con variables unidimensionales.

Ejemplo:

Lanzar dos monedas normales al mismo tiempo, que equivale a la lanzar una moneda dos veces, y sea X la variable aleatoria que representa el número de caras.

El espacio muestral $\Omega = \{CC, CS, SC, SS\}$ y la variable aleatoria X toma los valores siguientes:

$$X = \{0,1,2\}$$

asignando un valor numérico a los resultados de X:

$$P(0) = \frac{1}{4}$$

$$P(1) = \frac{2}{4} = \frac{1}{2}$$

$$P(2) = \frac{1}{4}$$

4.5. Distribuciones de probabilidad

Al introducir una variable aleatoria y haber calculado la probabilidad de sus puntos del espacio muestral, se ha determinado la distribución de probabilidad de la variable aleatoria, a la que se llamará su distribución de probabilidad.

Por ejemplo si el experimento es lanzar dos dados normales vimos que el espacio muestral consta de 36 puntos:

$$\Omega = \begin{Bmatrix} 1,1 & 2,1 & 3,1 & 4,1 & 5,1 & 6,1 \\ 1,2 & 2,2 & 3,2 & 4,2 & 5,2 & 6,2 \\ 1,3 & 2,3 & 3,3 & 4,3 & 5,3 & 6,3 \\ 1,4 & 2,4 & 3,4 & 4,4 & 5,4 & 6,4 \\ 1,5 & 2,5 & 3,5 & 4,5 & 5,5 & 6,5 \\ 1,6 & 2,6 & 3,6 & 4,6 & 5,6 & 6,6 \end{Bmatrix}$$

Si nuestra variable aleatoria X la definimos como la suma de los puntos que muestran las caras superiores del lanzamiento de los dos dados, tendremos que toma los valores:

$$X = \{2,3,4,5,6,7,8,9,10,11,12\}$$

calculemos las probabilidades de que sumen 2,3,4,5,6,7,8,9,10,11 y 12

$P(2) = \frac{1}{36}$

$P(3) = \frac{2}{36}$

$P(4) = \frac{3}{36}$

$P(5) = \frac{4}{36}$

$P(6) = \frac{5}{36}$

$P(7) = \frac{6}{36}$

$P(8) = \frac{5}{36}$

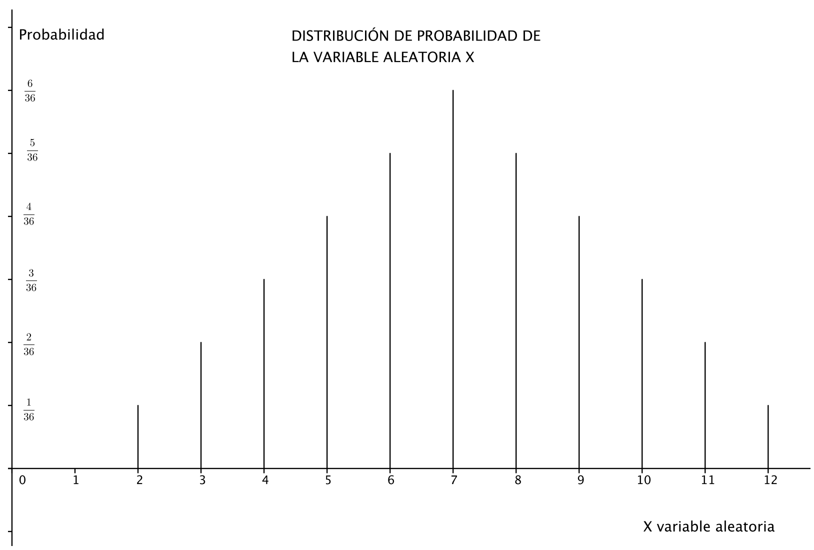
$P(9) = \frac{4}{36}$

$P10 = \frac{3}{36}$

$P(11) = \frac{2}{36}$

$P(12) = \frac{1}{36}$

esto es la distribución de probabilidad de la variable aleatoria X, si la representamos gráficamente, tendremos:



En capítulos anteriores hicimos histogramas dando una representación gráfica o empírica de un conjunto de datos, después se dio una representación aritmética con las medidas de tendencia central y de dispersión asignando valores numéricos, ahora haremos lo mismo pero en lugar de usar las frecuencias relativas usaremos la probabilidad, dado que cuando la población se hace más grande $\frac{f_i}{N} \approx P(x_i)$.

En el capítulo 2 vimos que la *media aritmética* o promedio para datos agrupados la podemos escribir:

$$\bar{x} = \sum \frac{fx}{N} = \sum_{i=1}^{K} x_i \frac{f_i}{N}$$

La *media teórica* que denotaremos con la letra griega μ la obtenemos sustituyendo la frecuencia relativa $\frac{f_i}{N}$ que corresponda al valor x_i por la respectiva probabilidad $P(x_i)$ y tendremos:

$$\mu = \sum_{i=1}^{k} x_i P(x_i)$$

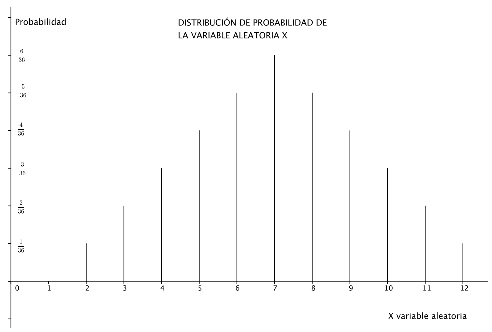
De la misma manera para la varianza empírica podemos pasar a la forma modificada.

$$s^2 = \sum_{i=1}^{K} (x_i - \bar{x})^2 \frac{f_i}{N}$$

La varianza teórica la denotaremos con $\sigma^2 = \sum_{i=1}^{k} (x_i - \mu)^2 P(x_i)$
Recordemos que la desviación estándar es la raíz cuadrada de la varianza.

Es importante remarcar que la media, varianza y desviación estándar de una variable aleatoria son también la media, varianza y desviación estándar de la distribución de probabilidad.

Para ejemplificar las fórmulas de la mediana, varianza y desviación estándar teóricas utilizaremos el experimento del lanzamiento de dos dados normales, donde observamos que la variable aleatoria X toma los valores que suman las caras superiores de los dados.



De la gráfica se puede adelantar el resultado de la media teórica, gracias a la simetría que presenta la gráfica, que la media teórica se localiza en el valor 7.

Cálculo de la media teórica:

x_i	$P(x_i)$	$x_iP(x_i)$
2	$\frac{1}{36}$	$\frac{2}{36}$
3	$\frac{2}{36}$	$\frac{6}{36}$
4	$\frac{3}{36}$	$\frac{12}{36}$
5	$\frac{4}{36}$	$\frac{20}{36}$
6	$\frac{5}{36}$	$\frac{30}{36}$
7	$\frac{6}{36}$	$\frac{42}{36}$
8	$\frac{5}{36}$	$\frac{40}{36}$
9	$\frac{4}{36}$	$\frac{36}{36}$
10	$\frac{3}{36}$	$\frac{30}{36}$
11	$\frac{2}{36}$	$\frac{22}{36}$
12	$\frac{1}{36}$	$\frac{12}{36}$
Total	$\frac{36}{36} = 1$	$\frac{252}{36} = 7$

Cálculo de la varianza de la distribución de probabilidad:

x_i	$P(x_i)$	$(x_i-\mu)$	$(x_i-\mu)^2$	$(x_i-\mu)^2P(x_i)$
2	$\frac{1}{36}$	2-7=-5	25	$\frac{25}{36}$
3	$\frac{2}{36}$	3-7=-4	16	$\frac{32}{36}$
4	$\frac{3}{36}$	4-7=-3	9	$\frac{27}{36}$
5	$\frac{4}{36}$	5-7=-2	4	$\frac{16}{36}$
6	$\frac{5}{36}$	6-7=-1	1	$\frac{5}{36}$
7	$\frac{6}{36}$	7-7=0	0	$\frac{0}{36}$
8	$\frac{5}{36}$	8-7=1	1	$\frac{5}{36}$
9	$\frac{4}{36}$	9-7=2	4	$\frac{16}{36}$
10	$\frac{3}{36}$	10-7=3	9	$\frac{27}{36}$
11	$\frac{2}{36}$	11-7=4	16	$\frac{32}{36}$
12	$\frac{1}{36}$	12-7=5	25	$\frac{25}{36}$
Total	1			$\frac{210}{36}=5.83$

$\therefore \quad \sigma^2 = 5.83 \quad \text{y} \quad \sigma = 2.41$

La *media aritmética* también recibe los nombres de *esperanza matemática* o *valor esperado*.

4.5.1. Distribución binomial

Una de las distribuciones de probabilidad más importantes para variables discretas es la distribución binomial. Tiene la característica de representar experimentos que solo tienen dos posibles resultados, cuando ocurre el suceso elegido se le llama éxito e y cuando no ocurre fracaso f, es un eufemismo no es que se desee que suceda algo bueno o malo, este tipo de experimentos reciben el nombre de experimentos de Bernoulli[16]. El espacio muestral consta de dos elementos $\Omega = \{e, f\}$, denotaremos la probabilidad de éxito $P(e)$ con p y en consecuencia por ser el evento complementario $P(f) = 1 - p$ y lo denotaremos con q.

La variable aleatoria X o variable binomial, representa el número de éxitos ocurridos al realizar n veces el experimento, donde cada vez que se realiza el experimento cada uno de ellos es independiente del anterior, ejemplos de estos son: el lanzamiento de una moneda, donde si llamamos éxito que caiga un sello (S), el otro posible resultado es que no salga sello, esto es cara (C), así no importando cuántas veces se realice el experimento de lanzar la moneda obtendremos solo dos posibles resultados, y cada uno de ellos independiente. Otro ejemplo es lanzar un dado normal, si llamamos éxito a que salga un número par, fracaso significa que sale un número impar. La inspección de calidad de algún producto, este puede funcionar o no funcionar. En un juego deportivo de futbol, podemos llamar éxito a que un equipo realice una anotación y fracaso que no realice anotaciones, o ganar y perder. Para calcular la probabilidad de un evento de tipo binomial utilizamos el siguiente teorema.

Teorema:

La probabilidad de que x éxitos ocurran en n pruebas repetidas se calcula mediante la fórmula:

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

donde n es el número de veces que se realiza el experimento,

p la probabilidad de que ocurra un evento, este es que se tenga éxito,

q es la probabilidad de que no ocurra el evento, es decir que sea un fracaso.

Observe que $\frac{n!}{x!(n-x)!} = nCx$ son las combinaciones de n objetos tomados x que se vio en el capítulo 3.

Por ejemplo, suponga que la probabilidad de que los padres con diabetes la hereden a los hijos es del 3% . En una familia de 8 hijos, ¿cuál es la probabilidad de que por lo menos 6 hijos hereden diabetes?

Aquí podemos considerar a cada hijo como una prueba independiente, por lo tanto n=8, llamaremos éxito cuando uno de los hijos herede la diabetes, y fracaso, cuando no la hereden, así

$$n = 8$$

$$x \geq 6 \text{ esto es } x = \{6,7,8\}$$

$$p = 3\% = 0.03$$

$$q = 1 - p = 1 - 0.03 = 0.97$$

Calculamos la probabilidad de que ocurran 6, 7 y 8 éxitos y sumamos sus probabilidades ya que cada evento es mutuamente excluyente.

utilizando la fórmula:

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

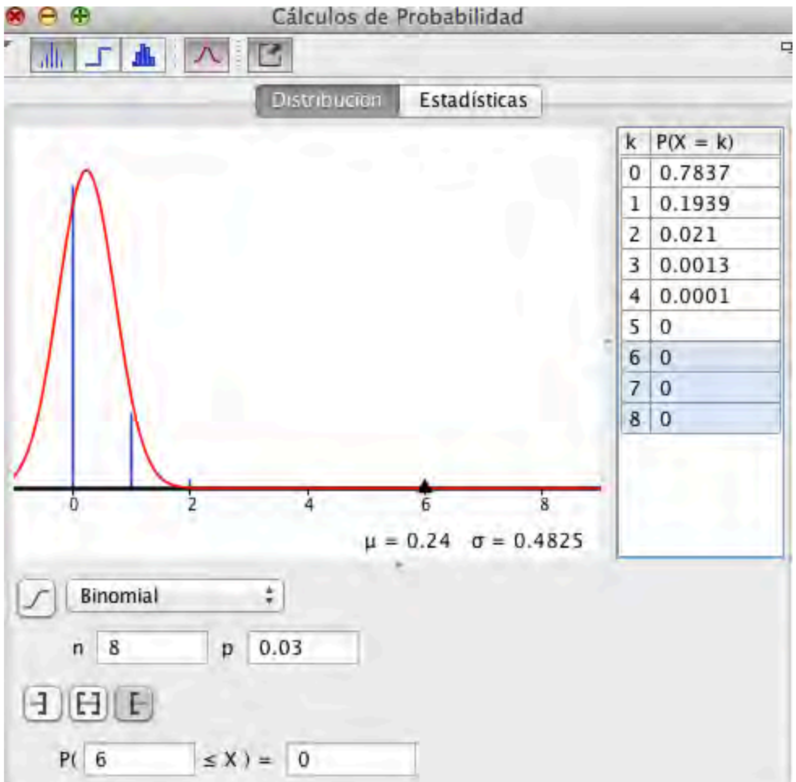
$$P(6) = \frac{8!}{6!\,(8-6)!}\,(0.03)^6(0.97)^2 = 1.9205\times10^{-08}$$

$$P(7) = \frac{8!}{7!\,(8-7)!}\,(0.03)^7(0.97)^1 = 1.6971\times10^{-10}$$

$$P(8) = \frac{8!}{8!\,(8-8)!}\,(0.03)^8(0.97)^0 = 6.561\times10^{-13}$$

$$P(x \geq 6) = P(6) + P(7) + P(8) = 1.9205{\times}10^{-08} + 1.6971{\times}10^{-10} + 6.561{\times}10^{-13}$$
$$= 0.0000000193760169 \approx 0\%$$

Como se puede observar bajo estas condiciones la probabilidad de que por lo menos 6 hijos hereden la diabetes en una familia con 8 hijos es casi cero o imposible.

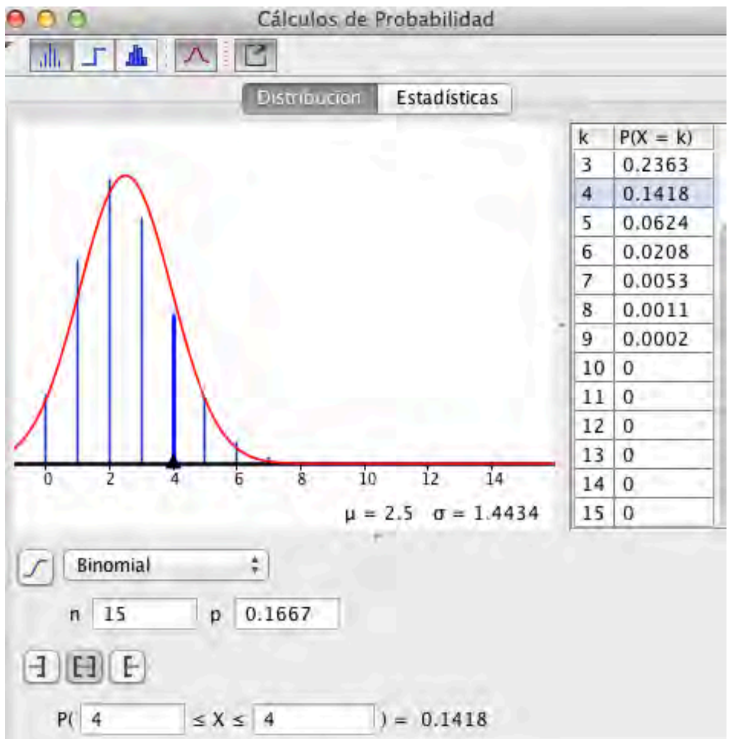


Un dado normal se lanza 15 veces, si llamamos éxito que caiga el número 5, ¿cuál es la probabilidad de que caiga el número 5 exactamente 4 veces?

El experimento es de tipo binomial o de Bernoulli, ofrece dos posibles resultados; éxito cuando cae el 5 y fracaso si sale cualquier otro número (1,2,3,4 o 6), por lo tanto la probabilidad de éxito $p = \frac{1}{6}$ y la de fracaso $q = 1 - p = 1 - \frac{1}{6} = \frac{5}{6}$, el experimento se realiza 15 veces por lo tanto $n = 15$, aplicando la fórmula:

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

$$P(4) = \frac{15!}{4!\,(15-4)!} \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^{11} = 0.141753549 \cong 14.18\%$$

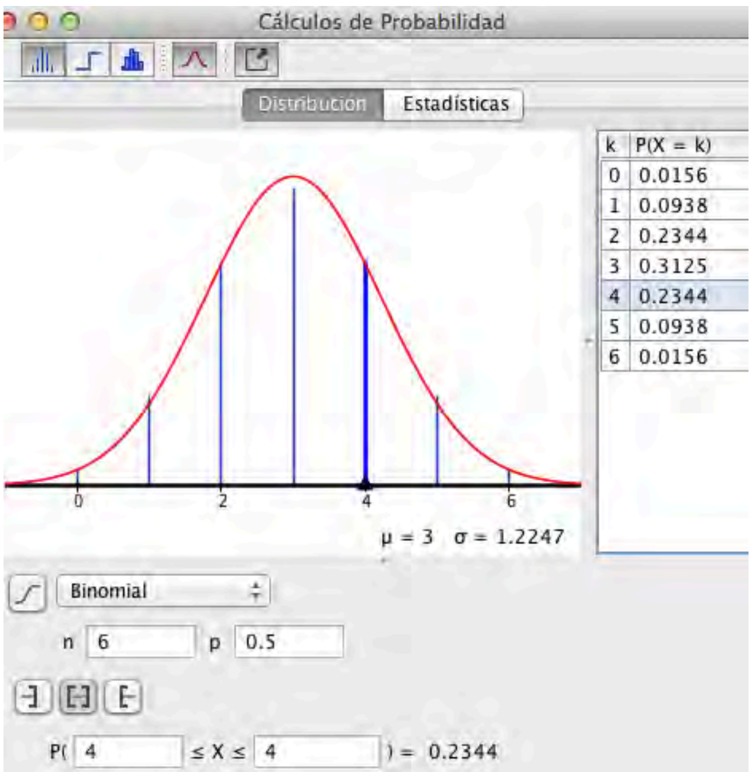


Una pareja desea tener 6 hijos, si la probabilidad de que sea niña es $\frac{1}{2}$. Calcular la probabilidad de que de los 6 hijos 4 sean niñas y 2 niños.

Estamos considerando que un éxito es que nazca niña, así que $p = \frac{1}{2} \;\therefore\; q = \frac{1}{2}$, suponemos que cada hijo es un evento independiente así que $n = 6$, el valor de la variable aleatoria, que es la que toma el número de éxitos, tiene por valor 4, si sustituimos lo anterior en la fórmula para una distribución binomial tendremos:

$$P(x) = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}$$

$$P(4) = \frac{6!}{4!\,(6-4)!} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^2 = 0.2343 \approx 23.43\%$$



Propiedades de la distribución binomial

Media teórica	$\mu = np$
Varianza teórica	$\sigma^2 = npq$
Desviación estándar	$\sigma = \sqrt{npq}$

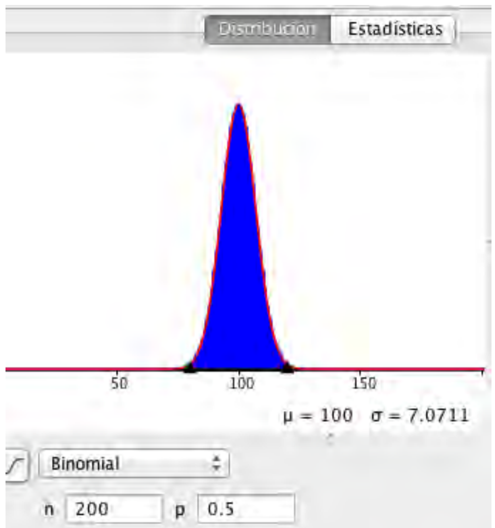
Calcular el valor esperado de caras que caerán, si se lanza una moneda normal 200 veces.

El número esperado de caras, esperanza matemática o media aritmética es:

$$E[X] = \mu = np = (200)\left(\frac{1}{2}\right) = 100$$

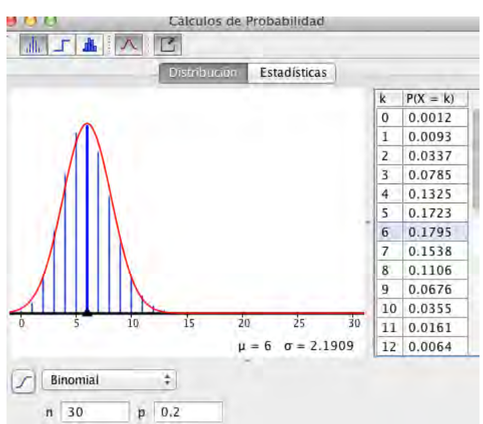
De este mismo experimento calculemos su varianza y desviación estándar:

$$\sigma^2 = npq = (200)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = 50 \quad \therefore \quad \sigma = \sqrt{50} = 7.07$$



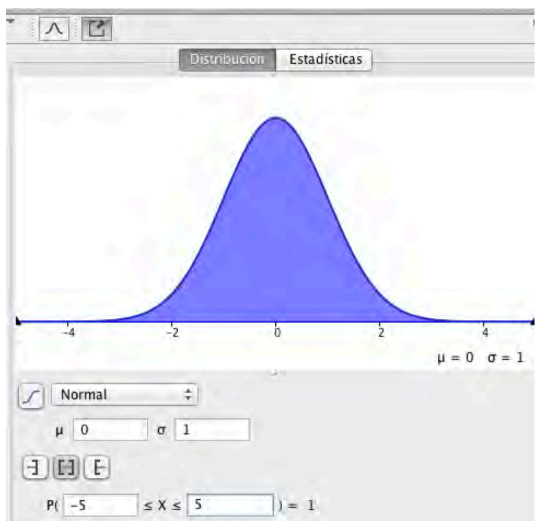
En un grupo escolar de 30 alumnos, cada uno está propenso a contraer la influenza estacional, asistieron a un evento donde se pusieron en contacto con un portador de la enfermedad. Si la probabilidad de que el portador contagie a cualquier alumno es del 20%. ¿Cuántos alumnos se espera que contraigan la enfermedad?

$$E[X] = \mu = np = (30)(.2) = 6$$



4.5.2. Distribución normal

Una de las distribuciones de variables discretas más importante es la *distribución normal o de Gauss*[17], esta ocurre en experimentos que se repiten un gran número de veces, donde los errores son puramente aleatorios, los resultados muestran tendencia a agruparse alrededor de la media de forma simétrica. Entre más veces se repita el experimento, los resultados llegan a tomar la forma de una curva de nombre *campana de Gauss* o distribución normal. Este tipo de distribución ayuda en modelar mucho fenómenos, sociales como el consumo de algún producto en cierta población, psicológicos como el coeficiente intelectual, etc..



El área bajo la curva representa la probabilidad

Observe en la gráfica anterior que la probabilidad es igual al área bajo la curva, el área total es aproximadamente del 100% o igual a 1, tiene forma de una campana, es simétrica respecto del valor central que es la media.

El cálculo de la probabilidad está dado por la siguiente fórmula dada en el teorema.

Teorema: La probabilidad de una distribución normal con media μ y desviación estándar σ esta dada por la función

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Dado que la curva es simétrica , su valor central que es la media, queda a la mitad de la gráfica.

Como la media y desviación estándar puede variar infinitamente, es conveniente estandarizar convirtiendo una distribución normal a una distribución normal estándar, haciendo la media $\mu = 0$ y $\sigma = 1$, la relación para pasar a una distribución normal estandarizada es mediante la fórmula:

$$z = \frac{x-\mu}{\sigma}$$

Así la fórmula anterior transformará cualquier variable x a una con media cero y desviación estándar 1. Para calcular el valor de la probabilidad se hace uso de una tabla que indica el área bajo la curva .

Ejemplo:

Calcular la probabilidad de que salgan 38 y 54 caras al lanzar una moneda normal 100 veces.

$$n = 200$$

$$p = \frac{1}{2} \therefore q = \frac{1}{2}$$

éxito salga cara así que $P(38 \leq x \leq 54)$

la media $\mu = np = 100\left(\frac{1}{2}\right) = 50\ y\ \sigma^2 = \sqrt{npq} = \sqrt{50\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{50} = 12.5$

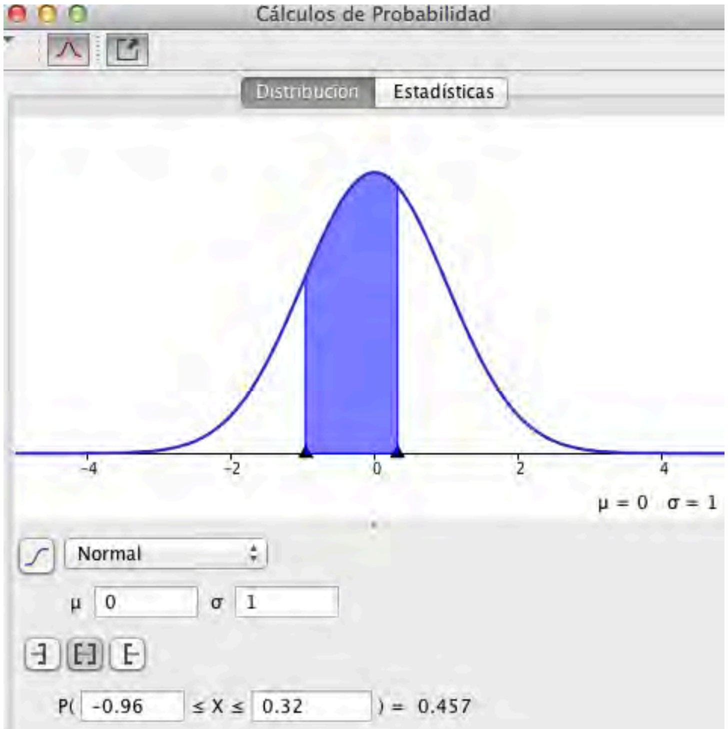
estandarizando tenemos:

$$z = \frac{x-\mu}{\sigma} = \frac{38-50}{12.5} = \frac{-12}{12.5} = -0.96$$

$$z = \frac{x-\mu}{\sigma} = \frac{54-50}{12.5} = \frac{4}{12.5} = 0.32$$

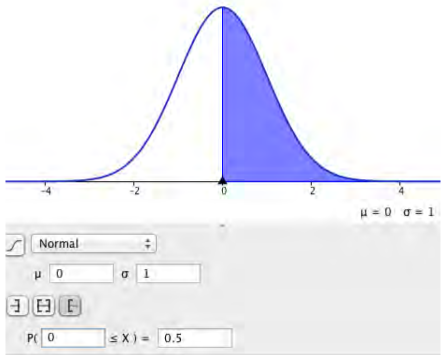
transformando la probabilidad:

$$P(-8 \leq x \leq 54) = P(-0.96 \leq z \leq 0.32) = 0.457 \approx 45.7\%$$

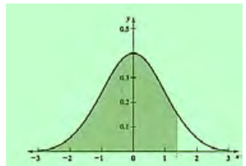


Como la gráfica es simétrica podemos hacer uso de tablas de probabilidad, calculando primeramente la probabilidad $P(0 \leq x \leq 0.96) + P(0 \leq x \leq 0.32) = 0.3315 + 0.1255 = 0.457$ se busca en la columna de z 0.9 y buscamos la intersección de la columna .06 igual a 0.3315, de la misma manera buscamos 0.3 y la columna .02 igual a .1255 sumando da la probabilidad 0.457.

Observe las tablas siguientes que muestran el área bajo la curva normal estándar, la primera de cero a 3.9=0.5, estos valores van del cero al valor buscado, en la parte positiva de la campana, y la segunda muestra el área desde -3.9 hasta el valor buscado, se puede usar indistintamente cualquier tabla, teniendo el cuidado de observar de cuál se trata. Los cálculos usando *geogebra* se muestran en las imágenes que acompañan cada ejercicio.



z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000



z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5754
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7258	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7518	0.7549
0.7	0.7580	0.7612	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7996	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

Ejemplo:

Un laboratorista trabaja con cuyos, sus estudios mostraron que tiene un promedio de vida de 50 meses cuando se ven expuestos a un exceso de stress, supongamos que la vida de los cuyos sigue una distribución normal con una desviación estándar de 5.9 meses, calcular la probabilidad de que un cuyo determinado viva:

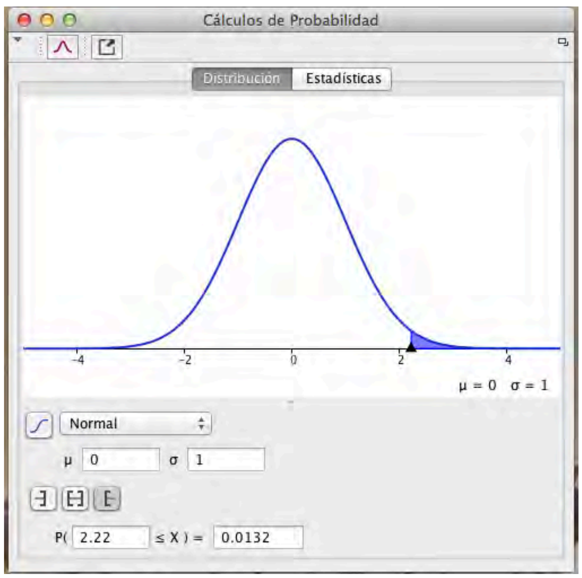
a) Más de 30 meses

Tenemos que $\mu = 50\ y\ \sigma = 9$

estandarizando

$$z = \frac{x-\mu}{\sigma} = \frac{30-50}{9} = \frac{20}{9} = 2.22$$

$$P(x \geq 30) = P(x \geq 2.2) = 0.5 - P(0 \leq x \leq 2.22) = 0.5 - 0.4868 = 0.0132$$



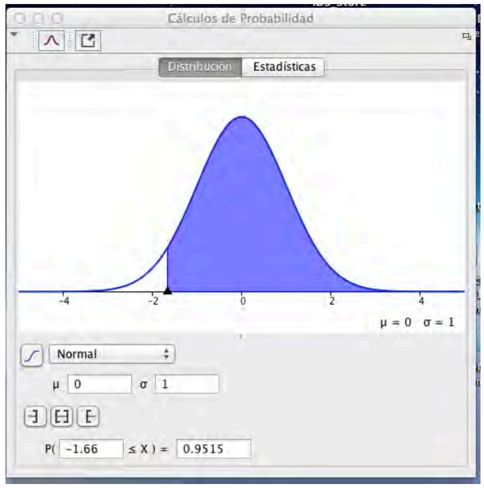
b) Entre 35 y 50 meses

estandarizando

$$z = \frac{x-\mu}{\sigma} = \frac{35-50}{9} = \frac{-15}{9} = -1.66$$

$$z = \frac{x-\mu}{\sigma} = \frac{50-50}{9} = \frac{0}{9} = 0$$

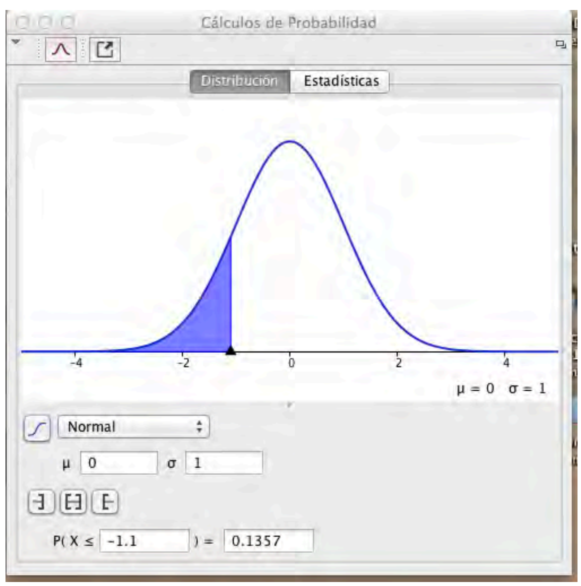
$$P(35 \leq x \leq 45) = P(-1.66 \leq x \leq 0) = 0.5 + P(0 \leq x \leq 1.66) = 0.5 + 0.4515 = 0.9515$$



c) Menos de 40 meses

$$z = \frac{x-\mu}{\sigma} = \frac{40-50}{9} = \frac{-10}{9} = -1.1$$

$$P(x \leq 40) = P(x \leq -1.1) = 0.5 - P(0 \leq x \leq 1.1) = 0.5 - 0.3643 = 0.1357$$



Un alumno de CONALEP va a la escuela en transporte urbano de su casa a la escuela, hace un promedio de 30 minutos en su recorrido, con una desviación estándar de 5 minutos, si suponemos que sigue una distribución normal el tiempo de su recorrido, calcular la probabilidad de que:

a) Su recorrido lo realice en 25 menos de minutos

Tenemos que $\mu = 30$ y $\sigma = 5$

queremos calcular $P(x < 25)$

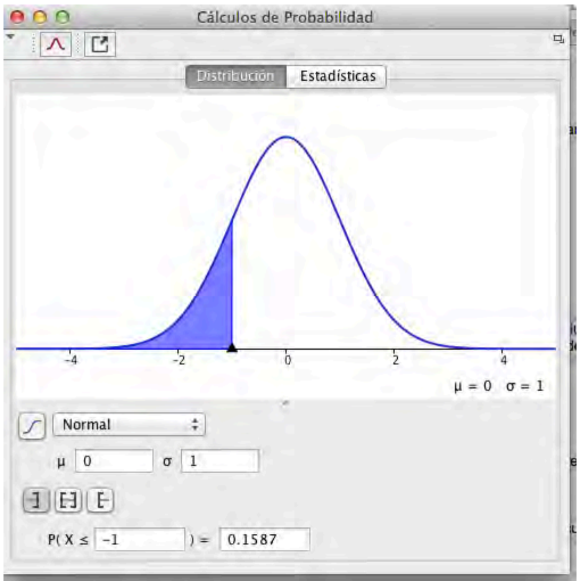
estandarizando

$$z = \frac{x-\mu}{\sigma} = \frac{25-30}{5} = \frac{-5}{5} = -1$$

esto es igual a:

$$P(x < 25) = P(x < -1) = 0.5 - P(0 < x < 1) =$$

$$0.5 - P(0 < x < 1) = 0.5 - 0.3413 = 0.1587$$



b) Su recorrido lo realice entre 20 y 25 minutos

queremos calcular $P(20 < x < 25)$

estandarizando

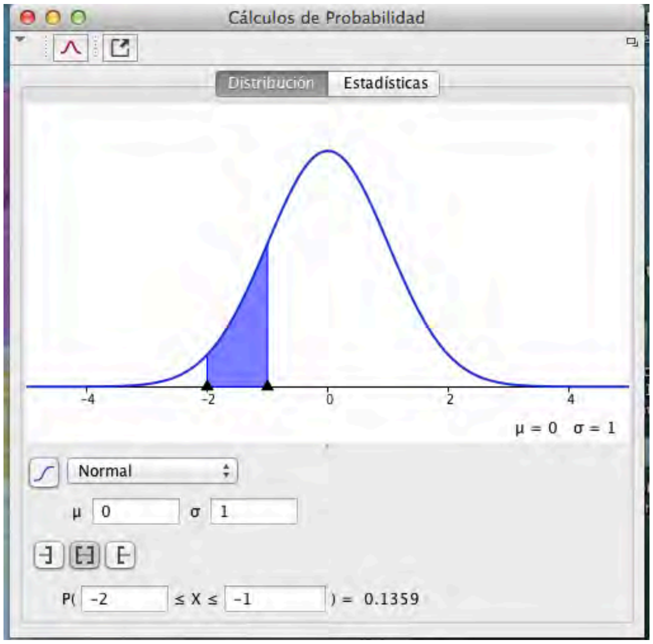
$$z = \frac{x-\mu}{\sigma} = \frac{20-30}{5} = \frac{-10}{5} = -2$$

$$z = \frac{x-\mu}{\sigma} = \frac{25-30}{5} = \frac{-5}{5} = -1$$

esto es igual a:

$$P(20 < x < 25) = P(-2 < x < -1) =$$

$$[0.5 - P(0 < x < 1)] - [P(0 < x < 2)]$$
$$= [0.5 - 0.3413] - [0.5 - 0.4772]$$
$$= 0.1587 - 0.0228 = 0.1359$$



Para resolver:

En parejas y utilizando las dos tablas para calcular la probabilidad de una distribución normal estándar, calcular la probabilidad de que el alumno del ejercicio anterior haga su recorrido en:

1) Un tiempo menor que 24 minutos

2) Entre 20 y 24 minutos

4.5.3. Distribución de Poisson

Otra distribución de probabilidad discreta es la distribución de Poisson, que sirve de modelo para experimentos en los que intervienen periodos de tiempo, como el número de pacientes que llegan en un determinado tiempo a una clínica, el número de llamadas que recibe una compañía, el número de clientes que recibe en ventanilla un banco, el número de glóbulos en un litro de sangre, etc..

La fórmula para calcular la distribución de Poisson se define como:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

donde $P(x)$= a la probabilidad de que x ocurra,

λ = promedio o valor esperado=varianza,

x = el valor de la variable aleatoria.

Propiedades de la distribución de Poisson[18]:

Media	$\mu = \lambda$
Varianza	$\sigma^2 = \lambda$
Desviación estándar	$\sigma = \sqrt{\lambda}$

Este tipo de distribución es una buena aproximación a una distribución binomial cuando x es muy pequeño, p pequeña y $\lambda = np$, como lo observamos en la siguiente tabla, con $n = 100$, $p = \frac{1}{100}$ y $\lambda = np = 1$

x	0	1	2	3	4	5	6
Poisson	0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005
Binomial	0.366	0.3697	0.1849	0.061	0.0149	0.0029	0.0005

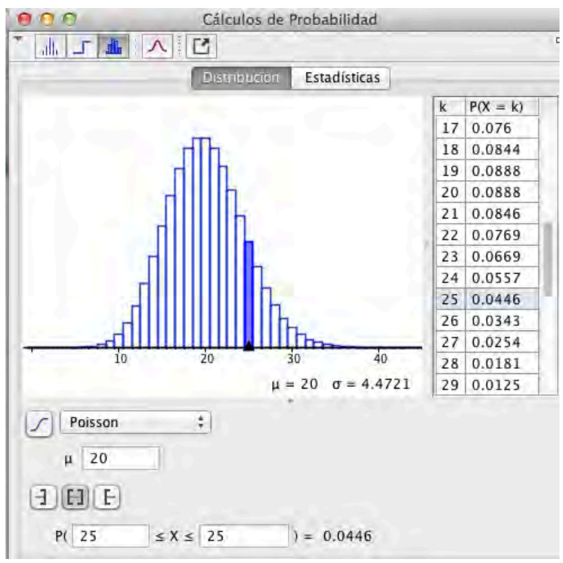
Ejemplo:

Un estacionamiento del centro de la ciudad de Morelia en su hora pico de 4 a 5 pm, recibe gran cantidad de automóviles, supongamos que siguen una distribución de Poisson con $\lambda = 20$. Calcular la probabilidad de:

a) El número de automóviles que reciban en la hora pico sea igual a 25

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$P(25) = \frac{20^{25} e^{-20}}{25!} = 0.0445$$



b) Lleguen entre 18 y 23 autos

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

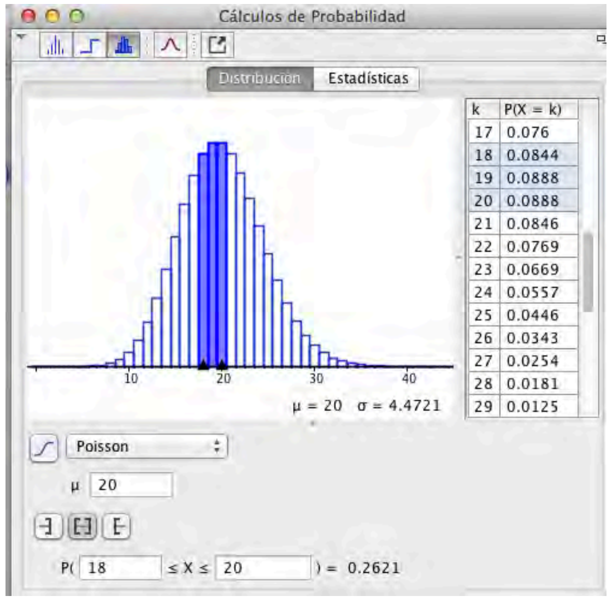
$$P(18 \leq x \leq 20) = P(18) + P(19) + P(20)$$

$$P(18) = \frac{(20)^{18}(e)^{-20}}{18!} = 0.0843$$

$$P(19) = \frac{(20)^{19}(e)^{-20}}{19!} = 0.08883$$

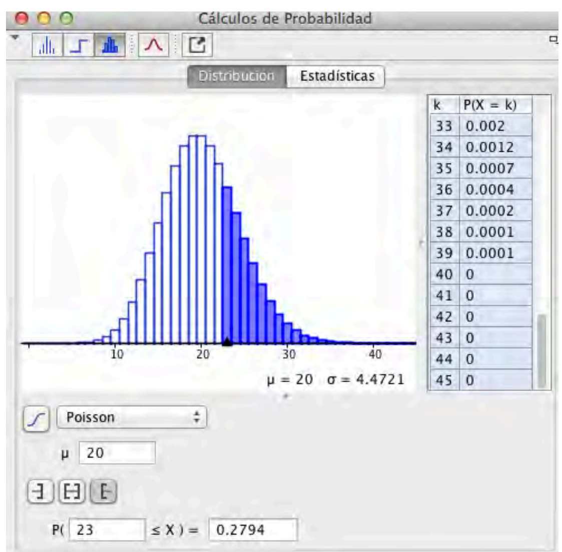
$$P(20) = \frac{(20)^{20}(e)^{-20}}{20!} = 0.08883$$

$$P(18 \leq x \leq 20) = 0.0843 + 0.08883 + 0.08883 = 0.2621$$



c) El número de automóviles que reciben en la hora pico sea mayor o igual a 23 autos.

$$P(x \geq 23) = 0.2794$$



Usando el programa geogebra puedes obtener los cálculos de la probabilidad así como la gráfica que indica la región calculada, te invitamos a descargarlo de la red, es gratuito y muy práctico, además ofrece tutoriales que facilitan su uso.

4.6. Problemario

1. Experimento: lanzar un dado normal y observar el número de puntos que sale en la cara superior, calcular la probabilidad de que caiga,

a) el 4

b) el 10

c) un número mayor o igual a 4

2. De una baraja inglesa (52 cartas) te reparten una mano de 5 cartas, calcula la probabilidad de que te salga:

a) una tercia

b) una corrida o escalera

c) una flor

d) full

e) un póker

f) una flor imperial

g) una "pancha"

3. Una caja contiene, 5 pequeñas pelotas de color rojo, 2 de color blancas. Se extraen dos esferas una tras otra sin regresar la primera que se extrajo, calcular la probabilidad de que las dos pelotas extraídas sean las blancas.

4. En un sorteo de lotería se hacen 100,000 boletos, si compras 5 boletos y solo un boleto es el premiado, calcula la probabilidad de ganar.

5. Se lanzar dos dados normales observando los punto de las caras superiores, ¿cuál es la probabilidad de que ambas caras sean iguales?

6. Se lanza un dado normal, calcular la probabilidad de que salga un número par o impar.

7. Un estudio realizado en México sobre padecimientos del corazón y la obesidad arrojó los siguientes resultados: mujeres de 40 años padecen del corazón un 15%, mujeres de la misma edad sufren de obesidad un 40% y mujeres de 40 años que padece obesidad y sobrepeso en un 4%. Calcular la probabilidad de que si escogemos una persona al azar y esta padece de obesidad , padezca del corazón.

8. De una bolsa con 5 esferas rosas, 2 moradas y 4 azules, se extraen 3 esferas una tras otra sin regresar la que se extrae. ¿Cuál es la probabilidad de sacar una esfera rosa, una morada y una azul?

9. Una escuela preparatoria estudió la probabilidad de que sus alumnos ingresen a una Universidad en particular y encontró que sus alumnos tiene un 65% de ingresar en la primera convocatoria, por su parte la Universidad recibe en cada licenciatura 100 alumnos. Calcular la probabilidad de que de un grupo de alumnos que presenten examen de admisión de dicha preparatoria 60 de ellos logran su ingreso. Suponga que siguen una distribución binomial.

10. Calcula la probabilidad de obtener 5 veces el número 2 al lanzar un dado 10 veces.

11. Las investigaciones del gen pelirrojo han mostrado que el nacimiento de un bebé pelirrojo es del 0.015 y siguen una distribución de Poisson. ¿Cuál es la probabilidad de que entre el nacimiento de 1000 recién nacidos haya 6 pelirrojos?

4.7. Autoevaluación

1. ¿Cuál es la probabilidad de que salga un número primo al lanzar un dado normal?

2. ¿Cuál es la probabilidad de que sumen 7 las caras superiores al lanzar un dado normal 2 veces?

3. ¿Cuál es la probabilidad de que te salgan dos pares en una baraja inglesa, si te entregan 6 cartas?

4. En una rifa se hacen 500 boletos, los premios son un auto de año, una bicicleta y un horno de microondas , ¿cuál es la probabilidad de que una persona se saque el auto?

5. Tres urnas contienen:

urna A, 3 votos para María y 5 para Laura

urna B, 2 votos para María y 1 para Laura

urna C, 2 votos para María y 3 para Laura

Si seccionamos la urna A y sacamos un voto. ¿Cuál es la probabilidad de que el voto elegido sea de Laura?

6. Se lanzan un par de dados normales, si los números que se muestran al caer son diferentes, hallar la probabilidad de que la suma sea un número par.

7. Se lanzan 6 monedas normales una vez, si llamamos éxito que caiga cara, ¿Cuál es la probabilidad de que salgan 2 caras?

8. En Michoacán la temperatura durante el mes de septiembre tiene una distribución normal, en promedio está en 28º y desviación estándar de 4º. Calcular la probabilidad de que la temperatura esté entre 30º y 40º

9. Una fábrica de fusibles conoce que el número de fusibles defectuosos es del 3%, calcular la probabilidad de que si tomamos una muestra de 150 fusibles haya 5 fusibles defectuosos. Como p es pequeña usar la distribución de Poisson.

4.8. Soluciones del problemario

1.

a) $P(4) = \frac{1}{6}$

b) $P(10) = 0\ evento\ imposible$

c) $P(x \geq 4) = \frac{3}{6} = \frac{1}{2}$

2.

a $P(tercia) = 0.0211$

b) $P(escalera) = 3.924646{\times}10^{-03}$

c) $P(flor) = 1.9654{\times}10^{-03}$

d) $P(full) = 1.4405{\times}10^{-03}$

e) $P(\text{póker}) = 2.400096{\times}10^{-04}$

f) $P(flor\ imperial) = 1.5390{\times}10^{-05}$

g) P("pancha")= 0.5011

3. $P(2\ blancas) = \left(\frac{2}{7}\right)\left(\frac{1}{6}\right) = \frac{2}{42} = \frac{1}{21} = 0.0476 \approx 4.8\%$

4. $P(ganar) = \frac{5}{100\,000} = 0.00005 \approx 0\%$

5. $P(2\ caras\ iguales) = P(11,22,33,44,55,66) = \frac{6}{36} = \frac{1}{6} = 0.1666 \approx 16.7\%$

6. Evento seguro:

$$A = \{2,4,6\} \quad P(par) = \frac{3}{6} = \frac{1}{2}$$

$$B = \{1,3,5\} \quad P(impar) = \frac{3}{6} = \frac{1}{2}$$

$$P(A \cup B) = \frac{1}{2} + \frac{1}{2} = 1 = 100\%$$

7.

A= mujeres de 40 años que padecen del corazón $P(A) = 15\% = 0.15$

B= mujeres de 40 años que padecen de obesidad $P(B) = 40\% = 0.40$

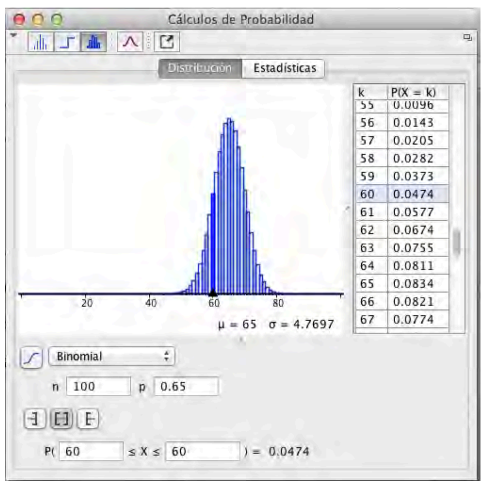
Una mujer padece obesidad y del corazón $P(A \cap B) = 4\% = 0.04$

$$P(B/A) = probabilidad\ de\ dado\ que\ padece\ obesidad\ padezca\ del\ corazón$$

$$P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.04}{0.15} = 0.2666 \approx 26.7\%$$

8. $P(rosa)P(morada)P(azul) = \frac{5}{11} \times \frac{2}{10} \times \frac{4}{9} = \frac{40}{990} = 0.04049 \approx 4\%$

9. $P(x = 60) = 0.0474$



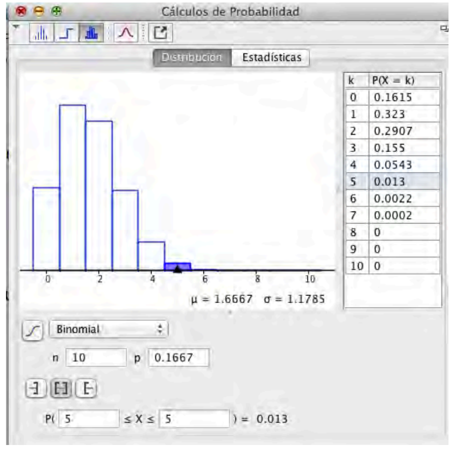
10.

$$n = 10$$

$$x = 5$$

$$p = \frac{1}{6} \therefore q = \frac{5}{6}$$

$$P(x = 5) = \frac{10!}{5!\,(10 - 5)!}\left(\frac{1}{6}\right)^5 \left(\frac{5}{6}\right)^5 = 0.13$$

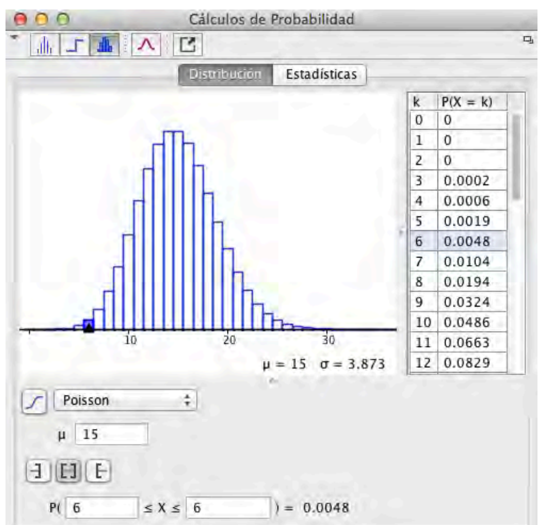


11.

$$\lambda = np = 1000(0.015) = 15$$

$$x = 6$$

$P(x = 6) = \frac{(15)^6 e^{-15}}{6!} = 0.004839 = 0.48\%$ de que sean bebés pelirrojos



4.9. Soluciones de autoevaluación

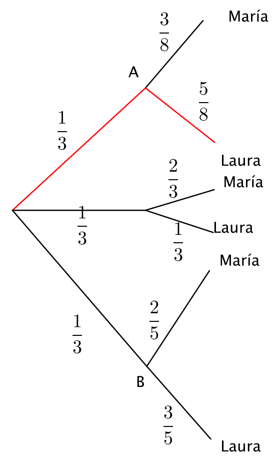
1. $P(primo) = P(2,3,5) = \frac{3}{6} = \frac{1}{2} = 50\%$

2. $P(1,6,2,5,3,4,4,3,5,2,6,1) = \frac{6}{36} = \frac{1}{6}$

3. $P(2pares) = 0.0000613$

4. $P(auto) = \frac{1}{500} = 0.002 \approx 0\%$

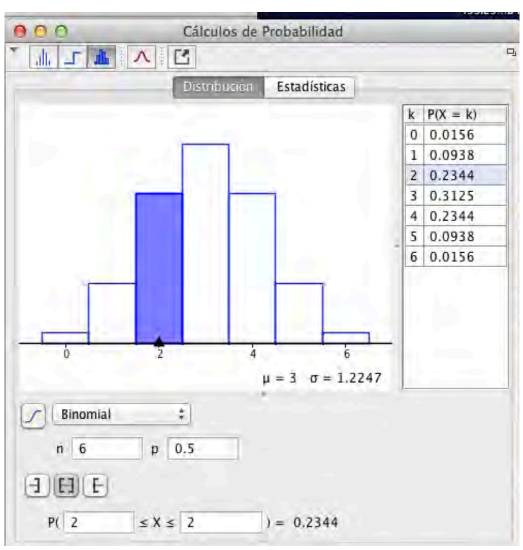
5. $P(M \,/\, A) = 0.2083$



6. $P = \frac{2}{5} = 0.4 \approx 40\%$

7. Tenemos que $n = 6$ y $p = \frac{1}{2}$ $\therefore$ $q = \frac{1}{2}$ y $x = 2$

$$P(x = 2) = \frac{6!}{2!\,(6-2)!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^4 = 0.2343$$



8. Estandarizando las unidades tenemos:

$$\mu = 28°, \sigma = 4°$$

$$z = \frac{x-\mu}{\sigma} = \frac{30° - 28°}{4°} = 0.5$$

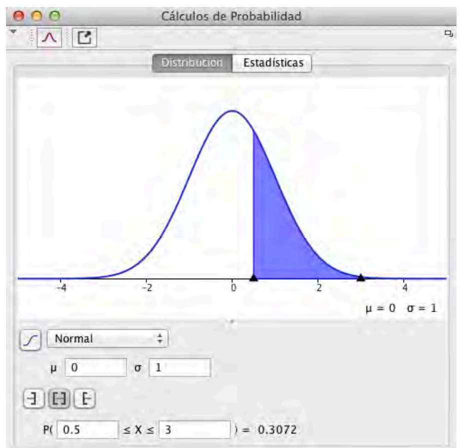
$$z = \frac{x-\mu}{\sigma} = \frac{40° - 28°}{4°} = 3$$

$$P(30° \leq X \leq 40°) = P(0.5 \leq X \leq 3) =$$

$$[0.5 - P(0 \leq x \leq 0.5)] - [0.5 - P(0 \leq x \leq 3)] =$$

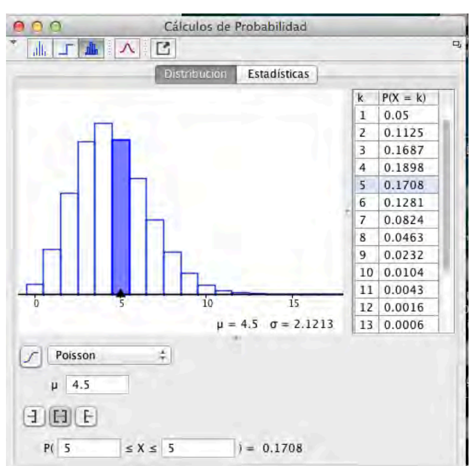
$$[0.5 - .1915] - [0.5 - 0.4987] =$$

$$0.3085 - 1.3 \times 10^{-03} = 0.3072$$



9. Tenemos que $\lambda = np = 150(.03) = 4.5$

$$P(x = 5) = \frac{(4.5)^5 (e)^{-4.5}}{5!} = 0.1708$$



4.10. Conclusión

La utilidad de la probabilidad es muy variada y extensa, esta es una breve introducción, que te sirve de apoyo para tus próximos estudios en nivel licenciatura, además que es una herramienta muy útil en la toma de decisiones, así como predecir los resultados en juegos de azar. Una herramienta muy útil en los cálculos probabilísticos y de las distribuciones de probabilidad es, usar programas que te ayudan a realizar los cálculos o verificar los resultados, geogebra es un programa muy útil y fácil de usar, con manuales sencillos.

Queda al lector seguir investigando el cálculo de probabilidades de otros tipos de eventos, así como otro tipo de distribuciones de probabilidad.

REFERENCIAS

[1] www.uam.es/personal_pdi/ciencias/barcelo/historia/Historia%20de%20la%20probabibilidad.pdf. Recuperado 19/07/2013
[2] Juez Pedro & Diez Javier (1997) Probabilidad y estadística en Medicina. España: Ediciones Díaz de Santo, S.A.
[3] Freund Max (2007) Lógica Jurídica. Costa Rica: Editorial Tecnológica de Costa Rica
[4] Miller Irwin (1973) Probabilidad y Estadística para ingenieros. España: Reverté
[5] Blanco Liliana (2004) Probabilidad. Colombia: Universidad Nacional de Colombia
[6] García Miguel (2008) Introducción a la teoría de la probabilidad. México: Fondo de Cultura Económica
[7] Wiley John (1992) Introducción a la teoría de probabilidades e inferencia estadística. México: LIMUSA
[8] Sánchez José (2006) Fundamentos para una teoría general de conjuntos. España: Editorial Crítica, S.L.
[9] Corona Rafael & Rodríguez Marisol et al. (2011) Manejo de espacios y cantidades. México: CONALEPMICH
[10] Pliego F. Javier & Ruiz-Maya Luis (2006) Fundamentos de Probabilidad. España: Thomson Editores
[11] García Miguel(2008) Introducción a la Teoría de La Probabilidad I. Primer Curso. México: Fondo de Cultura Económica
[12] Fornes ,Víctor(2009) Matemáticas: Unidades didácticas. 4º ESO. España: Gamma
[13] Cadoche liliane & Gastaldi Roque, et al. (2005) Elementos de matemática y estadística para las ciencias veterinarias. Argentina: Universidad Nacional del Litoral
[14] Sarabia José & Pascual Marta (2007) Curso básico de Estadística para Economía y Administración de Empresas. España: Publicaciones de la Universidad de Cantabria
[15] Hoel Paul (1990) Estadística Elemental. México: Compañía Editorial Continental, S.A. de C.V.
[16] Harold J. Larson (1992) Introducción a la teoría de probabilidades e inferencia estadística. México: LIMUSA, S.A. de C.V.
[17] Harris Daniel C. (2003) Análisis Químico Cuantitativo. España: Editorial Reverté
[18] Lipschutz Seymour (1982) Teoría y Problemas de Probabilidad. México: McGraw-Hill Book CO. S.A.