Ayudantía 10 - Mat044

22 de julio de 2021

Test de Hipótesis

Es una prueba que se realiza con el fin de comparar estadísticamente una hipótesis por sobre otra.
Por ejemplo, tenemos una muestra de datos las cuales provienen de 2 fuentes distintas, entonces nos gustaría saber si podemos determinar estadísticamente bajo un cierto nivel de confianza si efectivamente son distintas.

Considerando una muestra donde tenemos las observaciones de 2 fuentes distintas: una población $x \sim N(\mu_1, \sigma^2)$ y la otra $y \sim N(\mu_2, \sigma^2)$, entonces consideramos una muestra de tamaño n de cada población (no tienen que ser del mismo tamaño), $x_i \sim N(\mu_1, \sigma^2)$, $y_i \sim N(\mu_2, \sigma^2)$ para $i = 1, ...n$Entonces,

realizamos la pregunta si sus medias son distintas o no.

Considerando la hipótesis nula $H_0 : \mu_1 = \mu_2$ que nos dice que las dos poblaciones tienen igual media, mientras que nuestra hipótesis alternativa sería decir que no, son medias distintas $H_1 : \mu_1 \neq \mu_2$.

Luego, se requiere una regla de decisión que nos permita establecer si la hipótesis es verdadera o no.
Para esto, utilizamos un estadístico (una función de la muestra)

$$T(x_1, x_2, ...x_n, y_1, ..., y_n) = T \in \mathbb{R}$$

Luego, se rechazará la hipótesis nula si se tiene que $T(X)$ es muy grande, para esto comparamos contra un valor crítico.

$$se\ rechaza\ H_0\ si : T(X) \geq C$$

Por ejemplo, supongamos que $T \sim N(0, 1)$, entonces se tiene que

$$Z_{0,975} = 1{,}96$$
$$Z_{0,0,25} = -1{,}96$$

por lo tanto,

Por ejemplo, supongamos que $T \sim N(0,1)$, considerando que

$$Z_{0{,}975} = 1{,}96$$
$$Z_{0{,}025} = -1{,}96$$

por lo tanto, $P(|T| > 1{,}96) = 0{,}05$ entonces diremos a un nivel de confianza $\alpha = 5\,\%$ que rechazamos la hipótesis nula de que $\mu_1 = \mu_2$ si se tiene que $T > 1{,}96$.

Considerando un ejemplo ficticio para verificar la efectividad de una vacuna en la población.
A un total de 2000 individuos se le suministran 2 tipos de vacuna, a 1000 se le da un placebo y a los otros 1000 restantes se les inocula la vacuna.

Resultado	Placebo	Vacuna
N° contagiados	389	222
N° no contagiados	611	778

Entonces, se desea realizar un test de hipótesis para probar estadísticamente si la vacuna tiene algún efecto o no.
Para ello, considerar que cada observación sigue una distribución bernoulli:

$$X \sim Bernoulli(\theta_1)$$

donde $X = 1$ significa que el individuo X se contagió y a su vez $X = 0$ que no se contagió.

Por lo tanto tenemos que:

$$X_i \sim Ber(\theta_1) \quad \forall i = 1, \ldots, 1000$$
$$X_i \sim Ber(\theta_2) \quad \forall i = 1001, \ldots, 2000$$

Considerar la hipótesis nula:

$$H0 : \quad \theta_1 = \theta_2$$

versus la hipótesis alternativa:

$$H1 : \quad \theta_1 \neq \theta_2$$

Recordemos que para una distribución bernoulli(p):
$E(X) = p$ y $Var(X) = p(1-p)$.
Esto lo podemos reescribir como $\theta_1 - \theta_2 = 0$, es decir, la vacuna no tiene ningún efecto y ambas muestras son iid.

Por simplicidad realizaremos el test para las muestras por separado.
Para el grupo de los no vacunados queremos probar

$$H0 : \theta_1 = 0{,}4$$

Si la varianza es desconocida y que la muestra es suficientemente grande, tenemos el estadístico T:

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

$$T = \frac{0{,}389 - 0{,}4}{0{,}4878/\sqrt{1000}} = |-0{,}713| \sim t_{999}$$

como la distribución es simétrica, podemos ver una cola, entonces para una confianza del 95 % se tiene el valor crítico y la regla de rechazo:

$$|T| > t_{0{,}975}(999) = 1,962$$

$P - Value = 0{,}52$ Entonces, no rechazamos la hipótesis nula.

Para el caso de la segunda muestra, considerando $H0 : \theta_2 = 0{,}4$ tenemos

$$T = \frac{0{,}222 - 0{,}4}{0{,}4158/\sqrt{1000}} = |-13{,}53| \sim t_{0{,}975}(999) = 1,962$$

por lo tanto rechazamos la hipótesis nula, y la media es distinta a 0.4. Luego, si medimos la efectividad como:

$$eff = 1 - \frac{N\ contagiados\ con\ vacuna}{N\ contagiados\ sin\ vacuna}$$

Entonces, tendría una efectividad del 43 %.