1. Intervalos de confianza para muestras pequeñas

1.1. Para μ. Si Y_1, Y_2, ..., Y_n representa una muestra aleatoria tomada de una población normal, con $\bar{Y}$ la media de a muestra y S^2 la varianza de la muestra, queremos calcular un intervalo de confianza para μ, la media poblacional, con σ^2 desconocido y n pequeño. Bajo esas suposiciones

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

tiene distribución t con $n-1$ grados de libertad. Si utilizamos dicha cantidad como pivote, entonces para el intervalo

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) = 1-\alpha \Rightarrow P\left(\bar{Y} - t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right) \leq \mu \leq \bar{Y} + t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 1-\alpha.$$

Es importante recordar que $t_{\alpha/2}$ también depende de los $n-1$ grados de libertad. Los intervalos unilaterales serían $[\bar{Y} - t_\alpha(S/\sqrt{n}), +\infty)$ y $(-\infty, \bar{Y} + t_\alpha(S/\sqrt{n})]$.

<u>Ejemplo</u>. Los resultados de la prueba de conocimientos generales (SAT), que han disminuido lentamente desde el inicio de la prueba, ahora han comenzado a elevarase. Al principio se contempló un promedio de 500 puntos. En 1991 el promedio fue de aproximadamente 422 puntos para el examen de habilidad verbal y 474 puntos para el de matemáticas. Una muestra aleatoria de los resultados de 20 estudiantes de último año de una preparatoria urbana grande dio como resultado las medias y desviaciones estándar que aparecen en la siguiente tabla.

	Capacidad verbal	Matemáticas
Media de la muestra	419	455
Desviación estándar de la muestra	57	69

a Encuentre un intervalo de confianza de 90 % para la media de los resultados del examen de verbal de los estudiantes de último grado de la escuela preparatoria.
b ¿El intervalo que encontró en el inciso a) incluye el valor de 422, la media real de los resultados del examen de capacidad verbal aplicado en 1991?¿Qué concluye usted?
c Construya un intervalo de confianza del 90 % para la media de los resultados del examen de matemáticas que presentaron los estudiantes de último grado de la preparatoria. ¿Incluye este intervalo el valor 474, la media real de los resultados de matemáticas de 1991? ¿Qué puede concluir de esto?

<u>Solución</u>.
a) Para calcular el intervalo de confianza, tomando en cuenta que t tiene 19 grados de libertad, tenemos

$$\bar{Y} \pm t_{0{,}05}\left(\frac{s}{\sqrt{n}}\right) = 419 \pm 1{,}729\left(\frac{57}{\sqrt{20}}\right) = 419 \pm 22{,}04 \Rightarrow [396{,}96; 441{,}04].$$

b) Es notable que el valor real 422, esta incluido en el intervalo de confianza, eso nos ayuda a concluir que el 90 % de las veces el promedio real forma parte de la distribución de estos estudiantes.

c) Nuevamente, ahora para matemáticas

$$\bar{Y} \pm t_{0,05}\left(\frac{s}{\sqrt{n}}\right) = 455 \pm 1{,}729\left(\frac{69}{\sqrt{20}}\right) = 455 \pm 26{,}68 \Rightarrow [428{,}32; 481{,}68].$$

Nuevamente el valor real, 474, esta dentro del intervalo.

1.2. Para $\mu_1 - \mu_2$. Si se quiere comparar dos medias poblacioneales normales, la primera con media μ_1 y varianza σ_1^2; y la segunda con media μ_2 y varianza σ_2^2; y se tienen dos muestras independientes entonces se pueden construir intervalos de confianza para $\mu_1 - \mu_2$. Supongamos que $\bar{Y_1}$ y $\bar{Y_2}$ son las medias muestrales de las muestras aleatorias, entonces

$$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} = \frac{(\bar{Y_1} - \bar{Y_2}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}.$$

Utilicemos dicho Z como cantidad pivote, sabiendo que tiene que tener distribución normal estándar. Claro está que la varianza es desconocida para ambas muestras, así que supongamos que $\sigma_1^2 = \sigma_2^2 = \sigma^2$, sin perdida de generalidad, y por lo tanto nos quedaría

$$Z = \frac{(\bar{Y_1} - \bar{Y_2}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

Como σ es desconocido, entonces utilicemos

$$S_p^2 = \frac{\displaystyle\sum_{i=1}^{n_1}(Y_{1i} - \bar{Y_1})^2 + \sum_{i=1}^{n_2}(Y_{2i} - \bar{Y_2})^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

para construir

$$W = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} = \frac{\displaystyle\sum_{i=1}^{n_1}(Y_{1i} - \bar{Y_1})^2}{\sigma^2} + \frac{\displaystyle\sum_{i=1}^{n_2}(Y_{2i} - \bar{Y_2})^2}{\sigma^2}$$

que es la suma de dos variables aleatorias que se distribuyen χ^2 con $n_1 - 1$ y $n_2 - 1$ grados de libertad. Por lo tanto W también se distribuye χ^2 con $n_1 + n_2 - 2$ grados de libertad. Por lo tanto

$$T = \frac{Z}{\sqrt{\dfrac{W}{\nu}}} = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

y por construcción T se distribuye t con $n_1 + n_2 - 2$ grados de libertad. Esa T es nuestro pivote para construir el intervalo de confianza que queda:

$$\left((\bar{Y}_1 - \bar{Y}_2) - t_{\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, (\bar{Y}_1 - \bar{Y}_2) + t_{\alpha/2}S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

<u>Ejemplo</u>. ¿Las calificaciones que obtienen los estudiantes de preparatoria en el examen de conocimientos generales (SAT) son diferentes dependiendo de su área de especialización? Se compararon las calificaciones de 15 estudiantes que pretendían especializarse en ingeniería con las de 15 estudiantes que planeaban estudiar la especialidad de lengua y literatura. La siguiente tabla contiene las medias y las desviaciones estándar de las calificaciones que los dos grupos de estudiantes obtuvieron en el área de matemáticas y de habilidad verbal del examen SAT.

	Habilidad verbal		Matemáticas	
	$\bar{y}$	s	$\bar{y}$	s
Ingeniería	446	42	548	57
Lengua y literatura	534	45	517	52

a Construya un intervalo de confianza de 95 % para la diferencia de medias de calificaciones en habilidad verbal entre los estudiantes que piensan estudiar ingeniería y los que piensan especializarse en lengua y literatura.
b Construya un intervalo de confianza de 95 % para la diferencia de medias entre las calificaciones que obtuvieron en el examen de matemáticas los estudiantes que planean especializarse en ingeniería y las que obtuvieron los que se van a especializar en lengua y literatura.

<u>Solución</u>.
a) Para poder hacer el intervalo de confianza necesitamos el S_p estimado

$$S_p^2 = \frac{(15-1)(42)^2 + (15-1)(45)^2}{15+15-2} = 1894{,}5 \Rightarrow S_p = 43{,}53.$$

Luego, el intervalo de confianza, tomando en cuenta que la t tiene 28 grados de libertad, es

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{0{,}025}S_p\sqrt{\frac{1}{15} + \frac{1}{15}} = (534 - 446) \pm (2{,}045)(43{,}53)(0{,}37)$$

$$= 88 \pm 32{,}5 \Rightarrow [55{,}5; 120{,}5].$$

Podemos ver que claramente las notas de los que quieren lengua y literatura, con confianza de 95 %, son mayores que las de los que quieren ingenería en habilidad verbal.

b) Ahora seguimos la misma técnica para el intervalo de la diferencia en las notas de matemáticas

$$S_p^2 = \frac{(15-1)(57)^2 + (15-1)(52)^2}{15+15-2} = 2976{,}5 \Rightarrow S_p = 54{,}56.$$

Ahora para el intervalo de confianza

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{0{,}025} S_p \sqrt{\frac{1}{15} + \frac{1}{15}} = (548 - 517) \pm (2{,}045)(54{,}56)(0{,}37)$$

$$= 31 \pm 41{,}28 \Rightarrow [-10{,}28; 72{,}28].$$

En el caso de las notas de matemáticas, como el valor cero esta incluido en el intervalo de confianza, no podemos asegurar que realmente exista una diferencia entre los que quieren lengua y literatura o los que quieren ingenería.

1.3. Para σ^2. Nuevamente suponemos que $Y_1, Y_2, ..., Y_n$ es una muestra aleatoria con distribución normal con media y varianza desconocida. Como se dijo anteriormente $\frac{(n-1)S^2}{\sigma^2}$ tiene distribución χ^2 con $n-1$ grados de libertad. Utilizaremos esa cantidad como pivote, entonces:

$$P\left(\chi_L^2 \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi_U^2\right) = 1 - \alpha$$

Ahora, a diferencia de las distribuciones utilizadas hasta ahora, la normal y la t, la χ^2 no es simétrica, así que se eligen como límites a $\chi^2_{1-(\alpha/2)}$ y $\chi^2_{\alpha/2}$ que son puntos que límitan áreas iguales en las colas. Luego

$$P\left(\chi^2_{1-(\alpha/2)} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2)}}\right) = 1 - \alpha.$$

Ejemplo. En el trabajo de laboratorio se procura verificar con cuidado la variabilidad de las lecturas generadas por muestras estándar. En un estudio sobre la cantidad de calcio presente en el agua potable, que formaba parte de un programa de control de calidad del agua, se analizó en el laboratorio una misma muestra estándar seis veces, en intervalos aleatorios. Las seis lecturas, expresadas en partes por millón, fueron de 9.54, 9.61, 9.32, 9.48, 9.70 y 9.26. Estime la varianza poblacional σ^2 para las lecturas de esta muestra estándar mediante un intervalo de confianza del 90 %.

<u>Solución</u>.

Sabemos que $S^2 = \dfrac{\sum_{i=1}^{6}(Y_i - \bar{Y})^2}{n-1}$, por lo tanto calculemos $(n-1)S^2$. Para ello primero calculemos $\bar{Y}$

$$\bar{Y} = \frac{\sum_{i=1}^{6} Y_i}{6} = \frac{56{,}91}{6} = 9{,}485.$$

Luego

$$(n-1)S^2 = \sum_{i=1}^{6}(Y_i - \bar{Y})^2 = 0{,}14275.$$

Ahora calculamos los χ^2 para $n-1=5$ grados de libertad. Tenemos que $\chi^2_{\alpha/2} = \chi^2_{0{,}05} = 11{,}07$ y $\chi^2_{1-(\alpha/2)} = \chi^2_{0{,}95} = 1{,}145$.

Por lo tanto el intervalo de confianza es:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}}; \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2)}}\right) = \left(\frac{0{,}14275}{11{,}07}; \frac{0{,}14275}{1{,}145}\right) = (0{,}0129; 0{,}1247).$$

1.4. Para la razón de dos varianzas. Si s_1^2 y s_2^2 son las varianzas de muestras independientes de tamaño n_1 y n_2, respectivamente, de poblaciones normales, entonces un intevalo de confianza de $(1-\alpha)100\,\%$ para σ_1^2/σ_2^2 es tal que:

$$P\left(\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(v_1,v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(v_2,v_1)\right) = 1-\alpha.$$

donde $f_{\alpha/2}(v_1,v_2)$ es un valor f con $v_1 = n_1 - 1$ y $v_2 = n_2 - 1$ grados de libertad y $f_{\alpha/2}(v_2,v_1)$ es un valor f similar con $v_2 = n_2 - 1$ y $v_1 = n_1 - 1$ grados de libertad.