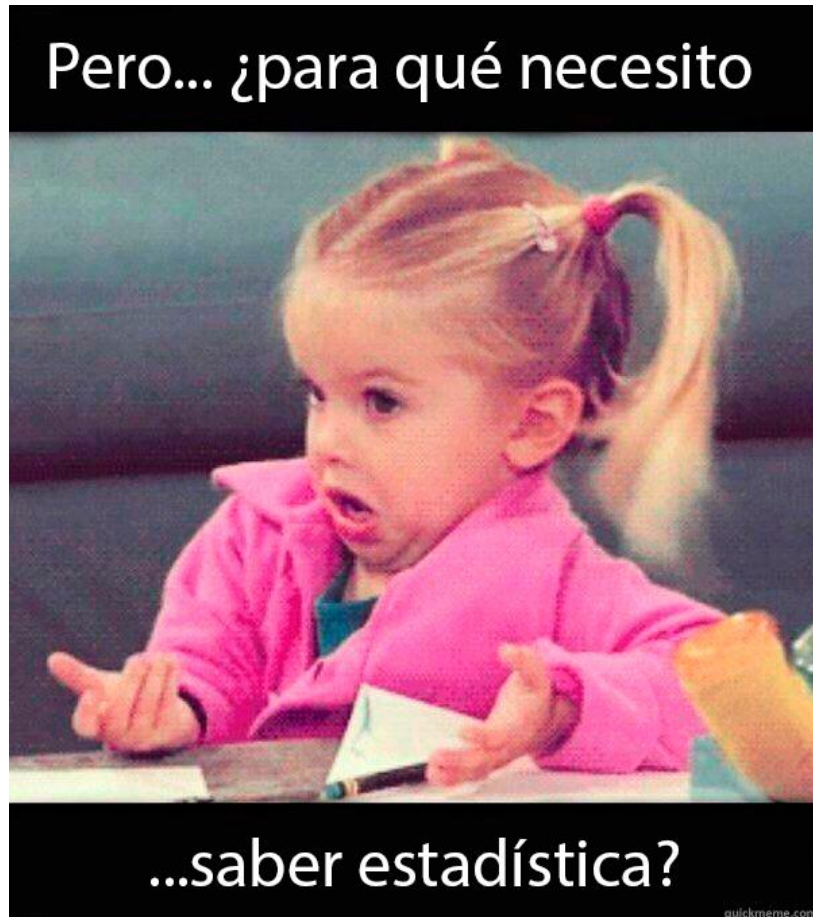


UNIDAD 1: ¿CÓMO TOMAR UNA DECISIÓN ESTADÍSTICA?

Estadística
Lic. en Nutrición
UCEL

ESTADÍSTICA Y MÉTODO CIENTÍFICO



La Estadística y el Método Científico proveen un conjunto de principios y procedimientos para obtener y resumir la información para la **toma de decisiones**.

ESTADÍSTICA Y MÉTODO CIENTÍFICO

Se sabe que en la población adulta argentina (mayores de 18 años) la **prevalencia de sobrepeso** es de 34%.

- ¿Cómo “se sabe”?
- ¿Qué es la prevalencia?
- ¿Cuál es la definición de sobrepeso?

La **población** es el grupo de objetos o individuos bajo estudio, sobre los cuales se requiere información.

- ¿Cuál es la población en el ejemplo?

ESTADÍSTICA Y EL MÉTODO CIENTÍFICO

Una Licenciada en Nutrición considera que en la ciudad de Rosario, la prevalencia de sobrepeso en la población adulta es mayor al 34%.

¿Cuál es la población en este caso?

Teoría 1: La prevalencia de sobrepeso en la población adulta de la ciudad de Rosario es de 34%

Teoría 2: La prevalencia de sobrepeso en la población adulta de la ciudad de Rosario es superior al 34%

¿Cómo decidir cuál de las teorías es cierta?

HIPÓTESIS ESTADÍSTICAS

En estadística, a las dos teorías en competencia acerca de una población se las denomina **Hipótesis Estadísticas**.

- **Hipótesis Nula:** es la creencia convencional, el status quo, aquello que “se sabe”. Es la afirmación de que nada está sucediendo, no hay cambios en la población. Se la simboliza H_0 .
- **Hipótesis Alternativa:** es la competencia a esa creencia, lo que se quiere demostrar. Es la afirmación que el investigador “espera” que sea cierta. Se la simboliza H_1 .

HIPÓTESIS ESTADÍSTICAS

- En el ejemplo:

H_0) La prevalencia de sobrepeso en la población adulta de la ciudad de Rosario es de 34%

H_1) La prevalencia de sobrepeso en la población adulta de la ciudad de Rosario es mayor a 34%

- Otra manera de expresar estas hipótesis es en términos matemáticos:

H_0) $p = 0,34$

H_1) $p > 0,34$

Donde p es la proporción de sobrepeso en la población adulta de la ciudad de Rosario.

PARÁMETROS

Las hipótesis estadísticas se escriben en lenguaje matemático en función del valor de un parámetro.

Un **parámetro** es una medida resumen que se calcula a partir de todas las unidades de la población.

- p : proporción. Este parámetro toma valores entre 0 y 1. Cuando los valores están expresados como porcentajes, se los divide por 100. Si el problema está planteado en relación a la prevalencia, también usamos este parámetro.
- μ : promedio, valor medio, media.

HIPÓTESIS ESTADÍSTICAS

Los ensayos de hipótesis estadísticas se pueden clasificar según la afirmación que se realice en la hipótesis alternativa:

- **Unilateral por derecha:** en este caso, la hipótesis alternativa afirma que el valor del parámetro es superior al establecido en la hipótesis nula
- **Unilateral por izquierda:** en este caso, la hipótesis alternativa afirma que el valor del parámetro es inferior al establecido en la hipótesis nula
- **Bilateral:** en este caso, la hipótesis alternativa afirma que el valor del parámetro es distinto al establecido en la hipótesis nula (podría ser mayor o menor)

MANOS A LA OBRA N° 1



LACTANCIA MATERNA

Los resultados de la 2° Encuesta Nacional de Nutrición y Salud indican que en Argentina, el 56% de los niños es puesto al pecho de su madre durante su primera hora de vida. Un investigador cree que en la provincia de Santa Fe, esa proporción es inferior a la nacional.

- 1) ¿Cuál es la población de interés del investigador?
- 2) Plantee las hipótesis nula y alternativa en sus palabras.
- 3) Identifique y defina el parámetro de interés.
- 4) Plantee las hipótesis nula y alternativa en términos matemáticos.
- 5) Clasifique el ensayo.

MANOS A LA OBRA N° 1



PESO DE CABRAS (Ejercicio de repaso 13)

Un estudio fue diseñado para averiguar si el ejercicio de la madre influye sobre el peso al nacimiento de cabras. Se entrenó a cabras preñadas para que corrieran en una cinta. Se sabe que el peso promedio al nacer de las cabras es de 1600 g.

- 1) ¿Cuál es la población de interés del investigador?
- 2) Plantee las hipótesis nula y alternativa en sus palabras.
- 3) Identifique y defina el parámetro de interés.
- 4) Plantee las hipótesis nula y alternativa en términos matemáticos.
- 5) Clasifique el ensayo.

MANOS A LA OBRA N° 1



SUPLEMENTACIÓN CON ZINC (Ejercicio 1.9)

Se realizó un estudio para verificar si tomar suplementos de zinc podría ayudar a ciertas mujeres, especialmente delgadas, a tener hijos más pesados. Mitad de las mujeres en el estudio tomaron vitaminas prenatales con 25 mg de zinc. El otro grupo recibió la misma vitamina prenatal, pero sin zinc. Los resultados fueron elevados al *Journal of the American Medical Association*. El estudio comparaba el peso medio al nacer, tasa de partos prematuros y por cesárea.

- 1) ¿Cuál es la población de interés de los investigadores?
- 2) Plantee las hipótesis nula y alternativa acerca del peso promedio al nacer en sus palabras.
- 3) Identifique y defina el parámetro de interés.
- 4) Plantee las hipótesis nula y alternativa en términos matemáticos.

ESTADÍSTICA Y EL MÉTODO CIENTÍFICO

¿Cómo decidir cuál de las teorías es cierta?

Inferencia estadística es el proceso de extraer conclusiones sobre la población basándose en la información de una muestra extraída de esa población.

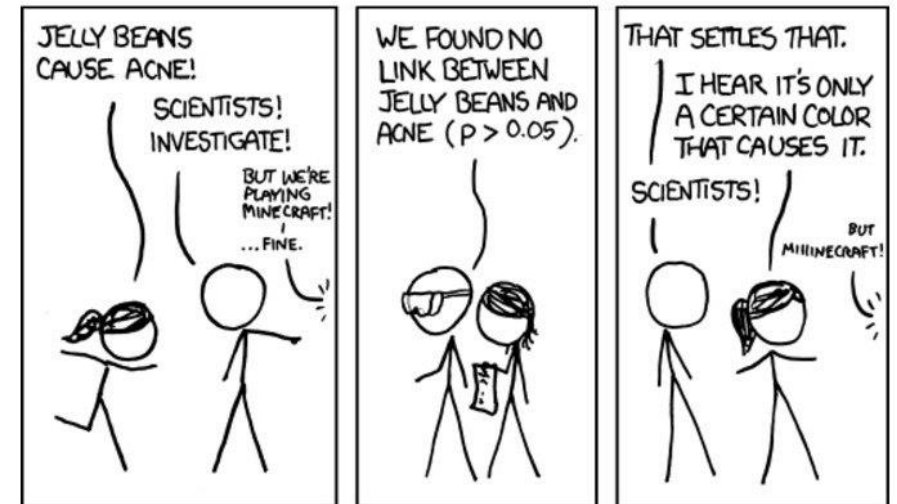
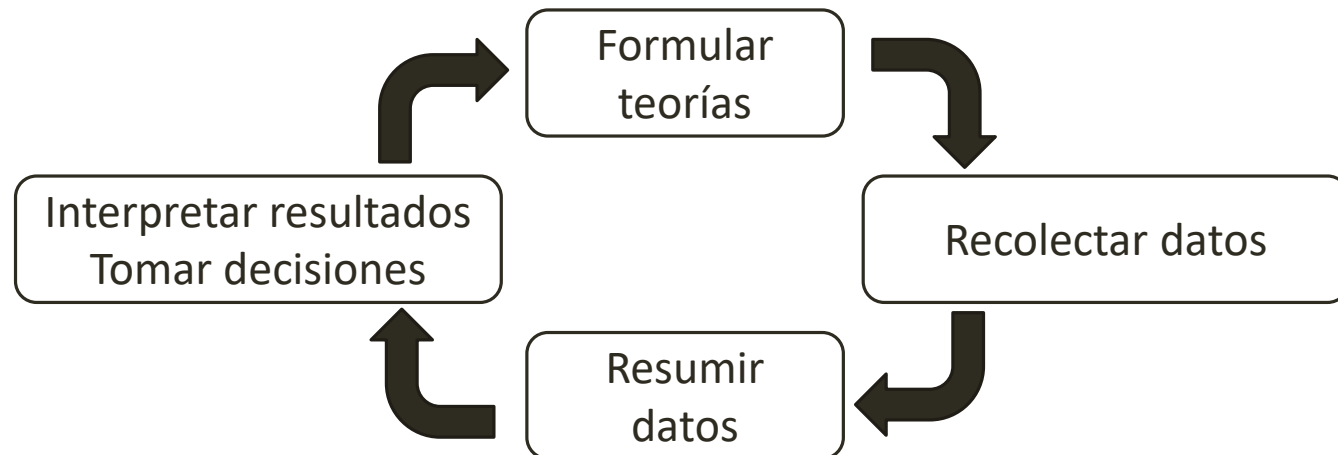
Una **muestra** es una parte de la población que se usa como información.

CUIDADO! No cualquier parte de la población representa una “buena” muestra. Veremos en la unidad 2 cómo obtener muestras adecuadas.

ESTADÍSTICA Y EL MÉTODO CIENTÍFICO

La Estadística y el Método Científico proveen un conjunto de principios y procedimientos para obtener y resumir la información para la **toma de decisiones**.

Los pasos del método científico:



HIPÓTESIS ESTADÍSTICAS

Mediante este proceso se puede elegir cuál de las hipótesis **se considera cierta**. Sin embargo, no es posible afirmar con total certidumbre cuál de las hipótesis es verdadera.

Una teoría será **rechazada** si se puede demostrar estadísticamente que los datos que observamos son **muy poco probables** de ocurrir si la teoría fuera cierta.

Una teoría es aceptada si no es rechazada por los datos.

- ¿Qué significa “poco probable”?

HIPÓTESIS ESTADÍSTICAS

En una población constituida por 6 niños, **se cree** que uno de ellos es obeso y los demás tienen peso normal .

H_0) La proporción de niños obesos en la población es de 0,17

H_1) La proporción de niños obesos en la población es distinta de 0,17

Se extrae una muestra de tamaño $n = 6$ de la siguiente manera: se arma una lista con los niños, identificando a cada uno con un número del 1 al 6, se lanza un dado y se elige al niño que corresponde al número obtenido para registrar si es obeso o no; esto se repite 6 veces.

HIPÓTESIS ESTADÍSTICAS

Resultado 1:

- Peso normal
- Peso normal
- Peso normal
- Peso normal
- Peso normal
- Peso normal

Resultado 2:

- Obeso
- Obeso
- Obeso
- Obeso
- Obeso
- Obeso

Resultado 3:

- Peso normal
- Peso normal
- Obeso
- Peso normal
- Obeso
- Peso normal

Con el Resultado 1 hay más evidencia a favor de la H_0 (no rechazamos H_0).

Con el Resultado 2 hay más evidencia a favor de H_1 (rechazamos H_0).

Con el resultado 3 la situación no es tan clara.

En cualquier caso, si decidimos rechazar o no rechazar H_0 podemos cometer un ERROR.

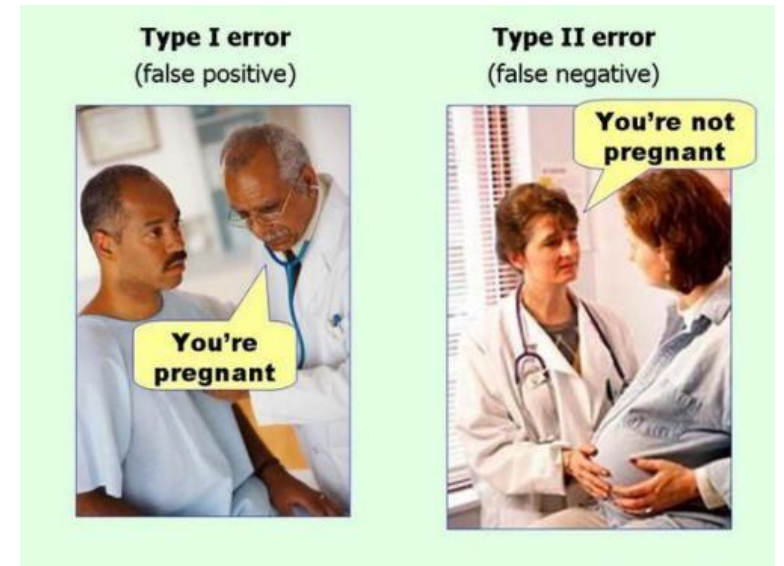
ERRORES

Error de tipo I (e_I): aquel que se comete cuando se decide rechazar la H_0 cuando en realidad esta es cierta.

Error de tipo II (e_{II}): aquel que se comete cuando se decide no rechazar la H_0 cuando en realidad es cierta H_1 .

En el ejemplo:

- Error de tipo I: pensar que la proporción de niños obesos en la población es mayor a 0,17 cuando en realidad es igual a 0,17.
- Error de tipo II: pensar que la proporción de niños obesos es la población es de 0,17 cuando en realidad es mayor.



MANOS A LA OBRA N° 2



ERRORES:

Para los ejemplos del Manos a la obra 1: LACTANCIA MATERNA, PESO DE CABRAS y SUPLEMENTACIÓN CON ZINC, escriba los errores que pueden cometerse al tomar una decisión, explicando cada uno de ellos en palabras del problema.

VERDADERO O FALSO:

Cuando se acepta una hipótesis nula, es posible que:

- a) Se haya tomado una decisión correcta
- b) Se haya cometido un error tipo I
- c) Haya ocurrido tanto a) como b)
- d) No haya ocurrido a) ni b)
- e) Ninguno de los anteriores

Aceptar una hipótesis nula cuando es falsa constituye:

- a) Un error de tipo I
- b) Un error de tipo II
- c) Ninguna de las anteriores

HIPÓTESIS ESTADÍSTICAS

$P(e_I)$: probabilidad de rechazar la H_0 cuando la H_0 es cierta (α).

$P(e_{II})$: probabilidad de no rechazar la H_0 cuando la H_1 es cierta (β).

Idealmente, queremos tomar siempre la decisión correcta y no equivocarnos, pero vamos a conformarnos logrando que la probabilidad de cometer un error sea pequeña.

Queremos minimizar ambas probabilidades de error. Lamentablemente, para un tamaño de muestra dado no pueden minimizarse las dos al mismo tiempo. Por eso, se decide fijar $P(e_I)$ en un valor pequeño, y minimizar $P(e_{II})$. Minimizar la $P(e_{II})$ es equivalente a maximizar la **potencia**.

La **potencia** de un test es la probabilidad de rechazar H_0 cuando H_1 es cierta, y se obtiene: $\text{potencia} = 1 - \beta$.

REGLA DE DECISIÓN

La **regla de decisión** de un ensayo de hipótesis es un criterio definido antes de extraer la muestra, que indica en qué casos debe rechazarse la hipótesis nula.

En el ejemplo de los niños, distintos ejemplos de reglas de decisión son:

- Rechazar H_0 si en la muestra se observa al menos un niño obeso
- Rechazar H_0 si en la muestra se observa más de un niño obeso

REGLA DE DECISIÓN

Debemos elegir una regla tal que fijada $P(e_I)$ se logre minimizar la $P(e_{II})$. Los valores más comunes de α son 0,01, 0,05 y 0,1, y los más pequeños se eligen cuando las consecuencias de rechazar H_0 equivocadamente son serias.

Al valor seleccionado de la probabilidad de error de tipo I se lo conoce como **nivel de significación**.

Si se incrementa el tamaño de muestra, se puede hacer disminuir las $P(e_I)$ y $P(e_{II})$.

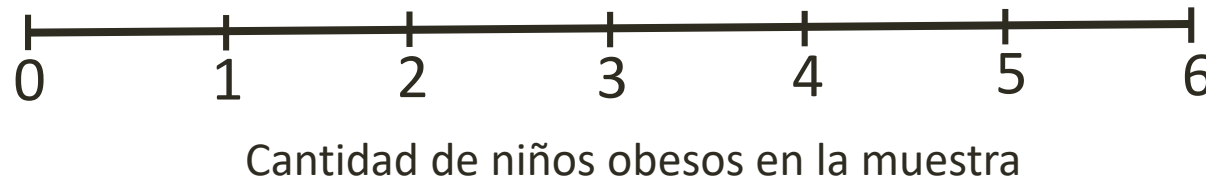
REGLA DE DECISIÓN

- **Región de rechazo o crítica:** conjunto de valores para los cuales rechazaríamos H_0 . Son valores contradictorios a H_0 y favorables a H_1 .
- **Región de aceptación:** comprende el conjunto de valores para los cuales aceptaríamos H_1 .
- **Valor crítico:** valor que marca el “punto inicial” del conjunto de valores que comprenden la región de rechazo o crítica.

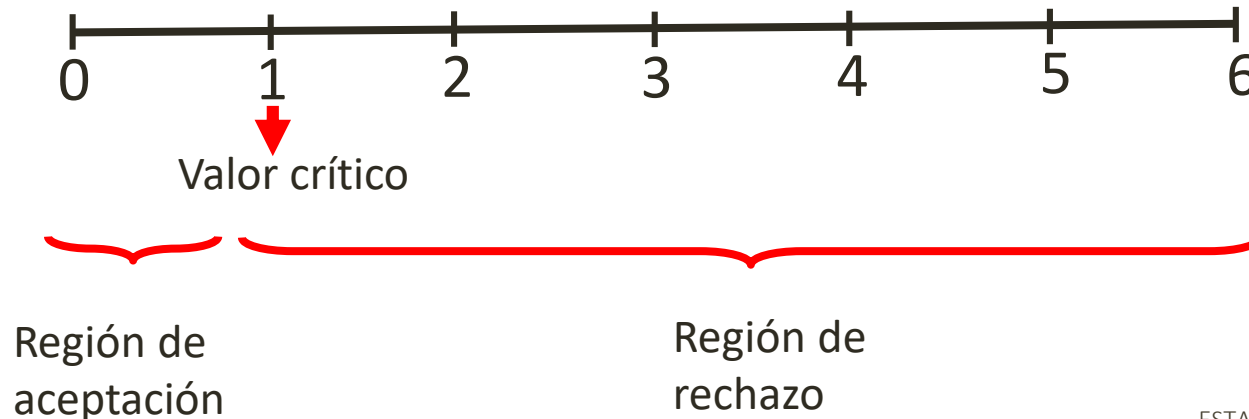


REGLA DE DECISIÓN

Para el ejemplo de la población de 6 niños, la cantidad de niños obesos en la muestra de tamaño $n = 6$ puede ser:



Si la regla de decisión es: rechazar H_0 si en la muestra se observa al menos un niño obeso:



MANOS A LA OBRA N° 3



POBLACIÓN DE 6 NIÑOS

Considerando la segunda regla de decisión planteada para este ejemplo: “Rechazar H_0 si en la muestra se observa más de un niño obeso”, esquematizar los valores posibles de la cantidad de niños obesos en una muestra de tamaño 6, marcar el valor crítico y las regiones de aceptación y de rechazo.

LACTANCIA MATERNA

Suponga que en este ejemplo se extrae una muestra de 10 niños y se calcula la proporción de ellos que recibieron leche materna en la primera hora de vida. La regla de decisión es: rechazar H_0 si la proporción en la muestra es menor a 0,5. Esquematizar los valores posibles de la proporción de niños que recibieron leche materna en la primera hora de vida, marcar el valor crítico y las regiones de aceptación y de rechazo.

MANOS A LA OBRA N° 3



PESO DE CABRAS (Ejercicio de repaso 13)

Suponga que en este ejemplo se analiza una muestra de 15 cabras y se registra su peso promedio. La regla de decisión es: rechazar H_0 si el peso promedio de las 15 cabras es mayor a 1700 g o menor a 1500 g. Esquematizar los valores posibles del peso promedio de las cabras, marcar el valor crítico y las regiones de aceptación y de rechazo.

REGLA DE DECISIÓN

Cuando se rechaza H_0 se dice que el ensayo de hipótesis es **estadísticamente significativo**.

Los datos observados conducen a un test estadísticamente significativos si son poco probables de ser observados bajo el supuesto de que H_0 es cierta.

CUIDADO! Un resultado estadísticamente significativo no necesariamente es un resultado de importancia práctica.

CUIDADO! Una vez tomada la decisión, puede haberse cometido un error o no y las probabilidades de error de tipo I y II son 0 o 1.

P-VALUE

El **p-value** o valor de **p-asociado** a un resultado es la chance de obtener el resultado observado más la chance de observar valores más extremos que el observado, computada asumiendo que la H_0 es la verdadera. El p-value puede tomar valores entre 0 y 1.

- Un p-value pequeño indica que hay una baja probabilidad de observar lo observado cuando H_0 es cierta, y por lo tanto puede rechazarse H_0 .
- Un p-value grande indica que hay una gran probabilidad de observar lo observado cuando H_0 es cierta, y por lo tanto no debe rechazarse H_0 .

CUIDADO! El p-value depende del tamaño de muestra.

P-VALUE



POCA EVIDENCIA
A FAVOR DE H_0

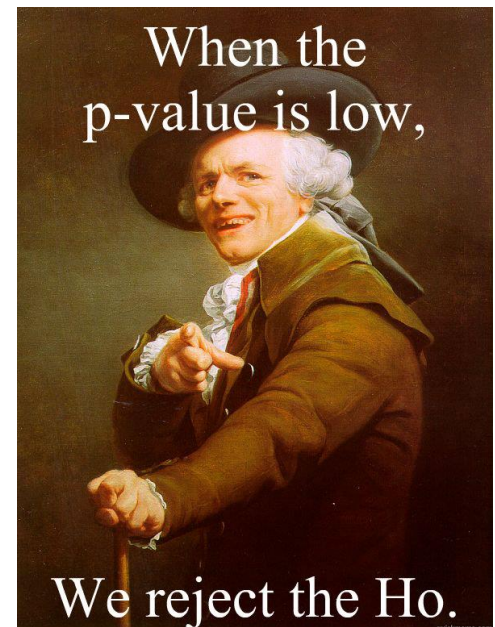
MUCHA EVIDENCIA
A FAVOR DE H_0

Otra forma de definir la regla de decisión de un ensayo de hipótesis es:

- Rechazar H_0 si $\text{p-value} < \alpha$

Es decir, el ensayo es estadísticamente significativo si $\text{p-value} < \alpha$

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE $P < 0.10$ LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥ 0.1	



MANOS A LA OBRA N° 4



POBLACIÓN DE 6 NIÑOS

Suponga que para este ejemplo se obtiene un p-value de 0,06.

- 1) ¿Es el ensayo estadísticamente significativo a un nivel del 10%?
- 2) ¿Es estadísticamente significativo a un nivel del 5%?

LACTANCIA MATERNA

Suponga que en este ejemplo se extrae una muestra de 10 niños y se calcula la proporción de ellos que recibieron leche materna en la primera hora de vida. Se obtiene un p-value de 0,027. Considerando un nivel de significación del 5%, plantear la regla de decisión y esquematizar los valores posibles del p-value, marcar el valor crítico y las regiones de aceptación y de rechazo. ¿Cuál es la decisión para este test?. Si la probabilidad de error de tipo II es 0,2, ¿cuál es la potencia del test?

MANOS A LA OBRA N° 4



PESO DE CABRAS (Ejercicio de repaso 13)

Suponga que en este ejemplo se analiza una muestra de 15 cabras y se registra su peso promedio. Se obtiene un p-value de 0,57. Considerando un nivel de significación del 1%, plantear la regla de decisión, esquematizar los valores posibles del p-value, marcar el valor crítico y las regiones de aceptación y de rechazo. Tome la decisión y explique su conclusión en palabras del problema.

PESO DE UN SUPLEMENTO VITAMÍNICO

Se realizó un test estadístico para probar las siguientes hipótesis, que resultó ser estadísticamente significativo:

H_0) El peso promedio de un suplemento vitamínico es 500 mg

H_1) El peso promedio de un suplemento vitamínico no es 500 mg

- 1) ¿Qué hipótesis fue sustentada?
- 2) ¿Cuál es la dirección de este test?
- 3) Suponga que el nivel de significación utilizado fue 0,05. Escriba dos valores posibles del p-value observado.

MANOS A LA OBRA N° 4

PARA RESOLVER 1.6. TRES ESTUDIOS (pagina 37)

La siguiente tabla resume los resultados de tres estudios diferentes.



Estudio	Hipótesis nula	Hipótesis Alternativa	p-value
A	La verdadera vida promedio es ≥ 54 meses	La verdadera vida promedio es < 54 meses	0,0251
B	El tiempo medio de supervivencia con el T_1 es igual que con el T_2	tiempo medio de supervivencia con el T_1 es distinto que con el T_2	0,0018
C	La verdadera proporción de personas que tienen dos empleos es $\leq 0,33$	La verdadera proporción de personas que tienen dos empleos es $> 0,33$	0,3590

- 1) ¿Para cuál estudio los resultados muestran mayor soporte para la hipótesis nula?
- 2) Suponga que en el estudio A se concluyó que los datos sustentan la hipótesis alternativa de que la verdadera vida promedio es menor a 54 meses pero, en realidad, ésta es mayor o igual a 54 meses. ¿Es este un error de tipo I o II?
- 3) Si los resultados del estudio C no son estadísticamente significativos, ¿cuál hipótesis aceptaría?