**ESTADÍSTICA**

Conjunto de herramientas que permiten recopilar, organizar, analizar e interpretar información. Puede ser descriptiva o inferencial.

- Descriptiva: Todas las herramientas que permiten presentar y resumir los datos a partir de tablas, gráficos o números.

- Inferencial: Permite, a partir de los datos de una muestra, caracterizar a una población bajo estudio.

**Conceptos**

. Población: conjunto de todos los objetos o individuos que tienen una característica común, la cual se quiere estudiar. Puede ser finita o infinita.

. Muestra: Es un subconjunto representativo de la población. Representativo significa que todos tienen que haber tenido exactamente la misma posibilidad de ser seleccionados a través del azar. El tamaño de la muestra depende de la población bajo estudio y su variabilidad. Si la población es muy variable, la muestra será más grande.

. Unidad elemental o de estudio: Objeto o sujeto que posee la cualidad o característica que se quiere evaluar. El número de unidades que conforman una muestra se denomina “n” y el número de la población “N”.

. **Variable**: Cualquier característica que se está evaluando en una unidad experimental y que puede variar de una unidad a otra. Se denomina aleatoria si el valor que adopta no se conoce hasta que ocurre el evento que se está esperando. Puede ser cualitativa o cuantitativa.

Variable Cualitativa: Las características de la unidad experimental están definidas en clases o categorías. Deben ser mutuamente excluyentes (no pueden ocurrir simultáneamente) y colectivamente exhaustivas (debe ocurrir al menos uno).

Variable Cuantitativa: Surgen de contar o medir las características bajo estudio. Pueden ser discretas (contar) o continuas (medir). En las variables discretas, el cero indica ausencia.

Escala de medición de la variable:

- Cualitativas: Nominal u ordinal.

- Cuantitativas: Intervalo (continuas) o razón (ambas).

Características para identificar la escala:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Orden | Distancia | Origen |
| Nominal | - | - | - |
| Ordinal | Sí | - | - |
| Intervalo | Sí | Sí | - |
| Razón | Sí | Sí | Sí |

**CÓMO ORGANIZAR Y RESUMIR LA INFORMACIÓN PARA VARIABLES CUALITATIVAS**

**Una variable**

Tabla de frecuencias:

|  |  |  |  |
| --- | --- | --- | --- |
| **xi** | **fi** | **fir** | **fi%** |
| x1 |  |  |  |
| x2 |  |  |  |
| … |  |  |  |
| xk |  |  |  |
|  |  |  |  |

xi= Posibles valores de la variable.

fi = Frecuencias absolutas = Número de veces que aparece xi en el conjunto de datos.

fir = = Frecuencias relativas.

fi% = = Frecuencias porcentuales.

Gráficas con barras, torta o bastones:

Barras: (x = xi; y = cualquier frecuencia)

Torta: 360° = 100%

**Dos variables**

*Ejemplo:*

*A = Tipo de local.*

*B = Tipo de propiedad.*

Tabla de doble entrada:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | *Tipo* | | | |
| *Propiedad* |  | *Alimentos* | *Ropa* | *Electrónica* | *TOTAL* |
| *Alquilada* | *10* | *30* | *5* | *45* |
| *Propio* | *25* | *50* | *3* | *78* |
| *TOTAL* | *35* | *80* | *8* | *123* |

Gráficas con barras múltiples, barras apiladas o pictogramas

**CÓMO ORGANIZAR Y RESUMIR LA INFORMACIÓN PARA VARIABLES CUANTITATIVAS DISCRETAS Y CONTINUAS**

**Discretas:**

•Si , realizar la tabla de frecuencias, agregando las frecuencias acumuladas:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **xi** | **fi** | **fir** | **fi%** | **Fi** | **Fir** | **Fi%** |
| x1 |  |  |  | f1 | fr1 | f%1 |
| x2 |  |  |  | f1+f2 | fr1+fr2 | f%1+f%2 |
| … |  |  |  | … | … | … |
| xk |  |  |  | f1+f2+...+fk = n | fr1+fr2+...+frk = 1 | f%1+f%2+...+f%k = 100 |
|  |  |  |  |  |  |  |

Gráficas con barras, bastones, escalones

**Discretas y continuas:**

•Si **n>15**, se utiliza una tabla con los datos agrupados en intervalos de clase:

Cantidad de intervalos: (Redondear al entero superior)

Rango:

Amplitud del intervalo: (Discretas: Redondear al entero superior | Continuas: Al decimal superior)

Intervalo: [Límite inferior – Límite superior]

1° [Mínimo – Mínimo+]

2° [Li – Li+]

…

Punto medio o marca de clase (PM o MC): Valor medio que representa al conjunto de todos los valores en ese intervalo.

Gráficas: Histograma de frecuencias, polígono de frecuencias u ojiva.

El polígono de frecuencias es para frecuencias sin acumular. Se puede hacer sobre el histograma o no. (Pasa por el centro de la barra)

La ojiva es para frecuencias acumuladas y también se puede hacer sobre el histograma o no. (Pasa por la esquina de la barra)

Diagrama de tallo y hoja (Funciona como tabla y como gráfico)

Tallo: Parte del número común a la mayoría de los datos. Generalmente, la parte entera.

Hoja: Resto del número que queda, ordenados de menor a mayor. Generalmente, la parte decimal.

No permitir más de 15 valores por hoja. Si se pasa, dividir cada valor del tallo en 2 o 5.

**MEDIDAS RESUMEN PARA VARIABLES CUANTITATIVAS**

.De posición: - Centrales: Media, mediana y modo.

- No centrales: Cuartiles.

.De dispersión: Rango, rango intercuartílico, desvío estándar, varianza y coeficiente de variación.

.De forma: Asimetría y curtosis.

**MEDIDAS DE POSICIÓN**

Indican algún punto en la distribución de los datos.

**Medidas centrales**

Indican el centro de la distribución.

**•Modo** (Mo)

Es el valor de la variable que más se repite. Se determina para datos agrupados o no. El modo no se ve afectado por la presencia de valores raros. Es la única medida resumen que se puede usar para una variable cualitativa.

Si hay dos valores con la misma frecuencia, hay dos modos. Si hay más de 2, no tiene modo.

Intervalo modal:

Para datos agrupados en intervalos.

Para determinar gráficamente el modo, se debe trazar en la barra más alta del histograma de frecuencias una cruz que vaya desde las esquinas de las barras laterales hasta arriba. El punto medio será el modo.

**•Mediana** (Mna)

Punto de equilibro del conjunto de datos. Es valor de la variable que separa en dos partes exactamente iguales al conjunto de datos. Los datos tienen que ordenarse de menor a mayor para poder determinarla. Al igual que el modo, la mediana no se ve influenciada por valores raros.

Para determinar la mediana se debe calcular la posición donde se encuentra el dato. Para ello hay que tener en cuenta si la cantidad de datos es par o impar:

Par: y (promedio entre los valores que estén en esas posiciones)

Impar:

Si los datos están agrupados, hay que buscar la posición de la mediana en las frecuencias acumuladas (), y el valor de la variable (o el promedio de los valores, si es par) que corresponda a esa frecuencia será la mediana.

En intervalos el valor de la mediana será:

Se puede hallar gráficamente en las gráficas que utilicen frecuencias acumuladas, donde la mediana será el 50% de los datos.

Propiedades:

- La sumatoria de las diferencias de cada valor de la variable con respecto de la mediana, dará un valor mínimo.

**•Media** ( o )

Es el promedio de los datos, es decir que surge de sumar todos los datos de la variable y dividir dicha suma por el total de los datos evaluados.

Si los datos están agrupados en tabla de frecuencias:

Si están agrupados en intervalos:

La media no se puede hallar gráficamente.

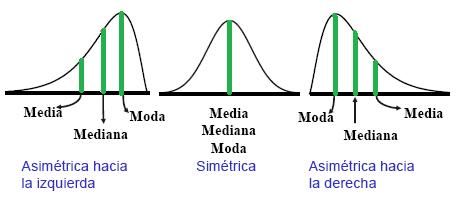
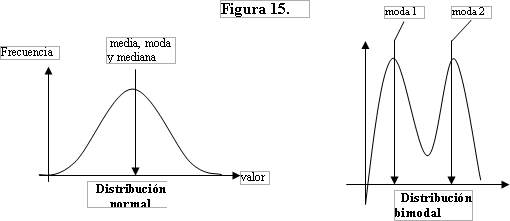
**Tipos de distribución**

Si entonces se tiene una distribución simétrica.

Si entonces la distribución es asimétrica negativa o a la izquierda.

Si entonces la distribución será asimétrica positiva o a la derecha.

En el caso de que haya dos modos, la distribución será bimodal.

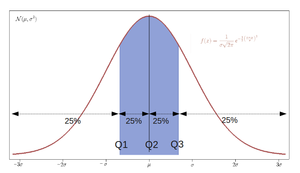
 

**Medidas no centrales**

Para caracterizar otros puntos en la distribución de los datos.

**•Cuartiles**

Dividen al conjunto de datos en 4 partes iguales.



En intervalos, la posición se busca en las frecuencias acumuladas () y se aplica la siguiente fórmula:

**MEDIDAS DE DISPERCIÓN**

Indican cuál es la variabilidad de un conjunto de datos. Nunca pueden ser negativas.

•**Rango**

Es una idea de la variabilidad en unidades de la variable. No se puede calcular en intervalos.

•**Rango intercuartílico**

Indica una medida de variabilidad del 50% central de los datos. Si es muy aproximado al valor del rango, significa que los datos están dispersos. De lo contrario, los datos están concentrados.

•**Varianza**

Es la suma de las desviaciones al cuadrado de los valores de la variable respecto de la media. Es consistente ya que usa todos los datos. Si los datos son de una población, se simboliza con .

Grado de libertad (n-1): Es el número de observaciones independientes para el cálculo de la varianza.

Si los datos están agrupados en tabla:

Si están en intervalos:

Propiedades:

- La varianza de una constante es cero.

- Jamás puede ser negativa.

- Si le sumamos una constante, la varianza se mantiene igual.

- Si se multiplica por una constante, la varianza será .

•**Desvío estándar**

Medida más potente para indicar la variabilidad.

•**Coeficiente de variación**

Es una medida adimensional. Permite determinar una medida de variabilidad de los datos en escalas o unidades diferentes.

**Tipificación**

Indica la distancia que tiene un valor de la variable con respecto a la media. Esta distancia está en unidades de desvío estándar.

**MEDIDAS DE FORMA**

Indican la forma de la distribución de los datos.

•**Coeficiente de asimetría**

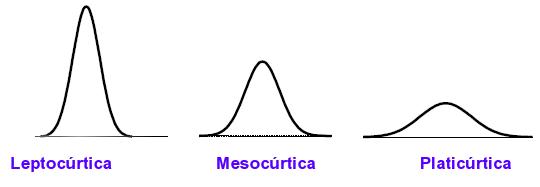
- Distribución simétrica:

- Distribución asimétrica derecha o positiva:

- Distribución asimétrica izquierda o negativa:

•**Curtosis**

Se puede calcular solo para distribuciones simétricas. Es el grado de apunamiento que tiene la distribución.

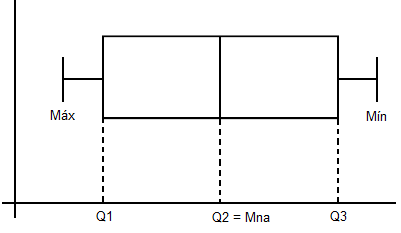
- Mesocúrtica:

- Leptocúrtica:

- Platicúrtica:

**DIAGRAMA DE CAJA**

Es una representación gráfica para visualizar en forma integral al conjunto de datos, a partir de las medidas resumen (Cuartiles). No se puede realizar con datos agrupados.



Si la cara resulta muy angosta, la distribución es leptocúrtica.

Si la mediana se encuentra en el centro, es una distribución mesocúrtica.

Para analizar la presencia de valores raros, se trazan vallas externas e internas. Los valores entre estas vallas serán atípicos y si hay valores antes de las vallas externas, serán valores raros extremos.

**PROBABILIDAD**

Es una incertidumbre, es decir, la falta de seguridad en la evaluación de un experimento.

- Experimento aleatorio: Operación cuyo resultado no se conoce hasta que el mismo no se realiza.

- Espacio muestral: Es el conjunto de todos los posibles resultados de un experimento aleatorio.

- Suceso: Es cualquier subconjunto del espacio muestral. Puede ser:

.Suceso elemental: Formado por un único elemento.

.Suceso compuesto: Formado por dos o más elementos del espacio muestral.

.Suceso seguro: Formado por todos los posibles elementos.

.Suceso imposible: Formado por elementos que no están definidos en el espacio muestral del experimento aleatorio que se está evaluando.

**Reglas de conteo**

Para definir cuántos son los resultados posibles de un experimento aleatorio, es decir, su espacio muestral.

- Regla de la multiplicación: Si se tienen dos tareas no excluyentes (pueden ocurrir en forma conjunta), donde la primera tarea se puede dar en *n1* formas diferentes y una segunda tarea en *n2* formas, entonces el total de resultados posibles será *n1.n2*.

- Regla de la división: Si se tienen dos eventos excluyentes entre sí, el espacio muestral se definirá como *n1+n2*.

Condiciones para determinar el tipo de conteo según orden y repetición:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Orden** | |
|  |  | Si | No |
| **Repetición** | Si | Permutación con repetición | Permutación con objetos indivisibles |
| No | Permutación, arreglo o variación  Si , | Combinaciones |

n = número de resultados posibles

r = tamaño del subgrupo

**Teorías de la probabilidad objetiva**

La probabilidad es inherente a los objetos o relaciones entre ellos.

Teoría clásica: Se basa en un principio matemático. Si no hay evidencia encontrada, todos los resultados posibles tienen que tener la misma probabilidad de ocurrencia.

: Número de resultados posibles que cumplen con la condición.

: Total de resultados posibles.

Teoría frecuentista: La única manera de determinar una probabilidad es repetir el experimento tantas veces como sea necesario hasta que una repetición más no cambie la probabilidad del experimento.

Enfoque subjetivo: Las probabilidades se van a determinar en función del conocimiento de un referente especializado.

**Principios o axiomas de la probabilidad**

•Ninguna probabilidad puede ser negativa.

•La probabilidad del espacio muestral es 1.

•La probabilidad de cualquier suceso está entre 0 y 1.

- Si dos sucesos definidos en el mismo espacio muestral son mutuamente excluyentes, la probabilidad de ocurrencia conjunta de esos sucesos va a ser igual a la suma de sus probabilidades individuales.

- Si no son mutuamente excluyentes, probabilidad de ocurrencia conjunta será:

- Probabilidad del complemento de A:

**Tipos de probabilidad**

Probabilidad simple

Probabilidad de un único suceso. (Puede tener más de un elemento)

.Probabilidad marginal: Dos sucesos en forma conjunta. Surge su cálculo de los márgenes de una tabla de doble entrada. Puede ser expresada como la suma de la intersección de las probabilidades:

Probabilidad conjunta

Intersección de dos sucesos.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **B** | | |
|  |  | B1 | B2 | T |
| **A** | A1 |  |  |  |
| A2 |  |  |  |
| T |  |  |  |

Probabilidad condicional

Consiste en determinar la probabilidad de que ocurra un evento A teniendo el conocimiento de que ya ocurrió un evento B. Está expresada como el cociente de una probabilidad conjunta definida por la condición sobre la probabilidad marginal de la condición.

P(Calcular/Condición)

- Regla de la multiplicación: Una probabilidad conjunta puede ser definida como el producto de una probabilidad condicional por la probabilidad marginal definida por la condición.

=

- Regla de la probabilidad total: Una probabilidad marginal o una simple puede ser expresada como la suma del producto de una probabilidad condicional por la marginal definida por la condición.

**Eventos independientes**

Dos eventos son independientes si la probabilidad de ocurrencia de uno de ellos, no afecta a la probabilidad de ocurrencia del otro suceso, es decir cuando la probabilidad de ocurrencia conjunta de los mismos es igual al producto de las probabilidades marginales en las que está definida.

La independencia o dependencia no influye en que sean mutuamente excluyentes.

**Teorema de Bayes**

Para calcular una probabilidad condicional a partir de otra probabilidad condicional (causa de ese evento). Es decir, vincula la probabilidad de A dado B con la probabilidad de B dado A.

Expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A.

Sean eventos mutuamente excluyentes y colectivamente exhaustivos, de tal forma que la probabilidad de cada uno de ellos es diferente de cero, y sea B un suceso cualquiera del que se conocen las probabilidades condicionales, entonces Bayes permite calcular la probabilidad de la condición inversa.

**VARIABLES ALEATORIAS**

La cuantificación de los resultados de un experimento aleatorio va a originar una variable aleatoria. Esta variable aleatoria es toda función que atribuye solo un número real a cada suceso elemental del espacio muestral. Puede ser discreta o continua.

**VARIABLE ALEATORIA DISCRETA**

Función de probabilidad

Si a cada uno de los resultados que puede adoptar una variable aleatoria se le asigna su probabilidad, se obtendrá una función de probabilidad, la cual es un conjunto de pares ordenados . Se representa con bastones o barras.

|  |  |
| --- | --- |
| ***xi*** | ***P*** |
| *x1* | *P1* |
| *x2* | *P2* |
| *…* | *…* |
| *xk* | *Pk* |

Función de distribución

Corresponde a la determinación de las probabilidades acumuladas. Se representa con escalones.

Esperanza o media

Propiedades:

.

.

.

.

Varianza

Propiedades:

.

.

.

Desvío estándar

**Distribuciones de probabilidad**

•Uniforme

Todos los resultados posibles que adopta la variable aleatoria tienen exactamente la misma probabilidad de ocurrencia.

a: Valor máximo de la variable.

b: Valor mínimo de la variable.

- Bernoulli

Tiene solo dos resultados posibles y se realiza una única vez.

p: Probabilidad de éxito del experimento, es decir, ocurre lo que se busca en las variables. (No significa que sea favorable)

q: Probabilidad de fracaso.

•Binomial - (x se distribuye en Binomial y tiene como parámetro n y p)

Surge de repetir *n* veces, en las mismas condiciones, el experimento Bernoulli. Cada experimento tiene ‘*p’* probabilidad de éxito y se mantiene constante en cada experimento, es decir que se trata de experimentos independientes.

n: Número de repeticiones

x: Número de éxitos del experimento. Tiene n+1 resultados posibles.

•Poisson -

Distribución de sucesos raros. Tiene las mismas características que la binomial. Presenta una probabilidad de éxito muy pequeña y el experimento se repite un número grande de veces. Se define en una unidad de espacio y tiempo. Al calcular, tener en cuenta que y estén en las mismas unidades de espacio y tiempo.

•Hipergeométrica -

También surge de repetir *n* veces un experimento aleatorio, pero la probabilidad de éxito se modifica de un experimento a otro, por lo que estos experimentos no son independientes entre sí. Surgen de muestreos son reposición o reemplazo y se sabe el tamaño que tiene esta muestra.

A: Número de casos que cumplen con la condición buscada.

N: Total de la población.

n: Tamaño de la muestra.

y

**VARIABLES ALEATORIAS CONTINUAS**

No existe una probabilidad puntual, siempre está definida en un intervalo.

Función de densidad

Permite calcular el área bajo la curva.

Si derivamos esta función, obtendremos la función de distribución, y al integrar la función de distribución, llegamos a la función de densidad.

**Distribuciones**

•Uniforme o rectangular

•Exponencial

Surge de evaluar una variable continua con distribución Poisson. Es la probabilidad del tiempo que se da entre dos sucesos.

•Normal

Forma simétrica perfecta. Es la base de la estadística inferencial. Como es complicado trabajar con valores continuos, tipificamos la variable.

→ Al tipificar:

**DISTRIBUCIONES DE PROBABILIDAD DERIVADAS DEL MUESTREO**

De una población tomamos *k* muestras sin reemplazo de igual tamaño (*n*). A cada una le calculamos la media y la varianza (el conjunto de medias y varianzas se puede considerar variables aleatorias). Pueden tener una distribución de probabilidad. El tipo de distribución depende de la variable que se evalúe y la cantidad de muestras (*k*).

Distribución en el muestreo de la media

* Si Siempre es así, no importa el tamaño de la muestra.

[Error estándar de las medias: error cometido al tomar las muestras]

* Si

Siempre y cuando

Teorema central del límite: la distribución muestral de la media es aproximadamente normal mientras el tamaño de la muestra sea grande.

[Tipifico y busco en la tabla de Z (normal)]

* Si Pero la varianza es desconocida.
* Si entonces
* Si entonces []

t de Student:

* Continua, simétrica, con forma de campana, toma valores de a .
* Su forma depende de los grados de libertad.
* Es más aplanada que la normal. Tiende a ser platicúrtica a medida que disminuyen los grados de libertad.
* Con grande aproxima a la normal.

Distribución de probabilidad para la diferencia de dos media

* Si ambas medias se distribuyen como normal y sus varianzas son conocidas e iguales, entonces la diferencia de las medias se distribuye normalmente.
* Si el desvío estándar de una, otra o ambas poblaciones es desconocido
  + Si y entonces la diferencia de las medias se distribuye como normal.
  + Si y/o entonces la diferencia de las medias se distribuye como

Distribuciones en el muestreo de muestras independientes

Media de las diferencias .

**Distribuciones en el muestreo relacionadas a la varianza**

Distribución –

Surge como la suma al cuadrado de variables estandarizadas a Z.

*Características:*

* Distribución continua (para variables continuas).
* Solo toma valores positivos.
* Distribución asimétrica positiva.
* Los grados de libertad son el *parámetro*
* Existen infinitas distribuciones.
* Tiende a ser simétrica a medida que aumentan los grados de libertad.

Distribución de Snedecor

Surge del cociente de dos varianzas muestrales. Se usa cuando interesa comparar dos varianzas para saber si son iguales o no.

*Características:*

* Definida entre y
* Conviene tomar a la varianza mayor como .
* Asimétrica positiva.
* Definida por dos grados de libertad:
* Existen infinitas distribuciones .
* A medida que ambas crecen en tamaño, se vuelve simétrica.

Distribución en el muestreo de la proporción de “…” (Objetos que cumplen con la condición deseada en la población. Ej: defectuoso)

probabilidad de la muestra.

Si y Uso Z probabilidad de la población.

Distribución en el muestreo de la diferencia de dos proporciones

[Condiciones]

**ESTADÍSTICA INFERENCIAL**

A partir del estudio de la muestra, permite estimar características de la población.

Formas de inferencia estadística:

* Inferencia Puntual
* Inferencia por Intervalo
* Prueba/Test de Hipótesis

**Inferencia Puntual**

Estimar el parámetro poblacional a partir de un único valor obtenido en la muestra.

Una estadística puntual es un buen estimador muestral del parámetro poblacional si cumple con cuatro propiedades:

* Insesgabilidad

Es insesgado si el valor esperado del estimador es igual al parámetro poblacional.

estimador de un parámetro poblacional

parámetro poblacional

Parámetro

Demostración:

* Eficiencia

Un estimador es más eficiente que un estimador si

* Suficiencia

Un estimador es suficiente si transmite toda la información posible que contiene la muestra (Ejemplo: S es más suficiente que R).

* Consistencia

Un estimador es consistente cuando su valor tiende a adoptar el valor del parámetro a medida que el tamaño de la muestra aumenta.

cuando

Metodologías para hallar el estimador muestral

1. Método de los mínimos cuadrados: busca reducir el término de error (diferencia entre lo que se estima y el valor del parámetro). Se basa en la distribución Normal.
2. Metodología de máxima verosimilitud: se basa en el supuesto de encontrar aquel estimador del parámetro que bajo una supuesta distribución de probabilidad, sea el valor más probable. Permite trabajar con cualquier distribución.

**Inferencia por intervalo**

Intervalo de confianza Cuando a la estimación puntual se le asocia un grado de incertidumbre o error, se determina un intervalo de confianza.

Se da un rango alrededor del cual se encuentra el verdadero valor del parámetro poblacional. La asignación del rango se calcula en función del error que estoy dispuesto a cometer: (riesgo de error tipo 1) indica la probabilidad de rechazar algo que es verdadero.

nivel de confianza del intervalo (probabilidad de contener al verdadero valor del parámetro)

a y b son los límites del intervalo. Están definidos en función del parámetro al cual se le construye el intervalo y la distribución de probabilidad que yo utilice para construir el intervalo.

**Prueba/Test de Hipótesis**

Prueba que permite testear si el parámetro poblacional que quiero evaluar cumple o no con una condición que quiero probar.

Hipótesis: supuesto que queremos testear. Jamás se plantean con estadísticas muestrales.

Existen dos tipos de hipótesis:

* Hipótesis Nula (H0): plantea o presenta lo que se conoce del parámetro poblacional a evaluar. Siempre su planteo va a contener el signo
* Hipótesis Alternativa (H1): va a plantear lo que se quiere probar o testear acerca del parámetro poblacional. Siempre va a contradecir a la nula.

Dos tipos de pruebas:

* Prueba bilateral (cuando la hipótesis nula tiene )

Testeo ambas posibilidades (que el parámetro sea mayor y menor)

* Prueba unilateral (cuando la hipótesis nula tiene )

Interesa probar si es mayor o menor.

Dos tipos de error:

* Error Tipo 1 o : error que se comete al rechazar una H0 que es verdadera.
* Error Tipo 2 o : error que se comete al aceptar una H0 que es falsa.

|  |  |  |
| --- | --- | --- |
|  | H0 Verdadera | H0 Falsa |
| Acepto | Decisión correcta | Error Tipo 2 o |
| Rechazo | Error Tipo 1 o | Decisión correcta |

Para un tamaño de muestra fijo, a medida que aumenta el riesgo de error , disminuye y viceversa.

Si aumenta el tamaño de la muestra, ambos errores disminuyen.

Se suele hablar de “potencia de la prueba” en vez de : (potencia más cercana a 1, prueba más exacta). Potencia de la prueba significa cuán potente es la prueba para llegar a una decisión correcta.

Estadístico de prueba: valor correspondiente a una distribución de probabilidad específica definida por la hipótesis a testear que va a permitir llegar a una conclusión. El estadístico de prueba se selecciona en función del parámetro planteado en las hipótesis. Permite afirmar si acepto o no la hipótesis nula.

Reglas para decidir si se acepta o no la hipótesis nula

Pasos de una prueba/test de hipótesis:

1. Formular las hipótesis
2. Definir el riesgo de error que se está dispuesto a cometer
3. Definir el estadístico de prueba
4. Definir las reglas para decidir si se acepta o no la hipótesis nula
5. Conclusión

**CONTROL ESTADÍSTICO DE LA CALIDAD**

Calidad: idoneidad para el uso. Un producto es de calidad si cumple con la función para la que fue diseñado.

Dos tipos de calidad:

* Calidad de diseño: resultado de las decisiones de ingeniería.
* Calidad de conformidad: calidad que permite decir que un producto o proceso cumple con las especificaciones dadas.

Control estadístico de la calidad (CEC): conjunto de herramientas estadísticas que permiten recopilar, estudiar y analizar la información de procesos repetitivos para poder tomar decisiones encaminadas a la mejora de los mismos.

Requisitos del proceso a evaluar para poder hacer CEC:

* Proceso repetitivo
* Proceso mesurable (observable o cuantificable)

Causas de variación de un proceso o producto:

* Controlables (o asignables): modificando el diseño, se puede llegar a controlarlas. Asignables porque se puede determinar su origen.
* No controlables: no son factibles de ser identificadas y deben su origen exclusivamente al azar.

El control estadístico de procesos va a trabajar para detectar las causas asignables de variación. Un proceso está bajo control si se ejercen sobre él causas de variación sólo debidas al azar.

Carta o Gráfico de control: herramienta para detectar causas de variabilidad controlables. Gráfico xy que representa a través del tiempo valores de algún topo de medición de un proceso continuo.

El proceso está bajo control estadístico si los puntos evaluados no salen de las líneas de control. Además esos puntos deben tener la característica de variar aleatoriamente alrededor de la línea media (es decir, sin seguir un patrón).

Pueden ser:

* Gráfico de control para variables: cuando se evalúa una variable cuantitativa generalmente continua. Siempre supone que la distribución de probabilidad de la variable es normal.
* Gráfico de control por atributos: se utiliza cuando la característica de calidad a evaluar tiene una respuesta binaria (dos respuestas posibles): aceptable o no aceptable. Puede ser:
  + Proporción de defectuosos
  + Número de defectuosos