Capítulo 13

Análisis de datos (I)

ESIC EDITORIAL

Investigación de mercados
Prof. Verónica Rosendo Ríos

"Es horrible hablar bien y estar equivocado"

SOPHOCLES

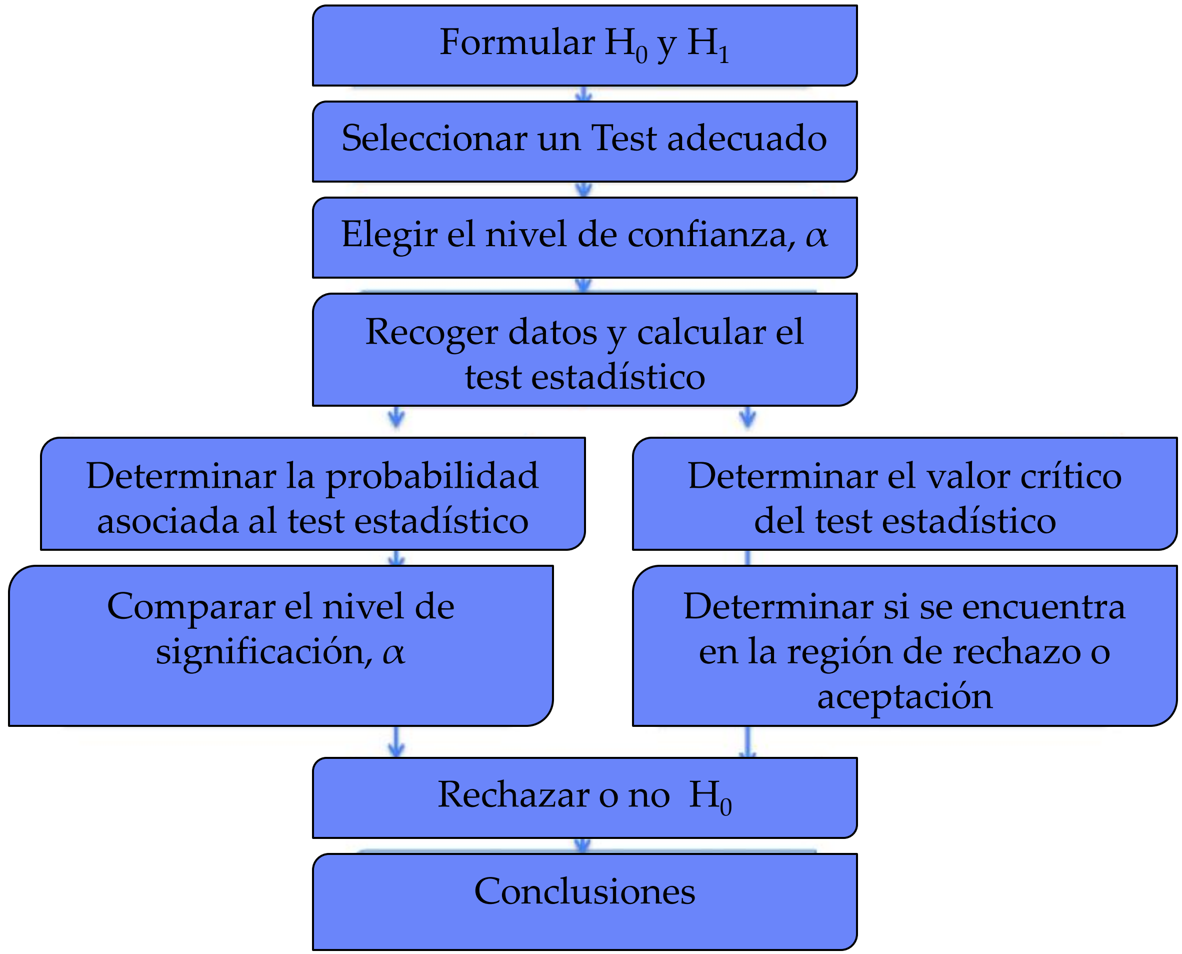
CONTENIDOS

★ DISTRIBUCIÓN DE FRECUENCIAS:

El análisis estadístico se puede dividir en varios grupos:

- **El análisis estadístico univariate** prueba hipótesis que involucran solo a una variable.
- **El análisis estadístico bivariante** prueba hipótesis que implica a <u>dos variables</u>.
- **El análisis estadístico multivariante** prueba hipótesis que implican a <u>múltiples (tres o más) variables o conjunto de variables.</u>

★ Procedimiento del contraste de hipótesis:



★ **Paso 1**: Formulación de la hipótesis

- La **hipótesis nula (o hipótesis de contraste)**

- La **hipótesis alternativa**

La hipótesis nula es siempre la hipótesis que se prueba. La hipótesis nula se refiere a un valor específico del parámetro de la población (por ejemplo, μ, σ, o π).

¡Una hipótesis nula puede ser rechazada, pero nunca puede ser aceptada en base a una sola prueba!

✓ Contrastes o tests estadísticos

Un contraste estadístico puede tener como resultado:

1. Que la hipótesis nula se rechaza y la hipótesis alternativa se acepta o…
2. Que la hipótesis nula NO se puede rechazar en base a la evidencia.

☑ Sería **incorrecto**, sin embargo, concluir que dado que la hipótesis nula no es rechazada, puede aceptarse como válida. En las pruebas de hipótesis clásicas, no hay forma de determinar si las hipótesis nulas son verdaderas.

✓ Contrastes estadísticos

En la investigación de mercado, la hipótesis nula está formulada de tal manera que *su rechazo conduce a la aceptación de la conclusión deseada*. *La hipótesis alternativa representa la conclusión para la cual se busca evidencia.*

Por ejemplo, una empresa de marketing industrial está considerando la introducción de un nuevo plan de servicio para piezas hidráulicas. El plan será presentado si es preferido por más del 40% de los clientes. La forma apropiada de formular la hipótesis es:

$$H_0 = \pi \leq 0.40$$

$$H_1 = \pi > 0.40$$

- ✓ Si se rechaza la hipótesis nula H_0, se aceptará la hipótesis alternativa H_1 y se introducirá el nuevo servicio. Pero por el contrario, si no se puede rechazar la hipótesis nula H_0, el nuevo plan de servicio no debería introducirse a menos que se obtenga evidencia adicional.

✓ Contraste estadístico

➢ Test de una cola

La hipótesis alternativa se expresa <u>direccionalmente</u>

➢ Test de dos colas

$$H_0: \ \pi = 0.40$$
$$H_1 \ \ \pi \neq 0.40$$

En la investigación comercial, el test de una cola es el más común.

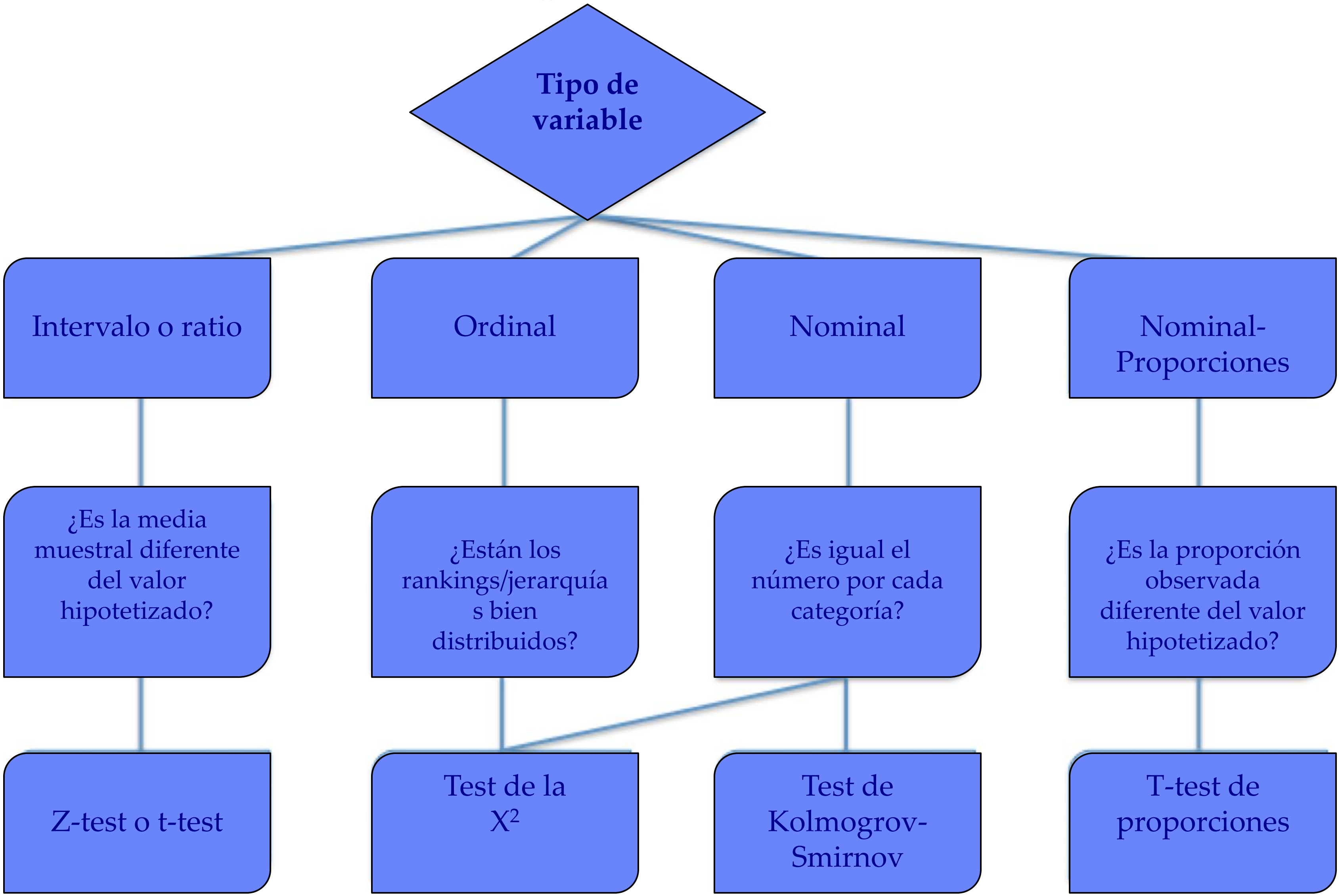
★ **Paso 2**: Selección de una técnica estadística apropiada

El test estadístico normalmente sigue una distribución conocida, tales como la distribución *normal*, *t*, o *chi-square distribution*.

Determinar cuándo usar cada método. Hacer la elección correcta se puede determinar considerando:

- ✓ El tipo de pregunta que se ha utilizado
- ✓ El número de variables implicadas
- ✓ El nivel de escalamiento/tipo de escalas utilizadas
- ✓ Tipo de distribución: tests paramétricos *vs.* no-paramétricos (o de distribución libre)

Opción univariante:



✓ La distritución T

- ✓ Distribución simétrica en forma de campana con una media = 0 y desviación estándar = 1.
- ✓ Cuando el tamaño de muestra (n) es mayor que 30, la distribución *t* y la distribución *z* son casi idénticas.
- ✓ La distribución *z* y la *t* son muy similares, por lo que el *z-test* y el *t-test* proporcionarán los mismos resultados en la mayoría de las situaciones.
- ✓ Sin embargo, cuando se conoce la desviación estándar de la población (σ), el Z-test es la más adecuado.
- ✓ Si se desconoce σ (lo más normal en la mayoría de los estudios), y el tamaño muestral es mayor de 30, el z-test también puede utilizarse.
- ✓ Cuando se desconoce σ y el tamaño muestral es pequeño (<30), el t-test es más apropiado.

★ **Paso 3**: Elección del nivel de significancia

Dos tipos de errores:

	Decision: Accept H_0	**Decision:** Reject H_0
Realitiy: H_0 is true	Correct-no error ☺	Type I error FALSE POSITIVE
Realitiy: H_0 is false	Type II error FALSE NEGATIVE	Correct-no error ☺

- ✓ Errores
- ➢ Error tipo I

FALSO POSITIVO (rechazar H_0 cuando es verdadera).

Ejemplos de error tipo I:

-H_0= El paciente/preso está sano/es inocente. Rechazamos H_0 y consideramos al paciente/preso como enfermo/culpable cuando en realidad está sano/es inocente.

La probabilidad del error de tipo I (α) se denomina el Nivel de significancia (i.e. el nivel de aceptación del error tipo I).

- ➢ Error tipo II

FALSO NEGATIVO (aceptar H_0 cuando es falsa)

-H_0= El paciente/preso está sano/es inocente. Aceptamos H_0 y consideramos al paciente/preso como sano/inocente cuando en realidad está enfermo/es culpable.

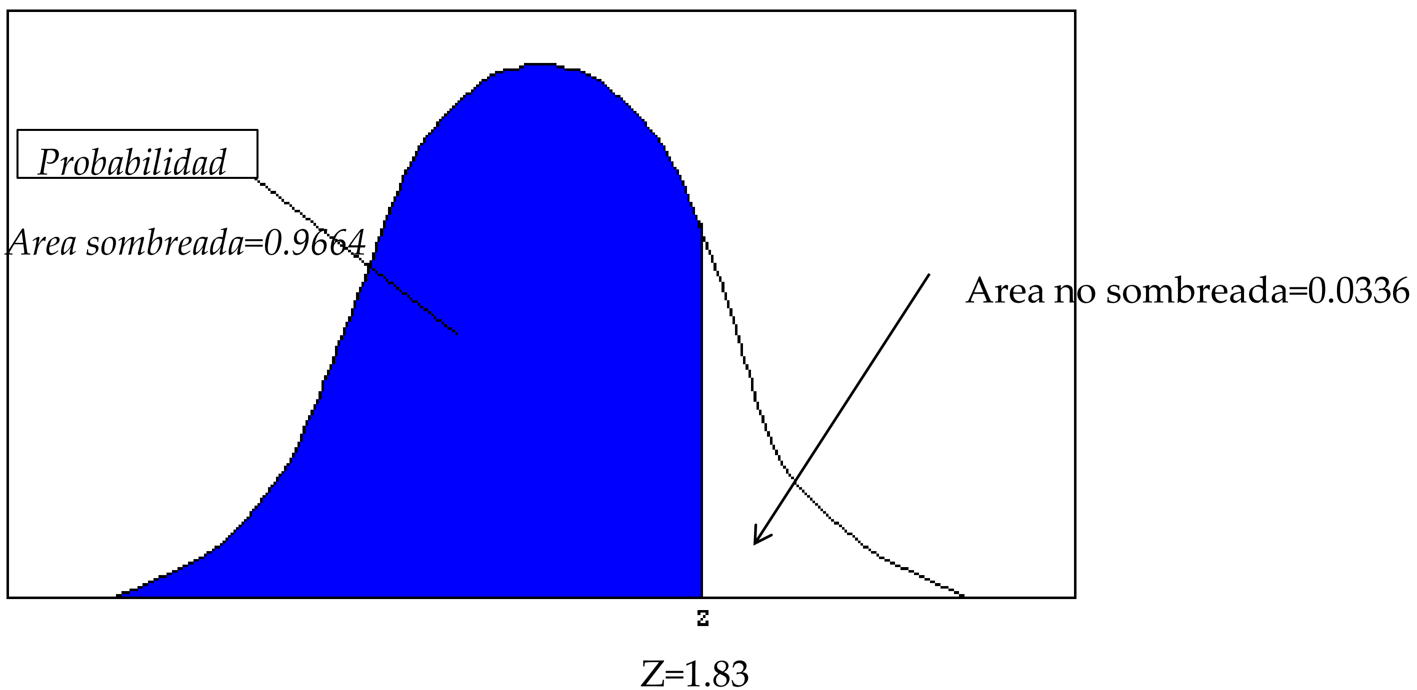
La probabilidad del error tipo II se denota como ß (la decisión incorrecta sería ß). A diferencia de α, que especifica el investigador, la magnitud de ß depende del valor real del parámetro poblacional (proporción). El complemento (1-ß) de la probabilidad de un error tipo II se denomina el poder del test estadístico.

★ Poder del test:

Probabilidad 1-ß

Aunque ß se desconoce, está en relación con α.

Probabilidad de Z con un test de una cola



★ **Paso 4**: Recogida de datos

- ✓ El tamaño de la muestra se determina después de tener en cuenta las consideraciones del α y otras, como las restricciones presupuestarias.
- ✓ Luego, se recogen los datos necesarios y se calcula el valor del test estadístico.
- ✓ Supongamos, siguiendo el ejemplo previo, que se encuestó a 500 clientes y 220 expresaron preferencia por el nuevo plan de servicio. Por tanto, el valor de la proporción muestral es p=220/500= 0.44 (44%).
- El valor de la desviación estándar de la proporción muestral $\boldsymbol{\sigma_p}$ se puede determinar como sigue:

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{(0.40)(0.6)}{500}} = 0.0219$$

El valor crítico se puede calcular como:

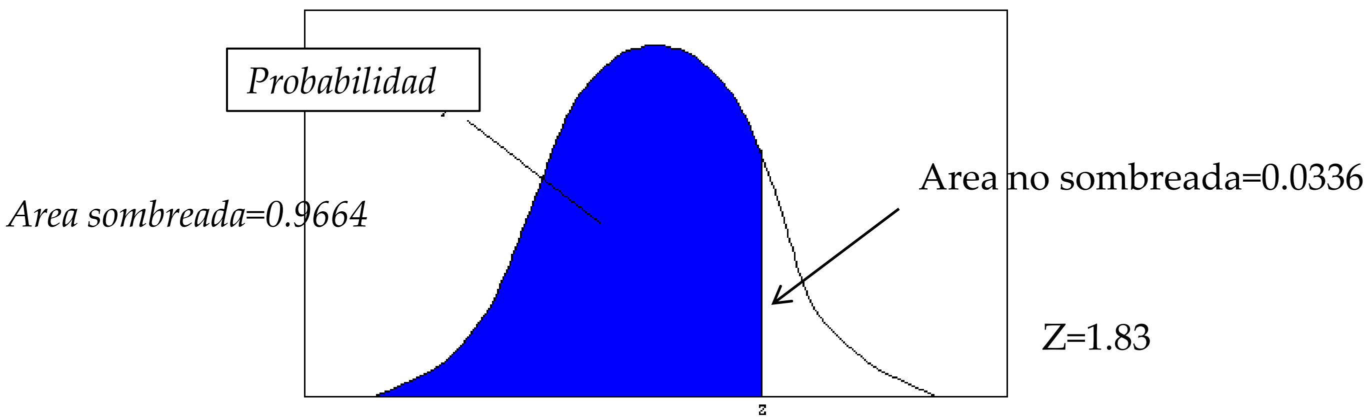
$$z\text{-valor} = \frac{p-\pi}{\sigma_p} = \frac{0.44-0.40}{0.0219} = 1.83$$

★ **Paso 5**: Determinar la probabilidad (valor crítico)

Existen dos formas:

1. Utilizando las tablas de la distribución normal: la probabilidad de obtener un *z-valor* de 1.83 se puede calcular utilizando las tablas de la normal (véase gráfico). El área sombreada entre -∞ y 1.83 es 0.9664 (véase tabla 1 en el apéndice). Por tanto, el área a la derecha de z= 1.83 es 1.0000-0.9664= 0.0336.

Probabilidad de Z con un test de una cola



- 2. Alternativamente, el valor crítico de z, que marcará el área a la derecha del valor crítico de $\alpha/2=0.05$ o $\alpha=0.10$ (90% intervalo de confianza), se encuentra entre 1.64 y 1.65 y sería 1.645. (Nótese que al determinar el valor crítico del test estadístico, el área a la derecha del valor crítico sería o α ó $\alpha/2$. Sería α para un test de una cola y $\alpha/2$ para un test de dos colas).

★ **Pasos 6 & 7:** Comparación de la probabilidad (valor crítico) y toma de decisiones

★ Existen dos formas de probar la hipótesis nula:

Ej.: Una empresa de marketing industrial está considerando la introducción de un nuevo plan de servicio. El plan será presentado si se prefiere por más del 40% de los clientes.

$$H_0: \ \pi \leq 0.40$$
$$H_1 \ \ \pi > 0.40$$

1. La probabilidad asociada con el valor calculado del test estadístico es 0.0336. Esta es la probabilidad de obtener un *p* valor de 0.44 cuando π=0.40. Sería menor que el nivel de significancia de 0.05. Por tanto, se rechaza la hipótesis nula. (0.0336<0.05; rechazar $H_{0;}$; se introduce el nuevo plan dado que se prefiere por más del 40% de los clientes).

2. Alternativamente, el valor calculado del test estadístico z= 1.83 se encuentra en la región de rechazo, más allá del valor de 1.645. De nuevo, se alcanza la misma conclusión de rechazar la hipótesis nula. (1.83>1.645; rechazar $H_{0;}$; se introduce el nuevo plan).

★ **Paso 8**: Conclusión de la investigación comercial

Ej.: Una empresa de marketing industrial está considerando la introducción de un nuevo plan de servicio. El plan será presentado si se prefiere por más del 40% de los clientes.

$$H_0: \ \pi \leq 0.40$$
$$H_1 \ \ \pi > 0.40$$

- ✓ La conclusión alcanzada en el contraste de hipótesis debe expresarse en términos del problema de investigación.
- ✓ En nuestro ejemplo, concluimos que hay evidencia de que la proporción de clientes que prefieren el nuevo plan de servicio es significativamente mayor a 0.4. Por tanto, la recomendación sería introducir el nuevo plan de servicio.

CONTENIDOS

1. Contraste de hipótesis (parte I)
2. Contrastes
    1. Tests de proporciones
    2. Tests de medias

★ El contaste de hipótesis se puede clasificar como: paramétrico *vs*. no-paramétrico

Estadísticos paramétricos

Ej.: t-test, z-test

Estadísticos no paramétricos (o de distribución libre)

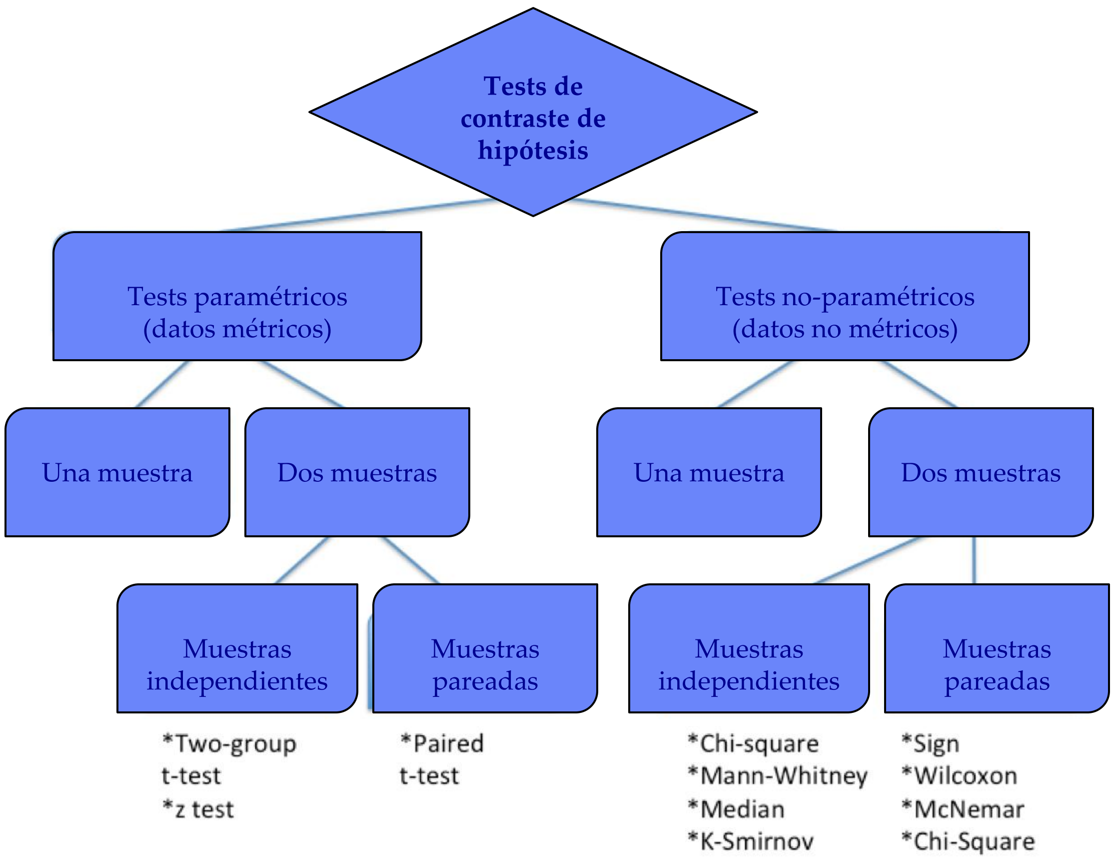
Se utilizan cuando el investigador *no sabe cómo están distribuidos los datos*.

Por tanto, las estadísticas no paramétricas se conocen como de *distribución libre*.

Ej.: Kolmogorov-Smirnov, Chi-Square ...

★ El contaste de hipótesis se puede clasificar como: paramétrico *vs.* no-paramétrico

Procedimientos de contrastes de hipótesis



★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Una muestra

Normalmente el investigador está interesado en hacer declaraciones sobre una sola variable en comparación con un criterio conocido o dado.

Ej.: Al menos el 65% de los clientes preferirán un nuevo diseño de embalaje o el 80% de los distribuidores preferirán la nueva política de precios.

★ Imagine que se introduce un nuevo componente si recibe una puntuación media de al menos 7 sobre una escala de Likert de 10 puntos. Se les muestra el nuevo componente a una muestra de 20 ingenieros para que la evalúen. El resultado indica una media de 7.9 con una desviación estándar de 1.6. Se selecciona un nivel de significancia de α=0.05. ¿Debería introducirse el nuevo componente?

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Una muestra

- ★ Imagine que se introduce un nuevo componente si recibe una puntuación media de al menos 7 sobre una escala de Likert de 10 puntos. Se les muestra el nuevo componente a una muestra de 20 ingenieros para que la evalúen. El resultado indica una media de 7.9 con una desviación estándar de 1.6. Se selecciona un nivel de significancia de α=0.05.
- ★ ¿Debería introducirse?

$$H_0 = \mu \leq 7.0$$

$$H_1 = \mu > 7.0$$

$$t = \frac{(X_{\text{high bar}} - \mu)}{S_{x\_\text{high bar}}}$$

$$S_{x\_\text{high bar}} = S/\sqrt{n}\ ;\ S_{x\_\text{high bar}} = 1.6/\sqrt{20} \quad 1.6/4.472 = 0.358$$

$$t = (7.9-7.0)/0.358 = 0.9/0.358 = \mathbf{2.514}$$

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Una muestra

★ ¿Debería introducirse?

t = (7.9-7.0)/0.358 = 0.9/0.358 = **2.514**

2.54>1.7291 se rechaza $H_{0.}$ Se introduce el nuevo componente

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Una muestra

- ★ ¿Debería introducirse?
- ★ Sin embargo, si se conoce la desviación estándar poblacionale debe utilizarse el **z test** en lugar del t test:

$$z = \frac{X_{\_high\ bar} - \mu}{\sigma_{X\_high\ bar}}$$

$$\sigma_{X\_high\ bar} = 1.5/\sqrt{20} = 1.5/\ 4.472 = 0.335$$

$$y\ z = 7.9 - 7.0\ /0.358 = 0.9\ /\ 0.358 = 2.514$$

Verónica Rosendo Ríos ©

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Una muestra

★ ¿Debería introducirse?

y z = 7.9 - 7.0 /0.358 = 0.9 / 0.358 = 2.514

1.645 < 2.514 por tanto, se rechaza H_0

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Dos muestras independientes

- Ciertas hipótesis se vinculan con parámetros de dos poblaciones diferentes:

Ej.:

Usuarios y no usuarios de una determinada marca pueden diferir en términos de sus percepciones sobre la marca, o la proporción de usuarios leales en el segmento I puede ser mayor que la proporción del segmento II.

- Las muestras tomadas al azar de diferentes poblaciones se denominan **muestras independientes.**
- Como en el caso de una muestra, las hipótesis podrían relacionarse con las medias o las proporciones.

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Dos muestras independientes

Medias

En el caso de medias para dos muestras independientes, las hipótesis serían como sigue:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

- Se muestrean las dos poblaciones, y las se calculan las medias y varianzas en base a muestras de tamaños n_1 y n_2.

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Dos muestras independientes

Medias

- ✓ Si se encuentra que ambas poblaciones ***tienen la misma varianza***, se calcula una estimación de varianza combinada (*pooled variance estimate*) a partir de las dos varianzas de las muestras.
- ✓ A continuación se estiman la desviación estándar del test estadístico y el t-valor. Los grados de libertad en este caso son (n_1+n_2-2).

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Dos muestras independientes

Medias

✓ Si las dos poblaciones tienen *varianzas desiguales* y no se puede calcular una *t* exacta para las diferencias en las medias muestrales, se calcula una aproximación a la *t*.

- El **F test** de varianza muestral puede utilizarse si no se sabe si las dos poblaciones tienen la misma varianza. En este caso las hipótesis serían:

$$H_0: \boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$$
$$H_1: \boldsymbol{\sigma}_1 \neq \boldsymbol{\sigma}_2$$

★ Tests paramétricos: medias y proporciones

Estadísticos paramétricos

✓ Dos muestras independientes

Medias

- El **F test** de varianza muestral puede utilizarse si no se sabe si las dos poblaciones tienen la misma varianza. En este caso las hipótesis serían:

$$H_0 : \boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$$
$$H_1 : \boldsymbol{\sigma}_1 \neq \boldsymbol{\sigma}_2$$

Si la probabilidad de F es mayor que el nivel de significación de α, H_0 no se rechaza y se estima una t basada en la varianza combinada (*pool variance estimate*).

Por el contrario, si la probabilidad de F es menor o igual que α, H_0 se rechaza y se utiliza la t basada en una varianza separada (*separate variance estimate*).

✓ Dos muestras independientes

Medias

Ej.: Imagine que el investigador desea determinar si los encuestados que están familiarizados con una tienda atribuyen una importancia diferente a las políticas de crédito y facturación de la tienda que aquellos que no están familiarizados con la tienda. Como antes, los encuestados se dividen en dos grupos de familiaridad basados en la mediana. Se lleva a cabo un test de dos muestras independientes y los resultados son como siguen:

	Resumen estadístico Número de casos	Media	Desviación estándar
Grupo no familiarizado	135	39.778	1.604
Grupo familiarizado	132	43.712	16.027

F Test for Equality of Variances

F Value	Two-tailed probability
1.03	0.871

t test

Pooled Variance Estimate			Separate Variance Estimate		
t value	Degrees of freedom	Two-tailed probability	t value	Degrees of freedom	Two tailed probability
-1.99	265	0.048	-1.99	264.56	0.048

Dado que 0.087>0.05, se usa el *pooled variance estimate*

✓ Dos muestras independientes

Medias

Resumen estadístico	Número de casos	Media	Desviación estándar
Grupo no familiarizado	135	39.778	1.604
Grupo familiarizado	132	43.712	16.027

Dado que 0.087>0.05, se usa el pooled variance estimate

F Test for Equality of Variances

F Value	Two-tailed probability
1.03	0.871

t test

Pooled Variance Estimate			Separate Variance Estimate		
t value	Degrees of freedom	Two-tailed probability	t value	Degrees of freedom	Two tailed probability
-1.99	265	0.048	-1.99	264.56	0.048

- El t-valor es -1.99, y con 265 grados de libertad, da una probabilidad de 0.048, que es menor que el nivel de significancia de 0.05. Por tanto, la hipótesis nula de igualdad de medias se rechaza.
- **Conclusión**: Dado que la media sobre la importancia de la política de crédito y facturación para el grupo no familiarizado es de 3.9778 y para el grupo familiarizado es de 4.3712, las personas familiarizadas dan significativamente más importancia a la política de crédito y facturación al seleccionar la tienda que aquellas que no están familiarizadas.

✓ Dos muestras independientes

Proporciones

Comparación de la proporción de usuarios de vaqueros en USA y Hong Kong

	Usuarios	No usuarios	Total
Estados Unidos	160	40	200
Hong Kong	120	80	200
Total	280	120	

¿Es la proporción de usuarios la misma en US y en Hong Kong? Las hipótesis nulas y alternativas serían:

$$H_0 : \ \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Se realiza un z test para comprobar la proporción. En este caso, sin embrgo, el test estadístico sería:

$$Z = \frac{P_1 - P_2}{S_{P1-P2 \text{ (\_high bars)}}}$$

✓ Dos muestras independientes

Proporciones

Comparación de la proporción de usuarios de vaqueros en USA y Hong Kong

	Usuarios	No usuarios	Total
Estados Unidos	160	40	200
Hong Kong	120	80	200
Total	280	120	

Se utiliza un z test como en el cálculo de proporciones de una muestra. En este caso, sin embargo, el estadístico sería:

$$Z = \frac{P_1 - P_2}{S_{P1-P2\ (\_high\ bars)}}$$

$S_{P1-P2\ (\_high\ bars)} = \sqrt{}P(1-P)\ (1/n_1 + 1/n_2)$; donde:

$P = n_1P_1 + n_2P_2 / n_1 + n_2$ Se selecciona $\alpha = 0.05$

$P_1 - P_2 = 0.8 - 0.6 = 0.2$

$$P = \frac{200x0.8 + 200x\ 0.6}{200+200}$$

$S_{P1-P2\ (\_high\ bars)} = \sqrt{}0.7x0.3\ (1/200 + 1/200) = 0.04583$

$$Z = \frac{0.2}{0.04583} = 4.36$$

✓ Dos muestras independientes

Proporciones

Comparación de la proporción de usuarios de vaqueros en USA y Hong Kong

	Usuarios	No usuarios	Total
Estados Unidos	160	40	200
Hong Kong	120	80	200
Total	280	120	

$$Z = \frac{0.2}{0.04583} = 4.36$$

4.36 > 1.96, se rechaza la hipótesis nula.

Conclusión: La proporción de usuarios (0.80 para USA, y 0.60 para Hong Kong) es significativamente diferente para las dos muestras.

✓ Muestras pareadas o dependientes

Ej.: Una muestra de encuestados puede valorar dos marcas competidoras, puede indicar la importancia relativa de dos atributos de un producto, o evaluar la misma marca en dos momentos temporales diferentes.

La diferencia en estos casos se examina mediante un *t* test de muestras pareadas.

Para calcular un *t* de muestras pareadas, se forma la variable de diferencia pareada, denominada D, y se calcula la media y la varianza.

✓ Muestras pareadas o dependientes

Para calcular un *t* de muestras pareadas, se forma la variable de diferencia pareada, denominada D, y se calcula la media y la varianza.

Posteriormente se calcula el *t* estadístico. Los grados de libertad son n-1, donde n es el número de pares.

Las hipótesis son:

$$H_0: \mu_D = 0$$

$$H_1: \mu_D \neq 0$$

https://cdn.pixabay.com/photo/2016/10/21/11/44/analytics-1757867_960_720.png

✓ Muestras pareadas o dependientes

Ej.: En el caso del centro comercial, un *t* test pareado se podría utilizar para determinar si los encuestados le dan más importancia a la calidad o a la política de precios.

Variable	Number of cases	Mean	Standard Deviation	Standard Error
Quality	271	5.6273	0.670	0.041
Store credit	271	4.1882	1.619	0.098

(Difference) Mean	Standard Deviation	Standard Error	Correlation	Two-tailed Probability	t value	Degrees of Freedom	Two-tail Probability
1.431	1.6087	0.0977	0.222	0.000	14.73	270	0.000

La diferencia de medias entre las vriables es 1.431, con una desviación estándar de 1.6087 y un error estándar de 0.0977. Esto da un t-valor de (1.4391/0.0977)=14.73, con 271-1=270 grados de libertad y una probabilidad de menos del 0.001.

Conclusión: la calidad es más importante que la política de precio al seleccionar una tienda comercial.

★ Tests no-paramétricos: medias y proporciones

Estadísticos no-paramétricos

Los tests no-paramétricos se utilizan **cuando las variables no son métricas**. Al igual que los tests paramétricos, los tests no paramétricos se utilizan para comprobar variables de una muestra, dos muestras independientes o dos muestras dependientes.

✓ Una muestra

1. El test Kolmogorov Smirnov (K-S) de una muestra. La decisión de rechazar la hipótesis se basa en el valor obtenido de la K.

Por ejemplo, en el contexto del departamento comercial, imagine que se quiere comprobar si la distribución de la importancia que se le da a la política de precios es una distribución normal. Se lleva a cabo un test de K-S de una muestra y se obtienen los siguientes datos:

★ Tests no-paramétricos: medias y proporciones

✓ Una muestra

1. Test Kolmogorov Smirnov (K-S) de una muestra.

Por ejemplo, en el contexto del departamento comercial, imagine que se quiere comprobar si la distribución de la importancia que se le da a la política de precios es una distribución normal. Se lleva a cabo un test de K-S de una muestra y se obtienen los siguientes datos:

	Test Distribution, Normal
Mean:	4.19
Standard Deviation:	1.62
Cases:	271

Most Extreme Differences				
Absolute	Positive	Negative	K-S z	Two-tailed p
0.19754	0.13150	-0.19754	3.252	0.000

K=0.1975. El valor crítico de K es 1.36/271=0.005.

Dos formas de comprobación:

1. Dado que el valor calculado de K (0.1975) >el valor crítico (0.005), se rechaza la hipótesis nula.
2. Alternativamente, en la tabla previa se indica que la probabilidad de obtener un valor K de 0.1975, es menor de 0.001 (p=0.000). Puesto que es menor que el nivel de significación establecido de 0.05, se rechaza la hipótesis nula.

Por tanto, la distribución de la variable "impotancia que se da a la política de precios" se desvía notablemente de una distribución normal.

★ Tests no-paramétricos: medias y proporciones

✓ Una muestra

2. El test de chi-cuadrado

3. Otros tests de una muestra no paramétricos incluyen el *run test* y el test binomial.

- El *runs test*
- El test binomial

★ Tests no-paramétricos: medias y proporciones

✓ Dos muestras independientes

1. El test de Mann-Whitney U.
2. El test de la mediana de dos muestras.
3. El test de dos muestras de Kolmogorov-Smirnov.

★ Tests no-paramétricos: medias y proporciones

✓ Dos muestras pareadas o dependientes

1. El test de Wilcoxon.

Por ejemplo, veamos si los encuestados le confieren más importancia a la calidad que a la política de precios. Supongamos que se asume que ambas variables se miden con escala ordinales y no de intervalo.

	Quality with Store Credit	
Mean Rank	**Cases**	**(Store Credit Quality)**
99.88	177	• Ranks
46.89	14	+ Ranks
	80	Ties
	271	Total
z=-11.1262		2-tailed p= 0.0000

La probabilidad asociada con el estadítico z es menor que 0.05, indicando que la diferencia es ciertamente significativa.

★ Tests no-paramétricos: medias y proporciones

✓ Dos muestras pareadas o dependientes

1. Otra prueba no paramétrica de muestra pareada es la prueba de signo (*sign test*).
2. El test de McNemar.
3. Alternativamente, el test de la chi-cuadrado.

Las diversas pruebas paramétricas y no paramétricas se resumen en:

Muestra	Aplicación	Nivel de escalamiento	Tests/Comentarios
Una muestra			
Una muestra	Distribuciones	No-métrico	K-S y Chi-cuadrado para bonad de ajuste; Run tests para aleatoriedad; Test binomial para bondad de ajuste con variables dicotómicas
Una muestra	Medias	Métrico	t-test, si la varianza se desconoce z-test si se conoce la varianza
Una muestra	Proporciones	Métrico	z-test
Dos muestras independientes			
Dos muestras independientes	Distribuciones	No-métrico	Test de K-S para dos muestras para examinar la equivalencia de dos distribuciones ; Test de dos grupos; F-test para igualdad de varianzas
Dos muestras independientes	Medias	Métrico	z-test
Dos muestras independientes	Proporciones	Métrico	Test de Chi-cuadrado
Dos muestras independientes	Rankings/medianas	No-métrico	El test de Mann-Whitney U es más robusto que el test de medianas
Muestras pareadas			
Muestras pareadas	Medias	Métrico	t-test pareado
Muestras pareadas	Proporciones	No-métrico	Test McNemar para variables binarias; test de la Chi-cuadrado
Muestras pareadas	Rankings/medianas	No-métrico	El test de Wilcoxon marched-pairs ranked-signs es más robusto que el test de signos

EJEMPLO

“Entonces, ¿están satisfechos o no?”

Rob Baer, director de operaciones, y Kathy Hahn, directora ejecutiva de PrecisionMetals solicitaron reunirse con Ed Ed Bond para debatir sobre la encuesta de satisfacción de los empleados realizada hacía un mes.

- "Ed, seguimos preocupándonos por la pérdida de empleados metalúrgicos en nuestra planta de Madison, pero nuestra planta de Richmond parece estar mejorando en términos de rotación", afirmó Rob. "¿Qué opina sobre la satisfacción de nuestros empleados?". Ed respondió: "Elaboramos un índice de tres preguntas sobre la satisfacción laboral. Hemos analizado los datos de la planta de Richmond, y nuestra satisfacción promedio es de 3.9 "
- Kathy preguntó: "¿Qué significa 3.9? ¿Qué se supone que significa eso? “

https://cdn.pixabay.com/photo/2016/04/19/21/10/frog-1339892_960_720.jpg

EJEMPLO

“Entonces, ¿están satisfechos o no?”

Ed respondió: "Lo siento, debería haberlo explicado mejor. Preguntamos a nuestros empleados en una escala de cinco puntos, siendo 1 "Muy en desacuerdo" y 5 ”Totalmente de acuerdo". Cuando las puntuaciones se promediaron para Richmond, nuestra satisfacción general fue de 3.9 sobre esta escala de 5.0 puntos ".

Rob continuó, "**¿Eso es bueno o malo?** Suena bien... supongo. ¿Qué pasa con Madison? “ Ed, dándose cuenta de que no estaba comunicando bien la información, respondió: "La puntuación de satisfacción de Madison es de 3.5. Históricamente, ambas plantas tenían una puntuación de satisfacción de 3.5”.

Kathy se dio cuenta de que Ed se estaba poniendo nervioso. "Ed, apreciamos lo que estás haciendo. Lo siento, pero no sé exactamente qué significan las puntuaciones **¿Es bueno 3.9? ¿Es bueno 3.5? ¿Es significativa la diferencia entre Richmond este año y las puntuaciones de años anteriores?** ¿La diferencia entre estas puntuaciones es suficiente para explicar la diferencia en la facturación? Solo quiero saber si la encuesta muestra si nuestros empleados están satisfechos o no". Ed volvió al departamento de investigación y mientras entraba en su oficina y cerraba la puerta silenciosamente, se decía a sí mismo: "**No puedo hablar de calificaciones de 3.9 y 3.5. Estoy aquí para ayudarlos a comprender lo que realmente significan las puntuaciones**". ¡Era hora de trabajar!

Fuente: Zikmund *et al.* (2013).

https://cdn.pixabay.com/photo/2017/12/23/23/28/businessman-3036181_960_720.jpg

REFERENCIAS BIBLIOGRÁFICAS

- Brown, T. J., y Suter, T. (2012): *MR*. South Western, Cenage Learning. USA.
- Hair, J.; Bush, R., y Ortinau, D. (2006): *Marketing research. Within a changing environment*. Revised International Edition (3rd ed.). McGraw-Hill, New York, USA.
- Malhotra, N. K. (1996): *Marketing Research. An Applied Orientation*. 2nd ed. Prentice-Hall International. USA.
- Rosendo-Ríos, V., y Pérez del Campo, E. (2013): *Business Research Methods: Theory and Practice*. ESIC Editorial. España.
- Rosendo-Ríos, V.; de Esteban, J., y Antonovica, A. (2012): *MR: Development of Theoretical Concepts for Market Research I and II*. South Western, Cenage Learning. USA.
- Zikmund, W. G.; Babin, B. J.; Carr, J. C., y Griffin, M. (2013): *Business Research Methods*. 9th Edition. South Western, Cenage Learning. USA.