UNIVERSIDAD DEL SURESTE

∑ S T A D Í ∫ T I C A

Ing. Aldo Irecta Nájera

Introducción

- Los procesos estadísticos son herramientas cuantitativas y cualitativas que permiten evaluar magnitudes de lo real. Por lo tanto diversas ciencias tanto naturales como sociales toman como disciplina de apoyo y complemento nociones estadísticas que de manera eficaz permitan la evaluación y descripción de fenómenos mediante el cálculo de operaciones matemáticas a través de las cuales se puede caracterizar y determinar los aspectos más significativos de una población o muestra.

Origen de la Estadística como Disciplina Científica

- El Origen de la estadística se remonta a los comienzos de la historia, ya desde el cuarto milenio a.C. los chinos, griegos y egipcios realizaban censos de población y tabulaciones de las actividades agrícolas.

- Las Primeras tentativas orientadas a sistematizar los diversos procedimientos matemáticos utilizados en esas civilizaciones surge en Alemania en el S. XVII, influenciadas por el estudio de los juegos de azar y el cálculo de probabilidades.

- La estadística pese a su desarrollo, aparece en la modernidad alrededor de 1850, con la definición derivada de la raíz “Status” (Estado) ligado a la actividad gubernamental abocada a conocer extensiones territoriales de cierta población, habitantes residentes en ella y cantidad de impuestos a obtener de ella.

- El término “estadística” proviene de la palabra italiana “statista”, utilizada por primera vez por Gottfried Achenwell (1719-1772). Su uso fue difundido por Sir Jhon Sinclair en su obra “Statistal Account of Scotland” (1719-1799), “Informe estadístico sobre Escocia”.

La Estadística en las Ciencias Sociales

- Representa específicamente para la Sociología, la realización de operaciones con números que expresan valores de mediciones para satisfacer ciertos supuestos.

- “La estadística expresa cierto estado del alma colectiva” Durkheim. Por ello es importante que el investigador social considere que no existe ningún sustituto estadístico apropiado para una correcta conceptualización teórica, base para lograr un buen uso de las técnicas estadísticas.

- En términos metodológicos la **“**operacionalización de conceptos” surge como fase de inducción de los métodos estadísticos en la investigación, convirtiéndose en el paso intermedio que une la formulación teórica de un problema y la medición de variables.

- Proporcionan al investigador social la posibilidad de resumir y extraer información relevante de las mediciones observadas.

- Para la aplicación de técnicas de medición es de suma importancia la definición del tipo de variable, cuantitativa o cualitativa. Así como sus niveles de medición; nominal, ordinal, intervalo y razón respectivamente.

Estadística Descriptiva

- La estadística descriptiva suministra los instrumentos que permiten el salto de las observaciones a la inferencia, siendo el resumen de las observaciones el paso previo.

- La Estadística Descriptiva se dedica a expresar regularidades propias de las observaciones ó conjunto de datos, a través de operaciones numéricas para permitir la cuantificación.

- La Estadística Descriptiva agrupa todas aquellas técnicas y procedimientos que permiten caracterizar una muestra y población, algunas de estas técnicas son las medidas de tendencia central, dispersión, posición, regresión y correlación.

Algunos métodos para organizar datos

- Matriz de Datos
  - Es una forma de sintetizar la información recogida de la realidad para investigar un problema y tratar de obtener conocimiento científico que intente explicar dicho problema.
- Composición: Dimensión, Unidades, Valores.

- Distribuciones de Frecuencias: tablas de datos referentes al número de veces en las que se repite la categoría de una variable que graficado, refleja la forma de la distribución construida.
  - Absolutas: Reflejan el número de observaciones del conjunto de datos que cae en cada una de las clases.
  - Relativas: Permite expresar la frecuencia de cada valor con una fracción o porcentaje del total del número total de observaciones.
    - Proporciones: Son cocientes que indican la relación existente entre una cantidad y el total de las unidades consideradas.
    - Porcentajes: Permite estandarizar en relación con el volumen calculando el número de objetos que habría en una categoría si el total de los casos fuese 100.

Medidas de Tendencia Central

Lugar donde se centra el conjunto de datos de una distribución particular en la escala de valores.

- **Media:** Es el valor típico o promedio, representativo del conjunto de datos considerados.
  - Ventajas: Toma en consideración la realidad de todo el conjunto de datos.
  - Desventajas: Puede verse afectada por valores extremos no representativos del resto de los datos.
- **Mediana:** Es un valor que divide la distribución de datos en 2 partes iguales, tal que, el conjunto de datos por encima de este sea igual al número de datos por debajo de la misma.
  - Ventajas: Los valores extremos no afectan a la mediana tan intensamente como a la media.
  - Desventajas: Ciertos procedimientos estadísticos que utilizan la mediana son más complejos que aquellos que utilizan la media, es por ello que, si deseamos utilizar una estadística de muestra para estimar un parámetro de población, la media es más cómoda.
- **Moda:** Es el valor que más se repite en una distribución de datos.
  - Ventajas: No se ve afectada por valores extremos dado que se escoge el valor más frecuente, puede emplearse aún cuando existan clases de extremo abierto.
  - Desventajas: Cuando los datos son multimodales resulta complejo interpretar y comparar

Medidas de Dispersión

Son aquellas que permiten reflejar la distancia entre los valores de la variable con respecto al valor central de la distribución.

Medidas de Dispersión Absolutas

Son aquellas no comparables entre diferentes muestras

- **Amplitud o Rango: Nos ofrece una visión de donde a donde se expresan los datos. Es la diferencia entre observaciones extremas.**
- **Varianza: Es la media de los cuadrados de las diferencias entre cada valor de la variable y la media aritmética de la distribución.**
  - **Desventajas: Sensibilidad con respecto a los valores extremos, sus unidades son al cuadrado por ello es difícil de interpretar.**
- **Desviación Típica: Refleja la distancia de cada valor con respecto a la media. Es la raíz cuadrada de la varianza.**
  - **Ventajas: Tiene las mismas unidades que la variable, es más estable que el rango, toma en consideración el valor de cada dato.**

Medidas de Dispersión Relativa

Son aquellas que nos permiten comparar muestras diferentes

- **Coeficientes de Variación de Pearson: Nos permite comparar el grado de dispersión de muestras cuyas unidades son diferentes o donde las medias son extremadamente desiguales.**
- **Coeficiente de Variación Medianal: Refleja el grado de dispersión de muestras diferentes con respecto a la mediana.**

Cuantiles

Son valores que dividen la distribución en partes iguales, es decir; en intervalos que comprenden el mismo número de valores.

Los cuantiles son las medidas de posición que determinan mediante operaciones matemáticas la ubicación de los valores, en la distribución.

- **Cuartiles: Son los tres valores que dividen al conjunto de datos ordenados en cuatro partes porcentualmente iguales.**
- **Deciles: Son los nueve valores que dividen al conjunto de datos en diez partes porcentualmente iguales.**
- **Percentiles: Son las medidas más utilizadas para propósitos de ubicación o clasificación, dividen la sucesión en cien partes porcentualmente iguales.**

INFERENCIA ESTADÍSTICA

Rama de la estadística que utilizando información a partir de muestras de población, se apoya en las teorías de la probabilidad para realizar suposiciones, de que en determinado momento y lugar, o bajos ciertas condiciones, sucederán fenómenos específicos en menor ó mayor medida, sin tener la certeza de ocurrencia de ellos.

Por lo tanto, se apoya en el cálculo de probabilidades para atender dos problemas fundamentales: La estimación y La Contrastación de Hipótesis. En ambas realizamos inferencias acerca de las características de población.

Manejar la incertidumbre que acompaña toda acción social para la toma de decisiones efectivas.

Términología básica

- **Población o Parámetro**

  Se refiere a la totalidad de posibles observaciones o elementos de la realidad que se estén considerando en una situación dada. Las características de una población se suelen tomar generalmente como sus parámetros (N, μ, σ,).

  Población Finita: Indica que la población tiene un tamaño establecido o limitado.

  Población Infinita: Hace referencia a una población en la que no es posible enumerar u observar todos los elementos que la conforman.

- **Muestra o Estadístico**

  Es una porción o parte de las observaciones o elementos tomados de una población dada. Toda característica de una muestra suele llamarse por lo general estadística. (n, x, s,).

  Con reemplazo: Alude a la no incorporación del elemento muestreado en la población después de haber sido escogido, y antes de elegir al próximo.

  Sin reemplazo: Indica que pronto agotaremos todos los elementos de la población

  Fracción de muestreo: Porción de la población contenida en una muestra

Estimaciones

Conjunto de técnicas que permiten dar un valor aproximado de un parámetro de una población a partir de los datos proporcionados por una muestra.

Estimación puntual

La estimación puntual utiliza solo un numero para estimar el parámetro de población desconocido. Sin embargo, es insuficiente debido a que sólo tiene dos opciones: es correcta o está equivocada.

Estimación de intervalos

La estimación de intervalo utiliza un rango de valores para estimar el parámetro de población desconocido.

Estimador

Se trata de un estadístico de la muestra utilizado para estimar un parámetro de la población.

Un Buen Estimador Debe Ser

- **Insesgado:** La media de la distribución muestral de las medias de la muestra tomadas de la misma población es igual a la media de la población misma
- **Eficiente:** Menor error y menor desviación estándar de la distribución muestral posible
- **Consistente:** Si al aumentar la muestra se tiene casi la certeza de que el valor de la estadística se aproxima bastante al parámetro poblacional buscado
- **Suficiente:** Si utiliza tanta información de la muestra que ningún otro estimador puede extraer, tal que, proporcione la mayor información adicional acerca del parámetro de población que se está estimando

Estimaciones de intervalo de la media: muestras grandes

$$P\left(x - z\,\alpha/2.\ \sigma\sqrt{n} < \mu < x + z\,\alpha/2.\ \sigma/\sqrt{n}\right) = 1 - \alpha$$

- Si $n \geq 30$, el teorema del límite central nos permite usar la distribución normal como distribución de muestreo.
- Cuando se conoce la desviación estándar de la población (σ). Si no se conoce la desviación estándar de la población, podemos estimarla a partir de la desviación estándar de la muestra $\sigma = s$. $\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$
- Si tenemos un tamaño de población finita sin reemplazo y nuestra muestra constituye más del 5% de la población, aplicamos el factor de corrección para derivar el error estándar. $\sqrt{\frac{N-n}{N-1}}$

Estimaciones de intervalo de la proporción: muestras grandes

$$P\left(\hat{p} - z\,\alpha/2.\ \sqrt{\frac{P.Q}{n}} < P < \hat{p} + z\,\alpha/2.\ \sqrt{\frac{P.Q}{n}}\right) = 1 - \alpha$$

- Teóricamente la distribución binomial es la distribución correcta a utilizar para estimar una proporción de población. Sin embargo, a medida que aumenta (n) la distribución binomial se aproxima a la normal. Se recomienda que n.p cómo n.q sean al menos 5 cuando se aproxime con la distribución normal.
- La media de la distribución de muestreo de la proporción de éxitos $\mu_{\hat{p}} = P$
- El error estándar de la proporción $\sigma_{\hat{p}} = \sqrt{\frac{P.Q}{n}}$

Estimaciones de intervalo con distribución t de Student

$$P\left(x - t_{\alpha/2}.\hat{\sigma}\sqrt{n} < \mu < x + t_{\alpha/2}.\hat{\sigma}/\sqrt{n}\right) = 1 - \alpha$$

- Si el tamaño de muestra (n) ≤ 30, se desconoce la desviación estándar de la población (σ), y la población es normal ó aproximadamente normal.
- Como primer paso debemos estimar la desviación estándar de la población (σ) a partir de la muestra (s). $\hat{\sigma} = s$ y calcular el Error estándar estimado de la media de población $\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$
- Si tenemos un tamaño de población finita sin reemplazo y nuestra muestra constituye más del 5% de la población, aplicamos el factor de corrección para derivar el error estándar.
- Grados de libertad (∂): Los utilizamos cuando elegimos distribución t para estimar una media de población. $\partial = n - 1$. Existe una distribución t para cada tamaño de muestra o grado de libertad posible.

Contraste de Hipótesis

Una suposición que hacemos con respecto a un parámetro de población. Para probar la validez de esta suposición:

1. Recolectamos datos de muestra.
2. Producimos estadísticas muéstrales.
3. Determinamos la diferencia entre nuestro valor hipotético y un parámetro hipotético de población. Mientras más pequeña la diferencia mayor será la probabilidad de que nuestro valor sea correcto.

Tipos de Hipótesis

Paramétricas: Hipótesis suceptibles de medición y tratamiento estadístico con referencia a un parámetro de población preestablecido.

No Paramétricas: No tenemos parámetros, se trabaja con frecuencias esperadas y observadas.

Sistema de Hipótesis

Hipótesis Nula (Ho): Simboliza la suposición que deseamos probar. El principio es rechazarla. Se llama nula porque tiene la igualdad, es decir; se reserva el cero y por tanto es más preciza. Su complemento es la hipótesis alternativa.

Hipótesis Alternativa (H^1): Simboliza el rechazo de nuestra suposición (Ho) y cumplimiento de algún otro evento. Su complemento es la hipótesis nula. Su función es orientar el contraste de hipótesis, (>) mayor que ó (<) menor que.

Procedimiento Básico para realizar el Contraste de Hipótesis

1. Formular la Hipótesis nula (Ho) y la Hipótesis alternativa (H^1)

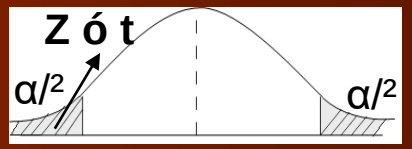
2. Seleccionar el tipo de distribución a usar o estadístico a contrastar en la prueba:

Distribución normal	Tamaño de muestra (n) es mayor que (>) 30	Se conoce la desviación estándar de la población σ
	Tamaño de muestra (n) es ≤ 30 y suponemos que la población es normal o aprox. Normal.	Se conoce la desviación estándar de la población σ
Distribución t	Tamaño de muestra (n) es ≤ 30 y suponemos que la población es normal o aprox. Normal a medida que aumentan los grados de libertad (∂)	No se conoce la desviación estándar de la población σ

3. Selección del nivel de significación: Consiste en decidir que criterio utilizar para confirmar si se acepta o no Ho. No existe un nivel estándar para probar hipótesis, todo depende de el error dispuesto a cometer. Los más usados 1%, 2%, 5% y 10%.

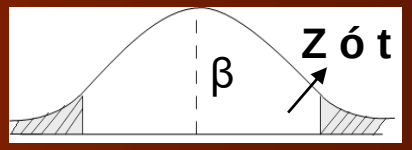
- Error tipo I: Rechazar la hipótesis nula cuando es cierta (α)

  1 - σ



- Error tipo II: Aceptar la hipótesis nula cuando es falsa (β)

  1 - β

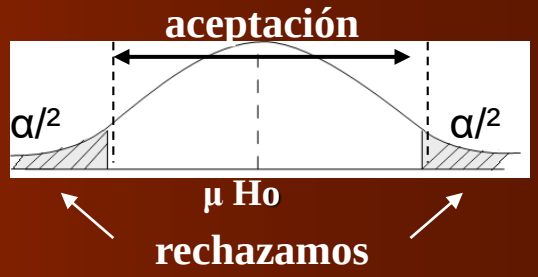


4. Definición de la región de aceptación o rechazo

Bilateral

Ho μ=μ Ho
H¹ μ≠μ Ho

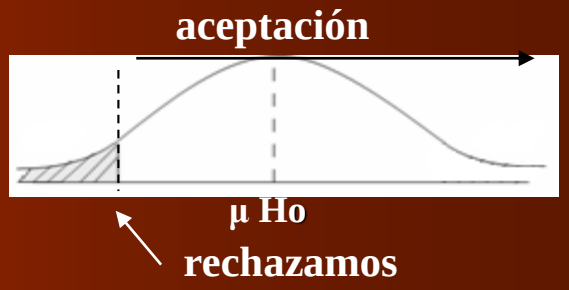
Se rechaza la hipótesis nula si la media de muestra es mayor o menor que la media hipotética de población



Unilateral

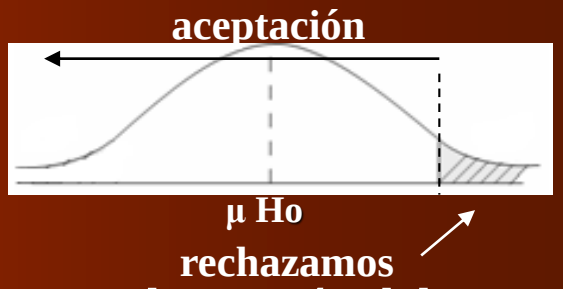
Ho μ=μ Ho
H¹ μ<μ Ho

Se rechaza la hipótesis nula si la media de la muestra es menor que la media hipotética de población



Ho μ=μ Ho
H¹ μ>μ Ho

Se rechaza la hipótesis nula si la media de la muestra es mayor que la media hipotética de población



5. Realizar los cálculos correspondientes: Error estándar y estandarización del estadístico de la muestra.

6. Interpretación de los resultados y toma de desiciones.

Contraste de Hipótesis: Prueba de una muestra.

Media

1. Establecemos las hipótesis, tipo de prueba y nivel de significación
•2. Elegimos las distribución apropiada.

Distribución Normal:
Desconocemos σ de la población
Muestra ≥ 30

Distribución " t " Student:
Desconocemos σ de la población
Muestra ≤ 30

3. Establecemos el nivel de significación y tipo de error a cometer.
4. Definición de la región de aceptación o rechazo.
5. Cálculo del **error estándar de la media** $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Como desconocemos σ de la población asumimos que $\hat{\sigma} = s$ entonces la fórmula quedaría de la siguiente manera:

Error estándar estimado de la media $\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$

Si conocemos el tamaño de la población y la fracción de muestro n / N es mayor a 0.05 aplicar el factor de corrección para poblaciones finitas $\sqrt{\frac{N-n}{N-1}}$

Estandarizamos los valores originales de x con la siguiente fórmula para distribución normal: $Z = \frac{\bar{x} - \mu_{Ho}}{\hat{\sigma}_{\bar{x}}}$ ó con la siguiente fórmula para distribución t: $t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}}$

6. Interpretación de los resultados y toma de decisiones

Proporción

1. Establecemos las hipótesis, tipo de prueba y nivel de significación

2. Elegimos las distribución apropiada. La distribución binomial es teóricamente la apropiada para trabajar con proporciones, porque los datos son discretos. Sin embargo, al aumentar la muestra, la distribución binomial se aproxima a la normal. Siempre que n.p y n.q cada una sea al menos 5, se puede utilizar la distribución normal como aproximación a binomial.

3. Establecemos el nivel de significación y tipo de error a cometer.
4. Definición de la región de aceptación o rechazo.
5. Cálculamos el error estándar de la proporción $\sigma_{\bar{p}} = \sqrt{\frac{PHo \cdot QHo}{n}}$
y estandarizamos $Z = \frac{\bar{p} - PHo}{\sigma_{\bar{p}}}$

6. Interpretación de los resultados y toma de desiciones

Contraste de Hipótesis: Prueba de dos muestras

- Para estudiar dos poblaciones la distribución que nos interesa es la **distribución muestral de la diferencia entre medias muéstrales.** Ésta se obtiene de la toma de muestras de distintas poblaciones y de su diferencia con respecto a las dos medias.

Diferencia postiva: Si $\bar{x}_1 > \bar{x}_2$ **Diferencia negativa**: Si $\bar{x}_1 < \bar{x}_2$

- La media de la distribución muestral de la diferencia entre las medias muéstrales se denota $\mu_{\bar{x}_1} - \mu_{\bar{x}_2}$. Si las medias muestrales son de la misma población entonces se anulan, de acuerdo al teorema del límite central.
- La desviación estándar de las diferencias entre medias muéstrales, se conoce como error estándar de la diferencia entre medias.

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}$$

- Si no conocemos las dos desviaciones estándar de la población, estimamos el error estándar utilizando el mismo método $\hat{\sigma} = s$

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(\hat{\sigma}_1)^2}{n_1} + \frac{(\hat{\sigma}_2)^2}{n_2}}$$

Prueba para diferencia entre dos medias: Muestras grandes

- Si la muestra es ≥ 30 procedemos a estandarizar los valores con distribución normal.

$$Z = \frac{(\bar{x}_1-\bar{x}_2) - (\mu_{\bar{x}_1} - \mu_{\bar{x}_2})}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}}$$

Prueba para diferencia entre dos medias: Muestras pequeñas e independientes entre sí

- Cuando los tamaños de las muestras son ≤ 30, no se conoce σ de la población y cada muestra se eligió de manera independiente de otra, usamos la distribución "t" student, pero con ciertos cambios técnicos:

1. Como no se conoce σ de la población debemos estimarla. Sabiendo que la muestra es pequeña y suponiendo que las desviaciones de población son iguales ($\sigma^1 = \sigma^2$), debemos usar un promedio ponderado de las desviaciones de ambas muestras (s^1, s^2). El peso de cada muestra son el número de grados de libertad (∂). Este promedio ponderado se le conoce como **estimación conjunta de σ.**

$$Sp = \frac{\sqrt{(n^1-1)\,.\,(s^1)^2 + (n^2-1)\,.\,(s^2)^2}}{n^1 + n^2 - 2}$$

2. Cálculamos el **error estándar estimado de la diferencia entre dos medias muestrales**

$$\hat{\sigma}_{\bar{x}^1-\bar{x}^2} = Sp\,.\,\sqrt{\frac{1}{n^1} + \frac{1}{n^2}}$$

- Luego procedemos con la estandarización de las diferencias de las medias de la muestra

$$t = \frac{(\bar{x}^1-\bar{x}^2) - (\mu_{\bar{x}^1} - \mu_{\bar{x}^2})}{\hat{\sigma}_{\bar{x}^1-\bar{x}^2}}$$

Prueba para diferencia entre dos medias: Muestras pequeñas y dependientes

- El uso de muestras dependientes (o apareadas), permite llevar a cabo análisis más precizos, porque permite controlar factores externos.
- Se sigue el procedimiento básico anterior de la prueba de hipótesis. Las únicas diferencias consisten en:

1. Se emplea la misma fórmula utilizada para el cálculo del error estándar estimado de la media para una sola muestra. $\hat{\sigma} = s$ $\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}}$ y estandarizamos $t = \frac{\bar{x} - \mu}{\hat{\sigma}_{\bar{x}}}$

   Conceptualmente, tenemos una muestra observada dos veces, es decir; dependientes.

2. Ambas muestras deber ser del mismo tamaño.

¿Cuándo tratar las muestras como dependientes o independientes?

Prueba entre proporciones: Muestras grandes

• El procedimiento general a seguir es muy parecido al de comparación de dos medias utilizando muestras independientes. La única diferencia importante se da en la forma de encontrar un estimación para el **error estándar de la diferencia entre dos proporciones de muestra,** para ello es necesario utilizar las proporciones combinadas de éxito de ambas muestras y obtener una **proporción global estimada de éxito en dos poblaciones.**

• Si la muestra es ≥ 30 usamos la distribución normal para aproximar a la binomial.

Prueba de dos colas (bilateral)

1. Establecemos las hipótesis, tipo de prueba
2. Seleccionar el tipo de distribución a usar o estadístico a contrastar en la prueba:
3. Establecemos el nivel de significación y tipo de error a cometer.
4. Definición de la región de aceptación o rechazo. En este caso es bilateral.
5. Cálculo del **error estándar de la diferencia entre proporciones**

$$\sigma_{\bar{p}^1 - \bar{p}^2} = \sqrt{\frac{p^1 . q^1}{} + \frac{p^2 . q^2}{}}$$

Si no se conocen los parámetros de la población, es necesario estimarlos a partir de la muestra, la fórmula quedaría:

$$\hat{\sigma}_{\bar{p}^1 - \bar{p}^2} = \sqrt{\frac{\bar{p}^1 . \bar{q}^1}{n^1} + \frac{\bar{p}^2 . \bar{q}^2}{n^2}}$$

• **Proporción global estimada de éxito en dos poblaciones.**

$$\hat{p} = \frac{n^1 . \bar{p}^1 + n^2 . \bar{p}^2}{n^1 + n^2} \qquad \hat{q} = 1 - \hat{p}$$

• **Error estándar estimado de la diferencia entre dos proporciones de muestra, usando estimaciones combinadas.**

$$\hat{\sigma}_{\bar{p}^1 - \bar{p}^2} = \sqrt{\frac{\hat{p}^1 . \hat{q}^1}{n^1} + \frac{\hat{p}^2 . \hat{q}^2}{n^2}}$$

Estandarizamos la diferencia $Z = \frac{(p^1 - p^2) - (P^1 - P^2)}{\hat{\sigma}_{p^1 - p^2}}$ → Se anula si suponemos que no hay diferencia entre las dos proporciones de población

Prueba de una cola (unilateral)

1. Establecemos las hipótesis, tipo de prueba. Con atención a las referencias sobre una proporción > ó <
2. Seleccionar el tipo de distribución a usar o estadístico a contrastar en la prueba:
3. Establecemos el nivel de significación y tipo de error a cometer.
4. Definición de la región de aceptación o rechazo. En este caso es unilateral.
5. Cálculo del **error estándar estimado de la diferencia entre proporciones,** para ello necesitamos calcular la **proporción global estimada de éxito.** Luego proceder a **estandarizar las proporciones de muestra (** $\bar{p}^1 - \bar{p}^2$)

Prueba de Independencia: Prueba Ji - cuadrada

- Nos permite probar si más de dos proporciones de población pueden ser consideradas iguales.
- Determinar si los atributos de una población clasificada en categorías, son independientes entre sí
- Los datos vienen dados a partir de una tabla de frecuencias.
- Partimos del principio de Independencia $P(A \cap B) = P(a) \cdot P(b)$

1. Establecemos las hipótesis.
2. Seleccionar el tipo de distribución a usar o estadístico a contrastar en la prueba. (∫) χ^2
3. Establecemos el nivel de significación, si no viene dado.
4. Definición de la región de aceptación y rechazo.
5. Calculamos la frecuencias esperadas para cada celda de la tabla de frecuencias
(fe) Total de filas x Total de columnas / Total general
6. Calculamos el estadístico Ji - cuadrado $\chi^2 = \frac{\Sigma (fo - fe)^2}{fe}$
7. Calculamos los grados de libertad (fila - 1) . (columna - 1)

Bosquejamos la distribución y observamos si los criterios Son Independientes o No Independientes.