Test de Hipótesis

Introducción a la Estadística

Fiona Franco Churruarín
fionafch96@gmail.com
Basado en notas de clase de Andrea Rotnitzky

UTDT

Febrero 2022

Motivación

Una herramienta extremadamente útil en estadística son las pruebas o *tests* de hipótesis.

Son una forma sistemática de poner a prueba declaraciones sobre al menos un parámetro poblacional desconocido en base a una muestra.

Una hipótesis podría ser:

- La tasa de desempleo del mes de Enero es superior al 32%; o
- La proporción de votantes a favor del candidato A es de al menos 40%; o tal vez mas interesante,
- La tasa de desempleo del mes de Enero 2020 es mayor que la de Diciembre 2019.

Motivación

Lo que abordaremos ahora es un procedimiento para poner en tela de juicio declaraciones como las de la filmina anterior, teniendo algún criterio.

En particular, tomaremos una afirmación y cuantificaremos que tan fuerte es la evidencia en la muestra a favor o en contra de dicha afirmación.

Motivación

Suponga que la efectividad de un vendedor se mide como la proporción de clientes que realizan una compra sobre el total de clientes atendidos por ese vendedor. Suponga que se sabe que esta proporción es una variable aleatoria con distribución Normal, con media 0.3 y varianza 0.04.

A fines de aumentar la efectividad de los vendedores, el gerente está considerando implementar una bonificación por cada venta concretada para el vendedor.

Para aprender sobre la eficacia de esta estrategia para aumentar la efectividad, el gerente le pide su ayuda y hacer una prueba piloto en 25 vendedores de una sucursal particular.

Motivación

¿Cómo haría usted para decidir si el bono es beneficioso para aumentar las ventas de cada vendedor o no?

Suponga por ejemplo que se concentra en analizar si el bono aumenta la **media poblacional** de la efectividad de los vendedores.

Usted debe construir una regla que le ayude a determinar si:

- el bono no es beneficioso: su implementación NO aumenta la media de efectividad de los vendedores; o
- el bono es beneficioso: su implementación aumenta la media de efectividad de los vendedores

Motivación

Como criterio, sabiendo que sin el bono ese parámetro es 0.3, **teniendo en cuenta que sólo trabaja con una muestra**, un analista decide que declarará que el bono es efectivo si la **media muestral** de efectividades es mayor a 0.35.

Notar que pueden haber cuatro situaciones:

1. El bono no es beneficioso, y la media muestral cae por debajo de 0.35.
2. El bono no es beneficioso, y aún así la media muestral resulta mayor que 0.35.
3. El bono es beneficioso, y aún así la media muestral cae por debajo de 0.35.
4. El bono es beneficioso, y la media muestral resulta mayor a 0.35.

¿Cúales situaciones constituyen un acierto, y cúales un error?

Motivación

Ninguna regla de decisión es infalible dado que se trabaja con una muestra, y la variabilidad muestral es inevitable:

- Los casos 1 y 4 son aciertos, la regla de decisión lleva a una conclusión acertada sobre la población.
- El caso 2 se califica como **Error de Tipo I**, y es típicamente considerado el tipo de error mas grave, y menos tolerable.
- El caso 3 se califica como **Error de tipo II**.

¿Cómo arribó el analista a su regla de decisión con un *valor de corte* 0.35? ¿Pueden reducirse las chances de caer en los casos 2 y 3? Para contestar estas preguntas debemos estudiar detenidamente el problema.

Hoja de ruta

El procedimiento lógico para llevar a cabo un test de hipótesis es el siguiente:

1. Se plantean dos hipótesis, nula y alternativa, que cubren todos los casos y tales que no hay caso que esté en las dos a la vez (son exhaustivas y excluyentes).
2. Se escoge un resumen de los datos, un estadístico, que sea útil para aprender sobre el parámetro que involucran las hipótesis.
3. Tomando como cierta la hipótesis nula, se delimita una región de rechazo, un umbral a partir del cuál diremos que rechazamos la hipótesis nula en favor de la alternativa. Por ejemplo, si el estadístico escogido supera un valor de corte k.
4. Se analizan los datos y se contrasta el valor obtenido para el estadístico con el valor de corte prefijado.

Definiciones

Notación y definiciones:

- Hipótesis nula y alternativa
- Error de tipo I y error de tipo II
- Probabilidad de rechazar H_0
- Potencia del test
- Nivel del test y regla de decisión
- P-valor

Hipótesis nula y alternativa

Se proponen dos alternativas para el parámetro poblacional bajo estudio:

1. La primera es la hipótesis nula, que representa el "status quo". Esta es la declaración que se pondrá en tela de juicio, y se considerará factible a menos que haya **evidencia significativa** en su contra.
2. La otra es la hipótesis alternativa, la que consideraremos mas factible si hay evidencia suficiente para descartar la nula. En tal caso diremos que se descarta la hipótesis nula en favor de la alternativa.

En ningún momento aceptamos o consideramos totalmente verdadera ninguna hipótesis. Originalmente el concepto de test de hipótesis hacía referencia únicamente a una hipótesis nula. La hipótesis nula se somente a una prueba para ver si la descartamos (en nuestro caso, a favor de la alternativa), o si simplemente no hay evidencia suficiente para descartarla.

Hipótesis nula y alternativa

En el ejemplo del bono por venta, las hipótesis serían:

- Hipótesis nula (H_0): el bono no es beneficioso, $\mu \leq 0.3$. Es decir, el bono no aumenta la media de efectividad.
- Hipótesis alternativa (H_1): el bono es beneficioso, $\mu > 0.3$. Es decir, el bono aumenta la media de efectividad.

Note que así como están planteadas, las hipótesis consituyen todos los casos posibles para la media (son *exhaustivas*), y además no es posible que ambas sean ciertas a la vez (son *excluyentes*).

Regiones críticas

Una **prueba de hipótesis** con respecto a una característica desconocida de cualquier población de interés es una regla para decidir si se rechaza o no la hipótesis nula.

La decisión se basa en un **estadístico** que depende solo de información muestral. Para ciertos valores de este estadístico la decisión será rechazar la hipótesis nula. Estos valores se conocen como los **valores críticos** y determinan una **región crítica**.

Regiones críticas

Si la H_0 sobre el parámetro de interés θ es del tipo:

$$H_0 : \theta \leq \theta_0 \qquad o \qquad H_0 : \theta \geq \theta_0$$

y si la H_1 es de la forma,

$$H_1 : \theta > \theta_0 \qquad o \qquad H_1 : \theta < \theta_0$$

se dice que es una **hipótesis alternativa unilateral**. La región crítica también recibe el nombre de región de rechazo unilateral.
De otro modo, debe establecerse una **hipótesis alternativa bilateral:**

$$H_0 : \theta = \theta_0 \qquad H_1 : \theta \neq \theta_0$$

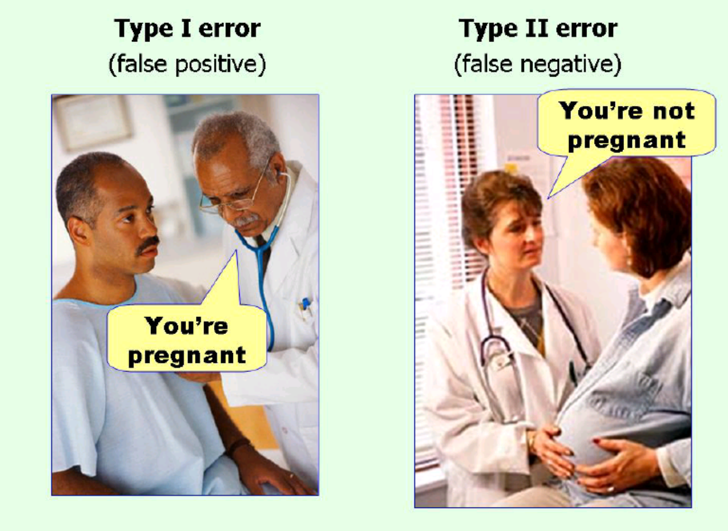
Una hipótesis alternativa bilateral implica la existencia de una región crítica bilateral (la cual es simétrica: las dos partes de la región se seleccionan de tal forma que el área bajo la curva de cada una de las regiones sea igual).

Tipos de errores

Posibles decisiones que pueden tomarse respecto a la hipótesis nula:

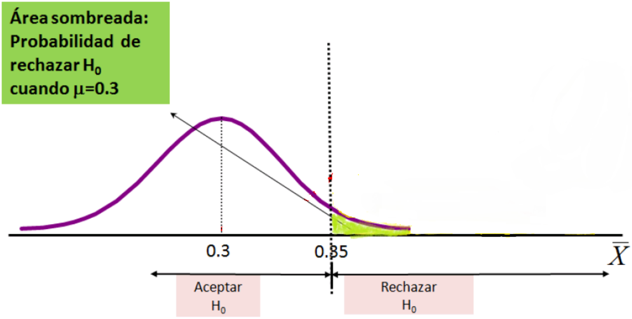
	Decisión respecto a H_0	
	Rechazar H_0	**No rechazar H_0**
H_0 **verdadera**	Error de Tipo I	Acierto
H_0 **falsa**	Acierto	Error de Tipo II

Tipos de errores



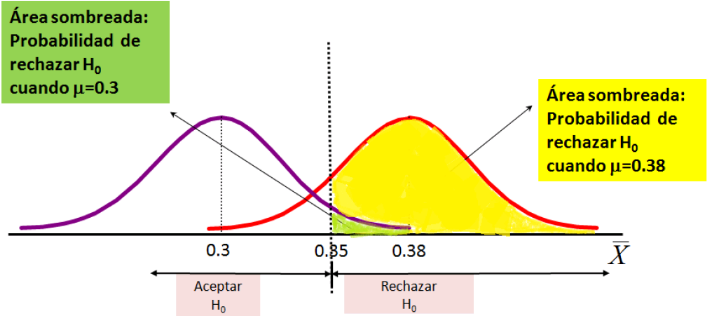
Probabilidad de rechazar H_0

Volviendo al ejemplo del bono por venta, supongamos que no tiene efecto y una vez implementado en la sucursal de prueba tomamos la muestra, que será (suponiendo que no tiene efecto, *bajo la hipótesis nula*) normal con media 0.3 y varianza 0.04. Nuestra regla de decisión es descartar $H_0 : \mu \leq 0.3$ si observamos una media muestral mayor o igual a 0.35. Así,



Probabilidad de rechazar H_0

Por el contrario, podría ocurrir que el bono sí sea beneficioso, y por ejemplo haga subir la media de la efectividad hasta 0.38, sin afectar la varianza. De esta forma cuando tomemos la muestra, lo haríamos de la distribución roja:



Potencia de un test

Definiremos la **potencia de un test** como la probabilidad de descartar la hipótesis nula, para un valor particular de la media μ que pertenezca a la hipótesis alternativa. Denotando cualquiera de estos valores como μ_1 y abusando de la notación de probabilidades condicionales, la potencia de un test es:

$$Potencia = P(\text{rechazar } H_0 | \mu = \mu_1)$$

Notar que mayor potencia es una propiedad deseable de un test: la potencia es la probabilidad de un tipo de acierto, detectar hipótesis alternativas.

¿Cuál es la potencia del test propuesto por el analista si la media verdadera fuera 0.38? ¿y si la media verdadera fuera 0.31? De infinitas repeticiones de ambos casos, ¿en cuál el test se da cuenta que la hipótesis nula es falsa mas veces?

Nivel de un test

El **nivel de un test** es la mayor probabilidad de rechazar incorrectamente la hipótesis nula. Representa de alguna forma "el peor de los casos", cuando en realidad H_0 es cierta pero el parámetro en cuestión está "muy cerca" de la región de H_1.

Mas formalmente:

$$\max_{\mu_0} \{P(\text{rechazar } H_0 | \mu = \mu_0)\}$$

Intuitivamente, el nivel de un test indica, en el peor de los casos, que probabilidad tenemos de cometer error de tipo I. Esta cantidad suele denotarse con α.

Nivel del test y regla de decisión

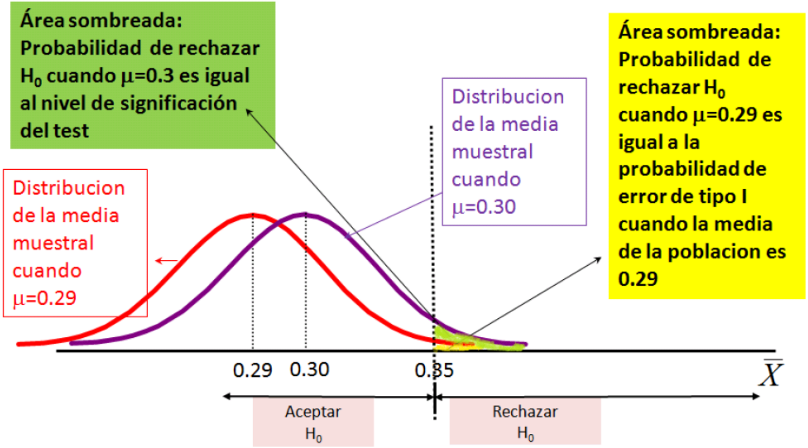
Una regla de decisión es una función que en base a datos de la muestra nos dice si descartamos o no la hipótesis nula.

En el caso de la efectividad de los vendedores, la regla era "se descarta H_0 si la media muestral es superior a 0.35".

Se denota con α a la probabilidad mas alta que puede haber de cometer error de tipo I al usar una regla de decisión particular. El valor α es conocido como el **nivel de significación del test**.

¿Cuál es el nivel de significación que escogió el analista para arribar a esa regla de decisión?

Nivel del test y regla de decisión



Eligiendo la regla de decisión

Como se considera que el error de tipo I es el mas grave de los dos, el procedimiento usado para determinar el valor de corte de la regla de decisión es fijar el nivel de significación del test en un valor que uno esté dispuesto a tolerar.

Se ancla α en un valor, por ejemplo 0.05. En base a ese nivel de tolerancia al error de tipo I se escoge una regla de decisión que no cometa error de tipo I mas del 5% de las veces en muchas muchas repeticiones del procedimiento.

Todo lo demás constante, si modificamos el valor de corte de la regla de decisión tenemos un trade-off entre potencia y error de tipo I.

Eligiendo la regla de decisión

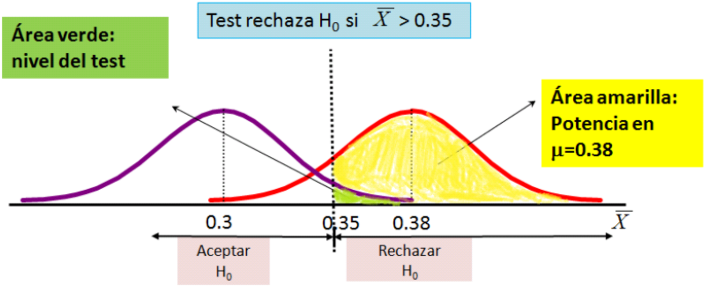
Supongamos que la población es normal y conocemos la varianza σ^2. Si nuestras hipótesis son del tipo

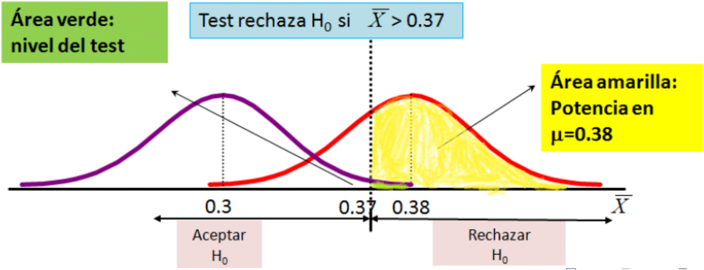
$$H_0 : \mu \leq \mu_0 \qquad H_1 : \mu > \mu_0$$

Y utilizamos un nivel de significación del $\alpha\%$, el valor de corte k para la media muestral $\bar{X}$ verifica:

$$\begin{aligned}
P(\text{Rechazar } H_0|\mu = \mu_0) &= \alpha \\
P(\bar{X} \geq k|\mu = \mu_0) &= \alpha \\
P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) &= \alpha \\
P\left(Z \geq \frac{k - \mu_0}{\sigma/\sqrt{n}}\right) &= \alpha \\
\Rightarrow \frac{k - \mu_0}{\sigma/\sqrt{n}} &= z_\alpha \\
k &= \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}
\end{aligned}$$

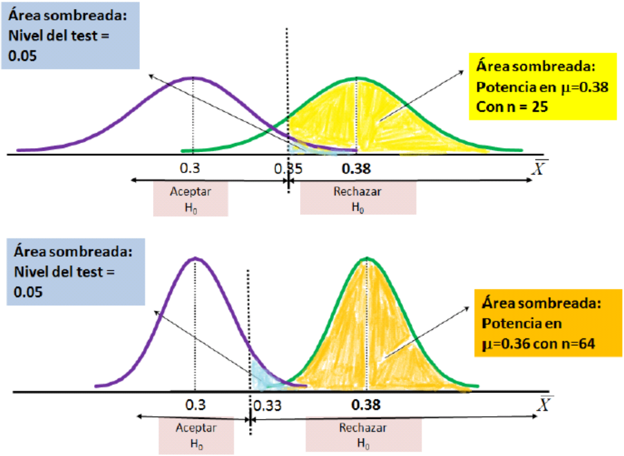
Potencia y nivel de significación





Potencia y nivel de significación

¿Cree que haya alguna forma de poder mejorar la potencia del test manteniendo el nivel de significación?



Ejercicio (muy conceptual)

Consideremos el siguiente juego que ayudará a fijar algunos conceptos. Hay dos bolilleros A y B. Cada bolillero contiene bolillas de 7 colores diferentes a las que denominamos con los números 1 al 7. Usted conoce las proporciones de bolillas de cada color que hay en cada bolillero las cuales quedan indicadas en la siguiente tabla:

	1	2	3	4	5	6	7
A	0.02	0.01	0.01	0.01	0.01	0.01	0.93
B	0.06	0.05	0.04	0.03	0.02	0.01	0.79

El juego consiste en que una persona saca una bolilla al azar de uno de los dos bolilleros y le informa el color de la bolilla que salió sorteada, pero no le indica de qué bolillero la sacó. Usted tiene que adivinar el bolillero de donde fue extraída la bolilla solamente conociendo su color (o número).

Ejercicio (muy conceptual)

Llamemos X a la variable aleatoria que consiste en el número que está en la bolilla que será extraida. Definamos un conjunto $\mathcal{C}$, que puede ser cualquier subconjunto de $\{1, 2, 3, 4, 5, 6, 7\}$, y tal que si el número de la bolilla extraida pertenece a $\mathcal{C}$ decidiremos que dicha bolilla proviene del bolillero B, mientras que de no ser así, diremos que viene del bolillero A.

- Indique todos los subconjuntos $\mathcal{C}$ tales que, cuando la bolilla realmente haya sido extraída del bolillero A, la probabilidad de que usted termine eligiendo incorrectamente el bolillero B sea igual a 0.02.
- De todos los subconjuntos encontrados en el inciso anterior indique aquel que maximiza la probabilidad de elegir el bolillero B cuando la bolilla efectivamente haya salido del bolillero B.

Hipótesis alternativa bilateral

Una fábrica de autopartes desea monitorear su proceso de fabricación de ciertos pistones para evaluar si sus diámetros cumplen con la especificación estipulada de 2 cm.

Cuando la máquina que produce los pistones está bien calibrada, los diámetros de los pistones producidos por la máquina siguen una distribución normal con una media de 2 cm y un desvío estandard de 0.06 cm. (aún bajo un correcto funcionamiento, la máquina no genera pistones de exactamente 2cm debido a pequenas fluctuaciones incontrolables en la temperatura ambiente, la tensión eléctrica, etc. que alteran ligeramente el proceso de perforación).

Se desea monitorear el proceso de fabricación, en particular, si la máquina está bien calibrada. Con este objetivo se elijen al azar 9 pistones y en cada uno de ellos se mide minuciosamente su diámetro.

Hipótesis alternativa bilateral

Como en los ejemplos anteriores, podemos estar en cualquiera de los siguientes dos universos:

- la máquina está bien calibrada: la media del diámetro de los pistones es igual a 2 cm; o
- la máquina NO está bien calibrada: la media del diámetro de los pistones no es igual a 2 cm.

Notar que este caso es algo distinto a los previos, en el sentido de que no queremos saber si un parámetro supera o no un umbral en una sola dirección, sino que buscamos saber si dista de un número. En este caso si la media es distinta (menor o mayor) de 2.

Una posible regla de decisión es declarar que la máquina no está bien calibrada si y sólo si:

$$|\bar{X} - 2| > 0.04$$

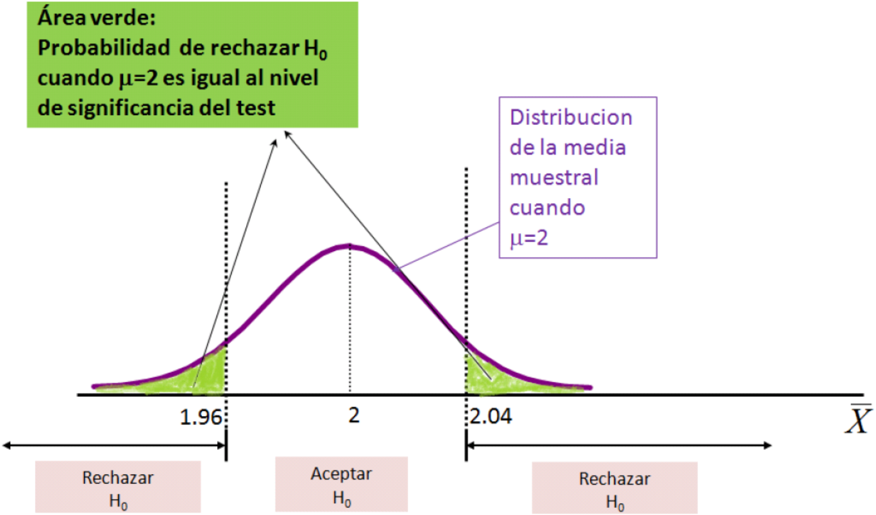
O lo que es lo mismo, si y sólo si:

$$\bar{X} > 2.04 \quad \text{ó} \quad \bar{X} < 1.96$$

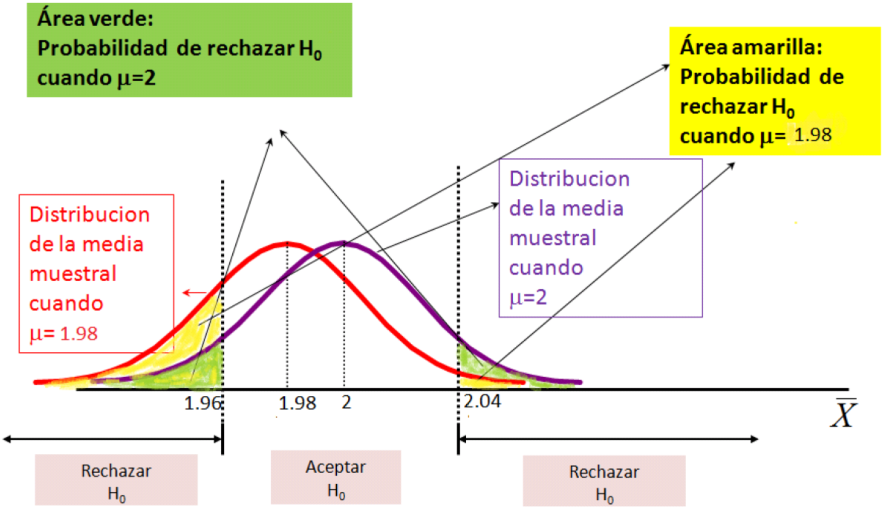
En el lenguaje que venimos desarrollando estableceríamos $H_0 : \mu = 2$ y $H_1 : \mu \neq 2$.

No es tan obvio como se extienden todos los conceptos que vimos hasta ahora para este tipo de hipótesis...

Hipótesis alternativa bilateral



Hipótesis alternativa bilateral



Verdadero o Falso

- El nivel de significación de un test es igual a la probabilidad de que la hipótesis nula sea cierta.
- Si un test con nivel de significación α rechaza la hipótesis nula, entonces la probabilidad de que la hipótesis nula sea cierta es igual a α.
- La probabilidad de que un test rechace la hipótesis nula incorrectamente (es decir, cuando es cierta) es igual a la potencia del test.
- Un error de tipo I ocurre cuando el test rechaza la hipótesis nula.

Ejercicio 1

Una empresa panificadora produce pan lactal de mesa en envases que anuncian contener un pan de 1kg. Según las reglamentaciones de defensa al consumidor, la empresa enfrentaría problemas si se detecta que los panes anunciados de 1kg pesan en promedio menos de 1kg.

Durante una inspección, el ente regulador supondrá que la empresa cumple con la reglamentación (es decir, la media de sus panes es mayor o igual a 1kg) evaluará si hay evidencia para descartar dicha afirmación en base a una muestra de 25 panes.

Se sabe por el fabricante de la máquina utilizada en la producción que el peso de los panes horneados se distribuye normal con varianza $\sigma^2 = 0.04$. Suponga que la media muestral obtenida es $\bar{x} = 0.94$

Ejercicio 1

- ¿Qué implicancias tiene en este caso cometer error de tipo I? ¿Y error de tipo II?
- Supongamos que si la panificadora logra demostrar su inocencia luego de ser declarada fraudulenta, el ente regulador enfrentaría un proceso judicial extremadamente costoso. Es por esto que el analista del ente regulador decide utilizar un nivel de significación del 1%. ¿Cuál es la regla de decisión en este caso? ¿Qué concluye el ente regulador?
- Con la regla de decisión establecida previamente, ¿cuál es la potencia del test si la media verdadera es $\mu_1 = 0.9$?
  Supongamos ahora que en caso de declarar incorrectamente que la panificadora es fraudulenta no hay graves consecuencias. Así, se escoge un nivel de significación del 10%. ¿Qué concluye el ente en este caso?

Ejercicio 2

Dada una muestra aleatoria tomada de una población con varianza $\sigma^2 = 625$, se desea testear la hipótesis nula $H_0 : \mu \leq 100$ frente a $H_1 : \mu > 100$ con un nivel de significación de $\alpha = 0.05$. Calcular el valor crítico $\bar{x}_c$ y la regla de decisión para los siguientes casos:

- Tamaño de muestra $n = 16$
- Tamaño de muestra $n = 25$
- Tamaño de muestra $n = 49$
- Comente como afecta n al valor crítico $\bar{x}_c$

Puede realizar el mismo ejercicio para un solo n pero suponiendo distintas varianzas poblacionales σ^2: 225, 400, y 900.

P-valor

Como notamos en el ejercicio anterior, una posible forma de contrastar una afirmación es escoger un nivel de significación y en base a ello utilizar una regla de decisión.

Así, distintas personas con distintos niveles de tolerancia al error de tipo I escogerían distintos umbrales a partir de los cuales descartar la hipótesis nula.

Como a medida que cae el nivel de significación uno está "subiendo la exigencia" para declarar falsa la hipótesis nula, podríamos pensar que ocurre con nuestra sentencia para una muestra particular a medida que modificamos α.

P-valor

Retomemos el ejemplo de la panificadora. Si $\sigma = 0.2$, $n = 25$, y en base a la muestra obtenida vemos $\bar{x} = 0.94$ calculamos que para un nivel de significación del 1% no se descarta la hipótesis nula, mientras que para un nivel de 10% si se descarta.

Pueden hacer los cálculos, pero para $\alpha = 0.9$ se descarta, para $\alpha = 0.7$ se descarta, para $\alpha = 0.6$ ya no se descarta (recuerden que a menor α nos volvemos mas prudentes/exigentes, porque nos gusta cada vez menos cometer error de tipo I)...

Un valor interesante es aquel nivel de significación tal que, con los datos obtenidos, si hubiéramos fijado ese nivel hubieramos descartado la hipótesis nula, pero de haber fijado uno levemente inferior, no.

P-valor

Llamamos *p-valor* al **mínimo nivel de significación** tal que, con los datos obtenidos, se descarta la hipótesis nula.

Notar que hasta ahora hablamos de propiedades de la herramienta que usamos, pero el p-valor es una propiedad de la *muestra obtenida*

Otra forma de entenderlo es como la probabilidad de obtener evidencia tan o más sorprendente en contra de H_0, es decir, la probabilidad de haber obtenido la muestra que nos tocó, o incluso una con mayor evidencia en contra de H_0, bajo la hipótesis nula.

NO es la probabilidad de que H_0 sea cierta. H_0 es cierta o no, pero nosotros no lo sabemos.

P-valor

Notar entonces que uno podría llevar a cabo un estudio y reportar el p-valor obtenido, para que luego cada persona que quiera utilizar los resultados compare ese p-valor con un nivel de significación que considere apropiado.

En la literatura suele usarse como valor 'estándar' que uno obtiene resultados 'estadísticamente significativos' si el p-valor del estudio es menor que 0.05. Es decir, tal que si fijáramos $\alpha = 0.05$ descartaríamos H_0.

Esto es sólo un valor de referencia comunmente usado. No tiene algún criterio detrás, es simplemente convención.

Ejercicio 3

Para una muestra aleatoria obtenida a partir de una población con varianza $\sigma^2 = 400$, n=16, se obtuvo una media muestral de $\bar{x} = 70$. Considere la hipótesis nula $H_0 : \mu \geq 80$ frente a $H_0 : \mu < 80$.

- ¿Cuál es el p-valor de esta muestra?
- Calcule nuevamente el p-valor para n=36.
- Calcule nuevamente el p-valor para n=64

Puede realizar el mismo ejercicio para un solo n pero suponiendo distintas varianzas poblacionales σ^2: 225, 400, y 900.

Ejercicio 4

A un fabricante farmacéutico le preocupa que la concentración de impurezas en las pastillas no supere el 3%. Se sabe que la concentraciones de impurezas sigue una distribuci ón normal con un desvío estándar de 0.4%. En una muestra aleatoria de 64 pastillas se encontró que la concentración de impurezas era 3.07%.

- Pruebe, considerando una probabilidad de error de Tipo I de a lo sumo 5%, la hipótesis nula de que la concentración media de impurezas es del 3% o menos, frente a la alternativa de que sea más de 3%.
- Encuentre el p-valor para esta prueba.

Ejercicio 4

- Considere ahora una hipótesis nula puntual y la hipótesis alternativa bilateral $H_1 : \mu \neq 3$. Discuta, sin hacer cálculos, si el p-valor de la prueba sería mayor que, menor que, o igual que el encontrado en la pregunta anterior. Realice una gráfico para ilustrar su razonamiento.
- En el contexto de este problema, explique por qué una hipótesis alternativa unilateral parecería ser más apropiada que una hipótesis alternativa bilateral. ¿Se le ocurre algún contexto similar donde pueda argumentarse que una alternativa bilateral sea mas adecuada?

Ejercicio 5

Los que están a favor de la energía eólica afirman que un nuevo molino de viento puede generar un promedio de al menos 800 kilovatios de potencia por día. Se puede suponer que la generación diaria de energía del molino de viento tiene distribución normal con un desvío estándar de 120 kilovatios. Para probar esta afirmación se toma una muestra aleatoria de 100 días*. La hipótesis alternativa sería que la verdadera media es inferior a 800 kilovatios. La afirmación no será rechazada si la media de la muestra es de 776 kilovatios o más y será rechazada en caso contrario.

- Utilizando está regla de decisión, ¿cuál es la probabilidad α de realizar un Error de Tipo I si la media poblacional es realmente 800 kilovatios por día?
- Utilizando la misma regla de decisión, ¿cuál es la probabilidad β de realizar un Error de Tipo II si la media poblacional es realmente 740 kilovatios por día?

Ejercicio 5

Supongamos ahora que se utiliza la misma regla de decisión, pero en lugar de trabajar con una muestra de 100 días, se trabaja con una muestra de 200 días.

- ¿El valor de α será mayor, menor, o igual al que se encontró en la parte anterior?
- ¿El valor de β será mayor, menor, o igual al que se encontró en la parte anterior?

Hasta ahora

Con lo que vimos hasta ahora uno estaría mas atento a la hora de leer titulares que digan que en base a una muestra de 1000 posibles votantes, el candidato A supera al candidato B. Querríamos saber si nos dió ese resultado de pura suerte o si a nivel poblacional también es así

En particular, podríamos construir un intervalo de confianza para ambas estimaciones puntuales, o ingeniarnosla para armar un contraste de hipótesis del estilo $H_0 : p_A \geq p_B$ vs $H_1 : p_A < p_B$ y testear.

Vimos conceptualmente (y en algunos casos desarrollamos matemáticamente) muchas ideas que están presentes en contrastes de hipótesis y son relevantes para entender el criterio utilizado a la hora de emitir juicios sobre afirmaciones acerca de un parámetro poblacional.

Ahora toca la parte de receta de cocina, y explorar distintos casos y ejemplos donde testeamos sobre parámetros que no sean la media, o sobre más de un parámetro a la vez.

Estadísticos

Hasta ahora procedimos contrastando la media muestral con un valor de corte, como en el caso del bono que dicho valor 0.35.

De manera análoga podríamos trabajar directamente con una versión *estandarizada* de este valor. Suponiendo que la población es normal, la varianza es conocida y la media bajo la hipótesis nula es μ_0:

$$\bar{X} > 0.35 \quad \iff \quad Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{0.35 - \mu_0}{\sigma/\sqrt{n}}$$

Donde $Z \sim N(0,1)$. En el ejemplo del bono entonces la región de rechazo utilizando el estadíatico estandarizado Z es

$$\text{Se rechaza la hipótesis nula si } Z > \frac{0.35 - 0.3}{0.2/\sqrt{25}} = 1,25$$

Estadísticos

En el ejemplo anterior estábamos realizando un test para la media μ suponiendo que la varianza es conocida. De no ser así (como en la mayoría de los casos), el procedimiento es análogo pero se utiliza el estadístico

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

Que tiene una distribución t de Student con $n - 1$ grados de libertad.