

Resumen sobre estadística y probabilidad.

Leandro Molina, Martin Borgo

29 de agosto de 2023



¹

¹Personaje perteneciente a [Twin-Sensei](#)

Índice

1. Capítulo 1. ¿Que es la estadística?	4
1.1. ¿Por qué se debe estudiar estadística?	4
1.2. ¿Que se entiende por estadística?	4
1.3. Tipos de estadística.	5
1.4. Tipos de variables.	6
1.5. Niveles de medición.	6
2. Capítulo 2. Descripción de datos: medidas numéricas.	9
2.1. Introducción	9
2.2. La media poblacional	9
2.3. Media de una muestra	10
2.4. Propiedades de la media aritmética	10
2.5. Media ponderada	11
2.6. Mediana	11
2.7. Moda	12
2.8. Posiciones relativas de la media, la mediana y la moda.	12
2.9. ¿Por que estudiar la dispersión?	13
2.10. Medidas de dispersión	13
2.11. Rango intercuartílico o rango intercuartil	13
2.12. Interpretación y usos de la desviación estándar	15
2.13. Coeficiente de variación	15
2.14. Media y desviación estándar de datos agrupados	16
2.15. Ética e informe de resultados	17
3. Capítulo 3. Descripción de datos	18
3.1. Sesgo	18
3.2. Percentiles, deciles y cuartiles	19
3.3. Diagramas de caja	20
4. Capítulo 4. Probabilidad: Conceptos, Variables Aleatorias y Distribuciones.	22
4.1. ¿Que es la probabilidad?	22
4.2. Enfoques para asignar probabilidades.	23
4.3. Reglas para calcular probabilidades	25
4.4. Principios de conteo	28
4.5. Teorema de Bayes	30
4.6. Análisis de Variables Aleatorias	33
4.6.1. Distribución Binomial de Probabilidad	33
4.6.2. Distribución de Probabilidad de Poisson	35
4.6.3. Distribución Hipergeométrica de Probabilidad	36
4.7. Distribuciones de probabilidad continuas	37

5. Capitulo 5. Planes Muestrales y Diseño Experimental	39
5.1. Estadística y Distribuciones Muestrales	41
5.1.1. Teorema del Límite Central	42
5.1.2. La Distribución Muestral de la Media Muestral	43
5.1.3. La Distribución Muestral de la Proporción Muestral	43

1. Capitulo 1. ¿Que es la estadística?

1.1. ¿Por qué se debe estudiar estadística?

Hay 3 motivos para el estudio de la estadística estos son:

1. La primera razón consiste en que la información numérica prolifera por todas partes. si revisas diarios o revistas contienen mucha cantidad de información numérica.
2. Una segunda razón, es que las técnicas de la estadística se emplean para tomar decisiones que afectan la vida diaria, es decir, que incluyen en su bienestar.
3. Una tercera razón, el conocimiento de sus métodos facilita la comprensión de la forma en que se toman las decisiones y proporciona un entendimiento mas claro de como le afectan.

Al encarar la necesidad de tomar decisiones en las que tenes que saber hacer un análisis de datos resultara de utilidad. Con el fin de tomar una decisión informada, sera necesario llevar a cabo lo siguiente para poder tomar una decisión informada:

1. Determinar si existe información adecuada o si requiere información adicional.
2. Reunir información adicional, si se necesita, de manera que no se obtengan resultados erróneos.
3. Resumir los datos de manera útil e informativa.
4. Analizar la información disponible.
5. Obtener conclusiones y hacer inferencias al mismo tiempo que se evaluá el riesgo de tomar una decisión incorrecta.

En resumen hay por lo menos tres razones para estudiar estadística: 1) los datos proliferan por todas partes; 2) las técnicas estadísticas se emplean en la toma de decisiones que influyen en su vida; 3) sin que importe la carrera que elija, tomara decisiones profesionales que incluyan datos.

1.2. ¿Que se entiende por estadística?

Posee dos significados: su aceptación más común, la estadística se refiere a información numérica. Una colección de información numérica recibe el nombre de **estadísticas**. La información estadística se presenta en forma gráfica, es útil porque capta la atención del lector e incluye una gran cantidad de información.

Estadística: Ciencia que recoge, organiza, presenta, analiza e interpreta datos con el fin de propiciar una toma de decisiones mas eficaz.

El primer paso en el estudio de un problema consiste en recoger datos relevantes. Estos deben organizarse de alguna forma y, tal vez, representarse en una gráfica.

1.3. Tipos de estadística.

El estudio de la estadística se divide en dos categorías: la estadística descriptiva y la estadística inferencial.

Estadística descriptiva.

Es la ciencia que recoge, organiza, presenta, analiza...datos". Esta parte de la estadística recibe el nombre de **estadística descriptiva**.

Estadística descriptiva: Métodos para organizar, resumir y presentar datos de manera informativa.

Se trata de estadística descriptiva si calcula el crecimiento porcentual de una década a otra. Sin embargo, no sería de naturaleza descriptiva si utiliza estos para el calcular con esos datos algo futuro. Una masa de datos desorganizados resulta de poca utilidad. Las técnicas de la estadística descriptiva permiten organizar esta clase de datos y darles significado. Los datos se ordenan en una **distribución de frecuencia** (mas adelante lo veremos). Se emplean diversas clases de **graficas** para describir datos.

Estadística inferencial.

La estadística inferencial, también denominada **inferencia estadística**. El principal interés que despierta esta disciplina se relaciona con encontrar algo relacionado con una población a partir de una muestra de ella. Ya que estas son inferencias relacionadas con una población, basadas en datos de la muestra, se trata de estadística inferencial. Se podría considerar a la estadística inferencial como la mejor conjetura que es posible obtener del valor de una población sobre la base de la información de una muestra.

Estadística inferencial: Métodos que se emplean para determinar una propiedad de una **población** con base en la información de una **muestra** de ella.

Atención a las palabras población y muestra en la definición de estadística inferencial. Una **población** puede constar de individuos, también puede consistir en objetos. Desde una perspectiva estadística, una población no siempre que tiene que ver con personas.

Población: Conjunto de individuos u objetos de interés o medidas que se obtienen a partir de todos los medios u objetos de interés.

Con el objeto de inferir algo sobre una población, lo común es que se tome una muestra de ella.

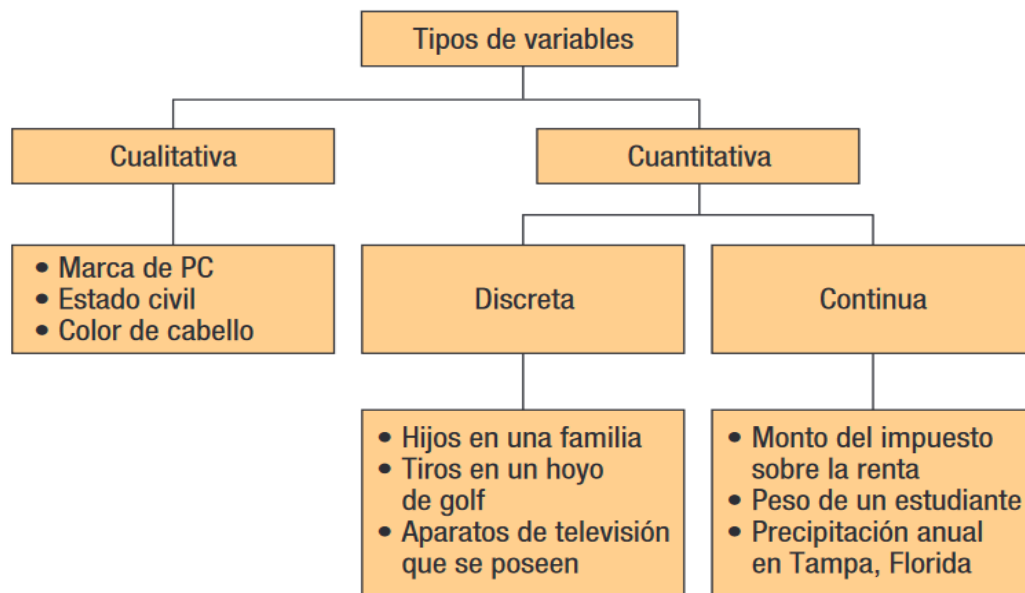
Muestra: Porción o parte de la población de interés.

La toma de muestras para aprender algo sobre una población es de uso frecuente en administración, agricultura, política y acciones de gobierno.

1.4. Tipos de variables.

Una variable es una característica observable en las unidades estadísticas y tiene, por lo menos, dos valores.

Dos tipos básicos de variables: 1)Cualitativas y 2)Cuantitativas, la característica que se estudia es de naturaleza no numérica, recibe el nombre de **variable cualitativa** o **atributo**. Cuando los datos son de naturaleza cualitativa, importa la cantidad o proporción que caen dentro de cada categoría. Los datos cualitativos se resumen en tablas o graficas de barras. Cuando la variable que se estudia aparece en forma numérica, se le denomina **variable cuantitativa**. Las variables cuantitativas pueden ser discretas o continuas. Las **variables discretas** adoptan solo ciertos valores y existen vacíos entre ellos. Las variables discretas son el resultados de una relación numérica, las observaciones de una **variable continua** toman cualquier valor dentro de un intervalo específico. Por lo general las variables continuas son el resultado de mediciones. Resumen de los tipos de variables:



1.5. Niveles de medición.

Los datos se clasifican por niveles de medición. El nivel de medición de los datos rige los cálculos que se llevan a cabo con el fin de resumir y presentar los datos. También determina las pruebas estadísticas que se deben realizar. De hecho, existen cuatro niveles de medición: nominal, ordinal, de intervalo y de razón. La medición mas baja, o mas primaria, corresponde al nivel nominal. La mas alta, o el nivel que proporciona la mayor información relacionada con la observación, es la medición de razón.

Datos de nivel nominal.

Las observaciones acerca de una variable cualitativa solo se clasifican y se cuentan. No existe una forma particular para ordenar las etiquetas, no existe un orden natural. Para el nivel nominal, la medición consiste en contar, a veces, para una mejor comprensión de lectura, estos conteos se convierten en porcentajes. Es necesario hacer que el porcentaje sume un total de 100 %, no existe un orden natural para los resultados. Para procesar datos a menudo se codifica la información en forma numérica. El nivel nominal tiene las siguientes propiedades:

1. La variable de interés se divide en categorías o resultados.
2. No existe un orden natural de los resultados.
3. Los datos solo se clasifican.

Datos de nivel ordinal.

El nivel inmediato superior de datos es el **nivel ordinal**. No es posible distinguir la magnitud de las diferencias entre los grupos, ¿la diferencia entre superior y bueno es la misma que entre lo malo e inferior? No es posible afirmarlo. Las propiedades del nivel ordinal de los datos son las siguientes:

1. Las clasificaciones de los datos se encuentran representadas por conjuntos de etiquetas o nombre (alto, medio, bajo), las cuales tienen valores relativos.
2. En consecuencia, los valores relativos de los datos se pueden clasificar u ordenar.

Datos de nivel de intervalo.

El nivel de intervalo de medición es el nivel inmediato superior. Incluye todas las características de nivel ordinal, pero, además, la diferencia entre valores constituye una magnitud constante. Si las distancias entre los números tienen sentido, aunque las razones no, entonces tiene una escala de intervalo de medición. Las propiedades de los datos de nivel intervalo son las siguientes:

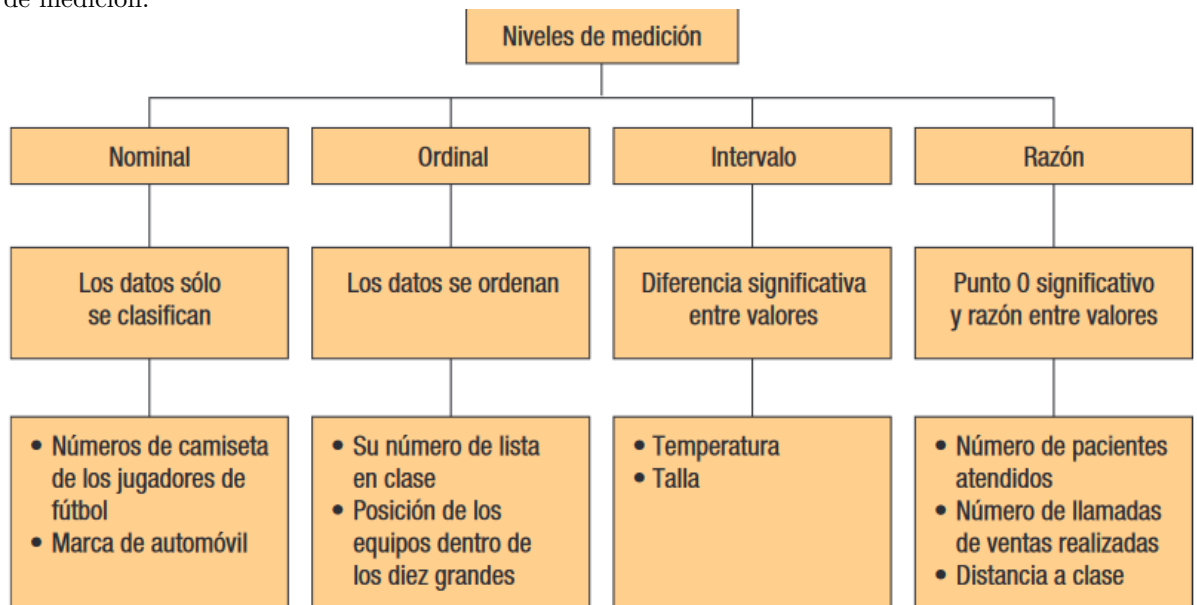
1. Las clasificaciones de datos se ordenan de acuerdo con el grado que posea de las característica en cuestión.
2. Diferencias iguales en la característica representan diferencias iguales en las mediciones (es decir la diferencia entre valores es significativa).
3. El cero es relativo (no indica ausencia de estado).

Datos de nivel de razón.

Todos los datos cuantitativos son registrados en el nivel de razón de la medición, el **nivel de razón** es el *mas alto*. Posee todas las características del nivel de intervalo, aunque, además, el punto 0 tiene sentido y la razón entre dos números significativa, si se encuentra en 0 significa la ausencia de algo (peso, dinero, etc). Las propiedades de los datos de nivel intervalo son las siguientes:

1. Las clasificaciones de datos se ordenan de acuerdo con la cantidad de características que poseen.
2. Diferencias iguales en la característica representan diferencias iguales en los números asignados a las clasificaciones.
3. El punto cero representa la ausencia de características y la razón entre dos números es significativa.
4. La razón entre valores es significativa.

La siguiente grafica resume las principales características de los diversos niveles de medición.



2. Capitulo 2. Descripción de datos: medidas numéricas.

2.1. Introducción

En este capítulo se presentan dos formas numéricas de describir datos cuantitativos: las **medidas de ubicación**² y las **medidas de dispersión**. A las medidas de ubicación a menudo se les llama **promedios**. El propósito de una medida de ubicación consiste en señalar el centro de un conjunto de valores. Vos ya estas conoces el concepto de promedio (media aritmética), medida de ubicación que muestra el valor central de los datos. Si solo tomas las medidas de ubicación en cuenta de un conjunto de datos o si compara varios conjuntos de datos utilizando valores centrales, llegara a una conclusión incorrecta. Además de las medidas de ubicación, debe tomar en consideración la **dispersión** (con frecuencia se le llama *frecuencia variación* o *propagación*) de los datos. Para describir la dispersión considere el rango, la desviación media, la varianza y la desviación estándar. En principio se explican las medidas de ubicación. No existe una única medida de dispersión; de hecho existen varias. Consideraremos cinco:

1. La media aritmética.
2. La media ponderada.
3. La mediana.
4. La moda.
5. La media geométrica.

La medida aritmética es la medida de ubicación que mas se utiliza y que se publica con mayor frecuencia, por lo cual se le considerará como parámetro para una población y como estadístico para las muestras.

2.2. La media poblacional

La media poblacional es la suma de todos los valores observados en la población dividida entre el numero de valores de la población. Para determinar la media poblacional, aplique la siguiente formula con la siguiente formula y formula con símbolos matemáticos:

$$\mu = \frac{\sum X}{N}$$

²La razón por la cual se llama *medida de ubicación* es porque estas medidas estadísticas representan la posición o ubicación de los datos dentro de un conjunto de datos. Es decir, indican dónde se encuentra la mayoría de los valores de un conjunto de datos y cómo están distribuidos alrededor de esta posición central. Por ejemplo, la media aritmética es una medida de ubicación porque representa el valor central alrededor del cual se agrupan los demás valores en un conjunto de datos. La mediana, por otro lado, es una medida de ubicación porque representa el valor que divide el conjunto de datos en dos partes iguales, es decir, la mitad de los valores están por encima de la mediana y la otra mitad están por debajo.

Cada símbolo representa:

- μ representa la media poblacional; se trata de la letra minúscula griega mu.
- N es la numera de valores en la población.
- X representa cualquier valor particular.
- \sum es la letra mayúscula griega sigma e indica la operación de suma.
- $\sum X$ es la suma de X valores en la población.

Cualquier característica medible de una población recibe el nombre de **parámetro**. La media de una población es un parámetro.

Parámetro: Característica de una población.

2.3. Media de una muestra

Con frecuencia se selecciona una muestra (Se menciona en el capítulo 1) de la población para estimar una característica específica de la población. La *media* es la suma de los valores de la muestra, divididos entre el número total de valores de la muestra. La media de una muestra se determina de la siguiente manera:

$$\bar{X} = \frac{\sum X}{n}$$

Cada símbolo representa:

- \bar{X} es la media de la muestra; se lee como "X barra."
- n es el número de valores de la muestra.
- X representa cualquier valor particular.
- \sum es la letra mayúscula griega sigma e indica la operación de suma.
- $\sum X$ es la suma de X valores en la población.

La media de una muestra o cualquier otra medición basada en una muestra de datos recibe el nombre de **estadístico**.

Estadístico: Característica de una muestra.

2.4. Propiedades de la media aritmética

La media aritmética es una medida de ubicación muy utilizada. Cuenta con algunas propiedades importantes:

1. **Todo conjunto de datos de intervalo -o de nivel de razón- posee una media.** Recordá que en el capítulo 1 los datos de nivel de razón incluyen datos como edades, ingresos y pesos, y que la distancia entre los números es constante.

2. **Todos los valores se encuentran incluidos en el calculo de la media.**
3. **La media es única.** Solo existe una media en un conjunto de datos. Mas adelante en el capitulo que un promedio que podría aparecer dos o mas veces en un conjunto de datos.
4. **La suma de las desviaciones de cada valor de la media es cero.** La media es un punto equilibrio de un conjunto de datos.

La media tiene un punto débil. Recuerde que el valor de cada elemento de una muestra o población, se utiliza cuando se calcula la media. Si uno de estos valores son extremadamente grandes o pequeños comparados con la mayoría de los datos, la media podría no ser un promedio adecuado para representar los datos. Es decir, la media se ve afectada en exceso por valores inusualmente grandes o pequeños.

2.5. Media ponderada

La media ponderada, que constituye un caso especial de la media aritmética, se presenta cuando hay varias observaciones con el mismo valor. Es una medida estadística que se utiliza para calcular el valor medio de un conjunto de datos, donde cada valor tiene un peso o una importancia específica. En la media ponderada, cada valor se multiplica por su peso y luego se divide por la suma de todos los pesos.

$$\overline{X_w} = \frac{\sum(wX)}{\sum w}$$

Observe que el denominador de una medida ponderada siempre es la suma de las ponderaciones.

2.6. Mediana

Ya se dijo que si los datos contienen uno o dos valores muy grandes o muy pequeños, la media aritmética no resulta representativa. Es posible describir el centro de dichos datos a partir de una medida de ubicación denominada **mediana**.

Mediana: Punto medio de los valores una vez que se han ordenado de menor a mayor o de mayor a menor.

Ordenas los datos y agarras el(o los) valor (valores) del medio (literalmente). ¿Como se determina la mediana en el caso de un numero par de observaciones? se ordenan las observaciones. En seguida, con el fin de obtener un único valor por convención, calcule la media de las dos observaciones medias. Así, en el caso de un numero par de observaciones, la mediana quizá no sea uno de los valores dados. Ejemplo ponele que tenes dos observaciones medias 5 y 7, la media de esos dos es 6. Las principales propiedades de la mediana son las siguientes:

- **No influyen en ella valores extremadamente grandes o pequeños.**
Por consiguiente, la mediana es una valiosa medida de ubicación cuando dichos valores se presentan.
- **Es calculable en el caso de datos de nivel ordinal o mas altos.** Recordá que en el capítulo 1 que los datos de nivel ordinal pueden ordenarse de menor a mayor

La mediana se determina para cualquier nivel de datos, excepto los nominales.

2.7. Moda

La moda es otra medida de ubicación (promedio).

Moda: Valor de la observación que aparece con mayor frecuencia.

La moda es de especial utilidad para resumir datos de nivel nominal. Es posible determinar la moda para todos los niveles de datos: nominal, ordinal, de intervalo y de razón. La moda también tiene la ventaja de que no influyen en ella valores extremadamente grandes o pequeños. No obstante, la moda tiene sus desventajas, por las cuales se le utiliza con menor frecuencias que a la media o mediana. En el caso de muchos conjuntos de datos no existe la moda porque ningún valor se presenta mas de una vez. Puede ocurrir que un agrupamiento de datos tenga dos modas se denomina bi-modal (tiene dos modas).

2.8. Posiciones relativas de la media, la mediana y la moda.

Se trata de una distribución simétrica que también tiene forma de campana. Esta distribución *posee* la misma forma a cualquier lado del centro. Si el polígono estuviera doblado a la mitad, las dos mitades serian idénticas. En cualquier distribución simétrica, la moda, la mediana, y la media siempre son iguales. Hay distribuciones simétricas que no tienen forma de campana (En una distribución en forma de campana la media, la mediana y la moda son iguales). Si una distribución no es simétrica, o **sesgada**, (Una distribución sesgada no es simétrica.) la relación entre las tres medidas cambia. En una distribución con sesgo positivo la media aritmética es la mayor de las tres medidas. ¿Por que? Porque en ella influyen, mas que sobre la mediana o la moda, unos cuantos valores extremadamente altos. Por lo general, la mediana es la siguiente medida mas grande en una distribución de frecuencias con sesgo positivo. La moda es la menor de las tres medidas. Si la distribución tiene un sesgo muy pronunciado, la media no seria una mediana adecuada. La mediana y la moda serian mas representativas.

Por el contrario, si una distribución tiene un **sesgo negativo**, la media es la menor medida de las tres. Por supuesto, la media es sensible a la influencia de una cantidad extremadamente pequeña de observaciones. La mediana es mayor

que la media aritmética y la moda es la mas grande de las tres medidas. De nuevo, si la distribución tiene un sesgo muy pronunciado, la media no se utilizaría para representar a los datos.

2.9. ¿Por que estudiar la dispersión?

Una medida de ubicación, como la media o la mediana, solo describe el centro de los datos. Desde este punto de vista resulta valiosa, pero no dice nada sobre la dispersión de los datos. Una medida de dispersión pequeña indica que los datos se acumulan con proximidad alrededor de la media aritmética. Por consiguiente, la media se considera representativa de los datos. Por el contrario, una medida grande de dispersión indica que la media no es confiable. Otra razón para estudiar la dispersión en un conjunto de datos consiste en comparar la propagación en dos o mas distribuciones. Una medida de dispersión sirve para evaluar la confiabilidad de dos o mas medidas de ubicación.

2.10. Medidas de dispersión

Consideraremos diversas medidas de dispersión. El rango se sustenta en los valores máximo y mínimo del conjunto de datos, es decir, solo se consideran dos valores. La desviación media, la varianza y la desviación estándar se basan en desviaciones de la media aritmética.

Rango

La medida mas simple de dispersión es el **rango**. Representa la diferencia entre los valores máximo y mínimo de un conjunto de datos. En forma de ecuación:

$$\text{Rango} = \text{Valor máximo} - \text{valor mínimo}$$

El rango se emplea mucho en aplicaciones de control de procesos estadísticos (CPE), debido a que resulta fácil de calcular y entender.

2.11. Rango intercuartílico o rango intercuartil

El rango intercuartílico o rango intercuartil (RIC o IQR), se obtiene como la diferencia entre los cuartiles superior e inferior: $\text{RIC} = Q_3 - Q_1$ mide el rango del 50 % central de la distribución. Gráficamente, está asociado a la “caja” del diagrama de caja.

Desviación media

Un problema que presenta el rango escriba en que parte de dos valores, el mas alto y el mas bajo, es decir, no los toma en cuenta a todos. La desviación media si lo hace; mide la cantidad media respecto de la cual los valores de una población o muestran varían. Expresado en forma de definición:

Desviación media: Media aritmética de los valores absolutos de las desviaciones con respecto a la media aritmética.

En el caso de una muestra, la desviación media, designada DM, se calcula mediante la formula:

$$DM = \frac{\sum |X - \bar{X}|}{n}$$

en donde:

- X es el valor de cada observación.
- \bar{X} es la media aritmética de los valores.
- n es el numero de observaciones en la muestra.
- $||$ indica el valor absoluto.

¿Por que ignorar los signos de las desviaciones de la media? De no hacerlo, las desviaciones positivas y negativas se compensarían con exactitud unas a otras y la desviación media siempre seria cero. Dicha medida (cero) resultaría un estadístico sin utilidad.

La desviación media posee dos ventajas, primero, incluye todos los valores de los cálculos. Recuerde que el rango solo incluye los valores máximo y mínimo. Segundo, es fácil de definir: es la cantidad promedio que los valores se desvían de la media. Sin embargo, su inconveniente es el empleo de valores absolutos. Por lo general, es difícil trabajar con valores absolutos, así que la desviación media no se emplea con tanta frecuencia como otras medidas de dispersión, como la desviación estándar.

Varianza y desviación estándar

La **varianza** y la **desviación estándar** también se fundamentan en las desviaciones de la media. Sin embargo, en lugar de trabajar con el valor absoluto de las desviaciones, la varianza y la desviación estándar lo hacen con el cuadrado de las desviaciones.

Varianza: Media aritmética de las desviaciones de la media elevadas al cuadrado.

La varianza es no negativa y es cero solo si todas las observaciones son las mismas. El principal problema de la varianza es que se mide el términos del cuadrado de las unidades originales de medición. Por ejemplo:

Si las mediciones están dadas en metros, entonces la varianza queda expresada en metros cuadrados. Tomando la raíz cuadrada de la varianza, obtenemos la desviación estándar, que regresa la medida de variabilidad a la unidades de medición originales

Desviación estándar: Raíz cuadrada de la varianza.

La **desviación estándar** de un conjunto de mediciones es igual a la raíz cuadrada de la varianza.

Desviación estándar poblacional: $\delta = \sqrt{\delta^2}$

Desviación estándar muestral: $s = \sqrt{s^2}$

2.12. Interpretación y usos de la desviación estándar

La desviación estándar normalmente se utiliza como medida para comparar la dispersión de dos o mas conjuntos de observaciones.

2.13. Coeficiente de variación

El coeficiente de variación (CV) nos informa acerca de la dispersión relativa de un conjunto de datos o mediciones.

Caso mediciones muestrales: $CV = \frac{S}{|\bar{x}|}$

Caso mediciones poblacionales: $CV = \frac{\delta}{|\mu|}$

Principales características del CV:

1. El coeficiente de variación no posee unidades, es decir es adimensional.
2. Para su interpretación se puede expresar como porcentaje, teniendo en cuenta que puede superar el 100 %.
3. Se utiliza para comparar conjuntos de datos pertenecientes a poblaciones distintas.
4. Cuanto más grande es el coeficiente de variación más dispersos están los datos respecto de la media, por lo que la media va perdiendo representatividad.

Teorema de Chebyshev

Ya se ha insistido en el hecho de que una desviación estándar pequeña de un conjunto de valores indica que estos se localizan cerca de la media. Por el contrario, una desviación grande revela que las observaciones se encuentran muy dispersas con respecto a la media. El pibe Chebyshev estableció un teorema que nos permite determinar la mínima porción de valores que se encuentran a cierta cantidad de desviaciones estándares de la media. El teorema del Cheby establece lo siguiente:

Teorema de Chebyshev: En cualquier conjunto de observaciones (muestra o población), la proporción de valores que se encuentran a k desviaciones estándares de la media es de por lo menos $1 - 1/k^2$, siendo k cualquier constante mayor que 1.

La regla empírica

El teorema Cheby se relaciona con cualquier conjunto de valores; es decir, que la distribución de valores puede tener cierta forma. Sin embargo, en cualquier distribución simétrica con forma de campana, es posible ser mas precisos en la explicación de la dispersión en torno a la media. Estas relaciones que implican la desviación estándar y la media se encuentran descritas en la **regla empírica**, a veces denominada **regla normal**. La regla empírica solo se aplica a distribuciones simétricas con forma de campana.

Regla empírica: En cualquier distribución de frecuencias simétrica con forma de campana, aproximadamente 68 % de las observaciones se encontraran entre mas y menos una desviación estándar de la media; cerca de 95 % de las observaciones se encontraran entre mas y menos dos desviaciones estándares de la media y, de hecho (99.7 %), estarán mas y menos tres desviaciones de la media.

Se ha observado que si una distribución es simétrica y tiene forma de campana, todas las observaciones se encuentran entre la media y menos tres desviaciones estándares.

2.14. Media y desviación estándar de datos agrupados

En la mayoría de los casos las medidas de ubicación, como la media y las medidas de dispersión, como la desviación estándar, se determinan utilizando valores individuales. Sin embargo, algunas veces solo se cuenta con la distribución de frecuencias y se desea calcular la media o la desviación estándar. La siguiente explicación explica como calcular la media y la desviación estándar a partir de datos organizados en una distribución de frecuencias. Hay que insistir en que una media o una desviación estándar de datos agrupados es una estimación de los valores reales correspondientes.

Media aritmética

Para aproximar la media aritmética de datos organizados en una distribución de frecuencia, comience suponiendo que las observaciones en cada clase se representan a través del punto medio de la clase. La media de una muestra de datos organizados en una distribución de frecuencias se calcula de la siguiente manera:

Media aritmética de datos agrupados:

$$\bar{X} = \frac{\sum fM}{n}$$

donde:

\bar{X} designa la media muestra.

M es el punto medio de cada clase.

f es la frecuencia en cada clase.

fM es la frecuencia en cada clase multiplicada por el punto medio de la clase.

$\sum fM$ es la suma de estos productos.

n es el numero total de frecuencias.

Desviación estándar

Para calcular la desviación estándar de datos agrupados en una distribución de frecuencias, necesita ajustar ligeramente la formula. Pondere cada una de las diferencias cuadradas por el numero de frecuencias en cada clase. La formula de la **Desviación estándar, datos agrupados** es:

$$s = \sqrt{\frac{\sum f(M - \bar{X})^2}{n - 1}}$$

donde:

s es el símbolo de la desviación estándar de la muestra.

M es el punto medio de la clase.

f es la frecuencia de clase.

n es el numero de observaciones en la muestra.

\bar{X} designa la media muestra.

2.15. Ética e informe de resultados

Esta aprendiendo a organizar, resumir e interpretar datos mediante la estadística, también es importante que comprenda esta disciplina con el fin de que se convierta en un consumidor inteligente de información. En este capítulo aprendió la forma de calcular estadísticas descriptivas de naturaleza numérica. En particular, la manera de calcular e interpretar medidas de ubicación de un conjunto de datos: la media, la mediana y la moda. También ha estudiado las ventajas y desventajas de cada estadístico. Conocer las ventajas y desventajas de la media, la mediana y la moda es importante al dar un informe estadístico y cuando se emplea información estadística para tomar decisiones. También aprendió a calcular medidas de dispersión: el rango, la desviación media y la desviación estándar. Cada uno de estos estadísticos tiene ventajas y desventajas.

Recuerde que el rango proporciona información sobre la dispersión total de una distribución. Sin embargo, no aporta información sobre la forma en que se acumulan los datos o se concentran entorno al centro de la distribución. Conforme aprenda más estadística, necesitará recordar que cuando emplea esta disciplina debe mantener un punto de vista independiente y basado en principios. Cualquier informe estadístico requiere la comunicación honesta y objetiva de los resultados.

3. Capítulo 3. Descripción de datos

En este capítulo continua el estudio de la estadística descriptiva. Estos diagramas y la estadística proporcionan una idea adicional del lugar en el que los valores se concentran, así como de la forma general de los datos. En seguida se consideran datos bivariados de cada una de las observaciones individuales o seleccionadas.

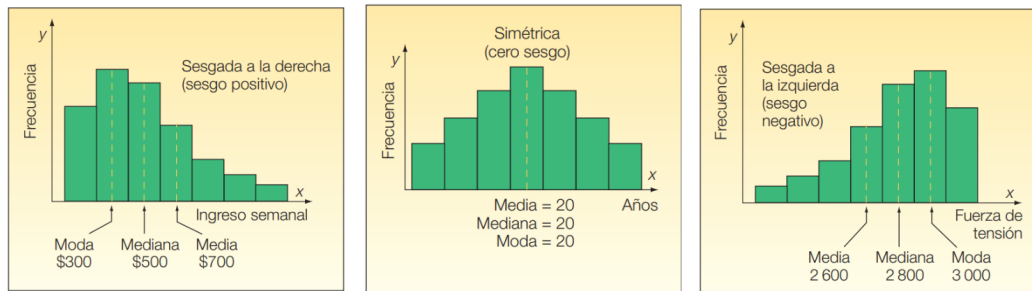
Parámetros: medidas descriptivas numéricas calculadas a partir de **mediciones poblacionales**. Por ejemplo, la media aritmética o promedio de una población, es decir, la media poblacional.

Estadísticos (o estadísticas): medidas descriptivas numéricas calculadas a partir de **mediciones muestrales**. Por ejemplo, la media aritmética o promedio de una muestra, es decir, la media muestral

Los parámetros se designan con **letras griegas** y los estadísticos con **letras latinas**, a veces con algún añadido como el caso de la media (con una barra arriba de la letra).

3.1. Sesgo

Hay cuatro formas: simétrica, con sesgo positivo, con sesgo negativo y bimodal. En un conjunto **simétrico** de observaciones la media y mediana son iguales, y los valores de datos se dispersan uniformemente en torno a estos valores. Un conjunto de valores se encuentra **sesgado a la derecha** o **positivamente sesgado** o **sesgado a la izquierda** si existe un solo pico y los valores se extienden mucho más allá a la derecha del pico que a la izquierda de este. En este caso **la media es más grande que la mediana**. En una distribución **negativamente sesgada** existe un solo pico, pero las observaciones se extienden más a la izquierda, en dirección negativa. En una distribución negativamente sesgada, **la media es menor que la mediana**. Una **distribución bimodal** tendrá dos o más picos. Con frecuencia este es el caso cuando los valores provienen de dos o más poblaciones

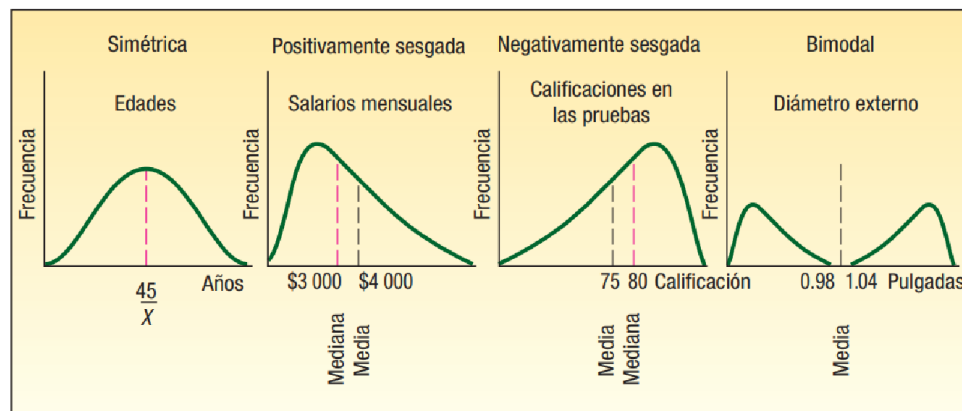


Sesgo positivo: Media > Mediana

Insesgada: Media = Mediana

Sesgo negativo: Media < Mediana

Análisis del **sesgo** a partir de *polígonos de frecuencias*



3.2. Percentiles, deciles y cuartiles

La desviación estándar es la medida de dispersión que mas se utiliza. No obstante, existen otras formas de describir la variación o dispersión de un conjunto de datos. Un método consiste en determinar la ubicación de los valores que dividen un conjunto de observaciones en partes iguales. Estas medidas incluyen los **cuartiles**, **deciles** y **percentiles**.

Los **cuartiles** dividen a un conjunto de observaciones en cuatro partes iguales, ordenadas de menor a mayor. Para explicarlo mejor, piense en un conjunto de valores ordenados de menor a mayor. Anteriormente se denominó a la *mediana* al valor intermedio de un conjunto de datos ordenados de menor a mayor. Es decir el 50 % de las observaciones son mayores que la mediana y 50 % son menores que la mediana. La mediana constituye una medida de ubicación, ya que señala el centro de los datos. De igual manera, los **cuartiles** dividen a un conjunto de observaciones en cuatro partes iguales. El primer cuartil, que se representa con Q_1 , es el valor debajo del cual se presenta el 25 % de las observaciones, y el tercer cuartil, que simboliza Q_3 , es el valor debajo del cual se presenta 75 % de

las observaciones. Lógicamente, Q_2 es la mediana. Q_1 puede considerarse como la mediana de la mitad inferior y Q_3 como la mediana de la parte superior de los datos.

Asimismo, los **deciles** dividen un conjunto de observaciones en 10 partes iguales y los **percentiles** en 100 partes iguales.

Un ejemplo de deciles es, si su promedio general en la universidad se encuentra en el octavo decil, usted podría concluir que 80 % de los estudiantes tuvieron un promedio general inferior al suyo y 20 %, un promedio superior.

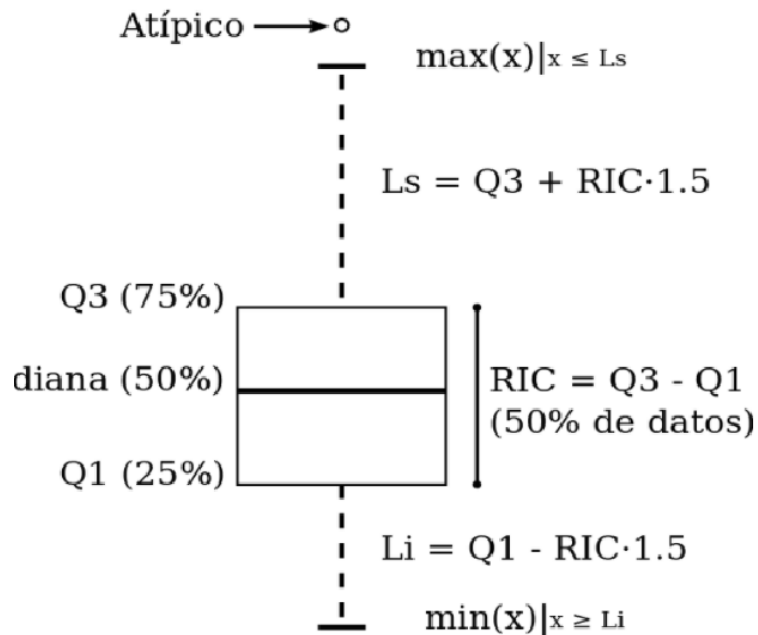
Un ejemplo de percentiles es, un promedio general ubicado en el trigésimo tercer percentil significa que el 33 % de los estudiantes tienen un promedio general mas bajo y que el 67 % un promedio mas alto.

3.3. Diagramas de caja

Un **diagrama de caja** o **box-plot** es una representación grafica, basada en cuartiles, que ayuda a presentar un conjunto de datos. Para construir un diagrama de caja, solo necesita cinco estadísticos:

1. El valor mínimo.
2. Primer cuartil.
3. La mediana (segundo cuartil).
4. Tercer cuartil.
5. El valor máximo.

Esquema



Construcción de un diagrama de caja

Cualquier medición a mayor distancia del límite superior o mayor distancia del límite inferior es un resultado o valor atípico; el resto de las mediciones, dentro de los límites, no son inusuales. Los resultados atípicos se suelen marcar con un asterisco o punto. El diagrama de caja marca el rango del conjunto de datos usando “bigotes” para conectar las mediciones más pequeñas y más grandes (excluyendo resultados atípicos) a la caja.

Límite inferior: $Q_1 - 1,5 \cdot (Q_3 - Q_1)$

Límite superior: $Q_3 + 1,5 \cdot (Q_3 - Q_1)$

Importante: Los límites son imaginarios, no se suelen trazar. A la diferencia $Q_3 - Q_1$ se le llama **rango intercuartílico** o **rango intercuartil**. Un diagrama de caja puede representarse de manera horizontal o vertical. Suele ser más sencillo de interpretar de manera horizontal.

4. Capítulo 4. Probabilidad: Conceptos, Variables Aleatorias y Distribuciones.

A la estadística descriptiva le concierne el resumen de datos recogidos de eventos pasados. Ahora se presenta la segunda faceta de la estadística, a saber, el *calculo de la probabilidad de que algo ocurra en el futuro*. Esta faceta de la estadística recibe el nombre de **inferencia estadística** o **estadística inferencial**. La inferencia estadística se relaciona con las conclusiones relacionadas con una población sobre la base de una muestra que se toma de ella. Dada la incertidumbre existente en la toma de decisiones, es importante que se evalúen científicamente todos los riesgos implicados. La teoría de la probabilidad, a menudo conocida como la ciencia de la incertidumbre, resulta útil para hacer esta evaluación. Su aplicación permite a quien toma decisiones y posee información limitada analizar los riesgos y reducir al mínimo el riesgo que existe, por ejemplo, al lanzar al mercado un nuevo producto o aceptar un envío que contenga partes defectuosas. Puesto que los conceptos de la probabilidad son importantes en el campo de la inferencia estadística, en este capítulo se introduce el lenguaje básico de la probabilidad, que incluye términos como experimento, evento, probabilidad subjetiva y reglas de la adición y de la multiplicación.

4.1. ¿Que es la probabilidad?

En general es un numero que describe la posibilidad de que algo suceda.

Probabilidad: Valor entre cero y uno, inclusive, que describe la posibilidad relativa (oportunidad o casualidad) de ocurra un evento.

Es común que una probabilidad sea expresada en forma décima, como 0.70, 0.27 o 0.50. No obstante, también se da forma de fracción, como $7/10$, $27/100$ o $1/2$. Se puede suponer cualquier numero de 0 a 1, inclusive. La probabilidad de 1 representa algo que seguramente sucederá, y la probabilidad de 0 representa algo que no sucederá. Cuanto mas próxima se encuentre una probabilidad a 0, mas improbable es que el evento suceda. Cuanto mas próxima se encuentre la probabilidad a 1, mas seguro que suceda. En el estudio de la probabilidad se utilizan tres palabras clave: **experimento**, **resultado** y **evento**.

Experimento: Proceso que induce a que ocurra una y solo una de varias posibles observaciones. Es el proceso mediante el cual se obtiene un resultado (observación o medición)



Respecto de la probabilidad, un experimento tiene dos o mas posibles resultados y no sabe cual ocurrirá. Un experimento es aleatorio cuando no se puede predecir el resultado que se va a obtener. Por ejemplo, lanzar una moneda, tirar un dado, jugar a la lotería, realizar una encuesta, etc.

Resultado: Resultado particular de un experimento.

Un ejemplo: lanzar una moneda al aire constituye un experimento. Usted puede observar el lanzamiento de una moneda, pero no esta seguro si caerá *cara* o *cruz*. Si se lanza una moneda, un resultado particular es *cara*. El otro posible resultado es *cruz*. Cuando se observan uno o mas resultados en los experimentos constituyen un evento.

Evento: Conjunto de uno o mas resultados de un experimento. Se lo suele denotar con una letra mayúscula.

Un ejemplo: En el caso del experimento del lanzamiento de un dado, hay seis posibles resultados, pero existen varios posibles eventos. Cuando se cuenta el numero de miembros de la junta directiva de las compañías de Fortune 500 que tienen mas de 60 años de antigüedad el numero posibles de resultados varia de cero al total de miembros. Hay un numero aun mayor de eventos posibles en este experimento.

		
Experimento	Lanzar un dado	Listado del número de miembros de la junta directiva de las compañías de <i>Fortune 500</i> , mayores de 60 años
Todos los posibles resultados	Se observa un 1 Se observa un 2 Se observa un 3 Se observa un 4 Se observa un 5 Se observa un 6	Ninguno tiene más de 60 Uno tiene más de 60 Dos tiene más de 60 ... 29 tiene más de 60 48 tiene más de 60 ...
Algunos posibles eventos	Se observa un número par Se observa un número mayor que 4 Se observa un 3 o un número menor	Más de 13 tiene más de 60 Menos de 20 tiene más de 60

4.2. Enfoques para asignar probabilidades.

Probabilidad clásica.

La **probabilidad clásica** parte del supuesto de que los resultados de un experimento son igualmente posibles. De acuerdo con el punto de vista clásico,

la probabilidad de un evento que se esta llevando a cabo se calcula dividiendo el numero de resultados favorables entre el numero de posibles resultados.

Probabilidad clásica:

$$\text{Probabilidad de un evento} = \frac{\text{Numero de resultados favorables}}{\text{Numero total de posibles resultados}}$$

El concepto de conjuntos mutuamente excluyentes, recordá que cuando creamos clases de tal manera que un evento particular se incluyera en una sola de las clases y que no hubiera superposición entre ellas. Por lo tanto, solo uno de varios eventos puede presentarse en cierto momento.

Mutuamente excluyente: El hecho de que un evento se presente significa que ninguno de los demás eventos puede ocurrir al mismo tiempo.

Unos ejemplos: La variable *genero* da origen a resultados mutuamente excluyentes: hombre y mujer. Un empleado seleccionado al azar es hombre o mujer, pero no puede tener ambos géneros. Una pieza fabricada es aceptable o no lo es. La pieza no puede ser aceptable e inaceptable al mismo tiempo. En una muestra de piezas fabricadas, el evento de seleccionar una pieza no aceptable y el evento de seleccionar una pieza aceptable son mutuamente excluyentes. Si un experimento incluye un conjunto de eventos con todo tipo de resultados posibles, como los eventos ün numero parz ün numero impar.^{en} el experimento del lanzamiento del dado, entonces el conjunto de eventos es **colectivamente exhaustivo**. En el experimento del lanzamiento del dado, cada resultado sera par o impar. Por consiguiente, el conjunto es colectivamente exhaustivo.

Colectivamente exhaustivo: Por lo menos uno de los eventos debe ocurrir cuando se lleva a cabo un experimento.

Si el conjunto de eventos es colectivamente exhaustivo y los eventos son mutuamente excluyentes, la suma de las probabilidades es 1. Resulta innecesario llevar a cabo un experimento para determinar la probabilidad de un evento mediante el enfoque clásico, ya que el numero total de resultados se sabe antes de realizar el experimento. Por lógica, es posible determinar la probabilidad de sacar una cruz al lanzar una moneda o tres caras al lanzar tres monedas.

Probabilidad empírica.

La **probabilidad empírica** o **frecuencia relativa**, el segundo tipo de probabilidad, se basa en el numero de veces que ocurre el evento como proporción del numero de intentos conocidos.

Probabilidad empírica: La probabilidad de que un evento ocurra representa una fracción de los eventos similares que sucedieron en el pasado.

En términos de una formula:

$$\text{Probabilidad empírica} = \frac{\text{Numero de veces que el evento ocurre}}{\text{Numero total de observaciones}}$$

El enfoque empírico de la probabilidad se basa en la llamada *ley de los grandes números*. La clave para determinar probabilidades de forma empírica consiste en que una mayor cantidad de observaciones proporcionaran un calculo mas preciso de la probabilidad.

Ley de los grandes números: En una gran cantidad de intentos, la probabilidad empírica de un evento se aproximara a su probabilidad real.

Para explicar la ley de los grandes números, supongamos que lanzamos una moneda común. El resultado de cada lanzamiento es cara o cruz. Si se lanza la moneda una sola vez, la probabilidad empírica de las caras es cero o uno. Si lanzamos la moneda una gran cantidad de veces, la probabilidad del resultado de las caras se aproximara a 0.5. Observa que conforme incrementamos el numero de intentos, la probabilidad empírica de que salga una cara se aproxima a 0.5, que es su valor de acuerdo con el enfoque clásico de la probabilidad. ¿Que demostramos? Que a partir de la definición clásica de probabilidad, la posibilidad de obtener una cara en un solo lanzamiento de una moneda común es de 0.5. Según el enfoque empírico de la frecuencia relativa de la probabilidad, la probabilidad del evento se aproxima al mismo valor determinado de acuerdo con la definición clásica de probabilidad. Este razonamiento permite emplear el enfoque empírico y de la frecuencia relativa para determinar una probabilidad.

4.3. Reglas para calcular probabilidades

Ahora, que definimos probabilidad y descrito sus diferentes enfoques, cabe atender al calculo de la probabilidad de dos o mas eventos aplicando las reglas de la adición y multiplicación.

Regla general de la adición

Cuando dos eventos ocurren al mismo tiempo, la probabilidad se denomina **probabilidad conjunta**.

Probabilidad conjunta: Probabilidad que mide la posibilidad de que dos o mas eventos sucedan simultáneamente.

Esta regla para dos eventos designados A y B se escribe:

$$\text{Regla general de la adición } P(A \text{ o } B) = P(A) + P(B) - P(A \text{ y } B)$$

En el caso de la expresión $P(A \text{ o } B)$, la conjunción o sugiere que puede ocurrir A o puede ocurrir B. Esto también incluye posibilidad de que A y B ocurran. Tal uso o a veces se denomina **inclusivo**. También es posible escribir $P(A \text{ o } B)$

B o Ambos) para hacer hincapié en el hecho de que la unión de dos eventos incluye la intersección de A y B. Si comparamos las reglas general y especial de la adición, la diferencia que importa consiste en determinar si los eventos son mutuamente excluyentes. Si lo son, entonces la probabilidad conjunta $P(A \text{ y } B)$ es 0 y podríamos aplicar la regla especial de la adición. De lo contrario, debemos tomar en cuenta la probabilidad conjunta y aplicar la regla general de la adición.

Por ejemplo, si tenemos los eventos:

A: obtener un 1 o un 4. En símbolos: $A = \{1, 4\}$

B: obtener un numero par. En símbolos: $B = \{2, 4, 6\}$ Entonces:

$$P(A \cup B) = \frac{2}{6} + \frac{3}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3} \text{ por lo que: } P(A \cup B) = \frac{2}{3}$$

Regla especial de la adición

Para aplicar la regla especial de la adición, los eventos deben ser *mutuamente excluyentes*. Recordá que significa que cuando un evento ocurre, ninguno de los otros eventos puede ocurrir al mismo tiempo. Si dos eventos A y B son mutuamente excluyentes, la regla especial de la adición establece que la probabilidad de que ocurra uno u otro es igual a la suma de sus probabilidades. Esta regla se expresa mediante la siguiente formula:

$$\text{Regla especial de la adición: } P(A \text{ o } B) = P(A) + P(B)$$

En el caso de los tres eventos mutuamente excluyentes designados A, B y C, la regla se expresa de la siguiente manera:

$$P(A \text{ o } B \text{ o } C) = P(A) + P(B) + P(C)$$

Por ejemplo, si tenemos los eventos:

A: obtener un 1 o un 3. En símbolos: $A = \{1, 3\}$

B: obtener un numero par. En símbolos: $B = \{2, 4, 6\}$ Entonces, como son eventos disjuntos:

$$P(A \cup B) = \frac{2}{6} + \frac{3}{6} = \frac{5}{6} \text{ por lo que: } P(A \cup B) = \frac{5}{6}$$

Regla del complemento

Sea el A y su complemento A^c entonces tenemos que:

$$P(A) + P(A^c) = 1 \text{ y } P(A^c) = 1 - P(A)$$

Por ejemplo, si para el experimento 'tirar un dado al aire' consideramos el suceso A 'obtener un 2'. Entonces, A^c es 'no obtener un 2'.

$$P(A) = \frac{1}{6} \text{ y } P(A^c) = \frac{5}{6}$$

Por lo tanto:

$$P(A) + P(A^c) = \frac{1}{6} + \frac{5}{6} \text{ y } P(A^c) = 1 - \frac{1}{6} = \frac{5}{6}$$

Tal es la regla **regla del complemento**. Se emplea para determinar la probabilidad de que un evento ocurra restando de 1 la probabilidad de un evento que no ha ocurrido. Esta regla es útil porque a veces es mas fácil calcular la probabilidad de que un evento suceda determinando la probabilidad de que no suceda y restando el resultado de 1. Observa que los eventos A y A^c son mutuamente excluyentes y colectivamente exhaustivos. Por consiguiente, las probabilidades de A y A^c suman 1.

Regla general de la multiplicación

La regla general de la multiplicación sirve para determinar la probabilidad conjunta de dos eventos cuando estos no son independientes. Por ejemplo, cuando el evento B ocurre después del evento A, y A influye en la probabilidad de que el evento B suceda, entonces A y B no son independientes. Si dos eventos no son independientes, se dice que son **dependientes**. Ejemplo de dependencia: Supongamos que hay 10 latas de refresco en un refrigerador, 7 de los cuales son normales y 3 dietéticos. Se saca una lata del refrigerador. La probabilidad de que sea una lata de refresco dietético es de 3/10, y la probabilidad de que sea una lata de refresco normal es de 7/10. Luego se elige una segunda lata del refrigerador sin devolver la primera. La probabilidad de que la segunda lata sea de refresco dietético depende de que la primera lo haya sido o no. La probabilidad de que la segunda lata sea de refresco dietético es:

2/9, si la primera bebida es dietética (solo dos latas de refresco dietético quedan en el refrigerador).

3/9, si la primera lata elegida es normal (los tres refrescos aun están en el refrigerador).

La denominación adecuada de la fracción 2/9 (o 3/9) es **probabilidad condicional**, ya que su valor se encuentra condicionado (o depende) del hecho de que un refresco regular o dietético haya sido el primer en ser seleccionado del refrigerador.

Probabilidad condicional: Probabilidad de que un evento en particular ocurra, dado que otro evento haya acontecido. Se denota por $P(A|B)$, la barra vertical se lee 'dado' y los eventos que aparecen a la derecha de la barra son aquellos que se saben que han ocurrido. La probabilidad condicional del evento A, dado que el evento B ha ocurrido, es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

La probabilidad condicional del evento B, dado que el evento A ha ocurrido, es:

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Estas fórmulas se pueden obtener despejando $P(A|B)$ y $P(B|A)$ de las fórmulas anteriores

La regla general de la multiplicación establece que en caso de dos eventos, A y B, la probabilidad conjunta de ambos eventos ocurran se determina multiplicando la probabilidad de que ocurra el evento A por la probabilidad de que ocurra el evento B, dado que A ha ocurrido. Simbólicamente, la probabilidad conjunta, $P(A \text{ y } B)$, se calcula de la siguiente manera:

Regla general de la multiplicación: $P(A \cap B) = P(A).P(B|A)$ o $P(A \cap B) = P(B).P(A|B)$

Regla especial de la multiplicación

La regla especial de la multiplicación requiere que dos eventos, A y B, sean independientes, y lo son si el hecho que uno ocurra no altera la probabilidad de que el otro suceda.

Independencia: Si un evento ocurre, no tiene ningún efecto sobre la probabilidad de que otro evento acontezca.

En el caso de dos eventos independientes A y B, la probabilidad de que A y B ocurran se determina multiplicando las dos probabilidades, tal es la **regla especial de la multiplicación**, cuya expresión simbólica es la siguiente:

Regla especial de la multiplicación: $P(A \cap B) = P(A).P(B)$

En el caso de tres eventos independientes, A, B y C, la regla especial de la multiplicación que se utiliza para determinar la probabilidad de que los tres eventos ocurran es:

Regla especial de la multiplicación para 3: $P(A \cap B \cap C) = P(A).P(B).P(C)$

4.4. Principios de conteo

Si la cantidad de posibles resultados de un experimento es pequeña, resulta relativamente fácil contarlas. Por ejemplo, existen seis posibles resultados del lanzamiento de un dado. Sin embargo, si hay un número muy grande de resultados, tal como el número de caras y cruces en un experimento de 10 lanzamientos de una moneda, sería tedioso contar todas las posibilidades. Todas podrían ser caras, una cruz y nueve caras, y así sucesivamente. Para facilitar la cuenta se analizarán tres fórmulas para contar: la **fórmula de la multiplicación** (no confundir con la regla de la multiplicación), la **fórmula de las permutaciones** y la **fórmula de las combinaciones**.

Formula de la multiplicación

Formula de la multiplicación: Si hay m formas de hacer una cosa y n formas de hacer otra, hay $m \times n$ formas de hacer ambas.

En términos de formula: Numero total de disposiciones = $(m) \cdot (n)$

Esta fórmula se puede extender a más de dos eventos. En el caso de tres eventos m , n y o :

$$\text{Numero total de disposiciones} = (m)(n)(o)$$

Formula de las permutaciones

La fórmula de la multiplicación se aplica para determinar el número de posibles disposiciones de dos o más grupos. En contraste, **la fórmula de las permutaciones** se aplica para determinar el número posible de disposiciones cuando solo hay un grupo de objetos. Ejemplo de permutacion:

- Tres piezas electrónicas (un transistor, un LED y un sintetizador) se van a montar en una unidad que se conecta a un aparato de televisión. Estas se pueden montar en cualquier orden. La pregunta es: ¿de cuantas formas pueden montarse?

Un orden seria: primero el transistor, en seguida el LED y en tercer lugar el sintetizador. A esta distribución se le conoce como **permutación**.

Permutación: Cualquier distribución de r objetos seleccionados de un solo grupo de n posibles objetos.

Formula de las permutaciones:

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

donde:

n representa el total de objetos;

r representa el total de objetos seleccionados.

La solución del ejemplo es que hay tres piezas electrónicas que van a montarse acá, así que $n = 3$. Como las tres se van a insertar en la unidad conectable, $r = 3$. Usa la formula y te da 6, que significa que hay seis formas en la que las tres piezas electrónicas, se pueden ordenar.

Formula de las combinaciones

Si el orden de los objetos seleccionados *no* es importante, cualquier selección se denomina **combinación**. La formula para contar el numero r combinaciones de objetos de un conjunto de n objeto es:

Formula de las combinaciones:

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

Por ejemplo, si los ejecutivos Juan, Pedro y Fernando van a ser elegidos para formar un comité de negociación de una fusión, solo existe una posible combinación con estos tres; el comité formado por Juan, Pedro y Fernando es el mismo comité que el que forman Pedro, Fernando y Juan. De acuerdo con la formula de las combinaciones:

$${}_3C_3 = \frac{3!}{3!(3-3)!} = 1$$

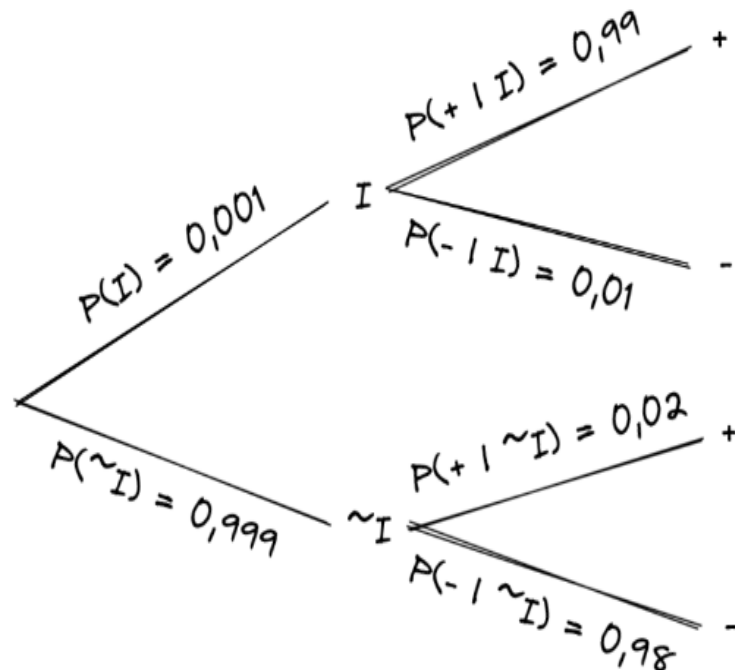
4.5. Teorema de Bayes

El teorema de Bayes es un caso de aplicación de la probabilidad condicional, el cual se puede aplicar cuando se cuenta con información a priori de los datos que se están analizando. Pero para comenzar a introducirnos a este teorema primero tenemos que tener en cuenta una serie de cosas.

1. Tener conocimiento de que es un diagrama de árbol, como leerlos y como este nos va a ayudar en el planteamiento del problema.
2. El concepto de probabilidad total de un evento.
3. La regla general de la multiplicación, que se utiliza para obtener la fórmula del teorema de Bayes.

Diagrama de Árbol

Los diagramas de árbol, son una forma de representar las distintas condiciones o alternativas a las que están sometidas las variables de un problema, para su mejor comprensión supongamos que nos encontramos con el siguiente problema: El 0,1 % de la población de un país está infectada con una bacteria, y la prueba para detectar la enfermedad tiene una efectividad del 99 % en las personas enfermas y para las personas que no están infectadas este porcentaje es de 2 %. Con fin de representar estos datos de forma más gráfica y ordenada, podemos poner toda esta información en un diagrama de árbol, el cual quedaría conformado de la siguiente manera.



Como se puede observar cada rama del árbol contiene los respectivos valores representativos para cada situación. Un dato importante a tener en cuenta a la hora de realizar este tipo de diagrama es que la suma de las probabilidades de las ramas pertenecientes al mismo padre deben dar siempre 1, por ejemplo, si la probabilidad de estar infectado es de 0,001 entonces la de no estar infectado debe ser si o si 0,999; ya que la sumas de estos dos siempre tiene que dar 1.

Teorema de la Probabilidad Total

Siguiendo con el ejemplo presentado anteriormente, supongamos que queremos saber cual es la probabilidad de que una persona de positiva en la prueba, independientemente de si está infectado con la bacteria o no. En ese caso se tendrán que sumar todos los caminos posibles del diagrama de árbol que nos lleve hacia ese evento específico. Aunque ya existe una fórmula que nos facilita encontrar la probabilidad total.

$$\sum_{j=1}^k P(S_j)P(A|S_j)$$

Nótese que en sí esta ecuación en primera instancia es la sumatoria de las intersecciones entre los demás eventos del experimento, con el evento que nosotros estamos analizando, en nuestro caso cuál es la probabilidad de que de positivo

una prueba, independientemente de si está infectado o no. Dicha sumatoria nos quedaría de la siguiente manera.

$$\sum_{j=1}^k S_j \cap A$$

Esta fórmula se transforma, en la del teorema de la probabilidad total, si aplicamos la regla general de la multiplicación, en la cual una intersección entre eventos se puede representar como la multiplicación de la siguiente manera:

$$P(S \cap A) = P(S)P(A|S)$$

$$P(S \cap A) = P(A)P(S|A)$$

Fórmula del Teorema de Bayes

Supongamos ahora que queremos saber la probabilidad de que la prueba de positiva siendo que estas infectado con la bacteria. En ese caso ese valor sería obtenido a través de la siguiente fórmula:

$$P(S_i|A) = \frac{P(S_i)P(A|S_i)}{\sum_{j=1}^k P(S_j)P(A|S_j)}$$

Como podemos ver, esta fórmula es la misma que para la probabilidad condicionada, con la diferencia que se realiza una serie de cambios leves, resultado de aplicar las propiedades antes mencionadas. Este teorema resulta muy útil para muchos campos de la ciencia. Pero tiene una serie de puntos débiles que lo convierten en una no muy buena opción para algunas situaciones en particular.

1. Si no se tiene un conocimiento del fenómeno, no es posible asignar las probabilidades iniciales (a priori) y por lo tanto no es posible aplicar el teorema de Bayes. Este problema se suele evitar, suponiendo que las probabilidades a priori tienen una distribución uniforme, lo cual es considerado por algunos autores como inaceptable, porque en muchas situaciones no conocemos la distribución inicial, pero sabemos claramente que no es uniforme. Otra posibilidad sería asignar subjetivamente las probabilidades iniciales, es decir asignar probabilidad sin contar con evidencias que la respaldan, pero esto llevaría a que dos investigadores con los mismos datos obtuviesen unas probabilidades finales diferentes.
2. El método de Bayes no permite calcular ni revisar "probabilidades a priori objetivas" de la hipótesis sino las "probabilidades a priori subjetivas" de dicha hipótesis, la cual es establecida por cada investigador. Es decir su grado personal de creencia en la veracidad de la hipótesis.

4.6. Análisis de Variables Aleatorias

Si consideramos un experimento aleatorio, como lo sería tirar un dado equilibrado, y realizamos un cierto número n de pruebas relativas al mismo, por ejemplo tiramos 100 veces el dado, obtenemos un conjunto de observaciones, que se llama una muestra aleatoria de tamaño n .

Ese conjunto de resultados dará lugar a una tabla en que a unos ciertos valores de la variable, en este caso los números de cada cara del dado, corresponden unas ciertas frecuencias, la cantidad de veces que salió cada uno. A tal variable, que representa únicamente los n resultados obtenidos la denominamos **variable estadística**. Puede pasar que en las 10 primeras tiradas que hicimos el número 3 no haya salido ninguna vez, en este caso la variable estadística toma los valores 1, 2, 4, 5 y 6.

Si imaginamos realizar infinitas pruebas de este experimento, la infinidad de resultados da origen a la noción de variable aleatoria. En este caso, a la variable que toma todos los valores que representan los sucesos elementales posibles de dicho experimento, con unas ciertas probabilidades de ocurrir, se llama variable aleatoria. En este caso la variable aleatoria puede tomar los valores 1, 2, 3, 4, 5 y 6, siendo sus probabilidades todas iguales a $1/6$. Al igual que las variables estadísticas cuantitativas, las variables aleatorias se clasifican en discretas, si los eventos toman valores exactos. O continuas si los eventos toman valores con decimales. Es importante distinguir entre variables aleatorias discretas y continuas, porque se usan técnicas diferentes para describir (calcular) sus distribuciones de probabilidad.

Distribución de las Probabilidades en Variables Discretas

Una distribución de probabilidad proporciona toda la gama de valores que se pueden presentar en un experimento y sus probabilidades de ocurrencia. Esas probabilidades representan los valores límites de las frecuencias relativas cuando el experimento se repite una y otra vez. La distribución de probabilidad se construye como un modelo para toda la población de mediciones.

Se pueden hallar ejemplos de variables aleatorias discretas en numerosas situaciones cotidianas y en casi todas las disciplinas académicas. No obstante, hay tres distribuciones discretas de probabilidad que sirven como modelos para un gran número de estas aplicaciones. El único requisito que tienen que cumplir es que todas las probabilidades deben estar entre 0 y 1, y que la suma de todas estas probabilidades de 1, de lo contrario el modelo se considera erróneo.

4.6.1. Distribución Binomial de Probabilidad

Un experimento se puede considerar como binomial si cumple con estas cinco características:

1. El experimento consiste en n intentos idénticos.
2. Cada intento resulta en uno de dos resultados. Por falta de un mejor nombre, el resultado uno se llama éxito, S , y el otro se llama fracaso, F .

3. La probabilidad de éxito en un solo intento es igual a p y es igual de un intento a otro. La probabilidad de fracaso es igual a $(1 - p) = q$.
4. Los intentos son independientes.
5. Estamos interesados en x , el número de éxitos observado durante los n intentos, para $x = 0, 1, 2, \dots, n$.

Para lograr distinguir que tipo de problemas cumplen estas características veamos dos ejemplos:

Ejemplo 1: Suponga que hay alrededor de un millón de adultos en un condado y una proporción desconocida p están a favor de limitar el periodo de función de políticos. Se escogerá una muestra de mil adultos en forma tal que cada uno del millón de adultos tenga igual probabilidad de ser seleccionado y a cada uno se le pregunta si él o ella está a favor de limitar el periodo.

1. Un “intento” es la selección de un solo adulto de entre el millón de adultos del condado. Esta muestra consta de $n = 1.000$ intentos idénticos.
2. Como cada adulto estará a favor o no estará a favor de limitar el periodo, hay dos resultados que representan los “éxitos” y “fracasos” del experimento binomial.
3. La probabilidad de éxito p , es la probabilidad de que un adulto esté a favor del límite del periodo. ¿Esta probabilidad sigue igual para cada uno de los adultos de la muestra? Para todos los fines prácticos, la respuesta es sí. Por ejemplo, si 500 mil adultos de la población están a favor de limitar el periodo, entonces la probabilidad de un “éxito” cuando se escoja al primer adulto es $1/2$. Cuando se escoja al segundo adulto, la probabilidad p cambia ligeramente, dependiendo de la primera selección. Esto es, habrá 499.999 o 500.000 éxitos que queden entre los 999.999 adultos. En cualquiera de estos casos, p es todavía más o menos igual a $1/2$.
4. La independencia de los intentos está garantizada debido al grupo grande de adultos del que se toma la muestra. La probabilidad de que un adulto esté a favor de limitar el periodo no cambia, dependiendo de las respuestas de las personas previamente escogidas.
5. La variable aleatoria x es el número de adultos de la muestra que estén a favor de limitar el periodo.

Debido a que el estudio satisface las cinco características razonablemente bien, para todos los fines prácticos se puede ver como un experimento binomial.

Ejemplo 2: Un paciente llena una receta para un régimen de 10 días de dos píldoras diarias. Sin que lo sepan el farmacéutico ni el paciente, las 20 pastillas están formadas por 18 píldoras del medicamento prescrito y dos píldoras que son el equivalente genérico del medicamento prescrito. El paciente selecciona dos píldoras al azar para la dosis del primer día. Si verificamos la selección y registramos el número de píldoras que son genéricas.

Para este problema en particular no se satisface la 4ª condición, porque la probabilidad de sacar una píldora genérica en el segundo intento depende del primer intento. Por ejemplo, si la primera píldora sacada es genérica entonces hay sólo una píldora genérica en las restantes 19. Por tanto, la probabilidad de que eso pase va a ser de $1/19$, de lo contrario si primero se saca una de las pastillas recetadas, la probabilidad de que la segunda sea genérica es de $2/19$. Los intentos son dependientes, por lo que el muestreo no representa un experimento binomial.

Regla practica: Si el tamaño muestral es grande con respecto al tamaño poblacional, en particular si $n/N \geq 0,5$, entonces el experimento resultante no es binomial, sino que se podría representar mediante un **modelo hipergeométrico**.

Un experimento binomial consta de n intentos idénticos con probabilidad p de éxito en cada intento. La probabilidad de k éxitos en n intentos se determina de la siguiente manera

$$P(x = k) = C_k^n p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Para esta distribución de probabilidad la media, varianza y desviación estándar se pueden calcular de la siguiente manera

Media: $\mu = np$

Varianza: $\sigma^2 = npq$

Desviación Estándar: $\sigma = \sqrt{npq}$

4.6.2. Distribución de Probabilidad de Poisson

Otra variable aleatoria discreta que tiene numerosas aplicaciones prácticas es la variable aleatoria de Poisson. Su distribución de probabilidad da un buen modelo para datos que representa el número de sucesos de un evento especificado en una unidad determinada de tiempo o espacio. Algunos experimentos que pueden ser modelados con la distribución de Poisson pueden ser, por ejemplo, El número de accidentes de tránsito en un cruce dado durante un tiempo determinado o El número de llegadas de clientes al mostrador de una caja de pago en un minuto determinado, entre muchos otros.

En cada uno de estos ejemplos, **x representa el número de eventos que ocurren en un periodo o espacio, durante el cual se puede esperar que ocurra un promedio μ de estos eventos.** Las únicas suposiciones necesarias, cuando uno usa la distribución de Poisson para modelar experimentos tales como éstos, son que las cuentas o eventos ocurren al azar e independientemente unos de otros.

Sea μ el número promedio de veces que ocurre un evento en cierto tiempo o espacio. La probabilidad de k sucesos de este evento está dada por

$$P(x = k) = \frac{\mu^k e^{-\mu}}{k!}$$

para valores de $k = 0, 1, 2, 3, \dots$. La media y desviación estándar de la variable aleatoria de Poisson x son

Media: μ

Desviación Estándar: $\sqrt{\mu}$

La distribución de probabilidad de Poisson da una aproximación sencilla, fácil de calcular y precisa a probabilidades binomiales cuando n es grande y $\mu = np$ es pequeña, de preferencia con $\mu < 7$.

Normalmente la probabilidad de Poisson se utiliza para crear tablas de probabilidad acumulada que facilitan el análisis de ciertas poblaciones dado distintos valores de μ , la manera de crear estas tablas es la siguiente

1. Encuentre el valor necesario de μ .
2. Haga una lista de valores de x en su evento. En donde se colocan los eventos en las filas y los distintos valores de μ en las columnas
3. Para cada valor de x , sustituya $x = k$ en la fórmula presentada anteriormente.
4. Sume las probabilidades individuales obtenidas del paso 3 para hallar la probabilidad de interés.

Al finalizar todo este proceso va a quedar conformada una tabla de probabilidad acumulada para todos los eventos para un valor de μ específico, lo que significa que para valores distintos de μ se tendrá que crear una nueva la tabla. Para poder calcular una probabilidad con estas tablas, primero se debe encontrar el valor de μ de su interés. Aísle la columna apropiada de la tabla, en esa celda encontrará la probabilidad de todos los eventos $x \leq k$, para calcular la probabilidad del evento k , se tendrá que realizar una resta entre esta probabilidad acumulada obtenida hasta llegado ese evento y la probabilidad acumulada de el evento anterior al que estamos analizando. Por ejemplo, supongamos que queremos encontrar la probabilidad del evento 3, en este caso el $P(x = 3) = P(x \leq 3) - P(x \leq 2)$.

4.6.3. Distribución Hipergeométrica de Probabilidad

Si el número de elementos de la población es grande con respecto al número en la muestra, como sucede en el **ejemplo 1 de la sección 5.6.1**, la probabilidad de seleccionar un éxito en un solo intento es igual a la proporción p de éxitos en la población. Debido a que la población es grande con respecto al

tamaño muestral, esta probabilidad permanecerá constante (para todos los fines prácticos) de un intento a otro y el número x de éxitos en la muestra seguirá una distribución binomial de probabilidad. No obstante, si el número de elementos en la población es pequeño con respecto al tamaño muestral, generalmente cuando $n/N \geq 0,5$, la probabilidad de un éxito para un intento determinado depende de los resultados de intentos precedentes. Entonces el número x de éxitos sigue lo que se conoce como una **distribución hipergeométrica de probabilidad**. Es fácil visualizar la variable hipergeométrica aleatoria x si se considera un tazón que contenga M esferas rojas y $N - M$ esferas blancas, donde N representa el total de esferas en el tazón. Usted selecciona n esferas del tazón y registra x , el número de esferas rojas que vea. Si ahora define un “éxito” como una esfera roja, tendrá un ejemplo de la variable aleatoria x hipergeométrica. Una población contiene M éxitos y $N - M$ fracasos. La probabilidad de exactamente k éxitos en una muestra aleatoria de tamaño n es

$$P(x = k) = \frac{C_k^M C_{n-k}^{N-M}}{C_n^N}$$

La media y la varianza de una variable aleatoria hipergeométrica son muy semejantes a las de una variable aleatoria binomial con una corrección para el tamaño finito de población:

Media: $\mu = n(\frac{M}{N})$

Varianza: $\sigma^2 = n(\frac{M}{N})(\frac{N-M}{N})(\frac{N-n}{N-1})$

Varianza: $\sigma = \sqrt{n(\frac{M}{N})(\frac{N-M}{N})(\frac{N-n}{N-1})}$

4.7. Distribuciones de probabilidad continuas

Las distribuciones de probabilidad continuas son una parte fundamental de la estadística y la probabilidad. A diferencia de las distribuciones de probabilidad discretas, que se utilizan para eventos que tienen resultados finitos y distintos (como lanzar una moneda), las distribuciones de probabilidad continuas se utilizan para eventos que tienen un número infinito de posibles resultados (como medir la altura de una persona). Las distribuciones de probabilidad continuas se describen a través de una función de densidad de probabilidad. El área bajo la curva de la función de densidad de probabilidad en un intervalo dado proporciona la probabilidad de que el resultado caiga en ese intervalo. En un histograma de frecuencia relativa se puede ver mientras mas barras haya se hace una curva mas suave. Esta curva se describe la **distribución de probabilidad de la variable aleatoria continua x** .

Ahora, ¿cómo podemos crear un modelo para una distribución de una variable continua? Para crear un modelo necesitamos crear una función de densidad de probabilidad (**PDF** en ingles, de **Probability Density Function**). Una función de densidad de probabilidad es una función que describe la probabilidad

relativa de que una variable aleatoria tome un valor dado. La probabilidad de que una variable aleatoria continua caiga en un intervalo particular se da por el área bajo la curva de la función de densidad en ese intervalo.

Una función $f(x)$ es una función de densidad de probabilidad (PDF) válida para una variable aleatoria continua si satisface las siguientes dos propiedades:

1. **No Negatividad:** Para todo x en el conjunto de números reales, la función $f(x)$ debe ser mayor o igual a cero. Esto tiene sentido ya que la probabilidad nunca puede ser un valor negativo. Matemáticamente, se representa como: $f(x) \geq 0$ para todo x .
2. **Normalización:** La integral de la función $f(x)$ sobre todo el espacio de la variable debe ser igual a 1. Esto se debe a que la suma de las probabilidades de todos los posibles resultados debe ser igual a 1. En términos matemáticos, se representa como: $\int_{-\infty}^{\infty} f(x)dx = 1$, la integral se realiza sobre todo el conjunto de números reales.

Estas son las propiedades básicas que toda función de densidad de probabilidad debe cumplir. Sin embargo, dependiendo de la distribución específica que se esté modelando, la PDF puede tener propiedades adicionales. Por ejemplo, la PDF de una distribución normal está definida por dos parámetros: la media y la desviación estándar, y tiene una forma simétrica en forma de campana.

Las distribuciones de probabilidad continua tienen diversas propiedades que son las siguientes:

Las distribuciones de probabilidad continuas, descritas mediante una función de densidad de probabilidad (PDF), tienen las siguientes propiedades:

1. **El área bajo una distribución continua de probabilidad es igual a 1:** Esto es una consecuencia de la definición de una PDF. Matemáticamente, para cualquier PDF $f(x)$, se cumple que:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Por lo tanto, el área total bajo la curva de la PDF debe ser igual a 1.

2. **La probabilidad de que x caiga en un intervalo particular, por ejemplo de a hasta b , es igual al área bajo la curva entre los dos puntos a y b :** Dado que la PDF proporciona la densidad de probabilidad en cada punto, para encontrar la probabilidad de que la variable aleatoria caiga en un intervalo, necesitamos integrar la densidad de probabilidad en ese intervalo. Matemáticamente, esto se representa como:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

3. **$P(X = a) = 0$ para variables aleatorias continuas:** Dado que hay infinitamente muchos posibles valores que la variable aleatoria podría tomar, la probabilidad de un valor específico es infinitesimalmente pequeña.

Por lo tanto:

$$P(X = a) = 0$$

4. $P(X \geq a) = P(X > a)$ y $P(X \leq a) = P(X < a)$: Dado que la probabilidad de un valor específico es cero, no hace ninguna diferencia si incluimos ese valor específico en el intervalo o no. Por lo tanto:

$$P(X \geq a) = P(X > a) \quad \text{y} \quad P(X \leq a) = P(X < a)$$

Es importante destacar que estas propiedades son específicas de las variables aleatorias continuas y no se aplican necesariamente a las variables aleatorias discretas.

5. Capítulo 5. Planes Muestrales y Diseño Experimental

La forma en que una muestra se selecciona recibe el nombre de **plan muestral** o **diseño experimental** y determina la cantidad de información de la muestra. Saber el plan muestral empleado en una situación particular permitirá medir la confiabilidad, también llamada bondad, de la inferencia.

El muestreo aleatorio simple es un plan muestral de uso común en el que cada muestra de tamaño n tiene la misma probabilidad de ser seleccionado. Por ejemplo, supongamos que se desea seleccionar una muestra de tamaño $n = 2$ de una población que contiene $N = 4$ objetos. Si los cuatro objetos están identificados por los símbolos x_1 , x_2 , x_3 y x_4 , hay seis pares distintos que podrían seleccionarse. Si la muestra de $n = 2$ observaciones se selecciona de modo que cada una de estas seis muestras tenga la misma probabilidad de selección, dada por $1/6$, entonces la muestra resultante se denomina **muestra aleatoria simple** o únicamente **muestra aleatoria**.

Definición: Si una muestra de n elementos se selecciona de entre una población de N elementos, usando un plan muestral en el que cada una de las posibles muestras tiene la misma probabilidad de selección, entonces se dice que el muestreo es aleatorio y la muestra resultante es una muestra aleatoria simple.

Un muestreo aleatorio perfecto es difícil de obtener en la práctica. Si el tamaño N de la población es pequeño, se podría escribir cada uno de los N números en una ficha, mezclar las fichas y seleccionar una muestra de n fichas. Los números que seleccione corresponden a las n mediciones que aparecen en la muestra. Como este método no siempre es práctico, un método más sencillo y confiable utiliza números aleatorios, es decir, dígitos generados de modo que los valores de 0 a 9 se presentan al azar y con igual frecuencia.

Los estudios que se realizan sobre datos que ya existen con posterioridad antes de que nosotros decidamos observarlos o describir características de ellos se denominan **estudios observacionales**. En esta categoría caen la mayor parte

de los estudios muestrales, en los que la información se capta con un cuestionario. Con la diferencia que para este tipo de casos hay que tener particular cuidado al efectuar el estudio muestral y hay que estar atento a los siguientes problemas que se pueden presentar:

- **No respuesta:** Usted ha seleccionado su muestra aleatoria y enviado sus cuestionarios, pero sólo 50 % de los entrevistados devolvió sus cuestionarios. ¿Las respuestas que usted recibió son representativas de toda la población o están sesgadas porque sólo quienes eran particularmente obstinados en el tema fueron escogidos para responder?
- **Cobertura demasiado baja:** Usted ha seleccionado su muestra aleatoria usando registros telefónicos como una base de datos. ¿La base de datos que usó sistemáticamente excluye ciertos segmentos de la población, quizá aquellos que no tienen teléfono?
- **Sesgo verbal:** El cuestionario de usted puede tener preguntas que son demasiado complicadas o tienden a confundir al lector. Posiblemente las preguntas son sensibles por naturaleza, por ejemplo, “¿Alguna vez ha consumido usted drogas?”.

Se han diseñado métodos para resolver algunos de estos problemas, pero sólo si usted sabe que existen. Si su encuesta está sesgada por cualquiera de estos problemas, entonces sus conclusiones no serán muy confiables, aunque haya seleccionado una muestra aleatoria.

Además del muestreo aleatorio simple, hay otros planes muestrales con carácter aleatorio y, por tanto, dan una base probabilística para hacer inferencias. Tres de esos planes están basados en muestreo estratificado, conglomerado y sistemático.

Cuando la población está formada por dos o más subpoblaciones, llamadas estratos, un plan muestral que asegura que cada subpoblación está representada en la muestra se denomina muestra aleatoria estratificada.

Definición: Un muestreo aleatorio estratificado comprende seleccionar una muestra aleatoria simple de cada uno de un número dado de subpoblaciones o estratos.

Se utiliza otra forma muestral aleatorio cuando las unidades muestrales disponibles son grupos de elementos, llamados conglomerados. Por ejemplo, una familia es un conglomerado de personas que viven juntas. Una manzana o vecindario de una ciudad podrían ser una cómoda unidad muestral y podría ser considerada un conglomerado para un plan determinado muestral. Cuando un conglomerado particular se incluye en la muestra, se toma un censo de cada uno de los elementos del conglomerado.

Definición: Una muestra de conglomerados es una simple muestra aleatoria tomada de los conglomerados disponibles en la población.

A veces la población a ser muestreada está ordenada, por ejemplo una lista de usuarios de la compañía de luz por direcciones de servicio o una lista de clientes por números de cuenta. En estas y otras situaciones, se escoge un elemento al azar de los primeros k elementos y, a continuación, cada k -ésimo elemento de ahí en adelante se incluye en la muestra.

Definición: Una muestra aleatoria sistemática 1 en k involucra la selección aleatoria de uno de los primeros k elementos de una población ordenada y luego la selección sistemática de cada k -ésimo elemento de ahí en adelante.

También existen otros tipos de muestras llamadas **muestras de conveniencia**, que son una forma fácil y sencilla, pero que no emplea la selección aleatoria. Dentro de estas muestras podemos encontrar al **muestreo de juicio** en donde la persona encargada de realizar el muestreo puede decidir quién estará incluido en la muestra y quien no. O el **muestreo de cuota**, en donde la composición de la muestra debe reflejar la composición de la población en alguna característica preestablecida. Cabe recalcar que en las muestras no aleatorias se pueden describir características de los datos, pero *no se pueden ser usadas para realizar inferencia*.

5.1. Estadística y Distribuciones Muestrales

Cuando se selecciona una muestra aleatoria de una población, las medidas numéricas descriptivas que se calculen de la muestra se denominan **estadísticas**. Las distribuciones de probabilidad para estadísticas se llaman **distribuciones muestrales** porque, en distintos muestreos repetidos, nos proporcionan información de qué valores de la estadística pueden presentarse y con qué frecuencia se presenta cada valor.

Definición: La distribución muestral de una estadística es la distribución de probabilidad para los posibles valores de la estadística, que resulta cuando muestras aleatorias de tamaño n se sacan repetidamente de la población.

Hay tres formas de hallar la distribución muestral de una estadística:

1. Deducir la distribución matemáticamente usando las leyes de probabilidad.
2. Usar una simulación para aproximar la distribución. Esto es, saque un gran número de muestras de tamaño n , calculando el valor de la estadística para cada muestra y tabular los resultados en un histograma de frecuencia relativa. Cuando el número de muestras es grande, el histograma será muy cercano a la distribución teórica muestral.
3. Usar *teoremas estadísticos* para obtener distribuciones muestrales exactas o aproximadas.

5.1.1. Teorema del Límite Central

El teorema del límite central dice que, bajo condiciones más bien generales, las sumas y medias de muestras aleatorias de mediciones tomadas de una población tienden a tener una distribución aproximadamente normal. El teorema es el siguiente:

Teorema: Si muestras aleatorias de n observaciones se sacan de una población no normal con media finita μ y desviación estándar σ , entonces, cuando n es grande, la distribución de muestreo de la media muestral \bar{x} está distribuida normalmente en forma aproximada, con media μ y desviación estándar

$$\frac{\sigma}{\sqrt{n}}$$

La aproximación se va volviendo cada vez más precisa cuando n se hace cada vez mas grande.

El teorema del límite central se puede expresar de otro modo para aplicar a la suma de las mediciones muestrales $\sum x_i$, que, cuando n se hace grande, también tiene una distribución aproximadamente normal con media $n\mu$ y desviación estándar $\sigma\sqrt{n}$.

La aportación más importante del teorema del límite central está en la inferencia estadística. Muchos estimadores que se usan para hacer inferencias acerca de parámetros poblacionales son sumas o promedios de las mediciones muestrales. Cuando el tamaño muestral es lo suficientemente grande, se puede esperar que estos estimadores tengan distribuciones muestrales que sean aproximadamente normales. Entonces se puede usar la distribución normal para describir el comportamiento de estos estimadores en muestreos repetidos y evaluar la probabilidad de observar ciertos resultados muestrales. Estas probabilidades se calculan usando la variable aleatoria normal estándar:

$$z = \frac{\text{Estimador} - \text{Media}}{\text{Desviación Estándar}}$$

El teorema menciona que la aproximación es válida mientras el tamaño de la muestra n sea grande. Desafortunadamente el tamaño de la muestra no se puede determinar de manera tan sencilla, ya que va a depender de la forma de la población de la cual se realiza el muestreo, y de cómo se desee usar la aproximación. No obstante los siguientes puntos pueden ayudar a determinar el tamaño de la muestra en la mayoría de los casos

- Si la población muestreada es normal, entonces la distribución muestral de \bar{x} también será normal, sin importar cuál sea el tamaño de la muestra que se escoja. Este resultado se puede demostrar en forma teórica, pero no debe ser demasiado difícil aceptarla sin prueba.
- Cuando la población muestreada es aproximadamente simétrica, la distribución muestral de \bar{x} se hace también aproximadamente normal para

valores relativamente pequeños de n . Como ejemplo tomemos el lanzamiento de un dado balanceado, vease que la probabilidad de que salga una cara de entre las demás es simétrica, pero si tomamos múltiples muestras de $n = 3$ veremos como esta distribución que parecía simétrica, toma forma de montículo o campana.

- Cuando la población muestreada está sesgada, el tamaño muestral n debe ser más grande, con n al menos 30 antes que la distribución muestral de \bar{x} se haga aproximadamente normal.

5.1.2. La Distribución Muestral de la Media Muestral

Si la media poblacional μ es desconocida, se pueden seleccionar varias estadísticas como estimador; la media muestral \bar{x} y la mediana muestral m son dos que con facilidad llegan a nuestra mente. Pero ¿Cuál debe usarse?, Considere estos criterios al seleccionar el estimador para μ , ¿Es fácil o difícil de calcular? ¿Produce estimaciones que sean demasiado altas o demasiado bajas de manera consistente? ¿Es más o menos variable que otros estimadores posibles? En muchas situaciones, la media muestral \bar{x} tiene propiedades deseables como un estimador que no son compartidas por otros estimadores competidores; por tanto, se emplea en forma más general:

- Si una muestra aleatoria de n mediciones se selecciona de una población con media μ y desviación estándar σ , la distribución muestral de la media muestral \bar{x} tendrá media μ y desviación estándar σ/\sqrt{n} .
- Si la población tiene una distribución *normal*, la distribución muestral de \bar{x} estará *exactamente* distribuida en forma normal, *cualquiera que sea el tamaño muestral n* .
- Si la distribución poblacional *no es normal*, la distribución muestral de \bar{x} estará distribuida normalmente *en forma aproximada* para muestras grandes, por el teorema del límite central.

Definición: La desviación estándar de una estadística empleada como estimador de un parámetro poblacional también se denomina **error estándar del estimador** (SE) porque se refiere a la precisión denominada. Por tanto, la desviación estándar de \bar{x} , dada por σ/\sqrt{n} , se conoce como **error estándar de la media** abreviado como $SE(\bar{x})$ o sólo SE.

5.1.3. La Distribución Muestral de la Proporción Muestral

Hay numerosos ejemplos prácticos de la variable aleatoria x binomial. Una aplicación común se relaciona con la preferencia del consumidor o encuestas de opinión, donde usamos una muestra aleatoria de n personas, para estimar la

proporción p de personas en la población que tienen una característica especificada. Si x de las personas muestreadas tienen esta característica, entonces la proporción muestral

$$\hat{p} = \frac{x}{n}$$

se puede usar para estimar la proporción poblacional p .

La variable aleatoria x binomial tiene una distribución de probabilidad que ya fueron descritas con anterioridad, poseyendo media np y desviación estándar \sqrt{npq} . Como \hat{p} es simplemente el valor de x , la distribución muestral de \hat{p} es idéntica a la distribución de probabilidad de x , excepto que tiene una nueva escala a lo largo del eje horizontal. Y debido a este cambio de escala, la media y desviación estándar de \hat{p} también tienen cambio de escala, de modo que la media de la distribución muestral de \hat{p} es p , y su error estándar es

$$SE(\hat{p}) = \sqrt{\frac{pq}{n}}, \text{ donde } q = 1 - p$$

Por último, así como podemos aproximar la distribución de probabilidad de x con una distribución normal cuando el tamaño muestral n es grande, podemos hacer lo mismo con la distribución muestral de \hat{p} .

La distribución muestral de la proporción muestral cuenta con un par de propiedades vale mencionar:

- Si una muestra aleatoria de n observaciones se selecciona de una población binomial con parámetro p , entonces la distribución muestral de la proporción muestral \hat{p} , denotada por la fórmula antes descrita, tendrá una media p y una desviación o error estándar, ya descrita anteriormente.
- Cuando el tamaño muestral n es grande, la distribución muestral de \hat{p} puede ser aproximada por una distribución normal. Dicha aproximación será adecuada si $np > 5$ y $nq > 5$.