Clase 1: Pruebas estadísticas y correlaciones

Denise Laroze

31 de agosto de 2018

CESS - Universidad de Santiago
denise.laroze@usach.cl

Resumen de contenidos

Pruebas de t ¿Cómo sabemos si dos poblaciones son distintas?

Co-varianza

Correlaciones

Pruebas de t ¿Cómo sabemos si dos poblaciones son distintas?

Pruebas de T

- La prueba de t (t-test o Student's T Test) compara dos medias y entrega un valor que permite cuantificar qué tan distintas son o si las diferencias podrían haber ocurrido por azar.
- En otras palabras prueba si es posible rechazar la **hipótesis nula** ($H_0 : \Delta = 0$) en favor de la hipótesis alternativa ($H_1 : \Delta \neq 0$).
- Las pruebas de t producen un valor de t que es igual a la razón entre las diferencias entre dos grupos (o medias) y las diferencias dentro de esos grupos.
- Mientras mayor es el valor de t, más grandes son las diferencias entre los grupos. Un t=3 significa que la diferencia entre grupos es tres veces mayor a las diferencias en los grupos.
- ¿Qué valor de t es suficientemente grande? Cada calor de t está asociada a un valor de p (de significancia estadística), que a su vez depende de los grados de liberad (el n-1).

Distintas pruebas de t

- Two sample t.test: compara dos muestras independientes.
- Paired sample t-test: Un t-test pariado (paired) que compara las medias para el mismo grupo o unidad en distintos momentos o lugares.
- One sample t-test: compara la media de una muestra con otra media conocida (ej. de la población o teórica)

Ejemplos en pizarra

- Poblaciones con poca varianza
- Poblaciones con harta varianza

Ecuación t-test

$$t = \frac{\widehat{\Delta}}{\widehat{se}(\widehat{\Delta})}$$

donde $\widehat{\Delta}$ es una diferencia entre estimadores de interés (ej. diferencia de medias) y $\widehat{se}(\widehat{\Delta})$ es el error estándar.

Para muestras grandes el valor de t para significancia estadística al 95 % de confianza es de 1.96.

Error Estándar

$$\widehat{se}(\widehat{\Delta}) = \frac{\widehat{\sigma}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

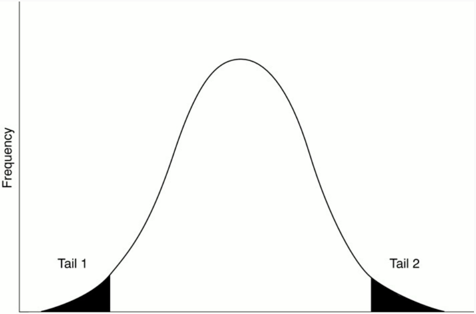
Varianza de la población es

$$\widehat{\sigma^2} = \frac{(n_2 - 1)s_2^2 + (n_1 - 1)s_1^2}{n_1 + n_2 - 2}$$

y la varianza es

$$s_1^2 = \frac{\sum (X_i - \overline{X_i})^2}{n_1 - 1}$$

Prueba de t



df	*0.1*	*0.05*	*0.01*	*0.001*
1	6.314	12.706	63.657	636.619
2	2.92	4.303	9.925	31.599
3	2.353	3.182	5.841	12.924
4	2.132	2.776	4.604	8.610
5	2.015	2.571	4.032	6.869
6	1.943	2.447	3.707	5.959
7	1.895	2.365	3.499	5.408

Página web con ejemplos

http://ci.columbia.edu/ci/premba_test/c0331/s7/s7_5.html

Co-varianza

Covarianza

La covarianza es una medida de la variabilidad conjunta de dos variables aleatorias.

Si los valores altos de una variable coinciden con los valores altos de una segunda variable, y lo mismo ocurre con los valores bajos (en otras palabras las variables presentan un comportamiento similar) la covarianza es positiva.

En el caso contrario, que los valores altos de una variable coincidan con los valores bajos de otra, la covarianza es negativa.

Ecuación Covarianza

$$\text{cov}(X, Y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - E(X))(y_i - E(Y))$$

donde

$$E(X) = \overline{X}$$

Correlaciones

Correlación

En el sentido más amplio, una correlación es una asociación estadística entre dos variables. Sin embargo, el uso habitual en estadística hace referencia a que tan cerca están dos variables de tener una relación lineal.

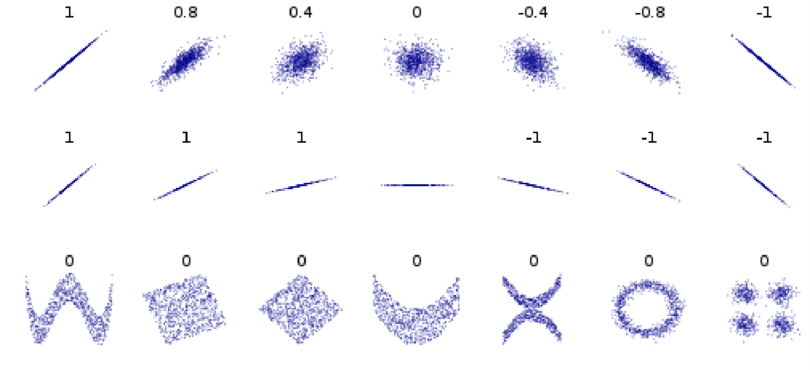
Por las propiedades de su ecuación, la correlación produce valores entre -1 y 1, donde 1 implica una correlación perfecta y positiva; -1 perfecta y negativa; y 0 no existe correlación.

Ecuación Correlación de Pearson

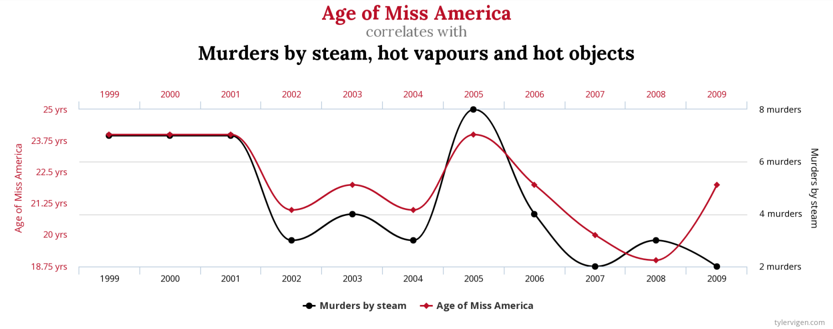
$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y},$$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Correlaciones



Correlaciones espurias I



http://www.tylervigen.com/spurious-correlations

Correlaciones espurias III