Intervalos de confianza

Universidad de Guayaquil

enero 08, 2019

Introducción

- $\hat{p}$ representa la estimación de una probabilidad de éxito p
- $\bar{X}$ representa la estimación de la media poblacional μ.
- S representa la estimación de la desviación estandar poblacional σ.

Son llamadas estimaciones puntuales, porque son sólo números, o puntos. Un hecho importante de las estimaciones puntuales es que casi nunca son exactamente iguales a los valores reales que están estimando.

Estimación por intervalos

Se denomina a la estimación de parámetros poblacionales en la que utilizando una muestra aleatoria $X^T = [X_1, X_2, ...X_n]$ se infiere acerca de θ pero en lugar de un solo punto se define un conjunto de puntos A, con cotas superior e inferior.

$$A = \{\theta \in \mid L_i(X) \leq \theta \leq L_s(X)\}$$

- Longitud del intervalo $\lambda(X)$
    - Finita o infinita
- Intervalos de confianza $[L_i(X), L_s(X)]$ unilaterales, bilaterales

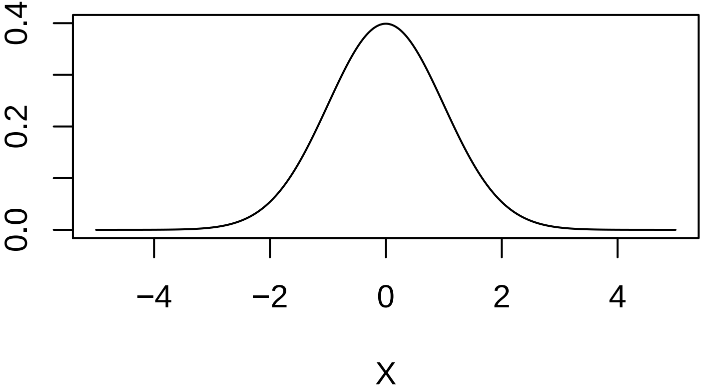
Interpretación de los intervalos de confianza

CI=(0.381 a 0.497)

Correcta: "Tenemos una confianza del 95 % de que el intervalo de 0.381 a 0.497 realmente contiene el valor verdadero de p". Esto significa que si seleccionamos muchas muestras diferentes de tamaño 280 y construimos los intervalos de confianza correspondientes, el 95 % de ellos incluirían realmente el valor de la proporción poblacional p. (Note que en esta interpretación correcta, el nivel del 95 % se refiere a la tasa de éxitos del proceso, utilizada para estimar la proporción, y no a la proporción de la población en sí).

Errónea: "Existe un 95 % de probabilidades de que el valor real de p esté entre 0.381 y 0.497".

Intervalo de confianza para la media poblacional

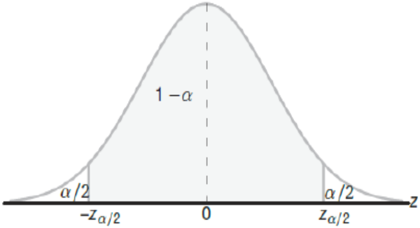


$$N = 100 \quad \bar{X} = 169 \quad \mu = ???$$

Intervalo de confianza para la media - Caso I

- CASO I

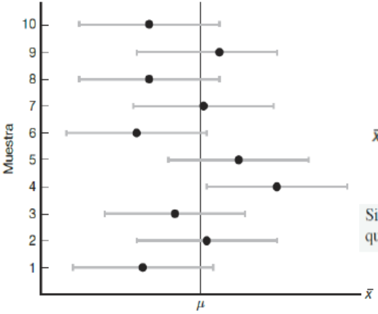
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

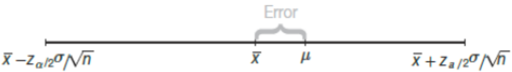


$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha.$$

- CASO I

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$





Si utilizamos $\bar{x}$ como una estimación de μ, podemos tener $100(1-\alpha)\%$ de confianza en que el error no excederá a $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Intervalo de confianza

$\overline{X} \pm \frac{s}{\sqrt{n}}$ es intervalo de confianza de 68% para μ.

$\overline{X} \pm 1.645 \frac{s}{\sqrt{n}}$ es intervalo de confianza de 90% para μ.

$\overline{X} \pm 1.96 \frac{s}{\sqrt{n}}$ es intervalo de confianza de 95% para μ.

$\overline{X} \pm 2.58 \frac{s}{\sqrt{n}}$ es intervalo de confianza de 99% para μ.

$\overline{X} \pm 3 \frac{s}{\sqrt{n}}$ es intervalo de confianza de 99.7% para μ.

Ejercicios I

El tiempo diario que los adultos de una determinada ciudad dedican a actividades deportivas, se aproxima por una variable aleatoria con distribución normal de desviación tipica $\sigma = 20$

1. Calcule un intervalo de confianza al 90 % para μ cuando se toma una muestra aleatoria simple de tamaño 250 y media 90
2. Que tamaño mínimo debe tener una muestra aleatoria simple para que le error máximo cometido en la estimación de μ por la media muestral sea menor que 1 minuto con el mismo intervalo de confianza

Ejercicios II

A partir de una muestra aleatoria simple de 64 frascos de mermelada se ha obtenido un intervalo de confianza para se ha obtenido un intervalo de confianza para μ con un error de estimación de 2.35 gramos

1. Determine el nivel de confianza utilizado para construir ese intervalo de confianza suponiendo un $\sigma = 10$

Ejercicios III

Las notas de un test de aptitud siguen una distribución Normal con desviación 28.2. Una muestra aleatoria de 9 alumnos nos da los siguientes resultados

$$\sum_{i=1}^{n} x_i = 1098 \quad \sum_{i=1}^{n} {x_i}^2 = 138148$$

1. Construya un intervalo de confianza al 90 % para la media poblacional.

```r
# mean and confidence interval
conf.level = 0.90
N=9
z = qnorm((1 - conf.level)/2, lower.tail = FALSE)
xbar = 1098/N
sdx = 28.2/sqrt(N)
c(xbar - (z * sdx), xbar + (z * sdx))
```

```
## [1] 106.5384 137.4616
```

Ejercicios III - continuación

2. Razonar sin hacer cálculos si la longitud de un intervalo al 95 % será menor, mayor o igual que el intervalo de confianza al 90 %.
3. ¿Cuál será el tamaño de la muestra mínimo necesario para obtener un intervalo al 90 % de nivel de confianza, con longitud 10?

```
# mean and confidence interval
conf.level = 0.90
z = qnorm((1 - conf.level)/2, lower.tail = FALSE)
error=5
n=(z*28.2/error)^2
n
```

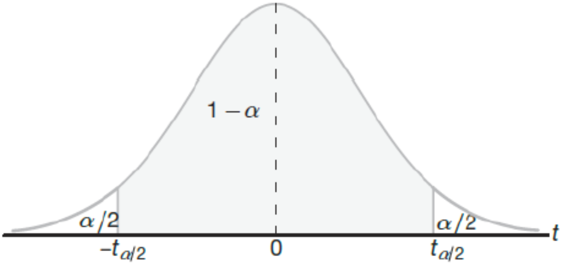
```
## [1] 86.06226
```

Diferentes intervalos de confianza para la media - Caso I

68 % 95 % y 99 %

- CASO II

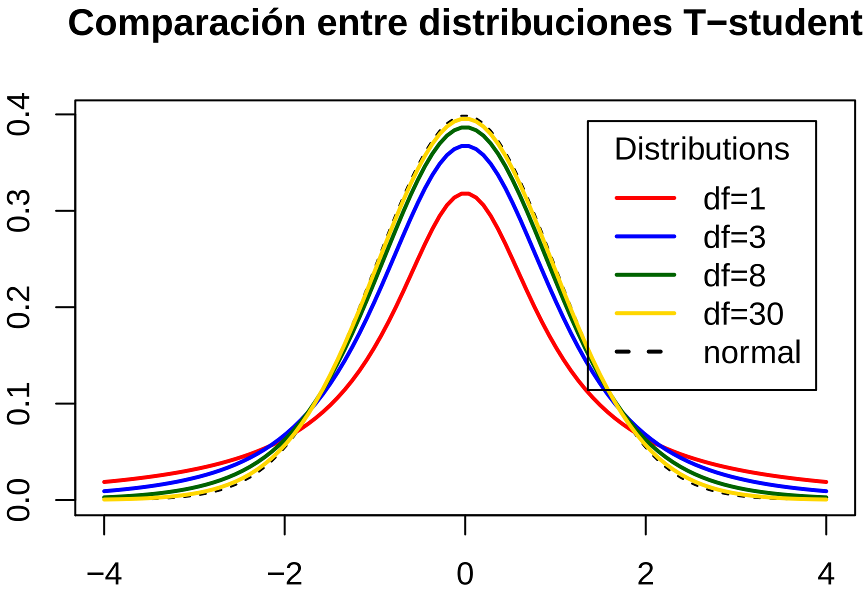
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$



$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

No use la estadística t de Student si la muestra contiene datos atípicos. Para que la estadística t de Student sea válida, la muestra debe provenir de una población que es aproximadamente normal. Tales muestras rara vez contienen datos atípicos. Por tanto, los métodos que implican la estadística t de Student no se deben utilizar en muestras que contienen datos atípicos.

Intervalo de confianza para la media - Caso II



Ejercicios - caso II

En un estudio de los efectos sobre los bebés del consumo de cocaína durante el embarazo, se obtuvieron los siguientes datos de pesos al nacer: $N = 19$, $\bar{X} = 2700gr$ y $s = 645gr$. Construya un estimado del intervalo de confianza del 95 % de μ, el peso medio al nacer de todos los niños

$$\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$2700 - 2{,}1\frac{645}{\sqrt{19}} < \mu < 2700 + 2{,}1\frac{645}{\sqrt{19}}$$

$$2389{,}256 < \mu < 3010{,}744$$

Ejercicios - caso II

```r
Datos=rnorm(19,mean = 2700, sd = 645)
t.test(Datos,alternative ="two.sided", conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  Datos
## t = 17.349, df = 18, p-value = 1.104e-12
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  2208.908 2817.599
## sample estimates:
## mean of x
##  2513.254
```

Ejercicios - caso II

Construya un intervalo que tenga 95 % de confianza para el tiempo medio de recarga de baterias que se utilizan para laptops y para ipads, si se conoce que ese tiempo de recarga es N(μ,σ^2) y al recargar 17 de estas baterías se obtiene una media $\bar{x} = 272$ minutos y una desviación típica de $s = 20$ minutos \ \

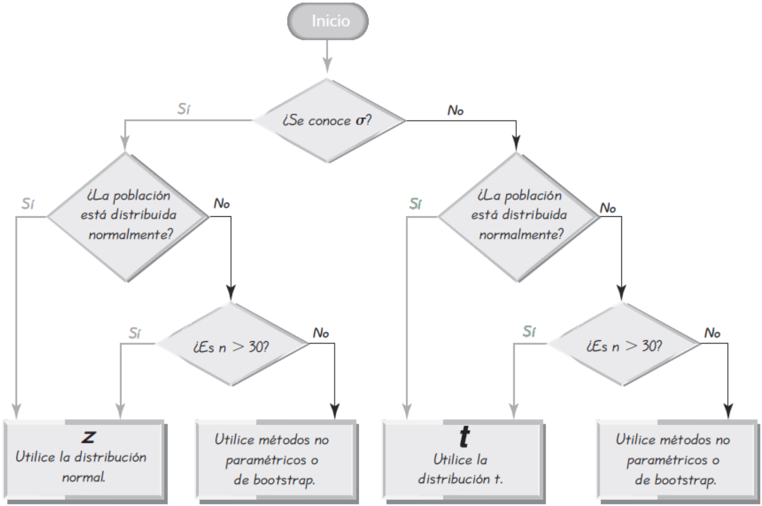
Tenemos $\bar{x} = 272$, suponiendo $X \sim N(\mu, \sigma)$ trabajamos con una distribución T con 16 grados de libertad $t_{0,025} = 2{,}120$

$$\bar{X} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$272 - 2{,}120\frac{20}{\sqrt{17}} < \mu < 272 + 2{,}120\frac{20}{\sqrt{17}}$$

$CI(261{,}717, 282{,}283)$

Intervalo de confianza para la media - RESUMEN



Intervalo de confianza para la media - RESUMEN

Tabla 7-1 Elección entre z y t

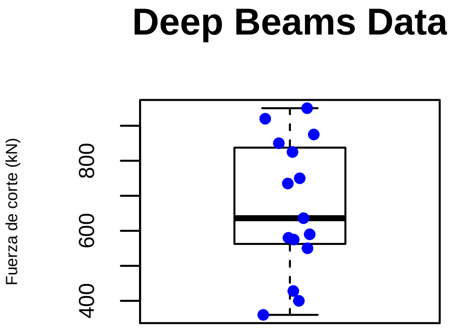
Método	Condiciones
Utilice la distribución normal (z).	σ conocida y población distribuida normalmente *o* σ conocida y $n > 30$
Utilice la distribución t.	σ desconocida y población distribuida normalmente *o* o σ desconocida y $n > 30$
Utilice un método no paramétrico o de *bootstrap*.	La población no está distribuida normalmente y $n \leq 30$

Notas:
1. **Criterios para decidir si la población está distribuida normalmente:** La población no necesita ser exactamente normal, pero debe tener una apariencia un tanto simétrica, con una moda y sin valores extremos.
2. **Tamaño muestral $n > 30$:** Éste es un lineamiento que se usa regularmente, pero tamaños muestrales de 15 a 30 son adecuados si la población parece tener una distribución normal y no existen valores extremos. Para algunas distribuciones poblacionales que estén extremadamente alejadas de la normal, puede requerirse que el tamaño muestral sea mayor de 50 o aun de 100.

Ejercicio ADICIONAL

El artículo "Direct Strut-and-Tie Model for Prestressed Deep Beams" (K. Tan, K. Tong y C. Tang, en Journal of Structural Engineering, 2001:1076-1084) presenta mediciones de la fuerza nominal de corte (en kN) para una muestra de 15 vigas de concreto. Los resultados son:

580 400 428 825 850 875 920 550 575 750 636 360 590 735 950



Ejercicio ADICIONAL

```r
datos=c(580,400,428,825,850,875,920,550,575,750,636,360,590,735,950)
t.test(datos,alternative ="two.sided", conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  datos
## t = 13.474, df = 14, p-value = 2.086e-09
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  520.6237 815.9097
## sample estimates:
## mean of x
##  668.2667
```

Intervalo de confianza para la varianza

$$X^2 = \frac{(n-1)S^2}{\sigma^2}.$$

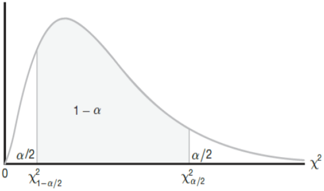


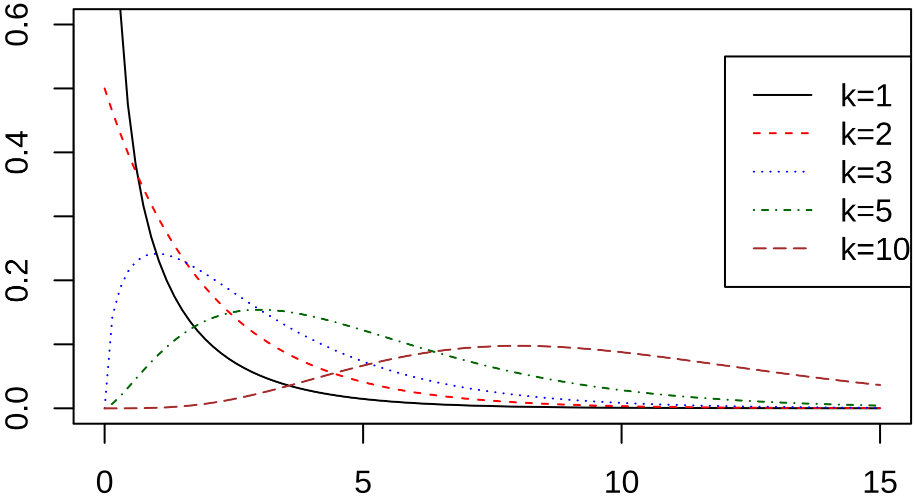
Figura 9.7: $P\left(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}\right) = 1-\alpha.$

Si s^2 es la varianza de una muestra aleatoria de tamaño n de una población normal, un intervalo de confianza del $100(1-\alpha)\%$ para σ^2 es

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}},$$

donde $\chi^2_{\alpha/2}$ y $\chi^2_{1-\alpha/2}$ son valores χ^2 con $v = n-1$ grados de libertad, que dejan áreas de $\alpha/2$ y $1-\alpha/2$, respectivamente, a la derecha.

Intervalo de confianza para la varianza

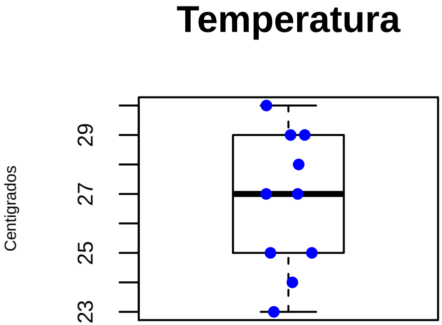


Intervalo de confianza para la varianza - Ejemplos

En la actualidad las monedas de 1 centavo de dólar se acuñan con una desviación estándar de 0.0165 g. Se prueba un nuevo equipo con la intención de mejorar la calidad al reducir la variación. Se obtiene una muestra aleatoria simple de 10 monedas de 1 centavo acuñadas con el equipo nuevo. Una gráfica cuantilar normal y un histograma indican que los pesos provienen de una población distribuida normalmente, y la muestra tiene una desviación estándar de 0.0125 g. Utilice los resultados muestrales para construir un estimado de un intervalo de confianza del 95 % de σ, la desviación estándar de los pesos de monedas de 1 centavo fabricadas con el equipo nuevo. Con base en los resultados, ¿parece que el equipo nuevo sirve para reducir la variación de los pesos?

Intervalo de confianza para media y varianza - EJERCICIO

Para verificar los efectos del calentamiento global del planeta, se efectuán medidas de la temperatura para un día cualquiera del mes de Junio de 2010 en diez lugares estratégicos de Guayaquil. Las lecturas, en grados centígrados, son los siguientes: 28 25 29 27 27 30 25 24 23 29.Construya intervalos con 99 % de confianza para la medida poblacional μ y de 95 % de confianza para la varianza σ^2



Intervalo de confianza para media y varianza - EJERCICIO

```r
datos=c(28,25,29,27,27,30,25,24,23,29)
t.test(datos,alternative ="two.sided", conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  datos
## t = 35.786, df = 9, p-value = 5.14e-11
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  24.27529 29.12471
## sample estimates:
## mean of x
##      26.7
```

```r
c(((10-1)*sd(datos)^2)/qchisq(c(.025),df=9, lower.tail=FALSE),
((10-1)*sd(datos)^2)/qchisq(c(.975),df=9, lower.tail=FALSE))
```

```
## [1]  2.633686 18.552879
```

Intervalos de confianza para proporciones

$$\hat{p} - E < p < \hat{p} + E \qquad \text{donde} \qquad E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

El intervalo de confianza suele expresarse en los siguientes formatos equivalentes:

$$\hat{p} \pm E$$

o

$$(\hat{p} - E, \hat{p} + E)$$

Intervalos de confianza para proporciones - Ejemplo

Tasa de éxito de la terapia de contacto. En 280 ensayos con terapeutas de contacto, la mano correcta fue elegida en 123 ensayos.

1. Calcule el margen de error aa un nivel de confianza del 95 %
2. Calcule el intervalo de confianza al 95 % de la proporción p

Intervalos de confianza para proporciones - Ejemplo

Las formas en las que nos comunicamos se han visto afectadas drásticamente por el uso de máquinas contestadoras telefónicas, máquinas de fax, correo de voz y correo electrónico. Suponga que un sociólogo quiere determinar el porcentaje actual de hogares en Estados Unidos que utilizan el correo electrónico. ¿Cuántos hogares deben encuestarse para tener una confianza del 95 % de que el porcentaje muestral es erróneo por no más de 4 puntos porcentuales?

1. Utilice el siguiente resultado de un estudio pionero: en 1997, el 16.9 % de los hogares estadounidenses usaban correo electrónico (según datos de The World Almanac and Book of Facts).

Intervalos de confianza para proporciones - Ejemplo

El artículo "High-Dose Nicotine Patch Therapy" de Dale, Hurt et al. (Journal of the American Medical Association, vol. 274, núm. 17) incluye esta afirmación: "De los 71 sujetos, el 70 % se abstuvo de fumar durante 8 semanas (intervalo de confianza [IC] del 95 %, del 58 % al 81 %)". Utilice esta afirmación para calcular el estimado puntual $\hat{p}$ y el margen de error E.

```
prop.test(71*0.7, 71, alt="two.sided",conf.level = 0.95)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  71 * 0.7 out of 71, null probability 0.5
## X-squared = 10.574, df = 1, p-value = 0.001147
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.5780822 0.8000717
## sample estimates:
##   p
## 0.7
```

Intervalos de confianza para la diferencia entre dos medias CASO I

Resumen

Sea $X_1, \ldots, X_{n_X}$ una muestra aleatoria *grande* de tamaño n_X de una población con media μ_X y desviación estándar σ_X, y sea $Y_1, \ldots, Y_{n_Y}$ una muestra aleatoria simple *grande* de tamaño n_Y de una población con media μ_Y y desviación estándar σ_Y. Si las dos muestras son independientes, con un intervalo de confianza de nivel $100(1-\alpha)\%$ para $\mu_X - \mu_Y$ es

$$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \qquad (5.16)$$

Cuando los valores de σ_X y σ_Y son desconocidos, se pueden sustituir con las desviaciones estándares muestrales s_X y s_Y.

Intervalos de confianza para la diferencia entre dos medias CASO II

Resumen

Sea $X_1, \ldots, X_{nX}$ una muestra aleatoria de tamaño n_X de una población *normal* con media μ_X, y $Y_1, \ldots, Y_{nY}$ una muestra aleatoria de tamaño n_Y de una población *normal* con media μ_Y. Suponga que las dos muestras son independientes.

Si las poblaciones no necesariamente tienen la misma varianza, un intervalo de confianza de nivel $100(1-\alpha)\%$ para $\mu_X - \mu_Y$ es

$$\overline{X} - \overline{Y} \pm t_{\nu,\alpha/2}\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}} \qquad (5.21)$$

El número de grados de libertad, ν, está dado por

$$\nu = \frac{\left(\dfrac{s_X^2}{n_X} + \dfrac{s_Y^2}{n_Y}\right)^2}{\dfrac{(s_X^2/n_X)^2}{n_X - 1} + \dfrac{(s_Y^2/n_Y)^2}{n_Y - 1}}$$ redondeado hacia abajo al entero más cercano.

Intervalos de confianza para la diferencia entre dos medias CASO III

Resumen

Sea $X_1, \ldots, X_{nX}$ una muestra aleatoria de tamaño n_X de una población *normal* con media μ_X y $Y_1, \ldots, Y_{nY}$ una muestra aleatoria de tamaño n_Y de una población *normal* con media μ_Y. Se supone que las dos muestras son independientes.

Si se sabe que las poblaciones tienen casi la misma varianza, un intervalo de confianza de nivel $100(1-\alpha)\%$ para $\mu_X - \mu_Y$ es

$$\overline{X} - \overline{Y} \pm t_{n_X+n_Y-2,\alpha/2} \cdot s_p \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \tag{5.22}$$

La cantidad s_p^2 es la varianza combinada, dada por

$$s_p^2 = \frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2} \tag{5.23}$$

Intervalos de confianza para observaciones pareadas

Resumen

Sea $D_1, \ldots, D_n$ una muestra aleatoria pequeña ($n \leq 30$) de diferencias de pares. Si la población de diferencias es aproximadamente normal, entonces un intervalo de confianza de nivel $100(1-\alpha)\%$ para la media de la diferencia μ_D está dado por

$$\overline{D} \pm t_{n-1,\alpha/2}\frac{s_D}{\sqrt{n}} \tag{5.24}$$

Observe que este intervalo es el mismo que el que se obtiene con la expresión (5.9).

Si el tamaño muestral es grande, un intervalo de confianza de nivel $100(1-\alpha)\%$ para la media de la diferencia μ_D está dado por

$$\overline{D} \pm z_{\alpha/2}\sigma_{\overline{D}} \tag{5.25}$$

En la práctica $\sigma_{\overline{D}}$ se aproxima con $s_D/\sqrt{n}$. Observe que este intervalo es el mismo que el que se obtiene con la expresión (5.1).