1. Teorema del Límite Central

<u>Teorema</u>: Sean Y_1, Y_2, ..., Y_n variables aleatorias independientes idénticamente distribuidas con $E(Y_i) = \mu$ y $V(Y_i) = \sigma^2 < \infty$. Definamos

$$U_n = \sqrt{n}\left(\frac{\overline{Y} - \mu}{\sigma}\right)$$

donde $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.
Entonces, la función de distribución de U_n converge a la función de distribución normal estándar conforme $n \to \infty$.

Como se dijo varias clases atras si tenemos n variables aleatorias, cada una de ellas $X_i \sim \text{Bern}(p)$, entonces $Y = \displaystyle\sum_{i=0}^{n} X_i \sim \text{Bin}(n,p)$, esto con el conocimiento de que las X_i son independientes, todas ellas con $E[X_i] = p$ y $V[X_i] = p(1-p)$. Cuando ese n es grande, entonces

$$\frac{Y}{n} = \frac{1}{n}\sum_{i=0}^{n} X_i = \bar{X}$$

Es claro que $E\left[\dfrac{Y}{n}\right] = p$ y $V\left(\dfrac{Y}{n}\right) = \dfrac{p(1-p)}{n}$ y esa nueva variable tiene distribución $\dfrac{Y}{n} \sim N\left(p, \dfrac{p(1-p)}{n}\right)$

2. Distribuciones muestrales relacionadas con la distribución normal

<u>Definición</u>: Un estadístico es una función de las variables aleatorias que pueden observarse en una muestra y las constantes conocidas.

<u>Teorema</u>: Si Y_1, Y_2, ..., Y_n es una muestra aleatoria de tamaño n de una distribución normal con media μ y varianza σ^2, entonces

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

tiene una distribución normal con media $\mu_{\bar{Y}} = \mu$ y varianza $\sigma^2_{\bar{Y}} = \sigma^2/n$.
Es claro entonces que

$$Z = \frac{\bar{Y} - \mu_{\bar{Y}}}{\sigma_{\bar{Y}}} = \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\bar{Y} - \mu}{\sigma}\right)$$

Que como sabemos por teorema del límite central converge a una distribución normal estándar.

Ejemplo. La Agencia para la Protección del Ambiente de EE.UU. desea establecer normas que regulen la cantidad de químicos tóxicos arrojados en río y lagos de agua dulce. Una medida común de la toxicidad de cualquier contaminante es la concentración de éste que mataría a la mitad de los especímenes de prueba en determinado tiempo (por lo general, 96 horas en el caso de los peces). Esta medida recibe el nombre de CL50 (concentración letal que mata 50 % de los especímenes de prueba). En muchos estudios los valores del logaritmo natural de las mediciones de CL50 adoptan una distribución normal, por lo que el análisis se basa en los datos del ln(CL50). Los estudios de los efectos del cobre en cierta especie de peces (A) demuestran que la varianza de las mediciones del ln(CL50) está alrededor de 0.4, con las mediciones de la concentración expresadas en miligramos por litro. Si se van a realizar 10 estudios de CL50 para el caso del cobre, encuentre la probabilidad de que la media muestral y la media real de ln(CL50) no difiera en más de 0.5.

Solución.
Tenemos como datos que $\sigma^2 = 0{,}4$ y $n = 10$. Queremos calcular $P(|\bar{Y} - \mu| < 0{,}5)$. Como los datos provienen de una distribución normal, entonces podemos aproximar con el TLC

$$P(|\bar{Y}-\mu| < 0{,}5) = P(-0{,}5 < \bar{Y}-\mu < 0{,}5) = P\left(\frac{-0{,}5\sqrt{10}}{\sqrt{0{,}4}} < \frac{(\bar{Y}-\mu)}{\sigma}\sqrt{n} < \frac{0{,}5\sqrt{10}}{\sqrt{0{,}4}}\right)$$

$$= P(-2{,}5 < Z < 2{,}5) = 0{,}9876.$$

Ejemplo. Ahora suponga que los efectos del cobre en otra especie (B) de peces indica que la varianza de las mediciones del ln(CL50) es 0.8. Si las medias de las poblaciones del ln(CL50) de las dos especies son iguales, calcule la probabilidad de que la media muestral de la especie A sea mayor a la media muestral de la especie B por lo menos una unidad, si se toman muestras aleatorias de 10 mediciones de cada especie.

Solución.
Se tiene que $\mu_A = \mu_B$, ya del ejercicio anterior sabiamos que $\sigma_A^2 = 0{,}4$ y ahora que $\sigma_B^2 = 0{,}8$. También tenemos que $n_A = n_B = 10$. Queremos conseguir $P(\bar{Y}_A - \bar{Y}_B > 1)$. Como es la resta de dos variables con distribución normal, entonces la resta de ellas dos también será normal. Como las medias poblacionales son iguales, entonces $\mu_A - \mu_B = 0$, y como son especies distintas, asumiremos que los efectos son independientes, por lo cual

$$V(\bar{Y}_A - \bar{Y}_B) = V(\bar{Y}_A) + V(\bar{Y}_B) = \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B} = \frac{0{,}4}{10} + \frac{0{,}8}{10} = \frac{1{,}2}{10}.$$

Luego

$$P(\bar{Y}_A - \bar{Y}_B > 1) = P\left(\frac{(\bar{Y}_A - \bar{Y}_B) - 0}{\sqrt{V(\bar{Y}_A - \bar{Y}_B)}} > \frac{1}{\sqrt{\frac{1{,}2}{10}}}\right) = P(Z > 2{,}88) = 0{,}0019.$$

<u>Teorema</u>: Si Y_1, Y_2, ..., Y_n se definen como en el teorema anterior, entonces $Z_i = \dfrac{Y_i - \mu}{\sigma}$ son variables aleatorias normales estándar e independientes, donde $i = 1, 2, ..., n$ y

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma}\right)^2$$

tienen una distribución χ^2 con n grados de libertad.

<u>Teorema</u>: Si Y_1, Y_2, ..., Y_n es una muestra aleatoria de una distribución normal con media μ y varianza σ^2, entonces

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$$

tiene una distribución χ^2 con $n-1$ grados de libertad. De la misma manera, $\bar{Y}$ y S^2 son variables aleatorias independientes.

<u>Ejemplo</u>. Para el mismo ejemplo anterior, suponga que se toman $n = 20$ observaciones respecto a las mediciones de ln(CL50) y que $\sigma^2 = 1{,}4$. Sea S^2 la varianza muestral de las 20 mediciones.

a Encuentre un número b para el cual $P(S^2 \leq b) = 0{,}975$.
b Encuentre un número a tal que $P(S^2 \geq a) = 0{,}975$.
c Si a y b son los valores anteriores, calcule $P(a \leq S^2 \leq b)$.

<u>Solución</u>.
a)

$$P(S^2 \leq b) = 0{,}975 \Rightarrow P\left(\frac{S^2(n-1)}{\sigma^2} \leq \frac{(n-1)b}{\sigma^2}\right) = 0{,}975 \Rightarrow P\left(\chi^2 > \frac{19b}{1{,}4}\right) = 0{,}025$$

Se busca en la tabla para $\chi^2_{0{,}025}$ con 19 grados de libertad, que es 32.9, luego

$$\frac{19b}{1{,}4} = 32{,}9 \Rightarrow b = 2{,}42.$$

b)

$$P(S^2 \geq a) = 0{,}975 \Rightarrow P\left(\frac{S^2(n-1)}{\sigma^2} \geq \frac{(n-1)a}{\sigma^2}\right) = 0{,}975 \Rightarrow P\left(\chi^2 \geq \frac{19a}{1{,}4}\right) = 0{,}975$$

Se busca para $\chi^2_{0,975}$ con 19 grados de libertad, y se obtiene 8.91, por lo tanto

$$\frac{19a}{1{,}4} = 8{,}91 \Rightarrow a = 0{,}65.$$

c) Ahora queremos

$$P(0{,}65 \leq S^2 \leq 2{,}42) = P(8{,}91 \leq \chi^2 \leq 32{,}9) = 0{,}975 - 0{,}025 = 0{,}9500.$$

<u>Definición</u>: Sea Z una variable aleatoria normal estándar, y W una variable con una distribución χ^2 con ν grados de libertad. Entonces, si Z y W son independientes,

$$T = \frac{Z}{\sqrt{W/\nu}}$$

se dice que tiene distribución t con ν grados de libertad (g.l.).

Ahora supongamos que Y_1, Y_2, ..., Y_n es una muestra aleatoria de una población normal de media μ y varianza σ^2, entonces $Z = \sqrt{n}(\bar{Y} - \mu)/\sigma$ se distribuye normal estándar y $W = (n-1)S^2/\sigma^2$ se distribuye χ^2 con $n-1$ grados de libertad. Como Z y W son independientes sigue que

$$T = \frac{Z}{\sqrt{W/\nu}} = \frac{\sqrt{n}(\bar{Y} - \mu)/\sigma}{\sqrt{[(n-1)S^2/\sigma^2]/(n-1)}} = \sqrt{n}\left(\frac{\bar{Y} - \mu}{S}\right).$$

tiene distribución t con $n-1$ grados de libertad.

<u>Definición</u>: Si W_1 y W_2 son variables aleatorias independientes que tienen una distribución χ^2 con ν_1 y ν_2 grados de libertad, respectivamente, entonces

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

se dice que tiene distribución F con ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador.

Si $W_1 = (n_1 - 1)S_1^2/\sigma_1^2$ y $W_2 = (n_2 - 1)S_2^2/\sigma_2^2$ tienen distribuciones χ^2 independientes con $\nu_1 = n_1 - 1$ y $\nu_2 = n_2 - 1$ grados de libertad, respectivamente; entonces

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} = \frac{[(n_1-1)S_1^2/\sigma_1^2]/(n_1-1)}{[(n_2-1)S_2^2/\sigma_2^2]/(n_2-2)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}.$$

tiene distribución F con $n_1 - 1$ grados de libertad en el numerador y $n_2 - 1$ grados de libertad en el denominador.

<u>Ejemplo</u>. Sea S_1^2 la varianza muestral de una muestra aleatoria de 10 valores del ln(CL50) para el cobre y sea S_2^2 la varianza muestral de una muestra aleatoria de 8 valores de ln(CL50) para el plomo. Se utilizaron muestras de la misma especie de peces. Suponga que la varianza de la población de mediciones respecto al cobre es el doble de la varianza de la población de mediciones correspondiente respecto al plomo. Suponga, además, que S_1^2 y S_2^2 son independientes.

a Encuentre un número b tal que $P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 0{,}95$.

b Encuentre un número a para el cual $P\left(a \leq \frac{S_1^2}{S_2^2}\right) = 0{,}95$.

<u>Solución</u>.
a) Queremos conseguir

$$P\left(\frac{S_1^2}{S_2^2} \leq b\right) = 0{,}95 \Rightarrow P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \leq b\frac{\sigma_2^2}{\sigma_1^2}\right) = 0{,}95$$

$$\Rightarrow P(F \leq \frac{b}{2}) = 0{,}95 \Rightarrow P(F > \frac{b}{2}) = 0{,}05.$$

Buscamos en la tabla de la distribución F para 0,05 con 9 grados de libertad en el numerador y 7 grados de libertad en el denominador, luego

$$\frac{b}{2} = 3{,}68 \Rightarrow b = 7{,}36.$$

b) Usaremos que $P(X/Y \leq k) = P(Y/X \leq 1/k)$.

$$P\left(\frac{S_1^2}{S_2^2} \geq a\right) = 0{,}95 \Rightarrow P\left(\frac{S_1^2}{S_2^2} < a\right) = 0{,}05$$

$$\Rightarrow P\left(\frac{S_2^2}{S_1^2} \geq \frac{1}{a}\right) = 0{,}05 \Rightarrow P\left(\frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \geq \frac{2}{a}\right) = 0{,}05$$

Se busca en la tabla de la distribución F para 0,05 con 7 grados de libertad en el numerador y 9 grados de libertad en el denominador, entonces

$$\frac{2}{a} = 3{,}29 \Rightarrow a = \frac{2}{3{,}29} \Rightarrow a = 0{,}61.$$