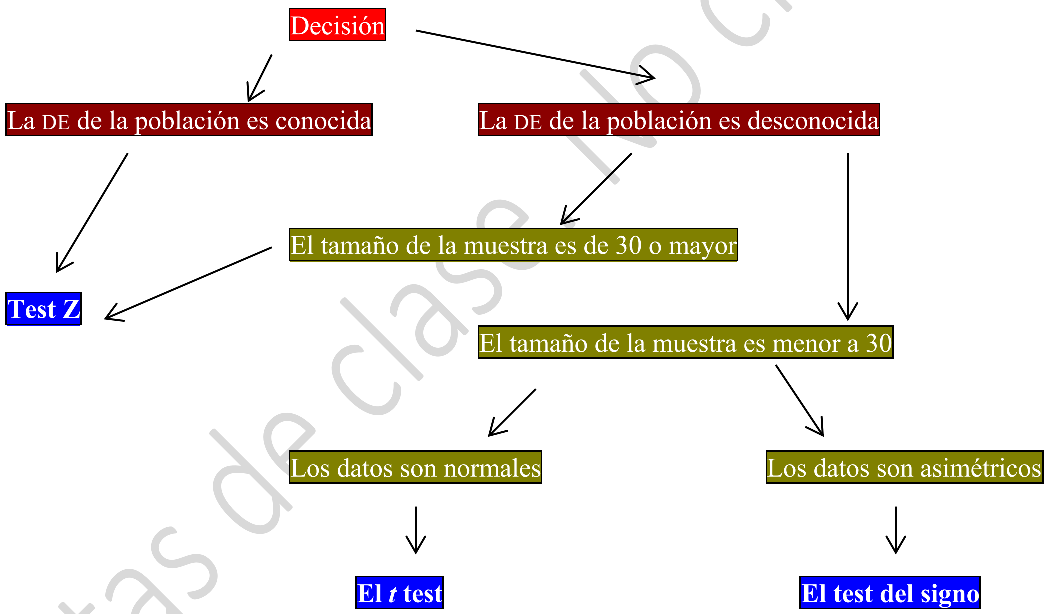
(3) EL *T TEST* DE *STUDENT*

Introducción

Hasta ahora hemos venido comparado una muestra con una población con los *tests* Z. Sin embargo, la desventaja es que sólo lo podemos hacer si conocemos la desviación estándar de la población, o si no, cuando el tamaño de la muestra es de 30 o mayor.

Para los casos en los cuales no conocemos la desviación estándar de la población y el tamaño de la muestra es menor a 30 existen dos métodos para muestras pequeñas y en las cuales se desconoce la DE de la población.

1. El ***T test*** de una muestra, el cual compara la media muestral con la media poblacional, al igual que el test Z. Es muy similar, sin embargo, se usan las tablas T en lugar de las tablas normales. El test T sólo puede usarse cuando los datos están distribuidos de forma normal.
2. El ***test del signo*** de una muestra, el cual compara la mediana de una muestra con la mediana de la población. Este test se puede usar cuando los datos no tienen una distribución normal. En otras palabras, es adecuado para datos asimétricos. También se usa si queremos comparar medianas en lugar de medias.



Ahora bien, hay que recordar que debido a que se trabaja con muestras pequeñas es más difícil rechazar la hipótesis nula con estos *tests* que con un test Z.

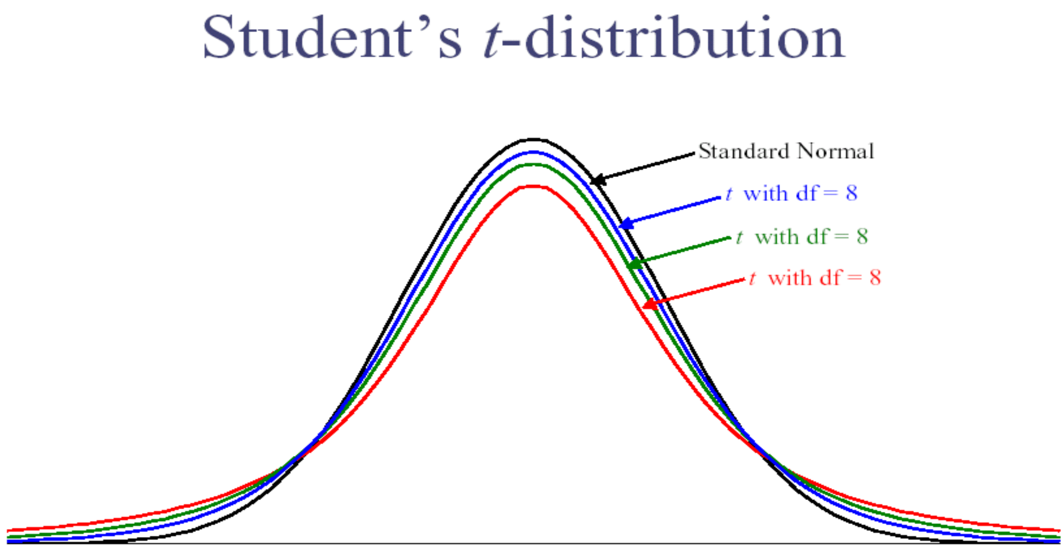
El *t* test[1]

Como mencionamos arriba, el *T test* sólo se usa cuando la DE es desconocida, el tamaño de la muestra es menor a 30 y los datos de la muestra están distribuidos normalmente. Asimismo, en lugar de usar la distribución normal, usamos la **distribución *t***, debido al tamaño pequeño de la muestra.

[1] También conocido como la ***prueba t*** o la ***prueba t de Student***.

La distribución *t*

La distribución *t* es similar a la distribución normal, pero hay una curva para cada valor de *t*. En la medida en que *t* se incrementa, la curva *t* se hace más alta y más delgada. Una vez que el tamaño de la muestra alcanza 30 o más la curva *t* es aproximadamente la misma que la distribución normal. Es por esto que podemos usar la distribución normal como una aproximación cuando $n \geq 30$.

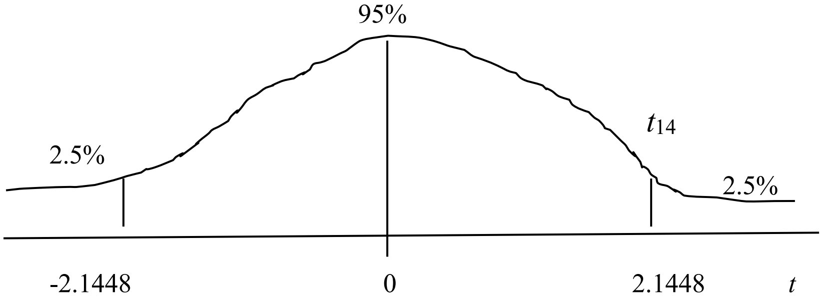


Procedimiento para el *t* test

En primer lugar, hay que escribir las hipótesis para proseguir a calcular la media estadística. La formula es igual que para el test Z:

$$t = \frac{\bar{x} - \mu}{DE / \sqrt{n}}$$

La diferencia más importante se encuentra en localizar los **valores críticos**. Esto lo hacemos usando las tablas *t*. La curva *t* a usar es la curva para *n*-1 grados de libertad. Así que, si el tamaño de la muestra es de 15, entonces t_{n-1} es t_{14}. Para encontrar el valor crítico de *t* para 14 grados de libertad hay que localizar 14 en la primera columna (que es la que nos da los grados de libertad) y buscar horizontalmente para ubicar el valor que estamos buscando. En caso de que se trate de un test de dos colas a un nivel de 5%, hay que buscar hacia abajo en la columna t_{df} (0.025). Esto es porque para este test habrá 95% en medio de la distribución y 2.5% (0.025) en cada lado, como se muestra a continuación.



Gráfica 1. Regiones críticas para un test de dos colas al nivel del 5% con 14 grados de libertad.

Así que para un nivel del 5% del test con 14 grados de libertad, el valor de t (a partir de la tabla) es de 2.1448 y los valores críticos serán de: -2.1448 y 2.1448.

De manera similar, para un test de dos colas en el nivel del 1%, hay que mirar debajo de la tabla en la columna t_{gl} (0.005) ya que habrá una probabilidad del 99% en la mitad de la distribución y 0.5% (0.005) en cada cola. ¿Cuáles columnas habría que ver para un test de una cola en el nivel del 5% y en el nivel del 1%?

EJEMPLO 1

En una investigación nacional se mostró que el consumo medio de alcohol semanal para mujeres en edades comprendidas entre los 16-24 es de 9.5 unidades. Usted se pregunta si sus compañeros estudiantes son diferentes en este aspecto y lleva a cabo una encuesta de los hábitos de bebida en la semana anterior. El número de unidades consumida por una muestra de mujeres estudiantes se muestran a continuación.

Cuadro 1. Unidades de alcohol consumido por 14 estudiantes en la semana previa a la encuesta.

11	0	12	23	17	6	1
17	11	14	0	18	14	5

1. Hipótesis

   H_0: $\mu = 9.5$ (el consumo de alcohol entre estudiantes mujeres es el **mismo** que el de la media nacional para mujeres entre 16-24 años).

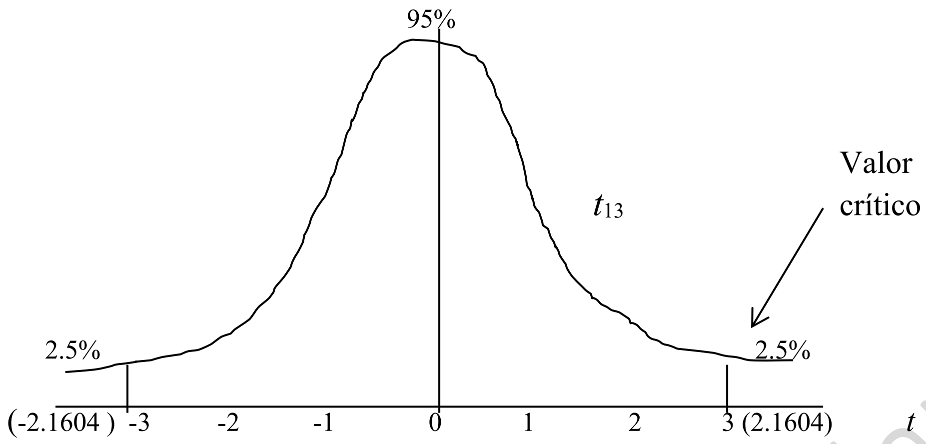
   H_1: $\mu \neq 9.5$ (el consumo de alcohol entre estudiantes mujeres es **diferente** de la media nacional para mujeres de entre 16-24 años).

2. Ubicar el estadístico muestral. De los datos podemos calcular que la media de unidades de alcohol en las semanas previas por la muestra fue de 10.64 con una desviación estándar de 7.26 unidades.

3. Calcular el test estadístico

$$t = \frac{\bar{x} - \mu}{DE / \sqrt{n}} = \frac{10.64 - 9.5}{7.26 / \sqrt{14}} = 0.59$$

4. Comparar el estadístico con los valores críticos. Con una muestra de 14, habrá 13 grados de libertad. De la tabla t, t_{13} (0.025) = 2.1604. Los valores críticos serán de -2.1604 y 2.1604. La hipótesis nula será rechazada si el estadístico es menos que -2.1604 o más grande que 2.1604. El estadístico 0.59 se ubica entre -2.1604 y 2.1604, así que no se puede rechazar la hipótesis nula. De esta muestra no tenemos evidencia de que el consumo de alcohol entre las mujeres estudiantes sea diferente de los niveles nacionales de alcohol para mujeres de entre 16-24.



EJEMPLO 2

Un grupo de estudiantes graduados están solicitando trabajos de dirección en una compañía que aplica un test psicométrico. Sus puntajes se muestran en el siguiente cuadro. Después descubren que el puntaje medio para el test entre todos los que lo presentaron es de 62. Son tan modestos que consideran que van a obtener el trabajo ya que tienen una inteligencia superior. ¿Existe evidencia alguna de que esto estudiantes graduados realmente sean mejores que la media en las pruebas psicométricas?

Cuadro 2. Puntajes de un examen psicométrico para 12 estudiantes graduados que están solicitando un puesto un trabajo a nivel de dirección.

71	63	62	74	69	67	59	65	68	65	66	67

1. Hipótesis
2. Calcular el test estadístico
3. Localizar los valores críticos y sacar una conclusión.

(3) La prueba del signo

Si la distribución de los datos es asimétrica, la mediana es una mejor medida que la media debido a que no es afectada por valores extremos. La prueba del signo compara la mediana de una muestra con la mediana de una población y se usa con muestras pequeñas de datos asimétricos. También se usa cuando se conoce la mediana de la población y no la media de esta.

La media tiene una distribución muestral tal y como la describe el *Teorema del Límite Central* lo que nos ayuda a comparar una media muestral con una media poblacional. Sin embargo, la mediana no tiene una distribución muestral similar a la de la media. Es por esto por lo que la prueba *del Signo* se ubica dentro de las pruebas no paramétricas; esto es “no paramétrico” nos dice que no hay una distribución normal.

Como sabemos, la mediana se ubica en la mitad de la distribución. En una población, sabemos que la mitad de las observaciones se ubican por arriba de la mediana y la otra mitad por debajo. Si no tenemos evidencia de que la muestra no es diferente a la población, entonces esperaríamos que la mitad de las observaciones en la muestra se ubicara por arriba de la mediana de la población y la otra mitad por debajo. Por el contrario, si la muestra es genuinamente diferente a la población, la proporción de observaciones por arriba de la mediana serán marcadamente más grandes o pequeñas que 0.5.

EJEMPLO 1

Los salarios hipotéticos de un grupo de estudiantes recién graduados se muestran en el siguiente cuadro. El salario mediano de los recién graduados (en general) en el año 2010 fue de $15,000. La pregunta es si los salarios de estos estudiantes son significativamente diferentes de la mediana poblacional.

Cuadro 1. Salario inicial de estudiantes recién graduados en el año 2010.

15 500	16 200	19 500	13 900	14 200
15 400	15 500	17 000	14 900	14 800
14 700	15 900	15 500	16 600	18 000

Para empezar, las hipótesis se escriben siempre en términos de la proporción de observaciones por arriba de la mediana. Aquí π_m es la proporción de observaciones por arriba de la mediana en la población y p_m es la proporción en la muestra. Así que podemos decir:

H_0: $\pi_m = 0.5$ (los salarios de los recién graduados vienen de una población con la mitad de las observaciones por arriba de $15 000).

H_1: $\prod_m \neq 0.5$ (los salarios de los recién graduados no vienen de una población con la mitad de las observaciones por arriba de $15 000).

Ahora bien, para encontrar el test estadístico p_m, la proporción de salarios por arriba de la mediana en la muestra debe ser calculada. Para hacerlo, debemos colocar un signo de más o menos por cada observación para indicar si se ubica arriba (+) o abajo (-) de la mediana. Así tenemos

15 500+	16 200+	19 500+	13 900-	14 200-
15 400+	15 500+	17 000+	14 900-	14 800-
14 700-	15 900+	15 500+	16 600+	18 000+

La proporción por arriba de la mediana será el número de signos + dividido por el total de números en la muestra:

$$p_m = \frac{\text{número de signos +}}{\text{número total}} = \frac{10}{15} = 0.67$$

Si tenemos una o más observaciones igual a la mediana, hay que considerar que la mitad se ubica arriba y la otra mitad por debajo.

Posteriormente, podemos calcular el test estadístico Z de la misma forma que el test de hipótesis para una proporción. Si la H_0 es verdadera, la proporción de la población por arriba de la media será de 0.5 y el error estándar de la población –si la H_0 fuese verdadera- sería:

$$EE = \sqrt{\left[\frac{0.5\ (1 - 0.5)}{15}\right]}$$

El test estadístico Z en este caso sería:

$$Z = \frac{p_m - \prod_m}{\sqrt{\left[\frac{\prod_m (1 - \prod_m)}{n}\right]}} = \frac{0.67 - 0.5}{\sqrt{\left[\frac{0.5\ (1 - 0.5)}{15}\right]}} = 1.34$$

Para un test de dos colas al 5% de nivel de confianza, los valores críticos serán de -1.96 y 1.96. El test estadístico se ubica entre -1.96 y 1.96, por lo tanto, no tenemos argumentos para rechazar la H_0. Debemos aceptar que los salarios de los recién egresados no son significativamente diferentes a la mediana de la población.

Aun cuando parecía (con sólo observar los datos) que los salarios de la muestra eran más altos que la mediana, no pudimos rechazar la H_0. La razón es que en las pruebas no paramétricas es menos probable encontrar una diferencia significativa entre la muestra y la población basándonos en la distribución normal (test Z). Recordemos que la prueba del signo no usa toda la información de los datos, solamente toma en cuenta si cada observación se ubica por arriba o por debajo de la mediana y su valor real. Por lo tanto, sólo se puede usar este test cuando no se satisfagan las condiciones para usar el test Z.

EJEMPLO 2

El costo mediano de los libros de texto en las librerías era de $13.99 dólares en 2012. A principios de 2013 se tomó una muestra de 25 libros de texto de una librería en particular y se vio que 18 de ellos costaba más de $13.99. ¿Hay evidencia que el precio de los libros de texto se haya incrementado en 2013?
Para resolver la pregunta, use un test de dos colas con un nivel de significación al 5%.

H_0:

H_1:

p_m =

Z =