Hypothesis Testing

Felipe José Bravo Márquez

December 10, 2020

Test de Hipótesis

- Cuando queremos probar si alguna **propiedad** asumida sobre una población se contrasta con una muestra estadística usamos un **Test de Hipótesis**
- El test se compone de las siguientes hipótesis:
  - **Hipótesis Nula** H_0: Simboliza la situación actual. Lo que se ha considerado real hasta el presente.
  - **Hipótesis Alternativa** H_a: es el modelo alternativo que queremos considerar.
- La idea es encontrar suficiente **evidencia estadística** para rechazar H_0 y poder concluir H_a
- Si no tenemos suficiente evidencia estadística **fallamos en rechazar** H_0

Test de Hipótesis (2)

Metodología para Realizar un Test de Hipótesis

- Elegir una hipótesis nula H_0 y alternativa H_a
- Fijar un nivel de significancia α del test
- Calcular un estadístico T a partir de los datos
- El estadístico T es generalmente un valor estandarizado que podemos chequear en una tabla de distribución
- Definir un criterio de rechazo para la hipótesis nula. Generalmente es un valor crítico c.

Test de Hipótesis (3)

- Ejemplo: Se sabe que la cantidad de horas promedio de uso de Internet mensual en Chile país es de 30 horas
- Supongamos que queremos demostrar que el promedio es distinto a ese valor.
- Tendríamos que $H_0 : \mu = 30$ y $H_a : \mu \neq 30$
- Fijamos $\alpha = 0.05$ y recolectamos 100 observaciones
- Supongamos que obtenemos $\overline{X_n} = 28$ y $s = 10$
- Una forma de hacer el test es construir un intervalo de confianza para μ y ver si H_0 está en el intervalo.

```
> 28-qt(p=0.975,99)*10/sqrt(100)
[1] 26.01578
> 28+qt(p=0.975,99)*10/sqrt(100)
[1] 29.98422
```

- El intervalo sería la zona de aceptación de H_0 y todo lo que esté fuera de éste será mi región de rechazo.
- Como 30 está en la región de rechazo, rechazo mi hipótesis nula con un 5% de confianza.

Test de Hipótesis (4)

- Otra forma de realizar el test es calcular el estadístico $T = \frac{\overline{X_n} - \mu_o}{\frac{s}{\sqrt{n}}}$
- En este caso sería

$$T = \frac{28 - 30}{\frac{10}{\sqrt{100}}} = -2$$

- Como $H_a : \mu \neq 30$, tenemos un test de dos lados, donde la región de aceptación es

$$t_{n-1,1-\alpha/2} < T < t_{n-1,\alpha/2}$$

```
 > qt(0.025,99)
[1] -1.984217
> qt(0.975,99)
[1] 1.984217
```

- Como T está en la región de rechazo, rechazamos la hipótesis nula.

Test de Hipótesis (5)

- Generalmente, además de saber si rechazamos o fallamos en rechazar una hipótesis nula queremos saber la evidencia que tenemos en contra de ella.
- Se define un **p-valor** como la probabilidad de obtener un resultado al menos tan extremo como el observado en los datos dado que la hipótesis nula es verdadera.
- "Extremo" significa lejos de la hipótesis nula.
- Si el **p-valor** es menor que el nivel de significancia α, rechazamos H_0
- Ejemplo:

```
> data(iris)
> mu<-3 # La hipótesis nula
> alpha<-0.05
> n<-length(iris$Petal.Length)
> xbar<-mean(iris$Petal.Length)
> s<-sd(iris$Petal.Length)
> se<-s/sqrt(n)
> t<-(xbar-mu)/(s/sqrt(n))
> pvalue<-2*pt(-abs(t),df=n-1)
> pvalue
[1] 4.94568e-07 # es menor que 0.05 entonces rechazamos H0
```

Test de Hipótesis (6)

- La forma elegante de hacerlo en R:

```
> t.test(x=iris$Petal.Length,mu=3)

One Sample t-test

data:  iris$Petal.Length
t = 5.2589, df = 149, p-value = 4.946e-07
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 3.473185 4.042815
sample estimates:
mean of x
    3.758
```

Test de Hipótesis (7)

- Tenemos dos tipos de errores cuando realizamos un test de hipótesis
- Error tipo I: es cuando rechazamos la hipótesis nula cuando ésta es cierta.
- Este error es equivalente al nivel de significancia α
- Error tipo II: es cuando la hipótesis nula es falsa pero no tenemos evidencia estadística para rechazarla.
- Para mitigar los errores tipo I generalmente usamos valores de α más pequeños.
- Para mitigar los errores tipo II generalmente trabajamos con muestras más grandes.
- Existe un trade-off entre los errores tipo I y tipo II.

	Retener H_0	Rechazar H_0
H_0 es verdadera	✓	error tipo I
H_1 es verdadera	error tipo II	✓

Statistical Power

Critics to Hypothesis Testing

FOUR CARDINAL RULES OF STATISTICS by Daniela Witten

- ONE: CORRELATION DOES NOT IMPLY CAUSATION. Yes, I know you know this, but it's so easy to forget! Yeah, YOU OVER THERE, you with the p-value of 0.0000001 — yes, YOU!! That's not causation.
- No matter how small the p-value for a regression of IQ onto shoe size is, that doesn't mean that big feet cause smarts!! It just means that grown-ups tend to have bigger feet and higher IQs than kids.
- So, unless you can design your study to uncover causation (very hard to do in most practical settings — the field of causal inference is devoted to understanding the settings in which it is possible), the best you can do is to discover correlations. Sad but true.
- TWO: A P-VALUE IS JUST A TEST OF SAMPLE SIZE. Read that again — I mean what I said! If your null hypothesis doesn't hold (and null hypotheses never hold IRL) then the larger your sample size, the smaller your p-value will tend to be.
- If you're testing whether mean=0 and actually the truth is that mean=0.000000001, and if you have a large enough sample size, then YOU WILL GET A TINY P-VALUE.
- Why does this matter? In many contemporary settings (think: the internet), sample sizes are so huge that we can get TINY p-values even when the deviation from the null hypothesis is negligible. In other words, we can have STATISTICAL significance w/o PRACTICAL significance.

FOUR CARDINAL RULES OF STATISTICS by Daniela Witten

- Often, people focus on that tiny p-value, and the fact that the effect is of **literally no practical relevance** is totally lost.
- This also means that with a large enough sample size we can reject basically ANY null hypothesis (since the null hypothesis never exactly holds IRL, but it might be “close enough” that the violation of the null hypothesis is not important).
- Want to write a paper saying Lucky Charms consumption is correlated w/blood type? W/a large enough sample size, you can get a small p-value. (Provided there’s some super convoluted mechanism with some teeny effect size. . . which there probably is, b/c IRL null never holds)
- THREE: SEEK AND YOU SHALL FIND. If you look at your data for long enough, you will find something interesting, even if only by chance! In principle, we know that we need to perform a correction for multiple testing if we conduct a bunch of tests.
- But in practice, what if we decide what test(s) to conduct AFTER we look at data? Our p-value will be misleadingly small because we peeked at the data. Pre-specifying our analysis plan in advance keeps us honest. . . but in reality, it’s hard to do!!!
- Everyone is asking me about the mysterious and much-anticipated fourth rule of statistics. The answer is simple: we haven’t figured it out yet.... that’s the reason we need to do research in statistics

References I