# CONTRASTE DE HIPÓTESIS

Teoría y cálculo para una muestra

Instituto CBQ
Educamos diferente

# FUNDAMENTACIÓN

- Muchas veces, al tomar una muestra nos interesa conocer si los resultados podrían proceder de una población de la que su parámetro es conocido.
- Por ejemplo, si queremos ver si una muestra de enfermos se diferencian en un parámetro de estudio de la población sana (de la cual conocemos su media), por ejemplo en el resultado de un test de laboratorio.

# HIPÓTESIS

- Son afirmaciones supuestas acerca de una población.
- En primer lugar, siempre se pone a prueba una afirmación llamada hipótesis nula (Ho), que establece que la muestra procede de la población. Dicho de otra manera, las diferencias entre la muestra y la población, aceptada la hipótesis nula se deben puramente al error de muestreo aleatorio. Solo se rechaza con evidencia convincente.
- La hipótesis alternativa ofrece la opción complementaria a la Ho y es la hipótesis de trabajo, que se aceptara solo si hay evidencia convincente.
- Entre las dos hipótesis se cubren todas las posibilidades, por lo que una de las dos siempre es verdadera.

# CONTRASTE DE HIPÓTESIS

- Una forma de abordar el problema de si rechazo o no la Ho es a través de los intervalos de confianza
- Queremos saber si la diferencia entre un parámetro en la Ho y su valor en la muestra se puede atribuir al error de muestreo o es muy poco probable de que sea por esto.
- El procedimiento por el cual decidimos si aceptar o rechazar un par de hipótesis se llama prueba o ensayo de la hipótesis, por el cual probamos si los valores de la muestra difieren significativamente de los resultados esperables bajo la hipótesis.

# DESARROLLO

- Si se sabe que el intervalo definido por:

$$\mu \pm 1{,}96\, {}^{S}/_{\sqrt{n}}$$

incluye al 95% de las medias muestrales, cuando se obtiene una media por fuera de este intervalo se dice que hay una diferencia significativa, con un nivel de 5% y un p menor a 0,05; siendo p la probabilidad de error al rechazar la hipótesis nula.

- Si la media muestral cae dentro del intervalo anterior no se puede rechazar la Ho a ese nivel de significación, porque la media muestral está dentro del intervalo en el que se encuentran la mayoría de las medias muestrales para ese valor de población.

# RECORDAR

- La Ho nunca se acepta, porque un resultado negativo nunca es evidencia de nada, es decir, el hecho de que no se observe una diferencia significativa no quiere decir que no exista, solo que con esa muestra no podemos comprobarla.
- De manera análoga, la Ha no se rechaza, sino que cuando hay diferencia significativa se acepta, y cuando no la hay no se acepta.

# NIVEL DE SIGNIFICACIÓN

- El nivel de significación, o alfa, viene de la construcción de los intervalos de confianza y en este caso representa la probabilidad de rechazar la hipótesis nula cuando esta es verdadera.
- Se elige antes de realizar la prueba
- Define los valores críticos que separan la zona de no rechazo y rechazo de la curva del intervalo de confianza

# SIGNIFICACIÓN OBSERVADA

- Al realizar el contraste obtengo un valor de probabilidad de obtener una muestra que sea mas distinta de la hipótesis nula que la nuestra.
- Se le llama valor p y representa la probabilidad de obtener una muestra todavía mas alejada solo por azar.
- Cuando p es menor que alfa se rechaza la Ho

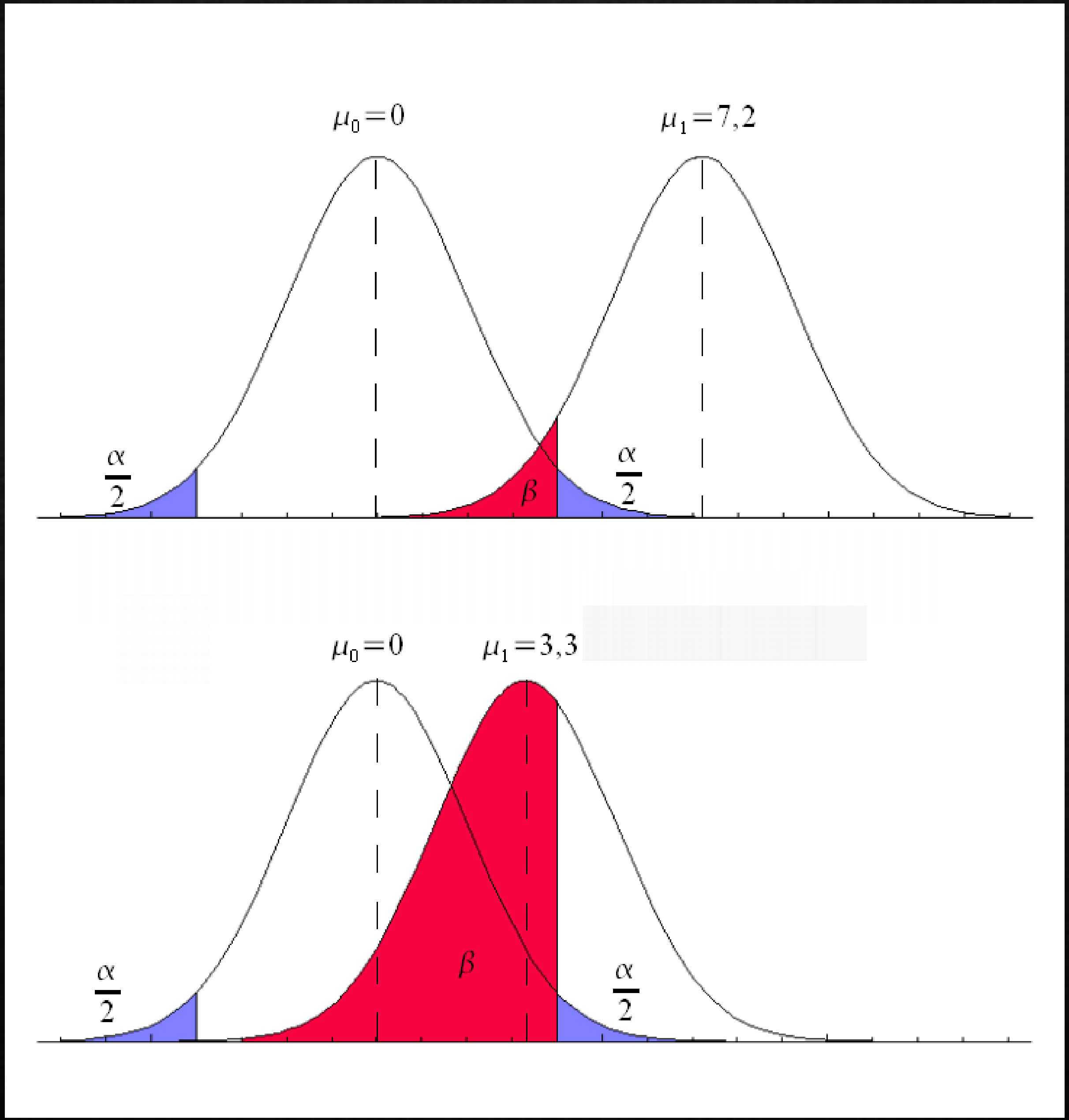
# ERRORES EN EL CONTRASTE

- Ya que todo esto se realiza a partir de intervalos de confianza siempre están presente dos posibilidades de errores:
  - Tipo I: Rechazar la hipótesis nula cuando es verdadera. Es igual al nivel de significación.
  - Tipo II: No rechazar la hipótesis nula cuando es falsa. Se le llama beta, y también se encuentra en relación al nivel de significación.

| | **Realidad** | |
|---|---|---|
| | Hipótesis nula es cierta | Hipótesis nula es falsa |
| No rechazo H nula | Correcto | Error de tipo II, hay una diferencia pero en la muestra obtenida no lo percibimos (Probabilidad Beta) |
| Rechazo H nula, acepto alternativa | Error de tipo I, no hay una diferencia, pero los datos de la muestra me llevan a decidir que sí (Probabilidad Alfa) | Correcto |

# PODER

- A partir de la probabilidad beta se define su complementaria, la probabilidad de rechazar la hipótesis nula cuando esta no es cierta; o de aceptar la hipótesis alternativa cuando es verdadera. Esto es el poder de la prueba
- Un mínimo aceptable en un diseño de pruebas es de un poder de 0,8.


$\mu_0 = 0$
$\mu_1 = 7,2$
$\frac{\alpha}{2}$
$\beta$
$\frac{\alpha}{2}$
$\mu_0 = 0$
$\mu_1 = 3,3$
$\frac{\alpha}{2}$
$\beta$
$\frac{\alpha}{2}$

# CONTRASTE DE HIPÓTESIS DE MEDIAS VARIANZA POBLACIONAL CONOCIDA

- Establecer las hipótesis.
- En este caso la distribución muestral es normal y se utiliza la siguiente prueba:

$$Z_c = \frac{(\bar{x} - \mu)}{(\sigma / \sqrt{n})}$$

- Calcular el estadístico.
- Comparar con valores críticos.
- Conclusiones.

# CONTRASTE DE HIPÓTESIS DE MEDIAS VARIANZA POBLACIONAL DESCONOCIDA

- Establecer las hipótesis.
- En este caso la distribución muestral sigue una distribución de t de student y se utiliza la siguiente prueba:

$$t_c = {(\bar{x} - \mu)} \Big/ {\left( \hat{s} \big/ \sqrt{n} \right)}$$

- Calcular el estadístico.
- Comparar con valores críticos.
- Conclusiones.

# CONTRASTE DE HIPÓTESIS DE PROPORCIONES

- Establecer las hipótesis.
- En este caso la distribución muestral sigue una distribución normal y se utiliza la siguiente prueba:

$$z = \frac{p - k}{\sqrt{k(1-k)/n}}$$

- Calcular el estadístico.
- Comparar con valores críticos.
- Conclusiones.

| Hipótesis nula: $H_0$ | Hipótesis alternativa $H_1$ | Tipo de contraste | Estadístico de contraste | Región de aceptación (El contraste se puede realizar utilizando una de las dos columnas; el resultado no varía) | |
|---|---|---|---|---|---|
| $\mu = \mu_0$ | $\mu \neq \mu_0$ | ***bilateral*** | $Z=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ | $\left(-z_{\frac{\alpha}{2}}\ ;\ z_{\frac{\alpha}{2}}\right)$ | $\left(\mu_0 - z_{\frac{\alpha}{2}} \cdot \dfrac{\sigma}{\sqrt{n}}\ ;\ \mu_0 + z_{\frac{\alpha}{2}} \cdot \dfrac{\sigma}{\sqrt{n}}\right)$ |
| $\mu \leq \mu_0$ | $\mu > \mu_0$ | ***Unilateral derecho*** | | $\left(-\infty\ ;\ z_{\alpha}\right)$ | $\left(-\infty;\ \mu_0 + z_{\alpha} \cdot \dfrac{\sigma}{\sqrt{n}}\right)$ |
| $\mu \geq \mu_0$ | $\mu < \mu_0$ | ***Unilateral izquierdo*** | $Si\ \sigma\ desconocida$ <br> $\sigma \rightarrow \hat{S}$ | $\left(-z_{\alpha};+\infty\right)$ | $\left(\mu_0 - z_{\alpha} \cdot \dfrac{\sigma}{\sqrt{n}}\ ;+\infty\right)$ |
| **El tipo de hipótesis nula debe plantearse de acuerdo al enunciado de cada problema** | | | | ***Ver si Z está dentro del intervalo*** | ***Ver si el valor media muestral está dentro del intervalo*** |

# EJERCICIO 1

Se cree que el tiempo medio de ocio al día que dedican los estudiantes universitarios sigue una distribución normal de media 350 minutos y tiene una desviación típica poblacional conocida de 1 hora.

En una muestra aleatoria de 100 alumnos se observa un tiempo medio de 320 minutos.

Con un nivel de significación de 90%, ¿es posible la media poblacional supuesta en base a los datos de la muestra?

# EJERCICIO 2

Un periodista afirma indignado que el 70% de los jóvenes de la ciudad usan las redes sociales como principal medio de información. Para ver la posibilidad de esta afirmación un liceo toma una muestra aleatoria de 500 adolescentes y cuentan que 340 de ellos usan efectivamente las redes sociales como principal medio de información.

¿Se puede aceptar con un 1% de error (alfa), que lo afirmado en el comienzo es cierto?

# EJERCICIO 3

Un sociólogo ha pronosticado, que si el voto no fuera obligatorio, el nivel de abstención en las próximas elecciones sería del 40% como mínimo. Se elige al azar una muestra aleatoria de 200 individuos, con derecho a voto, 75 de los cuales estarían dispuestos a votar. Determinar con un nivel de significación del 1%, si se puede admitir el pronóstico.

# EJERCICIO 4

- En una policlínica se quiere saber si la población local se ajusta a las tablas de IMC internacionales de la OMS para aplicar recomendaciones. Para el rango etario de la población local el IMC medio según la OMS sería 25.
- Se toma una muestra de 20 usuarios que arroja una media de 26 y desviación muestral de 4.
- ¿Es posible afirmar que la población local tiene un IMC medio de 25 con un 5% de significación?