Análisis Cuantitativo I

Introducción a las pruebas de hipótesis

Carlos Cardona Andrade

Universidad del Rosario

31 de marzo de 2017

Quiz

- Un laboratorio farmacéutico afirma que el número de horas que un medicamento de fabricación propia tarda en curar una determinada enfermedad sigue una distribución normal con desviación estándar igual a 8. Se toma una muestra de 100 enfermos a los que se les administra el medicamento y se observa que la media de horas que tardan en curarse es igual a 32.
- Encontrar un intervalo de confianza, con un nivel de confianza del 99 %, para la media del número de horas que tarda en curar el medicamento.

Introducción

- Usualmente es imposible (y poco práctico) para un investigador, observar cada uno de los individuos en una población.
- Considerando esto, se toma una muestra con el propósito de responder preguntas acerca de la población.
- Una prueba de hipótesis es un método estadístico para llevar a cabo inferencias sobre la población de interés.
- La prueba de hipótesis es uno de los procedimientos más comunes para realizar inferencia estadística. De hecho, en lo que resta del curso se harán pruebas de hipótesis en una variedad de situaciones y aplicaciones diferentes.

- Aunque ciertos detalles de una prueba de hipótesis cambia de una situación a otra, el proceso general se mantiene constante. En esta clase se introduce este proceso general.
- Para efectuar una prueba de hipótesis, es necesario combinar las técnicas estadísticas que se han desarrollado en clases anteriores (z-score, probabilidad y distribuciones muestrales).

Prueba de Hipótesis

Es un método estadístico que usa una muestra para evaluar una hipótesis acerca de una población.

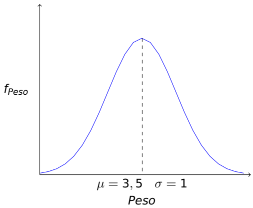
- En términos simples, la lógica detrás del procedimiento para una prueba de hipótesis es la siguiente:

1. Primero que todo, se establece una hipótesis acerca de una población. Normalmente, la hipótesis es sobre un parámetro poblacional. Por ejemplo, podríamos hipotetizar que los adultos colombianos ganan un promedio de $\mu = 3,5$ kgs entre el 24 y 31 de diciembre de cada año.
2. Antes de seleccionar una muestra, se utiliza la hipótesis para predecir las características que la muestra podría tener. Siguiendo el ejemplo anterior, al predecir que la ganancia de peso promedio para la población es $\mu = 3,5$ kgs entonces, se puede concluir que la muestra debería tener una media *cercana* a $3,5$.

3. Luego, se obtiene una muestra de la población. Por ejemplo, se podría seleccionar una muestra de $n = 200$ adultos colombianos y medir el cambio de peso promedio entre ambas fechas.
4. Finalmente, se comparan los datos de la muestra con la predicción que fue hecha de la hipótesis. Si la media muestral es consistente con la predicción, se puede concluir que la hipótesis es razonable. Sin embargo, si existe una gran discrepancia entre los datos y la predicción, se concluye que la hipótesis es errada.

La situación básica

Población Conocida

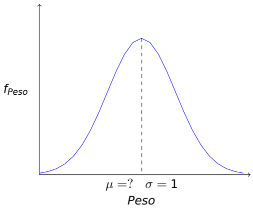


Tratamiento

$\rightarrow$

Condición

Población Desconocida



- Se asume que el parámetro poblacional es conocido para todos los individuos (condición) o antes de llevar a cabo el experimento (tratamiento).
- En el caso de la gráfica anterior, también se asume que el efecto del tratamiento es igual para todos los individuos.

National Cultures and Soccer Violence

Miguel, Saiegh & Satyanath (2008)

- ¿Pueden algunos actos de violencia ser explicados por la cultura de una sociedad?
- Los autores aprovechan el experimento natural ofrecido por la presencia de miles de jugadores internacionales en las 5 grandes ligas europeas.
- ¿Cómo medir comportamiento violento en el fútbol? $\Rightarrow$ Tarjetas amarillas y rojas.
- ¿Cómo medir la historia de violencia de un país? $\Rightarrow$ Años bajo conflicto civil.

- Supongamos que se quiere evaluar si los jugadores, cuyo país de origen ha sufrido más de 5 años de conflicto civil, se comportan de manera diferente dentro del campo de juego.
- Para la población general de jugadores, el número de tarjetas amarillas por temporada se distribuye normal con media $\mu = 4$ y desviación estándar de $\sigma = 2,5$.
- El plan es tomar una muestra de $n = 35$ jugadores que cumplan la condición de conflicto y comparar el promedio de tarjetas amarillas por temporada.
- Si la media de la muestra es, de manera notable, diferente de la media de la población de jugadores de las 5 grandes ligas europeas, se puede concluir que el conflicto tiene algún efecto sobre el comportamiento violento de los jugadores.

Pasos para una prueba de hipótesis

1. Establecer las hipótesis

- Previamente se establecieron dos hipótesis, las cuales están en términos de los parámetros poblacionales.
- La primera, y más importante, de las dos hipótesis es la *hipótesis nula*.
- Esta hipótesis afirma que el tratamiento/condición no tuvo efecto. En general, establece que no se presenta cambio, diferencia, efecto alguno (De ahí proviene el nombre *nulo*).
- Se identifica con el símbolo H_0 (La H es por hipótesis y el 0 indica que es la hipótesis de efecto cero).
- En nuestro caso:

$$H_0 \ : \mu_{conflicto} = 4$$

Pasos para una prueba de hipótesis

1. Establecer las hipótesis

- La segunda hipótesis es el opuesto de la hipótesis nula, y se denomina la *hipótesis alterativa* (H_1). Esta hipótesis indica que el tratamiento/condición tiene un efecto sobre la variable de interés.
- Para el ejemplo, la hipótesis alternativa es que el conflicto en el país de origen tiene un efecto sobre el número de tarjetas amarillas que recibe un jugador y, por lo tanto, habrá un cambio en su media.

$$H_1 \ : \mu_{conflicto} \neq 4$$

- Noten que no se está fijando una dirección del efecto.

Pasos para una prueba de hipótesis

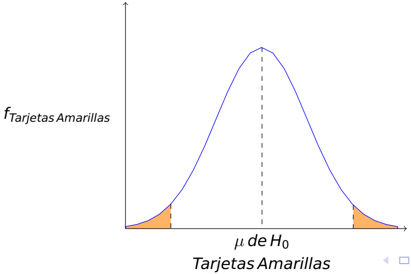
2. Fijar el criterio de decisión

- Para formalizar el proceso de decisión, se utiliza la hipótesis nula para predecir el tipo de media muestral resultante.
- Específicamente, se quiere evaluar cuáles medias muestrales son consistentes con la hipótesis nula y cuáles no lo son.
- Para determinar cuáles valores están cerca a 4 y cuáles son muy diferentes, se examinan todas las medias muestrales posibles si la hipótesis nula es cierta.
- Con respecto al ejemplo, esta es la distribución de medias muestrales para $n = 35$. De acuerdo a la hipótesis nula, la distribución está centrada en $\mu = 4$.

Pasos para una prueba de hipótesis

2. Fijar el criterio de decisión

- La distribución de medias muestrales se divide en dos secciones:
    1. Las medias muestrales que son probables de obtener si H_0 es cierta, es decir, las medias muestrales cercanas a la hipótesis nula.
    2. Las medias muestrales que son poco probables de obtener si H_0 es cierta, es decir, medias muestrales que son muy diferentes de la hipótesis nula.



Pasos para una prueba de hipótesis

2. Fijar el criterio de decisión

- Para encontrar los límites que separan las medias de alta probabilidad de las de baja probabilidad, se tiene que definir exactamente qué significa “alta” y “baja” probabilidad.
- Esto se logra al seleccionar una probabilidad específica , la cual se conoce como *nivel de significancia* (se denota con α), para la prueba de hipótesis.
- El valor α es una probabilidad pequeña que se utiliza para identificar muestras de poca probabilidad.
- Por convención, los valores alfa más comunes son $\alpha = 0{,}05$ y $\alpha = 0{,}01$. Por ejemplo, si usamos $\alpha = 0{,}05$, separamos el 5 % de las medias más improbables (valores extremos) del 95 % de las medias muestrales más probables (valores centrales).

Pasos para una prueba de hipótesis

2. Fijar el criterio de decisión

- Los valores extremos que son poco probables, definidos por el nivel de significancia, constituyen lo que se conoce como *región crítica*.
- Estos valores son inconsistentes con la hipótesis nula. También se pueden interpretar como valores muestrales que proveen evidencia convincente de que el tratamiento/condición tienen algún efecto.
- Al igual que con los intervalos de confianza, para determinar la ubicación exacta de los límites se utilizan el α y la tabla de la normal para encontrar el z-score *crítico*.
- Para un valor $\alpha = 0{,}05$, los límites separan el 5 % extremo del 95 % del medio. Este 5 % se divide entre las dos colas de la distribución, dejando 2.5 % en cada extremo ($Z = 1{,}96$).

Pasos para una prueba de hipótesis

3. Calcular los estadísticos muestrales

- Lo anterior debe hacerse antes de la recolección de los datos para asegurar una evaluación honesta y objetiva de los datos.
- Luego de la recolección es posible comparar la media muestral con la hipótesis nula. Esta comparación se logra al calcular el z-score que ubica la media muestral respecto a la media poblacional de H_0.
- El z-score para una media muestral se obtiene de la siguiente manera:

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

Pasos para una prueba de hipótesis

4. Tomar una decisión

- En este paso, se utiliza el z-score del paso anterior para tomar una decisión acerca de la hipótesis nula de acuerdo al criterio seleccionado en el paso 2. Existen dos posibles resultados:

1. La media muestral se ubica en la región crítica. Por definición, un valor muestral en esta región es poco probable si H_0 es cierta. Se puede concluir que la muestra no es consistente con la hipótesis nula, por lo tanto, nuestra decisión es rechazarla.

- Por ejemplo, si la media de nuestra muestra es $\bar{X} = 5,2$, el error estándar de la media muestral es:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2,5}{\sqrt{25}} = \frac{2,5}{5} = 0{,}5$$

Pasos para una prueba de hipótesis

4. Tomar una decisión

- Así, una media muestral $\bar{X} = 5,2$ produce un z-score de:

$$Z = \frac{5{,}2 - 4}{0,5} = \frac{1,2}{0,5} = 2,4$$

- Para un $\alpha = 0{,}05$, este z-score está por encima del valor crítico de $1,96$.
- Dado que el z-score se encuentra en la región crítica, la hipótesis nula es rechazada y se concluye que el conflicto en el país de origen tiene un efecto sobre el comportamiento de los jugadores.

Pasos para una prueba de hipótesis

4. Tomar una decisión

2. La segunda posibilidad es que la media muestral no se ubique en la región crítica. En este caso, la media se encuentra cerca a la media poblacional especificada por la hipótesis nula.

- Si la media muestral es $\bar{X} = 4,9$, se obtendría un z-score:

$$Z = \frac{4,9 - 4}{0,5} = \frac{0,9}{0,5} = 1,8$$

- Este z-score no se encuentra en la región crítica. Los datos no proveen suficiente evidencia de que H_0 es falsa, por lo tanto, **no se puede rechazar** la hipótesis nula y la conclusión es que el conflicto en el país de origen no parece tener efecto alguno.

Incertidumbre y errores en las pruebas de hipótesis

- Las pruebas de hipótesis son un procedimiento inferencial, lo cual significa que se utiliza información limitada.
- Específicamente, las muestras proveen al investigador de información incompleta de la población.
- Considerando que las pruebas de hipótesis utilizan muestras para concluir algo sobre una población, existe la posibilidad de que la conclusión final sea equivocada.
- En las pruebas de hipótesis se pueden cometer dos tipos de errores:
  1. Rechazar una hipótesis que es verdadera (Tipo I)
  2. No rechazar una hipótesis que es falsa (Tipo II)

Error Tipo I

- Es posible que los datos lleven a rechazar la hipótesis nula cuando en realidad no existe diferencia por la condición o el tratamiento no tuvo efecto.
- Si el investigador, de manera aleatoria, selecciona una de las muestras extremas de la distribución , la prueba de hipótesis concluirá que la diferencia es considerable aun cuando NO exista efecto alguno.
- En el ejemplo del comportamiento de los futbolistas, supongamos que se selecciona una muestra de $n = 35$ jugadores, los cuáles por otras razones fueran agresivos y tuviera más tarjetas amarillas que el promedio.
- Bajo este escenario, aún bajo la ausencia de conflicto en el país de origen, estos jugadores estarían por encima del promedio de amarillas.

- Esta selección de una muestra extrema, llevaría a concluir que el conflicto tiene relación con el comportamiento agresivo de los futbolistas cuando en realidad no es así.
- Esto puede conllevar a consecuencias serias. Por ejemplo, considerar que un medicamento fue efectivo cuando en realidad no lo es.
- El error Tipo I ocurre cuando el investigador, sin saberlo, selecciona una muestra extrema. Afortunadamente, las pruebas de hipótesis están diseñadas para minimizar el riesgo que esto ocurra.
- Al pensar en la distribución muestral de medias y con un $\alpha = 0,05$, se sabe que sólo 5 % de las muestras tienen medias en la región crítica.
- Por lo tanto, α **es la probabilidad de que ocurra el error Tipo I**.

Error Tipo II

- El error Tipo II ocurre cuando se falla en rechazar la hipótesis nula que es falsa.
- Esto puede suceder cuando la media muestral no se ubica en la región crítica a pesar de que la condición/tratamiento si tienen efecto.
- Se presenta cuando, por ejemplo, el tratamiento tuvo efecto pero la magnitud de ese efecto es muy pequeña. Por lo tanto, dado que la media muestral no es, sustancialmente, diferente de la población original se concluye que no existen diferencias.

- Las consecuencias de cometer el error Tipo II no son tan graves como las del Tipo I. En general, el error tipo II significa que los datos de investigación no presentan los resultados esperados.
- A diferencia del error Tipo I, no existe una probabilidad exacta de cometer el error Tipo II.
- En cambio, esta probabilidad es una función que depende de otras variables. Sin embargo, se representa por medio de la letra β.
- En resumen:

		Situación Actual	
		H_0 (V)	H_0 (F)
Decisión	Rechazar H_0	Tipo I	Decisión Correcta
	No Rechazar H_0	Decisión Correcta	Tipo II

Seleccionar α

- Como hemos visto α cumple dos propósitos:
  1. Define los límites de la región crítica al definir los resultados poco probables.
  2. Determina la probabilidad que suceda el error Tipo I.
- Al seleccionar un valor para α al inicio de la prueba, se influencian ambas funciones.
- Por este motivo, existe un dilema entre minimizar el riesgo del error Tipo I y rechazar la hipótesis nula. Se podría pensar en un α muy pequeño para evitar el error pero ¿qué sucede con la región crítica?
- Para un valor pequeño, la prueba requiere mucha evidencia para rechazar la hipótesis nula. Es necesario mantener un balance entre ambas funciones.

El valor p

- Como sabemos, un z-score tiene una probabilidad asociada dada la forma de la distribución normal.
- Por ende, la decisión de una prueba de hipótesis puede basarse tanto en el z-score como en su probabilidad asociada (p).
- En el ejemplo anterior, se encontró que el $Z = 1,8$, resultado que no es inusual, es decir, tiene una alta probabilidad relativa de ocurrir ($p > 0{,}05$).
- Recuerden que se rechaza la hipótesis nula con valores con baja probabilidad de ocurrencia (extremos), ubicados en la región crítica en las colas de la distribución.
- Considerando esto último, un resultado rechazará la hipótesis nula si $p \leq 0{,}05$.

- Por ejemplo, para un valor $z = 2$, sabemos que $p = 0{,}023$. Si $\alpha = 0{,}05$, se rechaza la hipótesis nula dado que $0{,}023 < 0{,}05$.
- Esto sucede para un nivel de significancia del 5 %, ¿qué sucede si este cambia a 1 %?
- El valor p es la probabilidad que el resultado ocurra si H_0 es verdadera, siendo también la probabilidad que se presente el error Tipo I.

$$\text{Si } Z_{\bar{X}} \geq Z_\alpha \to p \leq \alpha : \text{Se rechaza } H_0$$
$$\text{Si } Z_{\bar{X}} < Z_\alpha \to p > \alpha : \text{No se rechaza } H_0$$

- Un resultado es **estadísticamente significativo** si es poco probable que ocurra, es decir, es suficiente para rechazar la hipótesis nula.

Pruebas de hipótesis a una cola

- Usualmente, el investigador empieza con una predicción sobre la dirección en que la condición/tratamiento afecta los individuos.
- Por ejemplo, un programa especial de entrenamiento mejora el desempeño de los estudiantes o el consumo de alcohol aumenta el tiempo de reacción.
- En estas situaciones, es posible establecer la hipótesis estadística de una manera que incorpore la predicción en H_0 y H_1. El resultado es una prueba *direccional* o lo que se conoce comúnmente como prueba a una cola.

- Retomemos el ejemplo sobre violencia y fútbol. Las tarjetas amarilla se distribuye normal con $\mu = 4$ y $\sigma = 2,5$. Para esta stiuación en particular, se cree que los jugadores provenientes de países en conflicto son más violentos en el campo de juego.
- El investigador obtiene una media muestral $\bar{X} = 5$ para los $n = 25$ jugadores que seleccionó.¿Es este resultado suficiente para concluir que el conflicto tiene algún efecto?
- Al igual que para la prueba a dos colas, el primer paso es definir las dos hipótesis. En este caso:

$$H_0 : \mu_{conflicto} \leq 4$$

$$H_1 : \mu_{conflicto} > 4$$

- La distribución muestral de medias tendrá media $\mu = 4$ y error estándar $\sigma_{\bar{X}} = \frac{2{,}5}{\sqrt{25}} = \frac{2,5}{5} = 0,5$.
- La región crítica estará ubicada en su totalidad a la derecha de la distribución dado que se espera un efecto positiva del conflicto.
- El nivel de significancia α no se divide entre ambas colas, por lo tanto, es necesario buscar el 5 % a la derecha. Al buscar en la tabla de la normal, se encuentra que el $Z_\alpha = 1,65$.
- Para esta prueba se requieren dos cambios entonces:
  1. Incluir la predicción al definir las hipótesis.
  2. La región crítica se ubica totalmente en una de las colas de la distribución.

- El resto del procedimiento es exactamente igual al anterior. Se compara el z-score crítico con el z de la distribución.

$$Z = \frac{5-4}{0,5} = 2 \quad p = 0{,}023$$

- Como $Z > Z_\alpha$ o $p \leq 0{,}05$, se rechaza la H_0 y se concluye que provenir de un país en conflicto aumenta significativamente el comportamiento violento de los futbolistas.
- La principal diferencia entre ambos tipos de prueba es el criterio que usan para rechazar la H_0. La prueba a una cola permite rechazar la hipótesis nula cuando la diferencia entre la población y la muestra es pequeña relativamente.
- Por otro lado, la prueba a dos colas requiere una diferencia grande independientemente de la dirección.

Hace cinco años la tasa de peticiones promedio d e todos los 1567 neurocirujanos en la OMS fue 19.7 peticiones. Se obtienen los registros actuales de una muestra de 130 neurocirujanos y la tasa de peticiones es de $\bar{X}$ =20.9 con una desviación estándar de s_X =5.7 peticiones. ¿Ha aumentado la tasa de peticiones promedio de los neurocirujanos de la OMS desde hace cinco años a un nivel de confianza del 95 %?

Hace cinco años la tasa de peticiones promedio d e todos los 1567 neurocirujanos en la OMS fue 19.7 peticiones. Se obtienen los registros actuales de una muestra de 130 neurocirujanos y la tasa de peticiones es de $\bar{X}$ =20.9 con una desviación estándar de s_X =5.7 peticiones. ¿Ha aumentado la tasa de peticiones promedio de los neurocirujanos de la OMS desde hace cinco años a un nivel de confianza del 95 %?

$$H_0 \ : \mu_{hoy} \leq 19{,}7$$

$$H_1 \ : \mu_{hoy} > 19{,}7$$

$$s_{\bar{X}} = \frac{5{,}7}{\sqrt{130}} = 0{,}5 \quad Z_{\bar{X}} = \frac{20{,}9 - 19{,}7}{0{,}5} = 2{,}4 \quad p = 0{,}008$$

$$Z_{\bar{X}} = 2{,}4 > Z_{\alpha} = 1{,}64 \quad p < \alpha$$

Se rechaza H_0