Estadística Inferencial

Clifford Torres

Facultad de Ciencias (UNI)

September 24, 2019

Intervalos de confianza

Sea el tamaño de la muestra n:

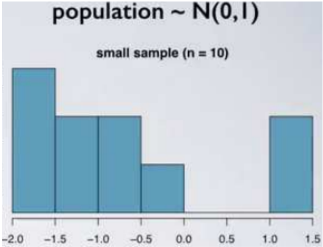
- $n > 30$ muestra de tamaño grande.
- $n \leq 30$ muestra de tamaño pequeño.

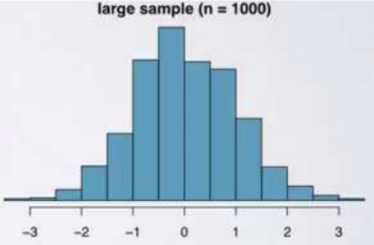
Intervalos de confianza

- No siempre se tiene la posibilidad de contar con una muestra grande.
- Vamos a focalizarnos en **muestras pequeñas** cuando el estadístico es la **media muestral**.
- Recordemos porque necesitábamos una muestra grande: *Siempre y cuando las observaciones sean i.i.d., y la distribución poblacional no demasiado asimétrica, una muestra grande nos aseguraba que*
  - La distribución muestral de la media se aproximaba a la normal a medida que n (tamaño de la muestra) crecía
  - Y el estimador del error estándar: $\sigma_{\hat{\theta}} = \frac{s}{\sqrt{n}}$ es confiable, donde s es el estimador de σ, el desvío poblacional que por lo general es desconocido.

¿Qué ocurre cuando $n < 30$?

- El TLC asevera que la distribución muestral de $\bar{X}$ es aprox. Normal cualquiera sea el tamaño de la muestra siempre y cuando la distribución poblacional sea aprox. normal .
- Sin embargo, no es fácil de verificar en muestras pequeñas la condición de normalidad. Ambas muestras ($n = 10$ y $n = 1000$) provienen de una $N(0, 1)$





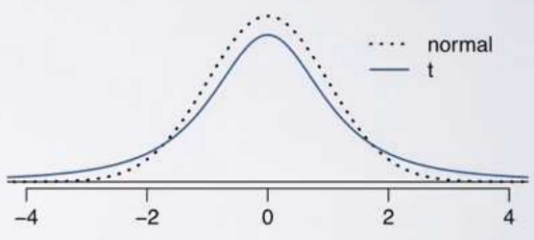
- Es difícil determinar a partir de una muestra pequeña cual es la distribución de la que provienen

¿Qué ocurre cuando $n < 30$?

- Vimos que si $n \geq 30$, y σ es desconocido, estimamos σ con s.
- Sin embargo, cuando $n < 30$ y σ es desconocido (casi siempre), también podemos utilizar s como el estimador natural de σ, pero el hecho de que n sea pequeño torna a s menos confiable.
- Para mitigar esta mayor incertidumbre en s y continuar reteniendo la confianza del 95% en la construcción de los intervalos de confianza, deberíamos entonces aumentar el ancho del intervalo.
- Luego deberíamos trabajar con una distribución que de cuenta de la necesidad de un intervalo más ancho.
- Por lo tanto, la distribución normal estandarizada Z es reemplazada por la t de Student.

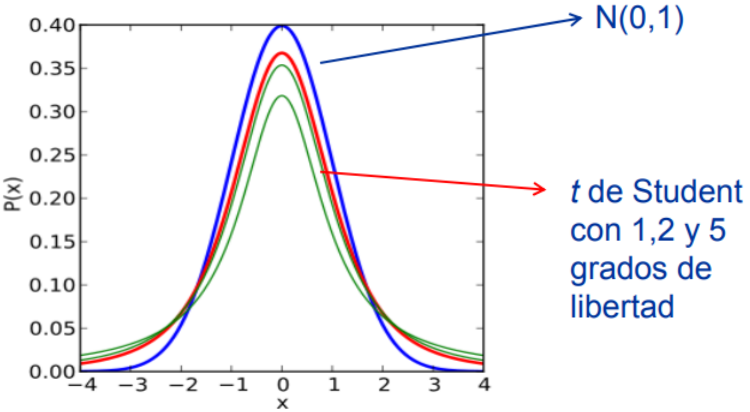
La distribución de t Student

- La t de Student también es simétrica alrededor de la media=0, con forma de campana, pero con colas más pesadas, i.e. es más probable tener más observaciones más allá de 2 desvíos estándar respecto de la media si se la compara con la distr. Normal estándar.
- Las colas más pesadas son las que van a mitigar la mayor incertidumbre originada en el cálculo del $\sigma_{\hat{\theta}}$



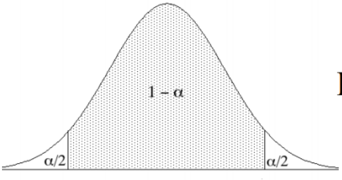
La distribución de t Student

- La t de Student tiene un solo parámetro, llamado **grados de libertad**, que determina cuan pesadas son las colas de la distribución.
- ¿Qué ocurre con la forma de la distribución cuando los grados de libertad se incrementan?



¿Cuándo y cómo se utiliza la t de Student?

- Cuando deseamos construir un intervalo de confianza para la media y el
  - σ desconocido
  - Tamaño de la muestra $n < 30$
- El intervalo de confianza (región de confianza$=1-\alpha$) se calcula de la misma manera pero en lugar de Z utilizamos t



$$\overline{X} \pm \mathrm{t}_{\text{nivel de confianza}} * \frac{S}{\sqrt{n}}$$

$$\mathrm{P}(|t| > \mathrm{t}_{\text{nivel de confianza}=\alpha/2}) = \alpha$$

$$\mathrm{P}(t > \mathrm{t}_{\text{nivel de confianza}=\alpha/2}) = \alpha/2$$

Comparando la $N(0,1)$ y la t

Calcular

- $P(Z \geq z_{0.025}) = 0.025$
- $P(t \geq t_{50,0.025}) = 0.025$
- $P(t \geq t_{10,0.025}) = 0.025$

La de t Student

- https://gallery.shinyapps.io/dist_calc/
- https://stattrek.com/online-calculator/t-distribution.aspx
- http://gauss.inf.um.es/feir/20/20A-distribuciones.html
- http://www.pwpamplona.com/wen/calcu/calcu2.htm#tv

Resumiendo...

Parámetro	Muestra	Distribución Poblacional	σ	Intervalo de confianza
Media	$n \geq 30$	Cualquiera	Conocida	$\overline{X} \pm z_c \dfrac{\sigma}{\sqrt{n}}$
Media	$n \geq 30$	Cualquiera	Desconocida	$\overline{X} \pm z_c \dfrac{S}{\sqrt{n}}$
Media	$n < 30$	Debe ser aproximadamente normal	Conocida	$\overline{X} \pm z_c \dfrac{\sigma}{\sqrt{n}}$
Media	$n < 30$	Debe ser aproximadamente normal	Desconocida	$\overline{X} \pm t_c \dfrac{S}{\sqrt{n}}$ $gdl = n - 1$
Proporción	$np \geq 10$ $n(1-p) \geq 10$			$\hat{p} \pm z_c \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Ejemplo

Cierta empresa está implementado un programa de adiestramiento por computadora para sus empleados. La empresa decide adiestrar a 15 empleados. La tabla muestra los tiempos de adiestramiento.

Tiempo de adiestramiento en dias							
Empleado	Tiempo		Empleado	Tiempo		Empleado	Tiempo
1	52		6	59		11	54
2	44		7	50		12	58
3	55		8	54		13	60
4	44		9	62		14	62
5	45		10	46		15	63

Calcular el intervalo de confianza al 95% para la media poblacional.