5 Distribuciones Muestrales

1. Introducción

Al definir la estadística se explicó que la probabilidad se trabaja desde la población hacia la muestra, mientras que la inferencia estadística se trabaja en sentido contrario, es decir, de la muestra hacia la población. Por lo tanto, para comenzar con la inferencia, necesitamos hacer un breve recuento del concepto de población y definir con claridad el concepto de muestra aleatoria.

Una población representa el estado de la naturaleza o la forma de las cosas con respecto a un fenómeno aleatorio, el cual puede identificarse a través de una característica medible. Por ejemplo, el nivel de colesterol en la sangre de una persona. Como la inferencia estadística se formula con base en una muestra de elementos de la población de interés, el proceso por el cual se obtiene la muestra será aquel que asegure una buena muestra. Una buena manera de escoger una muestra resulta cuando el proceso de muestreo proporciona a cada elemento de la población una oportunidad igual e independiente de ser incluido en la muestra. Si la población consta de N elementos y de éstos se toma una muestra de tamaño n, el proceso de muestreo debe asegurar que cada posible muestra de tamaño n tenga la misma posibilidad de ser seleccionada. Este procedimiento conduce a lo que se denomina como muestreo aleatorio simple.

La naturaleza de la información estadística requiere una total imparcialidad en la selección de la muestra. Al extraer una muestra aleatoria debemos analizar las características de la población. Muchas veces ésta no consta de objetos tangibles a partir de los cuales se selecciona un cierto número para formar la muestra. La población puede estar formada por un número infinito de posibles resultados para alguna característica de interés. Sea X esta característica, la cual puede estar definida por una función de densidad $f(x,\theta)$, correspondiente a la población.

Las siguientes son las formas de realizar el muestreo para esta población:

a) Se diseña un experimento y se lleva a cabo para extraer la primera observación X_1 de la característica medible. El experimento se repite bajo las mismas condiciones para extraer la segunda observación X_2. El proceso se continúa de igual manera hasta obtener n observaciones de la característica de interés $\{X_1, X_2,\ldots,X_n\}$. Las observaciones se obtienen a través de ensayos independientes que ocurren cada vez que el experimento se repite bajo las mismas condiciones. Cada una de las observaciones $X_1, X_2,\ldots,X_n$ es una variable aleatoria cuya distribución de probabilidad es idéntica a la de la población.

Situación diferente ocurre cuando la selección se lleva a cabo de objetos tangibles de una población que consta de un número finito de elementos. La característica medible puede ser un atributo o una medición cuantitativa como la duración de un servicio. En este caso existen dos formas de tomar la muestra.

b) Después de llevar a cabo una mezcla adecuada de cada uno de los elementos de la población, se extrae uno, se observa la característica medible. Sea X_1 esta observación. El elemento se repone a la población, ésta vuelve a mezclarse y se extrae el segundo elemento X_2. El proceso se continúa hasta que se obtengan n observaciones $X_1, X_2,\ldots,X_n$ de la característica X. Este proceso se denomina “muestreo con reemplazamiento”.

.

c) Después de llevar a cabo una mezcla adecuada de cada uno de los elementos de la población, se seleccionan n elementos uno después del otro, sin reemplazo. Este proceso se denomina “muestreo sin reemplazamiento”.

En el caso b) cada una de las observaciones $X_1, X_2,\ldots,X_n$ es una variable aleatoria cuya función de densidad es idéntica a la de la población original. En el caso c) las observaciones $X_1, X_2,\ldots,X_n$ son variables aleatorias cuyas distribuciones marginales son iguales a la de la población. La diferencia entre estas dos técnicas es el concepto de independencia. En el caso b) las observaciones $X_1, X_2,\ldots,X_n$ constituyen un conjunto de variables aleatorias independientes e idénticamente distribuidas, dado que por el proceso de reemplazo ninguna observación se ve afectada por las demás. En el caso c) las observaciones no son independientes.

En este capítulo presentaremos los elementos requeridos para analizar las distribuciones y los principales resultados obtenidos cuando se toma una muestra aleatoria de una población infinita, o de una población finita pero con reemplazamiento. En un capítulo posterior presentaremos las técnicas para analizar los resultados obtenidos al tomar una muestra de una población finita, sin reemplazamiento.

Muestreo
Proceso de toma de muestras, análisis y la obtención de conclusiones.

Muestra aleatoria
Sea $X_1, X_2,...,X_n$ un conjunto de variables aleatorias independientes e idénticamente distribuidas (iid). Se dice que $X_1, X_2,...,X_n$ forman una muestra aleatoria de tamaño n- ma(n)

Parámetro poblacional θ
Las características se refieren a la información básica de interés sobre las unidades muestreadas. Se denomina parámetro al valor que toma una característica de la población. Por ejemplo: El verdadero porcentaje de votantes a favor de determinado candidato, la resistencia media a la rotura de un envase, el porcentaje medio de defectuosos que da un proceso. Estos parámetros los denotaremos, por lo general, mediante la letra griega θ.

Generalmente el objetivo de los estudios de muestreo es estimar uno o más parámetros de la población. Generalmente estos parámetros se refieren o a la media poblacional, que denotaremos por μ, o a la proporción poblacional, que denotaremos por la letra θ.

Estadístico T

Se denomina "Estadístico T" a cualquier función de la muestra aleatoria, y que no depende de ningún parámetro desconocido. El estadístico lo denotaremos por T

$$T = t(X_1, X_2,\ldots, X_n)$$

Ejemplos de estadísticos son la media muestral, la proporción muestral o la varianza muestral, entre otros. Los estadísticos son variables aleatorias, y como tales tiene una función de densidad. Su función de densidad se denomina "distribución muestral".

2. Principales estadísticos

2.1. Media muestral

La media muestral se define como:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Se supone que la muestra aleatoria proviene de una población (finita o infinita), caracterizada por los siguientes parámetros: $E(X) = \mu$ y $V(X) = \sigma^2$

Algunas de las propiedades de la media muestral son:

a) Valor esperado: $E(\overline{X})$

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

b) Varianza muestral: $V(\overline{X})$

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right)$$

Como se trata de una muestra aleatoria compuesta por variables independientes, entonces la varianza de una suma es la suma de las varianzas, por lo cual:

$$V(\bar{X})=\frac{1}{n^2}\sum_{i=1}^{n}V(X_i)=\frac{1}{n^2}\sum_{i=1}^{n}V(X)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

donde $V(X_i) = V(X) = \sigma^2$ porque las variables son idénticamente distribuidas.

En resumen se tiene que:

$$E(\bar{X})=\mu \quad y \quad V(\bar{X})=\frac{\sigma^2}{n}$$

Recuérdese que la varianza es una medida de la variabilidad o dispersión de los datos alrededor de la media. Una varianza pequeña implica una alta probabilidad de que la variable aleatoria tome un valor cercano de la media.

Para un tamaño de muestra n grande la varianza de la media muestral $V(\bar{X})$ es muy pequeña lo cual implica que existe una alta probabilidad de que $\bar{X}$ tome un valor muy cerca de su media. Esto es parte de lo que se conoce como la Ley de los grandes números.

2.2. Proporción muestral P

Si se toma una muestra aleatoria $X_1, X_2,...,X_n$, donde cada X_i toma los valores 0 ó 1, (por ejemplo, cero si la i-ésima persona no está de acuerdo con determinada propuesta, y uno en caso de que lo este, ó cero si el i-ésimo artículo es bueno y uno si es defectuoso), la proporción muestral P se define como:

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

La variable aleatoria X_i sigue una distribución de Bernoulli con parámetro θ. Su valor esperado y su varianza están dados por

$$E(X_i) = \theta, \qquad V(X_i) = \theta(1 - \theta)$$

La variable aleatoria $X=\sum_{i=1}^{n}X_i$ tiene una distribución binomial, cuyas principales características son:

$$p(x)=\binom{n}{x}\theta^x(1-\theta)^{n-x}, \quad x=0, 1, 2,...,n$$

$$E(X) = n\theta, \qquad V(X) = n\,\theta(1 - \theta)$$

Algunas de las propiedades de la proporción muestral P son:

a) Valor esperado: E(P)

$$E(P)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}\theta=\frac{1}{n}n\theta=\theta$$

b) Varianza: V(P)

$$V(P)=V\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\frac{1}{n^2}V\left(\sum_{i=1}^{n}X_i\right)$$

Como se trata de una muestra aleatoria compuesta por variables independientes, entonces la varianza de una suma es la suma de las varianzas, por lo cual:

$$V(\overline{X})=\frac{1}{n^2}\sum_{i=1}^{n}V(x_i)=\frac{1}{n^2}\sum_{i=1}^{n}V(X)=\frac{1}{n^2}\sum_{i=1}^{n}\theta(1-\theta)=\frac{1}{n^2}n\theta(1-\theta)=\frac{\theta(1-\theta)}{n}=\frac{\sigma^2}{n}$$

donde $V(X_i) = V(X) = \sigma^2 = \theta(1-\theta)$ porque las variables son idénticamente distribuidas.

En resumen se tiene que:

$$E(P)=\theta\,,\quad V(P)=\frac{\sigma^2}{n}=\frac{\theta(1-\theta)}{n}$$

2.3. Varianza muestral S^2

La varianza muestral está definida como:

$$S^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$$

Para simplificar los cálculos la varianza muestral puede calcularse como:

$$S^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}=\frac{\sum_{i=1}^{n}(X_i^2-2\overline{X}X_i+\overline{X}^2)}{n-1}$$

$$=\frac{\sum_{i=1}^{n}X_i^2-2\overline{X}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\overline{X}^2}{n-1}=\frac{\sum_{i=1}^{n}X_i^2-2\overline{X}n\overline{X}+n\overline{X}^2}{n-1}$$

$$S^2=\frac{\sum_{i=1}^{n}X_i^2-n\overline{X}^2}{n-1}=\frac{n\sum_{i=1}^{n}\overline{X}_i^2-\left(\sum_{i=1}^{n}X_i\right)^2}{n(n-1)}$$

A la raíz cuadrada positiva de la varianza se la denomina “desviación estándar muestral”, es decir:

$$S=\sqrt{S^2}$$

Propiedades de la varianza muestral

a) Valor esperado $E(S^2)$

$$E(S^2)=E\left(\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}\right)=E\left(\frac{\sum_{i=1}^{n}(X_i^2-n\overline{X}^2)}{n-1}\right)$$

$$E(S^2)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i^2-n\overline{X}^2)\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}E(X_i^2)-nE(\overline{X}^2)\right)$$

Como se vio anteriormente, la varianza de una variable aleatoria puede expresarse en términos de su segundo y primer momentos como:

$$V(X) = E(X^2) - [E(X)]^2$$

Por lo cual el segundo momento $E(X^2)$ puede expresarse como:

$$E(X^2) = V(X) + [E(X)]^2 = \sigma^2 + \mu^2$$

De igual manera el segundo momento de la distribución de la media puede expresarse como:

$$E(\overline{X}^2)=V(\overline{X})+\left(E(\overline{X})\right)^2=\frac{\sigma^2}{n}+\mu^2$$

Por lo tanto el valor esperado de la varianza $E(S^2)$ puede expresarse como:

$$E(S^2)=\frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\right)=\frac{n\sigma^2+n\mu^2-\sigma^2-n\mu^2}{n-1}=\sigma^2$$

Se ha demostrado que el valor esperado de la varianza muestral es igual a la varianza poblacional. Esa es la razón por la cual al calcular la varianza muestral se divide por n-1 y no por el tamaño de la muestra (n), en cambio en el cálculo de la varianza poblacional se divide por el tamaño de la población (N).

b) Varianza

Se puede demostrar que la varianza de la varianza muestral está dada por:

$$V(S^2)=\frac{2\sigma^4}{n-1} \quad n>1$$

2.4. Estadísticos de orden

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de tamaño n de variables independientes e idénticamente distribuidas. Los valores de las variables corresponden al orden en que fue tomada la muestra. Suponga que la muestra aleatoria se ordena de menor a mayor. Sea $X_{(i)}$ la variable aleatoria que ocupa el puesto i en la muestra, donde $X_{(1)}$ corresponde al menor valor, y $X_{(n)}$ al mayor valor de la muestra. Estas nuevas variables aleatorias reciben el nombre de “estadísticos de orden”.

$X_{(1)}=Min\{X_i\}, i=1,2,\ldots,n$: Estadístico de orden 1

$X_{(n)}=Max\{X_i\}, i=1,2,\ldots,n$: Estadístico de orden n

3. Distribuciones Límites

3.1. Desigualdad de Chebyshev

Teorema. Sea X una variable aleatoria con E(X) = μ y varianza V(X) = σ^2. Sea ε cualquier número positivo, entonces

$$P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}$$

En muchas aplicaciones, el valor ε se expresa como múltiplo de la desviación estándar como $k\sigma$, entonces la desigualdad de Chebyshef se expresa como:

$$P(|X-\mu|\geq k\sigma)\leq\frac{1}{k^2}$$

Cuando la variable de interés no es una observación individual sino una media muestral $\overline{X}$, entonces el valor esperado está dado por E($\overline{X}$) = μ y varianza V($\overline{X}$) = σ^2/n. En este caso la desigualdad de Chebyshev estará dada por:

$$P(|\overline{X}-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{n\varepsilon^2}$$

$$P(|\overline{X}-\mu|\geq k\,\sigma_{\overline{x}})\leq\frac{1}{k^2}$$

Esto nos dice que cuando n es grande, la probabilidad de que haya alguna diferencia entre la verdadera media μ y su estimador $\overline{X}$ tiende a cero, es decir, que cuando $n\to\infty$, entonces $\overline{X}\to\mu$.

Ejemplo. Se tiene un lote grande de artículos y se desea estimar la fracción defectuosa usando muestreo aleatorio simple. Usando la desigualdad de Chebyshev, se desea encontrar el tamaño de muestra n tal que la probabilidad sea al menos del 95% de que la fracción defectuosa difiera de la verdadera fracción defectuosa en no más de 0.10.

3.2. La Ley de los Grandes Números

Suponga que X_1, X_2,...,X_n es una secuencia arbitraria de variables aleatorias con valores esperados EX_1), $E(X_2)$,...,$E(X_n)$. Suponga además que la variable aleatoria $\sum_{i=1}^{n} X_i$ tiene varianza para cada valor de n entero.

Teorema. Si $V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \to 0$ cuando $n\to\infty$ y ε es un número positivo, entonces,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))\right| \geq \varepsilon\right\} \to 0 \qquad \text{cuando } n\to\infty$$

o equivalentemente

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))\right| < \varepsilon\right\} \to 1 \qquad \text{cuando } n\to\infty$$

Definición. Una secuencia de variables aleatorias Z_n converge en probabilidad o converge estocásticamente a una constante "a" si para cada número positivo ε

$$P(|Z_n - a| \geq \varepsilon) \to 0 \quad \text{cuando } n\to\infty$$

Simbólicamente: $Z_n \xrightarrow{P} a$

El teorema enunciado anteriormente puede escribirse como:

$$\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i))\right|\right\} \xrightarrow{P} 0 \qquad \text{cuando } n\to\infty$$

El teorema anterior se le conoce como la “Ley débil de los grandes números”.

Colorario: Si $\overline{X}$ es la media muestral de una ma(n) de una población inducida por una variable aleatoria X con media μ y varianza σ^2, y si $\varepsilon>0$, entonces

$$P\left\{\left|\overline{X} - \mu\right| \geq \varepsilon\right\} \to 0 \quad \text{cuando} \quad n\to\infty$$

$$\text{ó } P\left\{\left|\overline{X} - \mu\right| < \varepsilon\right\} \to 1 \quad \text{cuando} \quad n\to\infty$$

$$\text{ó } \overline{X} \xrightarrow{P} \mu \quad \text{cuando} \quad n\to\infty$$

Conclusión: Si la muestra es grande existe una alta probabilidad de que la media muestral $\overline{X}$ esté cerca de la media poblacional μ.

Escogiendo un tamaño de muestra suficientemente grande podemos hacer que la probabilidad de que la media muestral tienda a la media poblacional sea tan alta (tan cerca de uno) como queramos.

3.3. Aplicación a la Distribución Binomial

Sea X el resultado de un ensayo de Bernoulli {0,1}, (por ejemplo, la inspección de un artículo) con $P(X=1) = \theta$, y $P(X=0) = q= 1 - \theta$.

Sea S_n el número de éxitos (artículos defectuosos) en los n ensayos de Bernoulli. El número medio de éxitos por ensayo $\overline{X}$ puede calcularse como:

$$\overline{X}=\frac{\sum_{i=1}^{n} X_i}{n}=\frac{S_n}{n}$$

Colorario. (Teorema de Bernoulli). Si S_n representa el número de éxitos en n ensayos independientes de un evento con probabilidad θ, y si $\varepsilon > 0$, entonces

$$P\left\{\left|\frac{S_n}{n}-\theta\right|\geq\varepsilon\right\}\to 0 \quad \text{cuando} \quad n\to\infty$$

$$\text{ó } P\left\{\left|\frac{S_n}{n}-\theta\right|<\varepsilon\right\}\to 1 \quad \text{cuando} \quad n\to\infty$$

$$\text{ó } P=\frac{S_n}{n}\xrightarrow{P}\theta \quad \text{cuando} \quad n\to\infty$$

Conclusión: Si la muestra es grande existe una alta probabilidad de que la proporción muestral P esté cerca de la verdadera proporción poblacional θ.

Combinación lineal de variables normales

Teorema. Sea X_1, X_2,..., X_n un conjunto de variables aleatorias distribuidas normalmente con valores esperados μ_i y varianzas σ_i^2, para i=1,2,...,n, entonces

$$Y=\sum_{i=1}^{n} a_i X_i$$

se distribuye normalmente con valores esperado $\mu_y=\sum_{i=1}^{n} a_i \mu_i$ y varianza $\sigma_y^2=\sum_{i=1}^{n} a_i^2 \sigma_i^2$, donde los a_i son valores constantes.

Ejemplo. Una estación de gasolina vende tres clases de combustibles: Diesel, gasolina corriente y gasolina extra, a precios de $2,100, $3,050 y $3,900 el galón, respectivamente. Suponga que la cantidad vendida diariamente de cada tipo se distribuye normalmente con medias 300, 500 y 1,000 galones, y desviaciones estándares de 80, 50 y 100 galones, respectivamente. Se pide calcular:

a) El ingreso medio diario
b) La desviación estándar
c) La probabilidad de que el ingreso diario supere los 6 millones de pesos? Los 7? Los 8?

4. Distribuciones muestrales

Se denomina distribución muestral a la función de densidad de un estadístico y esta función puede depender o no de parámetros desconocidos.

Sea X_1, X_2,...,X_n una muestra aleatoria de tamaño n de variables independientes e idénticamente distribuidas, con una función de densidad $f(x,\theta)$, donde θ es un parámetro desconocido (o un conjunto de parámetros).

Sea Θ el conjunto de todos los valores que θ puede tomar.

Sea $\Im = \{f(x, \theta), \theta\in\Theta\}$ el conjunto que representa la familia de todas las posibles funciones de densidad obtenidas para cualquier valor de θ.

La función de densidad conjunta de la muestra aleatoria X_1, X_2,...,X_n está dada por:

$$f_{X_1, X_2,\ldots, X_n}(x_1,x_2,\ldots,x_n)=f_1(x_1,\theta)f_2(x_2,\theta)\ldots f_n(x_n,\theta)=\prod_{i=1}^{n} f_i(x_i,\theta)$$

Como las variables son idénticamente distribuidas, la función de densidad conjunta puede expresarse como:

$$f(x_1,x_2,...,x_n,\theta)=\prod_{i=1}^{n} f(x_i,\theta)$$

Ejemplo. Si X ~Bernoulli(θ) ó Bin(1,θ) $\Rightarrow$ f(x) = $\theta^x(1-\theta)^{1-x} \Rightarrow$

$$f(x_1,x_2,...,x_n,\theta)=\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}=\theta^{\sum_{i=1}^{n}x_i}(1-\theta)^{n-\sum_{i=1}^{n}x_i}$$

Ejemplo: Si X~ N(μ, σ^2) $\Rightarrow \Theta = \{\theta_1, \theta_2\} \Rightarrow \theta_1 = \mu$, $\theta_2 = \sigma^2$.

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}, \qquad -\infty<x+\infty$$

$$f(x_1,x_2,...,x_n,\theta)=\prod_{i=1}^{n}\frac{1}{\sigma\sqrt{2\pi}}e^{-(x_i-\mu)^2/2\sigma^2}=\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma^2}$$

5. Teorema Central del Límite (Distribución de la media $\overline{X}$)

Importancia: El teorema central del límite (TCL) nos permite usar la distribución normal como la distribución de las medias de muestras grandes, sin interesar cual sea la distribución original de las variables aleatorias.

Teorema. Sea X_1, X_2,...,X_n una muestra aleatoria de tamaño n de variables independientes e idénticamente distribuidas tomadas de una población infinita, con media μ y varianza σ^2, entonces la distribución límite de

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

es la distribución normal estándar (0,1), cuando n $\to\infty$, (independiente de la distribución de X_1, X_2,...,X_n).

Otra forma de presentar el TCL es la siguiente:

Si X_1, X_2,...,X_n es una muestra aleatoria de tamaño n de variables independientes e idénticamente distribuidas tomadas de una población infinita, con media μ y varianza σ^2, y si $\overline{X}$ es la media muestral, entonces su distribución muestral tiende a una distribución normal con media μ y varianza σ^2/n cuando n$\to\infty$.

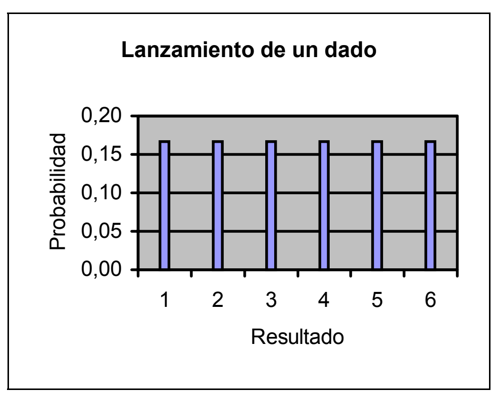
$$\overline{X}\Rightarrow N(\mu, \sigma^2/n) \quad \text{cuando } n\to\infty$$

Ejemplo gráfico

Con el fin de ilustrar gráficamente el TCL presentaremos la distribución de la media muestral obtenida al lanzar dos dados, en comparación con la distribución individual de cada dado.

Si X representa el resultado obtenido al lanzar un dado, entonces su función de probabilidad está dada por:

$$p(x)=\frac{1}{6}, \quad x=1,2,3,4,5,6$$

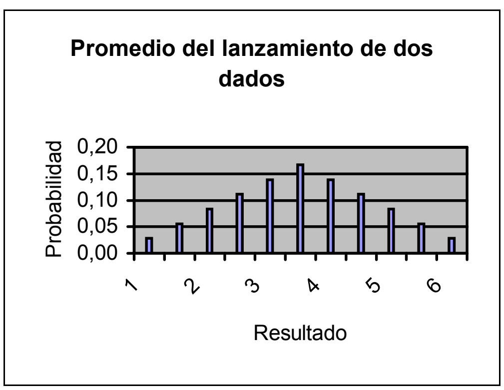


con la anterior representación gráfica.

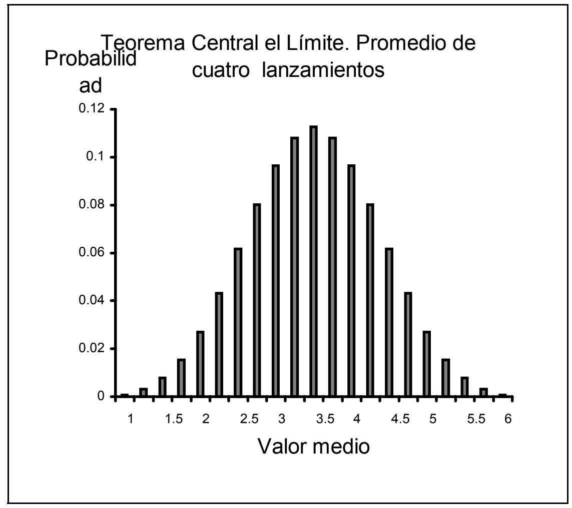
Consideremos ahora el lanzamiento de dos dados. Sean X_1 y X_2 los respectivos resultados. Sea $\overline{X}$ la media respectiva. La tabla siguiente presenta su respectiva distribución de probabilidad (la cual había sido analizada previamente al estudiar el concepto de variable aleatoria, y considerar la suma de los dos dados).

Distribución de la media de dos dados		
Suma	Media $\overline{X}$	P(X)
2	1.0	1/36
3	1.5	2/36
4	2.0	3/36
5	2.5	4/36
6	3.0	5/36
7	3.5	6/36
8	4.0	5/36
9	4.5	4/36
10	5.0	3/36
11	5.5	2/36
12	6.0	1/36

Su representación gráfica se presenta en la figura siguiente.



Como se puede observar, el cambio en la forma de la distribución es bastante notable, al pasar de una distribución completamente plana (uniforme discreta) a una distribución que, aunque no es normal, si tiende a



parecerse más a una distribución normal que a su distribución original. Si continuamos promediando más variables, la distribución resultante se aproximará aún más a una distribución. La siguiente gráfica presenta los resultados al promediar cuatro lanzamientos de la moneda.

Observación importante:

Debe tenerse en cuenta que si $n \to \infty$ entonces la varianza de $\overline{X}$ (= σ^2/n) tiende a cero, lo cual implica a su vez que $\overline{X} \to \mu$. Lo que el TCL dice es que cuando el tamaño muestral es grande, la media de una muestra aleatoria tiende a seguir la distribución normal. Cuándo n es lo suficientemente grande?. En general depende de la distribución original de la variable aleatoria X; sin embargo, para variables continuas y $n \geq 30$, la aproximación normal se aplica, no importa cual sea la distribución original. Para $n < 30$ la aproximación es válida según la forma de la distribución original.

Si la distribución original es continua y uniforme (por ejemplo el caso de los números aleatorios que van de cero a uno), para que el promedio tienda hacia una distribución normal, se requieren muestras de por lo menos 10 observaciones (esto se determinado mediante pruebas de bondad de ajuste).

Ejemplo: Una máquina vendedora de refrescos está programada para que la cantidad de refrescos que sirve sea una variable aleatoria con una media de 200 mililitros y una desviación estándar de 15 mililitros. Cuál es la probabilidad de que la cantidad media de refresco servido en una muestra aleatoria de 36 refrescos sea por lo menos 204 mililitros?. Realice los cálculos usando la desigualdad de Chebyshev y el TCL.

Se tiene lo siguiente: $\mu = 200$, $\sigma = 15$, $n = 36$

a) Usando la desigualdad de Tchevyshef, se tiene:

$$P(|\overline{X}-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \quad \text{donde } \varepsilon = 4,$$

$$P(|\overline{X}-\mu| \geq 4) \leq \frac{15^2}{364^2} = 0.39$$

Sin embargo, se tiene que

$$P(|\overline{X}-\mu| \geq 4) = P(\overline{X}-\mu \leq -4) + P(\overline{X}-\mu \geq 4) \leq 0.39 \Rightarrow P(\overline{X}-\mu \geq 4) \leq 0.195$$

suponiendo una distribución simétrica.

b) Usando el TCL:

$$P(\overline{X}-\mu \geq 4) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{204-200}{15/6}\right) \Rightarrow P(Z \geq 1.6) = 0.0548$$

Con la desigualdad de Chebyshev donde no se hace ninguna suposición sobre la distribución de la media, tenemos que la probabilidad es menor que 0.195, y usando el teorema central del límite se tiene que la probabilidad es .0548 (menor que 0.195)

Ejemplo: Continuando con el ejemplo anterior

a) Cuál es el error máximo que se está dispuesto a aceptar en la estimación de la media, si se especifica una probabilidad de 0.05?
b) Cuál debe ser el tamaño de la muestra de tal forma que el error máximo que se cometa en la estimación de la media sea de 4 mililitros con una probabilidad de 0.05 o menos?

Solución:

a) Queremos determinar el valor de ε de tal forma que

$$P(|\overline{X}-\mu| \geq \varepsilon) \leq 0.05$$

$$P(|\overline{X}-\mu| \geq \varepsilon) = P(\overline{X}-\mu \leq -\varepsilon) + P(\overline{X}-\mu \geq \varepsilon) = 0.05$$

$$P(|\overline{X}-\mu \geq \varepsilon|) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{-\varepsilon}{15/6}\right) + P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{\varepsilon}{15/6}\right)$$

$$P(|\overline{X}-\mu \geq \varepsilon|) = P\left(Z \leq \frac{-\varepsilon}{15/6}\right) + P\left(Z \geq \frac{\varepsilon}{15/6}\right)$$

Como la probabilidad del error es de 0.05, y la distribución es simétrica, entonces la probabilidad de un error menor de $-\varepsilon$ es 0.025 y la probabilidad de un error superior a $+\varepsilon$ es también de 0.025. Por lo tanto, los valores de $-\varepsilon/(15/6)$ y $+\varepsilon/(15/6)$ corresponden a los valores de la distribución normal que tienen áreas de 0.025 a la izquierda y +0.025 a la derecha, respectivamente. Es decir[1],

$$\frac{-\varepsilon}{15/6} = Z_{0.975} = -Z_{0.025}, \qquad \frac{+\varepsilon}{15/9} = Z_{0.025} = 1.96$$

Por lo tanto $\varepsilon = (1.96 \text{x} 15/6) = 4.9$

b) Queremos determinar el valor de n de tal forma que

[1] Notación: El valor Z_P corresponde al valor de la distribución normal que tiene un área de P a la derecha, o 1-P hacia la izquierda

$P(|\overline{X} - \mu| \geq \varepsilon) \leq \alpha$ ó equivalentemente $P(|\overline{X} - \mu| \leq \varepsilon) \geq 1-\alpha$, donde α = 0.05, y ε = 4.

De nuevo $|\overline{X} - \mu| \leq -\varepsilon \Rightarrow \varepsilon \leq \overline{X} - \mu \leq +\varepsilon$. Por lo tanto:

$$P(|\overline{X} - \mu| \leq \varepsilon) = P(-\varepsilon \leq \overline{X} - \mu \leq +\varepsilon) \geq 1 - \alpha$$

$$P\left(|\overline{X} - \mu \leq \varepsilon|\right) = P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{-\varepsilon}{\sigma/\sqrt{n}}\right) \geq 1-\alpha$$

La anterior expresión puede escribirse como:

$$P\left(|\overline{X} - \mu \leq \varepsilon|\right) = P(Z_{1-\alpha/2} \leq Z \leq Z_{\alpha/2}) = 1-\alpha$$

donde

$$\frac{-\varepsilon}{\sigma/\sqrt{n}} = Z_{1-\alpha/2} = -Z_{\alpha/2}, \qquad \frac{+\varepsilon}{\sigma/\sqrt{n}} = Z_{\alpha/2}$$

Por lo tanto el tamaño de muestra requerido está dado por:

$$n = \left(\frac{\sigma Z_{\alpha/2}}{\varepsilon}\right)^2 = \left(\frac{15x1.96}{4}\right)^2 = 54$$

La fórmula anterior es una fórmula general para determinar el tamaño de la muestra cuando se muestrea de una población infinita. Si se muestrea de una población finita(N) es necesario reemplazar la desviación estándar de la media $\sigma/\sqrt{n}$ por la desviación estándar de la media cuando se muestrea de la población finita, que está dada por $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$. El tamaño de muestra resultante está dado por:

$$n = \frac{Z^2_{1-\alpha/2}\sigma^2 N}{\varepsilon^2(N-1) + Z^2_{1-\alpha/2}\sigma^2}$$

Teorema de Moivre. Si S_n es el número de éxitos en n ensayos independientes de un evento con probabilidad θ, y si a y b son dos números reales con a<b, entonces

$$P\left\{n\theta - a\sqrt{n\theta 1-\theta)} < S_n < n\theta + a\sqrt{n\theta 1-\theta)}\right\} = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}dz \text{ , cuando } n \to \infty.$$

Es decir, el teorema de DeMoivre dice que cuando el tamaño de la muestra es grande, la distribución Binomial se puede aproximar por una distribución normal con media E(X) = μ = nθ y varianza V(X) = σ^2 = nθ(1-θ).

Problemas:

1) Suponga que el número de barriles de petróleo crudo que produce un pozo diariamente es una variable aleatoria con una distribución no especificada. Si se observa la producción de 64 días, seleccionadas en forma aleatoria, y se sabe que la desviación estándar del número de barriles por día es 16 barriles,

a) Determine la probabilidad de que la producción de un día no exceda de la media en más de 32 barriles.
b) Determine la probabilidad de que la producción media se encuentre a no más de 4 barriles del verdadero valor de la producción diaria. (Use la desigualdad de Chebyshef y el TCL).

2) Para realizar un experimento se le ha suministrado una resistencia. Como se desea estimar la resistencia muy exactamente, usted realiza 36 mediciones con el mismo método, cuya experiencia previa indica que tiene una varianza de 10 ohmios. La medida promedio de sus mediciones da 52 ohms.

a) Cuál es la probabilidad de que su promedio de 52 ohms esté en error por mas de 1 ohm?
b) Cuál es la probabilidad a priori de que se obtengan valores muestrales para los cuales la media muestral $\overline{X}$ difiera de la media poblacional μ por más de un ohm?

c) Cuál es el error máximo ε que se puede cometer si se desea que la probabilidad a priori de que se cometa dicho error no exceda de 0.05?
d) Cuántas mediciones deben realizarse con el fin de obtener una probabilidad a priori de 0.05 de que el promedio muestral difiera de la verdadera resistencia media por más de 1 ohm?

3) Considere el proceso de selección de una muestra de una distribución que tiene una varianza σ^2 = 10, pero con una media desconocida. Cuál debe ser el tamaño de la muestra para que la media $\overline{X}$ se encuentre dentro de un intervalo igual a dos unidades de la media poblacional, con una probabilidad de por lo menos 0.90
a) Usando la desigualdad de Chebyshef (n = 25)
b) Usando el TCL. (n = 7)

4) Un inspector de pesos y medidas visita una planta de empacado para verificar que el peso neto de las cajas sea el indicado en éstas. El gerente de la planta asegura al inspector que el peso promedio de cada caja es de 750 gr. con una desviación estándar de 5 gr. El inspector selecciona, al azar, 100 cajas y encuentra que el peso promedio es de 748 gr. Bajo estas condiciones, ¿qué tan probable es tener un peso de 748 o menos? ¿Qué actitud debe tomar el inspector?

5) La probabilidad de que un basquetbolista anote en un lanzamiento que realice es 0.5. Si realiza 20 lanzamientos, cuál es la probabilidad de que anote al menos en 9 lanzamientos?
a) Usando la distribución exacta.
b) Usando la aproximación normal sin factor de corrección de continuidad.
c) Usando la aproximación normal con factor de corrección de continuidad
d) Es aplicable la aproximación de Poisson?

Distribución de la media para poblaciones finitas

Si un experimento consiste en seleccionar uno o más valores de un conjunto finito de números ($C_1, C_2, \ldots, C_N$), este conjunto recibe el nombre de “población finita de tamaño N”.

Si se realiza un muestreo de esta población sin reemplazo, es decir, sin sustitución de los elementos muestreados previamente, entonces, los diferentes elementos de la muestra ($X_1, X_2, \ldots, X_n$) no son independientes.

Definición. Si X_1 es el primer elemento de la muestra, X_2 es el segundo valor tomado,...,X_n es el n ésimo valor tomado, la distribución de probabilidad conjunta de estas variables aleatorias está dada por:

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{N(N-1)(N-2)\ldots(n-n+1)}$$

y $X_1, X_2, \ldots, X_n$ constituyen la muestra aleatoria (n).

La probabilidad de tomar una muestra cualquiera de tamaño n de una población de N elementos está dada por $1 \Big/ \binom{N}{n}$

La distribución marginal de cualquier X_j está dada por:

$$f(x_j) = \frac{1}{N}, \quad \text{para } x_j = C_1, C_2, \ldots, C_N$$

La distribución marginal conjunta de dos variables aleatorias cualesquiera X_k y X_s está dada por:

$$g(x_k, x_s) = \frac{1}{N(N-1)}$$

Se puede demostrar que la covarianza entre X_k y X_s está dada por:

$$COV(X_k, X_s) = -\frac{\sigma^2}{N-1}$$

Teorema: Si $\overline{X}$ es la media de una muestra aleatoria de tamaño n tomada de una población finita de tamaño N con media μ y varianza σ^2, entonces

$$E(\overline{X}) = \mu \quad y \quad V(\overline{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}$$

El término $\frac{N-n}{N-1}$ recibe el nombre de "factor de corrección para población finita". Si N>>n este factor tiende a uno, y la varianza muestral será igual a σ^2/n (como si se tratara de una población infinita).

6. Distribución de la proporción

Sea X_1, X_2,...,X_n una muestra aleatoria tomada de una población con una distribución de Bernoulli con parámetro θ. Por lo tanto su función de probabilidad, su media y su varianza están dadas por:

$$p(x_i) = \theta^{x_i}(1-\theta)^{1-x_i}, \qquad x_i = 0,1$$

$$E(X) = \theta, \qquad V(X) = \theta(1 - \theta)$$

Si P es la proporción muestral, definida como

$$P = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum\limits_{i=1}^{n} X_i}{n} = \overline{X}$$

con $E(P) = \theta$, $V(P) = \frac{\sigma^2}{n} = \frac{\theta(1-\theta)}{n}$

En virtud del Teorema Central del Límite, como P es la media muestral de los diferentes valores de X_i, entonces P tiende a seguir una distribución normal con los parámetros dados anteriormente, es decir, P ~ $\text{Normal}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$. Ó también la variable aleatoria

$$Z = \frac{P-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$$

sigue una distribución normal estándar con media cero y varianza unitaria, cuando el tamaño de la muestra n es grande.

7. Distribución de la diferencia entre proporciones

Sea $X_{11}, X_{12}, \ldots, X_{1n_1}$ una muestra aleatoria (n_1) tomada de un proceso de Bernoulli con parámetro θ_1. Sea $X_{12}, X_{22}, \ldots, X_{2n_2}$ una muestra aleatoria (n_2) tomada de un proceso de Bernoulli con parámetro θ_2. Estamos interesado en conocer la distribución de la diferencia de proporciones muestrales $P_1 - P_2$.

Sabemos que $P_1 = \frac{X_{11} + X_{12} + X_{1n_1}}{n_1}$ se distribuye normalmente con una valor esperado θ_1 y una varianza $\frac{\theta_1(1-\theta_1)}{n_1}$ cuando n_1 es grande. De forma similar $P_2 = \frac{X_{21} + X_{22} + X_{2n_2}}{n_2}$ se distribuye normalmente con una valor esperado θ_2 y una varianza $\frac{\theta_2(1-\theta_2)}{n_2}$ cuando n_2 es grande.

Tenemos que:

$$E(P_1 - P_2) = \theta_1 - \theta_2$$

$$V(P_1 - P_2) = \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}$$

Como tanto P_1 como P_2 se distribuyen normalmente, entonces su diferencia también se distribuyen normalmente con los parámetros arriba mencionados. Es decir,

$$P_1 - P_2 \to \text{Normal}\left(\theta_1 - \theta_2,\ \frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}\right).$$

Ó también la variable aleatoria Z definida como

$$Z = \frac{P_1 - P_2 - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}}$$

tiene una distribución normal cuando n_1 y n_2 son grandes.

Si se desea verificar si las dos distribuciones son iguales, se tendría entonces que analizar si $\theta_1 = \theta_2$, es decir, $\theta_1 - \theta_2 = 0$.

Ejemplo: Un artículo del New York Times en 1987 reportó que se puede reducir el riesgo de sufrir ataques al corazón ingiriendo aspirina. Para llegar a esta conclusión el cronista se basó en los resultados de un experimento diseñado, en donde participaron dos grupos de personas. A un grupo de 11,034 personas se le suministró una dosis diaria de una pastilla que no contenía ninguna droga, y de estos 189 sufrieron posteriormente ataques al corazón, mientras que al otro grupo de 11,037 se les suministró una aspirina, y sólo 104 lo sufrieron.

Considera Usted que el cronista del New York Times estaba en lo correcto?.

8. Distribución Chi Cuadrado

Definición. Una variable aleatoria X tiene una distribución Chi Cuadrado ó Ji dos si su función de densidad de probabilidad está dada por:

$$f(x) = \begin{cases} \dfrac{1}{2^{\upsilon/2}\Gamma(\upsilon/2)} x^{\upsilon/2-1} e^{x/2}, & x \geq 0 \\ 0 \quad , & x < 0 \end{cases}$$

donde ν es el número de grados de libertad, o simplemente “grados de libertad”.

La distribución Chi Cuadrado es un caso particular de la distribución Gamma, cuya función de densidad está dada por:

$$f(t) = \lambda\, e^{-\lambda t} \frac{(\lambda t)^{k-1}}{\Gamma(k)} \qquad t \geq 0$$

donde $\Gamma(k)$ es la función gamma de k. Los valores correspondientes de los parámetros λ y k son $\lambda = 1/2$ y $k = \nu/2$.

El valor esperado y la varianza de la distribución Chi cuadrado están dados por:

$E(X) = \nu$, $\quad V(X) = 2\nu$

Notación: Si una variable tiene una distribución Chi cuadrado con ν grados de libertad, lo denotaremos como Chi(ν) ó χ^2_ν

Teoremas

1) Si Z es una variable aleatoria con distribución normal (0, 1), entonces Z^2 tiene una distribución Chi Cuadrado con un grado de libertad ($\nu = 1$).
2) Si $Z_1, Z_2, \ldots, Z_n$ es un conjunto de n variables independientes e idénticamente distribuidas con distribución normal (0,1), entonces

$$Z = Z_1^2 + Z_2^2 + \ldots + Z_n^2$$

tiene una distribución Chi Cuadrado con n grados de libertad.
3) Si $X_1, X_2, \ldots, X_k$ es un conjunto de k variables independientes con distribuciones Chi Cuadrado con $\nu_1, \nu_2, \ldots, \nu_k$ grados de libertad, respectivamente, entonces la variable aleatoria

$$X = X_1 + X_2 + \ldots + X_k$$

tiene una distribución Chi Cuadrado con $\nu = \nu_1 + \nu_2 + \ldots + \nu_k$ grados de libertad.

4) Sean X_1 y X_2 dos variables aleatorias. Si X_1 tiene una distribución ji dos con ν_1 grados de libertad, y $X_1 + X_2$ tiene otra distribución Chi cuadrado con $\nu > \nu_1$ grados de libertad, entonces X_2 tiene una distribución Chi cuadrado con $\nu - \nu_1$ grados de libertad.

Teorema. Si $\overline{X}$ y S^2 son la media y la varianza de una ma(n) tomada de una población normal con media μ y varianza σ^2, entonces
a) $\overline{X}$ y S^2 son independientes.
b) La variable aleatoria $\dfrac{(n-1)S^2}{\sigma^2}$ tiene una distribución Chi Cuadrado con n-1 grados de libertad.

La parte a) no se demostrará. A continuación se presenta la demostración de la parte b. Para ello considere la siguiente suma: $\sum_{i=1}^{n}(X_i-\mu)^2$. Esta expresión puede representarse de la siguiente manera, sumando y restando $\overline{X}$.

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left[(X_i-\overline{X})+(\overline{X}-\mu)\right]^2 = \sum_{i=1}^{n}\left[(X_i-\overline{X})^2+2(X_i-\overline{X})(\overline{X}-\mu)+(\overline{X}-\mu)^2\right]$$

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2+2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\overline{X})+\sum_{i=1}^{n}(\overline{X}-\mu)^2$$

$$\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-\mu)^2$$

dado que $\sum_{i=1}^{n}(X_i-\overline{X})$ es igual al cero. Si dividimos ambas expresiones por σ^2, se tiene:

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{I=1}^{N}(X_i-\overline{X})^2}{\sigma^2}+n\left(\frac{\overline{X}-\mu}{\sigma}\right)^2$$

Si multiplicamos y dividimos la primera parte de la expresión de la derecha por (n – 1) tenemos que:

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2}+\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

Observando la expresión anterior tenemos que el término $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$ sigue una distribución Chi Cuadrado con n grados de libertad, y el término $\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ sigue también una distribución Chi Cuadrado pero con un grado de libertad. Por lo tanto, la expresión $\frac{(n-1)S^2}{\sigma^2}$ sigue una distribución Chi Cuadrado con n-1 grados de libertad.

Tabulación. La función de distribución no puede calcularse en forma analítica; sin embargo, ha sido tabulada para diferentes valores de la probabilidad acumulada, y para varios grados de libertad. En algunas tablas se presenta la cola hacia la izquierda (probabilidad acumulada), y en otras la cola hacia la derecha. Para aquellos valores que no se encuentren en la tabla, se puede usar interpolación lineal, mediante la siguiente fórmula:

$$P = P_{i-1} + \frac{P_i - P_{i-1}}{X_i - X_{i-1}}\left(X - X_{i-1}\right)$$

Notación. Usaremos la notación $\chi^2_{\nu,P}$ para denotar el valor de la distribución Chi cuadrado con ν grados de libertad y una cola de P hacia la derecha (o una probabilidad acumulada de 1-P hacia la izquierda).

Problema: Haciendo uso de la relación existente entre las distribuciones gama y chi cuadrado, demuestre que la varianza de la varianza poblacional está dada por

$$V(S^2) = \frac{2\sigma^4}{n-1}, \quad n > 1$$

Ejemplo. Suponga que el espesor de un componente de un semiconductor es una dimensión crítica. El proceso de producción de tal característica se distribuye normalmente con una desviación estándar de 0.6 milésimas de pulgada. Para controlar el proceso se toman muestras periódicas de veinte piezas, y se define un límite de control con base en una probabilidad de 0.01 de que la varianza muestral exceda dicho límite, si el proceso está bajo control.

Qué se puede concluir si para una muestra dada la desviación estándar es 0.84 milésimas de pulgada?

Solución. La variable aleatoria de interés para nuestro caso es $X^2 = (n-1)S^2/\sigma^2$. Si denotamos por LSC el límite superior de control, entonces tenemos que se debe cumplir que:

$$P(X^2 > LSC) = P\left(\frac{(n-1)S^2}{\sigma^2} > LSC\right) = 0.01 \;\Rightarrow\; P\left(\frac{(n-1)S^2}{\sigma^2} \le LSC\right) = 0.99$$

Por lo tanto, debemos buscar en la tabla de la distribución Chi Cuadrado, con 19 grados de libertad, el valor que tenga una probabilidad hacia la derecha de 0.01 (ó hacia la izquierda de 0.99), denotado por $\chi^2_{19,0.01}$, correspondiente a 36.19, el cual debe satisfacer la siguiente desigualdad:

Se acepta si $\frac{(n-1)S^2}{\sigma^2} \le \chi^2_{19,0.01} = 36.19$

O también se acepta si $S^2 \le \frac{\chi^2_{19,0.01}\sigma^2}{(n-1)} = 0.6857$

Por lo tanto el criterio de decisión se puede expresar en una de las dos formas siguientes:

a) Se calcula = $X^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{19 \times 0.84^2}{0.6^2} = 37.24$

Como $X^2 = 37.24 > 36.19 \Rightarrow$ la muestra no proviene de un proceso con una desviación estándar de 0.60.

b) Se calcula $S^2 = .84^2 = 0.7056$. Como 0.7056 > 0.6857 se llega a la misma conclusión de que no es probable que la muestra tomada provenga de una población con una desviación estándar de 0.60 milésimas de pulgada.

Ejemplo. Un fabricante de baterías para automóviles garantiza que sus baterías duran, en promedio tres años con una desviación estándar de un año. Si cinco de sus baterías tienen duraciones de 1.9, 2.4 , 3.0, 3.5 y 4.3 años, puede asegurarse que las baterías tiene una duración estándar de un año?. Suponga que la duración de las baterías sigue una distribución normal.

Tenemos que: $\overline{X} = 3.0, \quad S^2 = 0.815$

$$X^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{4 \times 0.815}{1} = 3.26$$

Buscando en la tabla chi cuadrado para cuatro grados de libertad encontramos para $X^2 = 2.75$ una probabilidad P de 0.4 y para $X^2 = 4.04$ la probabilidad es P = 0.6. Usando interpolación lineal, tenemos una probabilidad aproximada de 0.479. Si la muestra proviene de una población con desviación estándar de uno, la probabilidad de que la muestra no pertenezca a esta distribución es de 1-0.479 = 0.521, la cual es excesivamente alta. Por lo tanto, no hay evidencia para concluir que la muestra no pertenece a una población con una varianza de uno.

9. Distribución t

Se sabe que $\overline{X}$ se distribuye normalmente con una media μ y una varianza σ^2/n, o la variable $Z = \frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}}$ se distribuye normalmente con media cero y varianza unitaria. Sin embargo, para calcular Z se requiere que σ sea conocido. Por lo tanto, se requiere una distribución para el caso en que σ sea desconocido y se pueda reemplazar por un estimativo, tal como S. Tal distribución es la distribución t.

Teorema. Sean Y y Z dos variables aleatorias independientes, Y con una distribución Chi cuadrado con ν grados de libertad, y Z con una distribución normal estándar (0,1), entonces la distribución de la variable

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

está dado por:

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad -\infty < t < +\infty$$

y se denomina “distribución t ó distribución de Student, con ν grados de libertad.

Origen: WS Gosset publicó inicialmente la distribución bajo el seudónimo de “Student”.

Propiedades generales

a) El valor esperado es cero $\Rightarrow E(T) = 0$
b) Distribución simétrica con respecto a cero.
c) La varianza de T está dada por $V(T) = \frac{v}{v-2}, \; v > 2$
d) La varianza de T es ligeramente mayor de 1.0, es decir, es ligeramente mayor que la de la distribución normal estandarizada.
e) Para $n \geq 30$ la distribución t tiende hacia la distribución normal.

Tabulación. La función de distribución no puede calcularse en forma analítica; sin embargo, ha sido tabulada para diferentes valores de la probabilidad acumulada, y para varios grados de libertad. Como la distribución es simétrica, solamente se presentan probabilidades acumuladas para valores positivos de t ($t \geq 0$). Los valores que se presentan en los encabezamientos de las columnas de la tabla corresponden a las probabilidades de exceder los respectivos valores de t, es decir, presentan las colas a la derecha de los valores respectivos de t. Para encontrar probabilidades correspondientes a valores negativos de t hay que hacer uso de la propiedad de simetría de la distribución t que nos dice que $\Phi(-t) = 1 - \Phi(t)$.

Notación. Usaremos la notación $t_{\nu,P}$ para denotar el valor de la distribución t con ν grados de libertad y una probabilidad acumulada de P hacia la derecha (o una probabilidad de 1-P hacia la izquierda).

La aplicación fundamental para la cual se usa esta distribución se presenta en el siguiente teorema.

Teorema. Si $\overline{X}$ y S^2 son la media y la varianza de una muestra aleatoria de tamaño n tomada de una población normal con media μ y varianza σ^2, entonces la variable

$$T = \frac{(\overline{X} - \mu)}{S/\sqrt{n}}$$

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

Demostración: Sabemos que

$$Y = \frac{(n-1)S^2}{\sigma^2}$$ tiene una distribución Chi Cuadrado con n-1 grados de libertad, y que

$$Z = \frac{(\overline{X} - \mu)}{\sigma/\sqrt{n}}$$ tiene una distribución normal (0,1). Entonces si aplicamos la definición de la distribución t tenemos:

$$T = \frac{Z}{\sqrt{Y/\nu}} = \frac{\dfrac{(\overline{X} - \mu)}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2} \Big/ (n-1)}} = \frac{(\overline{X} - \mu)}{S/\sqrt{n}}$$

Ejemplo. En un recorrido de prueba de una hora cada uno, el consumo promedio de gasolina de un motor fue 16.4 galones, con una desviación estándar de 2.1 galones. Se quiere saber si es cierta la afirmación de que "el consumo promedio de gasolina es 12 galones/hora".

Respuesta. Tenemos la siguiente información: n = 16, $\overline{X}$ = 16.4, s = 2.1 y μ = 12.0
Para responder la pregunta debemos verificar que tan probable es que una muestra de 16.4 galones pertenezca a una distribución con una media de 12. Por lo tanto, debemos calcular la probabilidad de que la media muestral sea mayor o igual que 16.4 si la verdadera media de donde proviene dicha muestra es 12 galones. Esto es:

$$P(\overline{X} \geq 16.4) = P\left(\frac{\overline{X} - \mu}{S/\sqrt{n}} \geq \frac{16.4 - 12.0}{2.1/\sqrt{16}}\right) \Rightarrow P(T \geq 8.38)$$

Buscando en la tabla de la distribución t con 15 grados de libertad, tenemos que para una probabilidad de 0.005 el respectivo valor de t es 2.947, lo cual implica que la probabilidad para t = 8.38 es cero). Por lo tanto, concluimos que la probabilidad de obtener una muestra con una media de 16.4 de una población cuya media es 12.0 es cero, es decir, que "el consumo promedio de gasolina no es 12 galones/hora", sino que es superior.

10. Distribución F

Es la distribución muestral aplicable para la relación de dos varianzas.

Teorema. Si U y W son dos variables aleatorias independientes, cada una con distribución Chi Cuadrado con ν_1 y ν_2 grados de libertad, respectivamente, entonces la distribución de la siguiente variable aleatoria

$$F = \frac{U/\nu_1}{W/\nu_2}$$

está dada por:

$$g(f) = \begin{cases} \dfrac{\Gamma\left(\dfrac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\dfrac{\nu_1}{2}\right)\Gamma\left(\dfrac{\nu_2}{2}\right)} \left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} f^{\,\nu_1/2-1} \left(1+\dfrac{\nu_1}{\nu_2} f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}, & f > 0 \\ 0 & \text{e o c} \end{cases}$$

y se denomina "distribución F con ν_1 y ν_2 grados de libertad" (ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador).

Notación. Usaremos la notación $F_{\nu1,\nu2,P}$ para denotar el valor de la distribución F con ν_1 grados de libertad en el numerador, ν_2 grados de libertad en el denominador y una probabilidad acumulada de P hacia la derecha (o una probabilidad de 1-P hacia la izquierda). Puede demostrarse que $F_{\nu1,\nu2,P} = \dfrac{1}{F_{\nu2,\nu1,1-P}}$, si se invierte la definición de la distribución F.

La aplicación principal para la cual se desarrolló la distribución F es la comparación de dos varianzas (de poblaciones normales).

Sea $X_{11}, X_{12},..., X_{1n_1}$ una muestra aleatoria (n_1) tomada de una población normal con varianza σ_1^2, y sea $X_{12}, X_{22},..., X_{2n_2}$ otra muestra aleatoria (n_2) tomada de una población normal con varianza σ_2^2. Si queremos realizar alguna inferencia sobre la igualdad o no de las varianzas, nos podemos basar en el hecho que las siguientes relaciones

$$X_1^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2} \qquad \text{y} \qquad X_2^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2}$$

son variables aleatorias con distribuciones Chi cuadrado con ν_1 y ν_2 grados de libertad, respectivamente, y con las cuales podemos construir la distribución F. El siguiente teorema clarifica este aspecto.

Teorema. Si S_1^2 y S_2^2 son las varianzas muestrales de dos variables aleatorias independientes de tamaños n_1 y n_2 tomadas de poblaciones normales con varianzas σ_1^2 y σ_2^2, entonces, la relación

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2\sigma_2^2}{S_2^2/\sigma_1^2}$$

tiene una distribución F con $n_1 - 1$ y n_2 -1 grados de libertad.

Demostración. Si llamamos U y W a los dos relaciones X_1^2 y X_2^2 dadas anteriormente, y aplicamos la definición de la distribución F, llegamos directamente al resultado deseado.

Tabulación. De nuevo, la función de distribución no puede calcularse en forma analítica; sin embargo, ha sido tabulada para diferentes valores de la probabilidad acumulada, y para varios grados de libertad en el numerador y en el denominador. Para cada valor de la probabilidad debe calcularse una tabla diferente. Los valores de las probabilidades dados en las tablas corresponden a las probabilidades de exceder los respectivos valores de F, es decir, presentan las colas a la derecha del valor respectivo de F. Las tablas están construidas bajo la suposición de que la distribución original de las variables aleatorias es normal.

11. Distribución de la diferencia entre dos medias

Sean X_1 y X_2 dos variables aleatorias con valores esperados μ_1 y μ_2 y varianzas σ_1^2 y σ_2^2, respectivamente. Por ejemplo, X_1 puede ser la duración de una batería para carro de una marca, y X_2 la duración de una batería de otra marca diferente. Si los medias μ_1 y μ_2 son desconocidas, podríamos estar interesados en conocer si ambas baterías tienen la misma duración media. En forma similar, si las varianzas son desconocidas, podríamos estar interesados en saber si son iguales o no. Para realizar estas inferencias, se pueden someter a pruebas idénticas diferentes baterías, controlando los factores externos, de tal forma que las diferencias se deban exclusivamente a la clase de marca probada

Inicialmente estaremos interesados en verificar si ambas distribuciones tienen la misma media poblacional, es decir si $\mu_1 = \mu_2$ ó equivalentemente $\mu_1 - \mu_2 = 0$.

Suponga que $X_{11}, X_{12}, \ldots, X_{1n_1}$ es una muestra aleatoria de tamaño n_1 tomada de una población con media μ_1 y varianza σ_1^2, y $X_{12}, X_{22}, \ldots, X_{2n_2}$ es otra muestra aleatoria de tamaño n_2 tomada de una población con media μ_2 y varianza σ_2^2. Si deseamos realizar alguna inferencia sobre $\mu_1 - \mu_2$, nos podemos basar en la distribución de la diferencia de las medias muestrales $\overline{X}_1 - \overline{X}_2$. Por el TCL sabemos que tanto $\overline{X}_1$ como $\overline{X}_2$ se distribuyen normalmente con los siguientes parámetros:

$$\overline{X}_1 \rightarrow \text{Normal}(\mu_1, \sigma_1^2 / n_1), \quad \overline{X}_2 \rightarrow \text{Normal}(\mu_2, \sigma_2^2 / n_2)$$

Ahora bien, para la diferencia de las medias muestrales $\overline{X}_1 - \overline{X}_2$ se tiene:

$$E(\overline{X}_1 - \overline{X}_2) = \mu_1 - \mu_2$$

$$V(\overline{X}_1 - \overline{X}_2) = V(\overline{X}_1) + V(\overline{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Para conocer la distribución muestral de las diferencias entre las medias se debe saber si las varianzas poblacionales son conocidas o desconocidas, y en caso de que sean desconocidas, se debe saber si son iguales o diferentes. Cada uno de estos tres casos se analizará por separado.

a) Distribución de la diferencia entre dos medias cuando las varianzas son conocidas.

Si las varianzas σ_1^2 y σ_2^2 son conocidas, tanto $\overline{X}_1$ como $\overline{X}_2$ se distribuyen normalmente. Por lo tanto la distribución de la diferencia entre las medias muestrales $\overline{X}_1 - \overline{X}_2$ es normal con el valor esperado y la varianza dados anteriormente, es decir,

$$\overline{X}_1 - \overline{X}_2 \rightarrow \text{Normal}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

De acuerdo con lo anterior la siguiente variable aleatoria tiene una distribución normal estándar:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Por lo tanto, con base en la expresión anterior se pueden realizar inferencias con respecto a la diferencia de medias poblacionales, bajo el supuesto de que las varianzas sean conocidas. Si además, son iguales, la expresión anterior se puede expresar como:

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

b) Distribución de la diferencia entre dos medias cuando las varianzas son desconocidas pero iguales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

Cuando las varianzas son desconocidas, se debe realizar previamente una prueba estadística para verificar si éstas son iguales o diferentes. Para realizar esta prueba debemos hacer uso de la distribución F para verificar si la relación de varianzas es igual a uno o diferente de uno.

Para cada una de las dos muestras se definen sus respectivas varianzas como:

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(X_{1j} - \overline{X}_1)^2, \qquad S_1^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(X_{2j} - \overline{X}_2)^2$$

Además $\dfrac{(n_1 - 1)S_1^2}{\sigma^2}$ y $\dfrac{(n_2 - 1)S_2^2}{\sigma^2}$ tienen distribuciones chi cuadrado con n_1 - 1 y n_2 -1 grados de libertad respectivamente. Por lo tanto su suma también sigue otra distribución chi cuadrado con $n_1 + n_2 - 2$ grados de libertad. Es decir:

$$Y = \frac{(n_1 - 1)S_1^2}{\sigma^2} + \frac{(n_2 - 1)S_2^2}{\sigma^2} \Rightarrow \chi^2(n_1 + n_2 - 2)$$

Ahora bien, si Z es una variable normal (0,1) y Y tiene una distribución chi cuadrado con ν grados de libertad, entonces la variable $T = \dfrac{Z}{\sqrt{Y/\nu}}$ tiene una distribución t con ν grados de libertad. Para nuestro caso la variable Z corresponde a la distribución de la diferencia de las dos medias, con varianzas conocidas, y la variable chi cuadrado corresponde a la variable Y acabada de definir. Por lo tanto

$$T = \frac{Z = \dfrac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}}{\sqrt{\left(\dfrac{(n_1 - 1)S_1^2}{\sigma^2} + \dfrac{(n_2 - 1)S_2^2}{\sigma^2}\right)/(n_1 + n_2 - 2)}} = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \approx t(n_1 + n_2 - 2)$$

donde $S_p^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ es un estimador ponderado de la varianza poblacional σ^2 obtenida ponderando las varianzas poblacionales por sus respectivos grados de libertad.

c) Distribución de la diferencia entre dos medias cuando las varianzas son desconocidas y diferentes ($\sigma_1^2 \neq \sigma_2^2$)

Cuando las varianzas son diferentes se puede demostrar que la siguiente variable aleatoria T sigue una distribución t con ν grados de libertad, donde

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \approx t_\upsilon$$

y el número de grados de libertad ν está dado por:

$$V = \frac{\left(S_1^2 / n_1 + S_2^2 / n_2\right)^2}{\frac{\left(S_1^2 / n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2 / n_2\right)^2}{n_2 - 1}}$$

Ejemplo. El gerente de una refinería piensa modificar el proceso para producir gasolina a partir de petróleo crudo. El gerente hará la modificación sólo si la gasolina promedio que se obtiene por este nuevo proceso (expresada como un porcentaje del crudo) aumenta su valor con respecto al proceso en uso. Con base en un experimento de laboratorio y mediante el empleo de dos muestras aleatorias de tamaño 12, una para cada proceso, la cantidad de gasolina promedio del proceso en uso es de 24.6 con una desviación estándar de 2.3, y para el proceso propuesto fue de 28.2 con una desviación estándar de 2.7. El gerente piensa que los resultados proporcionados por los dos procesos son variables aleatorias independientes normalmente distribuidas con varianzas iguales. Con base en esta evidencia, ¿debe adoptarse el nuevo proceso?

12. Problemas

Ejercicios varios[2]

5-65. Si X y Y son variables aleatorias normales independientes con *E(X)* **= 0,** *V(X) = 4, E(Y)* = 10 *y V(Y) = 9.* Calcule lo siguiente:

a. *E(2X +3Y)* *b.* *V(2X + 3Y)*
c. *P(2X +3Y < 30)* *d.* *P(2X + 3Y < 40)*

Respuestas : a) 30 b) 97c) 0.5 d) 0.846

5-66. Suponga que la variable aleatoria X representa la longitud, en pulgadas, de una pieza perforada. Sea *Y* la longitud de la pieza en milímetros. Si *E(X)* = 5 pulgadas, ¿cuál es la media de Y?

5-67. La envoltura de plástico para un disco magnético está formada por dos hojas. El espesor de cada una tiene una distribución normal con media 1.5 milímetros y desviación estándar de 0.1 milímetros. Las hojas son independientes.

a. Determine la media y la desviación estándar del espesor total de las dos hojas.
b. ¿Cuál es la probabilidad de que el espesor total sea mayor que 3.3 milímetros?
Respuestas :a) μ = 3 mm, σ = 0.1414 b) 0.017

5-68. El ancho del marco de una puerta tiene una distribución normal con media 24 pulgadas y desviación estándar de 1/8 de pulgada. El ancho de la puerta tiene una distribución normal con media 23 y7/8 pulgadas y desviación estándar de 1/16 pulgadas. Suponga independencia.

a. Determine la media y la desviación estándar de la diferencia entre el ancho del marco y el de la puerta.
b. ¿Cuál es la probabilidad de que la diferencia entre el ancho del marco y el de la puerta sea mayor que 1/4 de pulgada?
c. ¿Cuál es la probabilidad de que la puerta no quepa en el marco?

5-69. Un componente en forma de U está formado por tres piezas, *A, B* y *C.* La figura ilustra el componente. La longitud de *A* tiene una distribución normal con media de 10 milímetros y desviación estándar de 0.1 milímetros. El espesor de las piezas *B* y *C* está distribuido normalmente con media de 2 milímetros y desviación estándar de 0.05 milímetros. Suponga que todas las dimensiones son independientes.

a. Determine la media y la desviación estándar de la longitud del hueco *D.*
b. ¿Cuál es la probabilidad de que el hueco *D* sea menor que 5.9 milímetros?

[2] Ejercicios tomados de "Probabilidad y Estadística Aplicada la Ingeniería". Douglas C. Montgomery y George C. Runger. McGraw Hill, 1997.

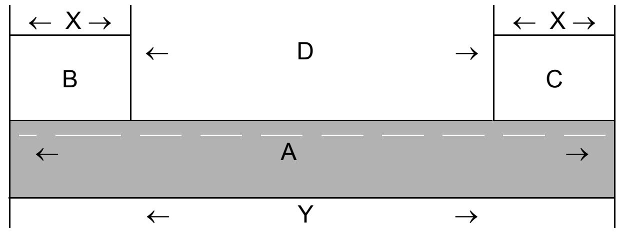


Figura 5-21 Figura para el ejercicio *5-69.*

Respuestas: a) E(D) =6 mm, σ_D = 0.1225 mm b) 0.206

5-70. El llenado de las latas de una bebida suave lo hace una máquina de llenado automática, con una desviación estándar de 0.5 onzas de líquido. Suponga además que los volúmenes con que se llenan las latas son variables aleatorias normales independientes.
a. ¿Cuál es la desviación estándar del volumen de llenado promedio de 100 latas?
b. Si el volumen de llenado promedio es 12.1 onzas, ¿cuál es la probabilidad de que el volumen de llenado promedio de 100 latas sea menor que 12 onzas de líquido?
e. ¿Cuál debe ser el valor del volumen de llenado promedio para que la probabilidad sea 0.005 de que el promedio en 100 latas sea menor que 12 onzas de líquido?

5-71 El espesor de la película fotoprotectora en un proceso de fabricación de semiconductores tiene una media de 10 micrómetros y una desviación estándar de 1 micrómetro. Supóngase que el espesor tiene una distribución normal, y que el espesor entre diferentes obleas es independiente.
a. Calcule la probabilidad de que espesor promedio de 10 obleas sea mayor que 11 o menor que 9 micrómetros.
b. Determine el número de obleas que es necesario medir para que la probabilidad sea 0.01 de que el espesor promedio sea mayor que 11 micrómetros.

Desigualdad de Chebychev y distribución normal[3]

5-72. El espesor de la película fotoprotectora en un proceso de fabricación de semiconductores tiene una media de 10 micrómetros y una desviación estándar de 1 micrómetro. Acote la probabilidad de que el espesor sea menor que 6 o mayor que 14 micrómetros.

5-73. Suponga que X tiene una distribución uniforme continua dentro del rango 0< x < 10. Utilice la regla de Chebychev para acotar la probabilidad de que X difiera de su media por más de dos desviaciones estándar y compare el resultado con el valor real de la probabilidad. Respuestas: 0.25 contra 0

5-74. Suponga que X tiene una distribución exponencial con media 20. Utilice la regla de Chebychev para acotar la probabilidad de que X difiera de su media por más de dos y tres desviaciones estándar, y compare los resultados con el valor real de la probabilidad en cada caso.

5-75. Suponga que X tiene una distribución Poisson con media λ. Utilice la regla de Chebychev para acotar la probabilidad de que X difiera de su media por más de dos y tres desviaciones estándar, y compare los resultados con el valor real de la probabilidad en cada caso.
Respuestas: ¼ y ½ contra 0.136 y 0.000046

5-76. Considere el proceso de taladrar agujeros en tarjetas de circuito impreso. Suponga que la desviación estándar de los diámetros es 0.01 y que éstos son independientes. Suponga además que se utiliza el promedio de 500 diámetros para estimar la media del proceso.
a. Si la probabilidad de que el promedio medido se encuentre dentro de alguna cota alrededor de la media del proceso, es al menos de 15/16, ¿cuál es el valor de la cota?
b. Si se supone que los diámetros tienen una distribución normal, determine una cota tal que la probabilidad de que el promedio medido esté más cercano a la media del proceso que la cota, sea al menos de 15/16.

5-97. El peso de un caramelo pequeño tiene una distribución normal con media 0.1 onzas desviación estándar de 0.01 onzas. Suponga que se colocan 16 caramelos en un paquete y que los pesos de éstos son independientes.
a. ¿Cuáles son la media y la varianza del peso neto del paquete?

[3] "Probabilidad y Estadística Aplicada la Ingeniería". Douglas C. Montgomery y George C. Runger. McGraw Hill, 1997.

b. ¿Cuál es la probabilidad de que el peso neto del paquete sea menor que 1.6 onza
c. Si se colocan 17 caramelos en cada paquete, ¿cuál es la probabilidad de que peso neto de un paquete sea menor que 1.6 onzas?. Respuestas: a) E(X) = 1.6, V(X) = 0.0016 b) 0.5 c) 0.008

5-98. El tiempo para que un sistema automatizado localice una pieza en un almacén, tiene u distribución normal con media de 45 segundos y desviación estándar de 30 segundos. Suponga que se hacen pedidos independientes por 10 piezas.
a. ¿Cuál es la probabilidad de que el tiempo promedio necesario para localizar las piezas sea mayor que 60 segundos?
b. ¿Cuál es la probabilidad de que el tiempo total necesario para localizar las piezas sea mayor que 600 segundos?

5-99. El ensamble mecánico empleado en el motor de un automóvil tiene cuatro componentes importantes. Los pesos de los componentes son independientes y están distribuidos normalmente con las siguientes medias y desviaciones estándar (en onzas):

componente	Media	Desviación estándar
tapa izquierda	4.0	0.4
tapa derecha	5.5	0.5
ensamble de cojinetes	10.0	0.2
ensamble de tornillos	8.0	0.5

a. ¿Cuál es la probabilidad de que el peso de un ensamble sea mayor que 29.5 onzas?
b. ¿Cuál es la probabilidad de que el peso promedio de ocho ensambles independientes sea mayor que 29 onzas? Respuestas: a) 0.008 b) 0

Ejercicios para TCL[4]

6-18. Se fabrica tubería PVC con un diámetro promedio de 1.01 in y desviación estándar de 0.003 in. Encuentre la probabilidad de que en una muestra aleatoria de *n* =9 secciones de tubería, el diámetro promedio de la muestra sea mayor que 1.009 in y menor que 1.012 in.

6-19. Suponga que se toman muestras aleatorias de tamaño *n* = 25 de una población normal con media 100 y desviación estándar 10. ¿Cuál es la probabilidad de que la media muestral se encuentre dentro del intervalo de -1.8 y +1 desviaciones estándares ? Respuestas: 0.805

6-20. En la fabricación de una alfombra se utiliza una fibra sintética con una resistencia a la tensión que tiene una distribución normal con media 75.5 psi y desviación estándar 3.5 psi. Encuentre la probabilidad de que en una muestra aleatoria de *n* = 6 especímenes de fibra, la media de la resistencia a la tensión en la muestra sea mayor que 75.75 psi.

6-21. Considere la fibra sintética del ejercicio anterior. ¿Cómo cambia la desviación estándar de la media muestral cuando el tamaño de la muestra aumenta desde *n* = 6 hasta *n* = 49?

6-22. La resistencia a la compresión del concreto tiene una media de 2500 psi y una desviación estándar de 50 psi. Encuentre la probabilidad de que la media muestral de una muestra aleatoria de *n* = 5 especímenes esté en el intervalo de 2499 a 2510 psi.

6-23. Considere los especímenes de concreto del ejemplo anterior. ¿Cuál es el error estándar de la media muestral? Respuestas: 22.361

6-24. Una población normal tiene una media de 100 y una varianza de 25. ¿De qué tamaño debe ser la muestra aleatoria que se tome de esta población para que el error estándar del promedio de la muestra sea 1.5?

6-25. Suponga que la variable aleatoria X tiene la distribución uniforme continua

$$f(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0 & \text{en cualquierotro caso} \end{cases}$$

Suponga que se toma una muestra aleatoria de *n* = 12 observaciones de esta distribución. ¿Cuál es la distribución de probabilidad de $X - 6$?. Encuentre la media y la varianza de esta cantidad. Respuestas:-11/2, 1/144

[4] "Probabilidad y Estadística Aplicada la Ingeniería". Douglas C. Montgomery y George C. Runger. McGraw Hill, 1997.

6-26. Suponga que X tiene una distribución uniforme discreta

$$f(x) = \begin{cases} 1/3, & x = 1,2,3 \\ 0 & \text{en cualquierotro caso} \end{cases}$$

De esta población se toma una muestra aleatoria de tamaño n = 36. Encuentre la probabilidad de que la media muestral sea mayor que 2.1 pero menor que 2.5. Suponga que la media muestral puede medirse hasta la décima más cercana.

6-27. El tiempo que un pasajero invierte esperando en un punto de revisión de un aeropuerto es una variable aleatoria con media de 8.2 minutos y desviación estándar de 1.5 minutos. Suponga que se observa una muestra aleatoria de *n* =49 pasajeros. Encuentre la probabilidad de que el tiempo de espera promedio en la fila para estos clientes sea

a. Menor que 10 minutos
b. Entre 5 y 10 minutos
c. Menor que 6 minutos
Respuestas: a) 1 b) 1 c) 0

6-28. Se toma una muestra aleatoria de tamaño n = 16 de una población normal que tiene una media de 75 y una desviación estándar de 8. De otra población normal se toma una muestra aleatoria de tamaño n_2 = 9; esta población tiene una media de 70 y una desviación estándar de 12. Sean $\overline{X}_1$ y $\overline{X}_2$ las medias de cada muestra, respectivamente. Encuentre

a. La probabilidad de que $\overline{X}_1 - \overline{X}_2$ sea mayor que cuatro.
b. La probabilidad de que $3.5 \le \overline{X}_1 - \overline{X}_2 \le 5.5$.

6-29. Una compañía que vende artículos electrónicos compara la brillantez de dos tipos diferentes de cinescopios para su uso en televisores. El cinescopio de tipo A tiene una brillantez promedio de 100 con una desviación estándar de 16, mientras que el cinescopio de tipo B tiene una brillantez promedio desconocida, pero se supone que la desviación estándar es la misma que la del cinescopio de tipo A. Se toma una muestra aleatoria de *n* = 25 cinescopios de cada tipo y se calcula $\overline{X}_A - \overline{X}_B$. Si μ_B es igual o mayor que μ_A el fabricante adoptará el cinescopio de tipo B para utilizarlo en los televisores que fabrica. La diferencia observada es $\overline{X}_A - \overline{X}_B = 3.5$. ¿Qué decisión tomará el fabricante y por qué?

6-30. La elasticidad de un polímero es afectada por la concentración de un reactivo. Cuando se utiliza una concentración baja, la elasticidad promedio verdadera es 55, mientras que cuando se emplea una concentración alta, la elasticidad promedio es 60. La desviación estándar de la elasticidad es 4, sin importar cuál sea la concentración. Si se toman dos muestras aleatorias de tamailo 16, encuentre la probabilidad de que $\overline{X}_{alta} - \overline{X}_{baja} > 2$

Ejercicios par TCL y distribución de diferencias[5]

1. Verifique la siguiente fórmula de cálculo del valor de la varianza de una muestra:

$$S^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

5. Si $X_1, X_2, \ldots, X_n$ son variables aleatorias independientes que tienen distribuciones de Bernoulli idénticas con el parámetro θ,entonces $\overline{X}$ es la proporción de aciertos en *n* ensayos, que representamos como $\hat{\theta}$. Verifique que $E(\hat{\theta}) = \theta$, $Var(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$

8 La que sigue es una condición suficiente para el teorema de límite central: *si las variables aleatorias* X_1,

[5] “Estadística Matemática con Aplicaciones”. John E. Freund y Ronald E. Walpole. Prentice Hall Hispanoamericana S.A., *página 287-289*

X_2, ...,X_n son *independientes y están uniformemente limitadas (es decir, hay una constante k positiva tal que la probabilidad de que cualquiera de los términos* X_i *tome un valor mayor que k o menor que -k sea* 0), *entonces si la varianza de* $Y_n = X_1 + X_2 + \ldots + X_n$ *se vuelve infinita cuando* $n \to \infty$*, la distribución de la media estandarizada de las* X_i *tiende a la distribución normal estándar.* Demuestre que esta condición suficiente se cumple para una secuencia de variables aleatorias independientes X_i que tienen las distribuciones de probabilidad:

$$f_i(x_i) = \begin{cases} \frac{1}{2} & para\ x_i = 1 - (\frac{1}{2})^i \\ \frac{1}{2} & para\ x_i = (\frac{1}{2})^i - 1 \end{cases}$$

13. Si se selecciona una muestra aleatoria de tamaño *n* de la población finita que consta de los primeros *N* enteros positivos, demuestre que
    a) la media de la distribución de $\overline{X}$ es $\frac{(N+1)}{2}$
    b) la varianza de la distribución de $\overline{X}$ es $\frac{(N+1)(N-n)}{12n}$
    c) la media y la varianza de la distribución de Y = *n* $\overline{X}$ son

$$E(Y) = \frac{(N+1)}{2} \quad \text{y} \quad V(Y) = \frac{n(N+1)(N-n)}{12}$$

Distribución normal y aproximaciones[6]

14. ¿Cuántas muestras diferentes de tamaño *n* = 3 se pueden tomar de una población finita de tamaño a) *N* = 12; b) *N* = 20; c) *N* = 50?

15. Si se toma una muestra aleatoria de tamaño *n* = 4 de una población finita de tamaño *N* = 200, ¿cuál es la probabilidad de cada muestra posible? (R/0.000000015)

16 Si se toma una muestra aleatoria de tamaño *n* = 3 de una población finita de tamaño *N* = 50, ¿cuál es la probabilidad de que un elemento en particular de la población sea incluido en la muestra?

17. En relación con una muestra aleatoria de una población infinita, ¿qué le sucede al error estándar de la media si el tamaño de la muestra
    a) se aumenta de 30 a 120;
    b) se aumenta de 80 a 180;
    c) se disminuye de 450 a 50;
    d) se disminuye de 250 a 40?
    R/a) Se divide entre 2 b) se divide entre 1.5 c) se multiplica por 3 d) se multiplica por 2.5)

18. Determine el valor del factor de corrección de la población finita para
    a) n = 5 y *N* = 200;
    b) n = 50 y N = 300;
    c) n = 200 y N = 800.

19. De una población infinita con la media μ = 75 y la varianza σ^2 = 256, se toma una muestra al azar de tamaño *n* = 100. Si aplicamos el teorema de Chebyshev, ¿con qué probabilidad podemos afirmar que el valor que se obtenga de $\overline{X}$ estará entre 67 y 83? (R/ Cuando menos 24/25)

20 Utilice el teorema de límite central en vez del de Chebyshev para resolver otra vez el ejercicio anterior.

21 De una población infinita con la media μ = 128 y la desviación estándar σ = 6.3, se toma una muestra aleatoria de tamaño *n* = 81. Con qué probabilidad podemos afirmar que el valor que se obtenga de $\overline{X}$ no estará entre 126.6 y 129.4, si utilizamos
    a) el teorema de Chebyshev;

[6] “Estadística Matemática con Aplicaciones”. John E. Freund y Ronald E. Walpole. Prentice Hall Hispanoamericana S.A., *página 289-291.*

b) el teorema de límite central.
(R/ a) Cuando mas 1/4 b) 0.0456

22. Resuelva otra vez el inciso b) del ejercicio anterior, suponiendo que la población no es infinita, sino finita y de tamaño $N = 400$.

23. De una población normal con $\mu = 51.4$ y $\sigma = 6.8$, se toma una muestra al azar de tamaño 64. ¿Cuál es la probabilidad de que la media de la muestra
a) exceda de 52.9;
b) esté entre 50.5 y 52.3;
c) sea menor que 50.6?
(R/ a) 0.0388 b) 0.7100 c) 0.1736

24. De una población normal con $\sigma = 25$, se toma una muestra aleatoria de tamaño 100. ¿Cuál es la probabilidad de que la media de la muestra difiera de la media de la población en 3 o más en una u otra forma?

25. De cada una de dos poblaciones con medias iguales y las desviaciones estándar $\sigma_1 = 20$ y $\sigma_2 = 30$, se toman muestras aleatorias independientes de tamaño 400. Mediante el uso del teorema de Chebyshev, ¿qué podemos afirmar, con una probabilidad de cuando menos 0.99, acerca del valor que obtendremos de $\bar{X}_1 - \bar{X}_2$?. (R/ El valor está entre -18.25 y 18.025)

26. Suponiendo que las dos poblaciones del ejercicio anterior son normales, utilice el resultado del ejercicio 4 para obtener *k* al que $P(-k < \bar{X}_1 - \bar{X}_2 < k) = 0.99$.

27. De dos poblaciones normales con las medias $\mu_1 = 78$ y $\mu_2 = 75$, y las varianzas $\sigma_1^2 = 150$ y $\sigma_2^2 = 200$, se toman muestras aleatorias independientes de tamaño $n_1 = 30$ y $n_2 = 50$. Use los resultados del ejercicio 4 para obtener la probabilidad de que la media de la primera muestra excederá la de la segunda muestra cuando menos en 4.8. (R/ 0.05)

28. La proporción real de familias de cierta ciudad, que son dueñas, no arrendatarios, de su casa es 0.70. Si al azar se entrevistan a 84 familias de esta ciudad y sus respuestas a la pregunta de si son dueñas o no de su casa- se consideran valores de variables aleatorias independientes que tienen distribuciones de Bernoulli idénticas con el parámetro $\theta=0.70$, ¿con qué probabilidad podemos afirmar que el valor que se obtenga de la proporción de la muestra -$\hat{\theta}$ - estará entre 0.64 y 0.76, usando
a) el teorema de Chebyshev;
b) el teorema de límite central?

29. La proporción real de hombres partidarios de cierta propuesta de impuestos es 0.40 y la proporción correspondiente de mujeres es 0.25; $n_1 = 500$ hombres y $n_2 = 400$ mujeres son entrevistados al azar y sus respuestas individuales se consideran como los valores de variables aleatorias independientes que tienen distribuciones de Bernoulli con los parámetros respectivos $\theta_1 = 0.40$ y $\theta_2 = 0.25$. ¿Qué podemos afirmar, según el teorema de Chebyshev, con una probabilidad de cuando menos 0.9975 acerca del valor que obtendremos para $\hat{\theta}_1 - \hat{\theta}_2$, la diferencia entre dos proporciones de la muestra de respuestas favorables? Utilice el resultado del ejercicio 6. (R/ t = -1.35; los datos fundamentan la afirmación)

Ejercicios Distribuciones Ji-Cuadrado, t y F[7]

6-31. Para una distribución ji-cuadrada, encuentre los siguientes valores:

a. $\chi^2_{0.95,8}$ **b.** $\chi^2_{0.0.5,10}$ c. $\chi^2_{0.99,18}$

Respuestas: a) 2.73 b) 9.34 c) 23.8

6-32. Para una distribución ji-cuadrada, encuentre los siguientes valores:

a. $\chi^2_{00.025,10}$ **b.** $\chi^2_{0.01,15}$ c. $\chi^2_{0.99,18}$

[7] "Probabilidad y Estadística Aplicada la Ingeniería". Douglas C. Montgomery y George C. Runger. McGraw Hill, 1997.

6-33. Para una distribución ji-cuadrada, encuentre $\chi^2_{\alpha,v}$ tal que

a. $P(X_{10} \le \chi^2_{\alpha,10}) = 0.975$ b. $P(X_{15} \le \chi^2_{\alpha,15}) = 0.025$

c. $P(26.296 \le X_{16} \le \chi^2_{\alpha,16}) = 0.045$

Respuestas: a)20.48 b) 6.26 c) 34.3

6-34. Para una distribución ji-cuadrada, encuentre $\chi^2_{\alpha,v}$ tal que

a. $P(X_5 \le \chi^2_{\alpha,5}) = 0.95$ b. $P(X_{10} \le \chi^2_{\alpha,10}) = 0.20$

c. $P(12.549 \le X_{10} \le \chi^2_{\alpha,10}) = 0.20$

6-35. Se toma una muestra aleatoria de *n* =25 observaciones de una población normal que tiene una varianza σ^2 = 10. Encuentre la probabilidad de que la varianza de la muestra sea mayor que 16.4. R/ 0.025

6-36. Para una distribución t, encuentre los siguientes valores:
a. $t_{0.025,10}$ b. $t_{0.015,15}$ c. $t_{0.01,20}$

6-37. Para una distribución t, encuentre los siguientes valores:
a. $t_{0.01,10}$ b. $t_{0.05,10}$ c. $t_{0.01,11}$
Respuestas: a) 2.76 b) 1.73 c) 2.72

6-38. Para una distribución t, encuentre $t_{\alpha,v}$ tal que
a. $P(T_{10} \le t_{\alpha,10}) = 0.95$ b. $P(T_{15} \le t_{\alpha,15}) = 0.01$ c. $P(T_8 > t_{\alpha,8}) = 0.90$

6-39. Para una distribución t, encuentre $t_{\alpha,v}$ tal que
a. $P(T_{20} \le t_{\alpha,20}) = 0.90$ b. $P(T_{15} > t_{\alpha,15}) = 0.95$ c. $P(1.476 \le T_8 \le t_{\alpha,5}) = 0.075$
Respuestas: a) 1.325 b) –1.753 c) 2.571

6-40. Una población normal tiene una media conocida igual a 10, y una varianza desconocida. De esta población se toma una muestra aleatoria de tamaño 25. Los resultados en la muestra son una media de 11 y una desviación estándar muestral de 4.2. ¿Cuán inusuales son estos resultados?

6-41. Para una distribución *F,* encuentre lo siguiente:
a. $f_{0.025,4,9}$ b. $f_{0.05,15,10}$ c. $f_{0.95,6,8}$ d. $f_{0.90,24,24}$
Respuestas: a) 1.63 b) 2.85 c) 0.241 d) 0.588

6-42. Si S_1^2 *y* S_2^2 son las varianzas muestrales de muestras aleatorias independientes de tamaños $n_1 = 10$ y $n_2 = 20$, tomadas de poblaciones normales que tienen las mismas varianzas, encuentre $P(S_1^2/S_2^2 \le 2.42)$.

6-43. Suponga que se tiene una variable aleatoria distribuida normalmente con media μ y varianza σ^2, y que de esta distribución se toma una muestra aleatoria de cinco observaciones. ¿Cuál es la función de densidad de probabilidad conjunta de la muestra?

6-44. Los transistores tienen una vida útil que está distribuida de manera exponencial, con parámetro λ. Se toma una muestra aleatoria de *n* transistores. ¿Cuál es la función de densidad de probabilidad conjunta de la muestra?

6-45. Suponga que X tiene una distribución uniforme desde 0 hasta 1. Se toma una muestra aleatoria de tamaño 4 de X. ¿Cuál es la función de densidad de probabilidad conjunta de la muestra?

6-46. Un especialista en adquisiciones compra 25 resistores del vendedor 1, y 30 del vendedor 2. Sean $\mathbf{X_{1,1}}$, $X_{1,2}$,. .. , $X_{1,25}$ las resistencias observadas del vendedor 1, las cuales se supone que están distribuidas de manera normal e independiente, con media 100 Ω y desviación estándar 1.5 Ω. De manera similar; sean $X_{2,1}$, $X_{2,2}$,... , $X_{2,30}$ las resistencias observadas del vendedor 2, las cuales se supone que están distribuidas de manera normal e independiente, con media 105 Ω y desviación estándar 2.0 Ω. ¿Cuál es la distribución de muestreo de $\overline{X}_1 - \overline{X}_2$

6-47. Considere el problema de los resistores del ejercicio 6-46. ¿Cuál es el error estándar de $\overline{X}_1 - \overline{X}_2$?
Respuestas: 0.4726

6-48. De una distribución normal con media 50 y desviación estándar 12, se toma una muestra aleatoria de 36 observaciones. Encuentre la probabilidad de que la media muestral esté en el intervalo 47 < $\overline{X}$ < 53.

6-49. ¿Es importante la hipótesis de normalidad en el ejercicio 6-48 ¿Por qué? R/ No

6-50. Se realizan pruebas en una muestra aleatoria de *n* =9 elementos estructurales para determinar su resistencia a la compresión. Se sabe que la media verdadera de la resistencia a la compresión es μ = 5500 psi y que la desviación estándar es σ = 100 psi. Encuentre la probabilidad de que la resistencia promedio a la compresión de la muestra sea mayor que 4985 psi.

6-51. Una población normal tiene una media conocida de 50 y una varianza desconocida. De esta población se toma una muestra aleatoria de tamaño *n* = 16; los resultados obtenidos de la muestra son $\overline{x}$ = 52 y *s* = 1.5. ¿Cuán inusuales son estos resultados?

6-52. Se toma una muestra aleatoria de tamaño *n* = 16 de una población normal que tiene una varianza σ^2 = 5. Encuentre la probabilidad de que la varianza de la muestra sea menor o igual que 7.44.

6-53. Un fabricante de dispositivos semiconductores toma una muestra aleatoria de 100 chips y los prueba; cada chip se clasifica como defectuoso o no defectuoso. Sea X_i = 0 si el chip no es defectuoso, y X_i = 1 si el chip es defectuoso. La fracción de chips defectuosos de la muestra es

$$\hat{p} = \frac{X_1 + X_2 + \ldots + X_{100}}{100}$$

¿Cuál es la distribución de muestreo de la variable aleatoria $\hat{p}$? R/ Aproximadamente normal

Distribuciones Ji Cuadrada, t y F[8]

3. Verifique la identidad

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

utilizada en clase

5. Si el intervalo de X es el conjunto de todos los números reales positivos, demuestre que para *k* > O

a) la probabilidad de que la variable aleatoria $\sqrt{2X} - \sqrt{2n}$ tome un valor menor que *k* es igual a la probabilidad de que la variable aleatoria $\frac{X-n}{\sqrt{2n}}$ tome un valor menor que $k + \frac{k^2}{2\sqrt{2n}}$

b) Si X tiene una distribución ji cuadrada con n grados de libertad, entonces para *n* grande la distribución de $\sqrt{2X} - \sqrt{2n}$ puede determinarse, aproximadamente, con la distribución normal estándar.

Utilice también el resultado del inciso b) y el teorema 8.10 para demostrar que paran grande la varianza de la distribución muestral de S es, aproximadamente, $\frac{\sigma^2}{2(n-1)}$

6. Determine valores aproximados de la probabilidad de que una variable aleatoria X que tiene una distribución ji cuadrada con 50 grados de libertad, tome un valor mayor que 68.0

a) tratando a $\frac{X-\nu}{\sqrt{2\nu}}$ con ν = 50 como una variable aleatoria con distribución normal estándar;

[8] “Estadística Matemática con Aplicaciones”. John E. Freund y Ronald E. Walpole. Prentice Hall Hispanoamericana S.A., *página 302-306*

b) tratando a $\sqrt{2X} - \sqrt{2\nu}$ con $\nu \geq 50$ como una variable aleatoria con distribución normal estándar.
Asimismo, juzgue los méritos de estas aproximaciones, dado que el valor real de la probabilidad redondeado a cinco decimales) es 0.04596.

9. Si *x* y y son independientes, x tiene una distribución normal con μ = 5 y la varianza σ^2 = 15 y Y tiene una distribución ji cuadrada con cinco grados de libertad, determine *P(X - 5* > 3√Y) (R/ 0.0763)

11. Demuestre que para ν > 2 la varianza de la distribución t con ν grados de libertad es $\dfrac{\nu}{\nu - 2}$

12. Utilice la fórmula de Stirling, del ejercicio 3 de la página 19, para demostrar que cuando $\nu \to \infty$ la distribución t se aproxima a la distribución normal estándar. (Fórmula de Stirloing $n! = \sqrt{2\pi n}\left(\dfrac{n}{e}\right)^n$ donde e es la base de los logaritmos naturales.

14. Pruebe que la distribución *t* con un grado de libertad es una distribución de Cauchy.

15. Demuestre que la distribución *F* con 4 y 4 grados de libertad está dada por

$$g(F) = \begin{cases} 6F(1+F)^{-4} & \textit{para } F > 0 \\ 0 & \textit{en cualquier otra parte} \end{cases}$$

y utilice esta densidad para obtener la probabilidad de que para muestras aleatorias independientes de tamaño 5, tomadas de dos poblaciones normales con la misma varianza, S_1^2 / S_2^1 tome un valor menor que ½ o mayor que 2. (R/ 14/27)

16. Si X tiene una distribución *F* con ν_1 y ν_2 grados de libertad, demuestre que Y = $\dfrac{1}{X}$ tiene una distribución *F* con ν_2 y ν_1 grados de libertad. (Este teorema es de importancia en la obtención de valores de *F* que no aparecen en la tabla.

17. Utilice el resultado del ejercicio 16 para probar que $F_{1-\alpha,\nu_1,\nu_2} = \dfrac{1}{F_{\alpha,\nu_2,\nu_1}}$

18. Pruebe que para ν_2 > 2 la media de la distribución *F* es $\dfrac{\nu_2}{\nu_2 - 2}$, e Investigue lo que sucede cuando ν_2 es igual a 1 y 2.

19. Si X tiene una distribución beta con $\alpha = \dfrac{\nu_1}{2}$ y $\beta = \dfrac{\nu_2}{2}$, demuestre que

$$F = \frac{\nu_2 X}{\nu_2 (1 - X)}$$

tiene una distribución *F* con ν_1 y ν_2 grados de libertad.

20. Si X tiene una distribución *F* con ν_1 y ν_2 grados de libertad, pruebe que cuando $\nu_2 \to \infty$ la distribución de $\nu_1 X$ tiende a la distribución ji cuadrada con ν_1 grados de libertad.

21. Demuestre que si X tiene la distribución t con ν grados de libertad, entonces X^2 tiene la distribución *F* con 1 y ν grados de libertad.

22. Integre la densidad ji cuadrada apropiada para obtener la probabilidad de que la varianza de una muestra aleatoria de tamaño 5 de una población normal con σ^2 = 25 quede entre 20 y 30.

23. La afirmación de que la varianza de una población normal es $\sigma^2 = 25$ se rechazará si la varianza de una muestra aleatoria de tamaño 16 excede 54.668 o es menor que 12.102. ¿Cuál es la probabilidad de que se rechace esta afirmación aunque $\sigma^2 = 25$? *(R/ 0.055)*

24. La afirmación de que la varianza de una población normal es $\sigma^2 = 4$ se rechazará si la varianza de una muestra aleatoria de tamaño 9 excede 7.7535. ¿Cuál es la probabilidad de que se rechace esta afirmación aunque $\sigma^2 = 4$?

25. Una muestra aleatoria de tamaño *25,* tomada de una población normal, tiene la media $\bar{x} = 47$ y la desviación estándar s = 7. Basando nuestra decisión en el valor estadístico t ¿podemos decir que la información dada soporta la conjetura de que la media de la población es $\mu = 42$? (R/ 3.57, los datos no soportan la conjetura)

26 Tomada de una población normal, una muestra aleatoria de tamaño 12 tiene la media $\bar{x} = 27.8$ y la varianza $s^2 = 3.24$. Basando nuestra decisión en el valor estadístico t, ¿podemos decir que la información dada soporta la afirmación de que la media de la población es $\mu = 28.5$?

27. Si s^2_1 y s_2^2 son las varianzas de variables aleatorias independientes de tamaño $n_1 = 61$ y $n_2 = 31$ de poblaciones normales con $\sigma^2_1 = 12$ y $\sigma^2_2 = 18$, determine $P(s^2_1/s_2^2 > 1.16)$. (R/ 0.05)

28. Si s^2_1 y s_2^2 son las varianzas de variables aleatorias independientes de tamaño $n_1 = 10$ y $n_2 = 15$ tomadas de poblaciones normales con varianzas iguales, obtenga $P(s^2_1/s_2^2 < 4.03)$.

DISTRIBUCIONES MUESTRALES

1) Distribución de la media: a) $Z = \dfrac{(\overline{X} - \mu)}{\sigma / \sqrt{n}} \rightarrow N(0,1)$, b) $T = \dfrac{(\overline{X} - \mu)}{S / \sqrt{n}} \rightarrow t_{n-1}$

2) Distribución de la proporción P ~ $\text{Normal}\left(\theta, \dfrac{\theta(1-\theta)}{n}\right)$. Ó $Z = \dfrac{P - \theta}{\sqrt{\dfrac{\theta(1-\theta)}{n}}} \rightarrow N(0,1)$

3) Diferencia de proporciones: $P_1 - P_2 \rightarrow \text{Normal}\left(\theta_1 - \theta_2, \dfrac{\theta_1(1-\theta_1)}{n_1} + \dfrac{\theta_2(1-\theta_2)}{n_2}\right)$.

4) **Distribución Chi Cuadrado:** $f(x) = \begin{cases} \dfrac{1}{2^{\upsilon/2}\Gamma(\upsilon/2)} x^{\upsilon/2-1} e^{x/2}, & x \geq 0 \\ 0 \quad , & x < 0 \end{cases}$

ν = número de grados de libertad, o "grados de libertad". $X^2 = \dfrac{(n-1)S^2}{\sigma^2} \rightarrow \chi^2_{n-1}$

5) Distribución t ó de Student, con ν grados de libertad.

$$T = \frac{Z}{\sqrt{Y/\nu}}, \quad f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \qquad -\infty < t < +\infty$$

6) Distribución F (ν_1 grados de libertad en el numerador y ν_2 grados de libertad en el denominador).

U y W son dos va´s independientes con distribuciones Chi Cuadrado con ν_1 y ν_2 grados de libertad, $F = \dfrac{U/\nu_1}{W/\nu_2}$,

$$g(f) = \begin{cases} \dfrac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\dfrac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} f^{\,\nu_1/2-1} \left(1 + \dfrac{\nu_1}{\nu_2} f\right)^{-\frac{1}{2}(\nu_1+\nu_2)}, & f > 0 \\ 0 & \text{e o c} \end{cases}$$

$F = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \dfrac{S_1^2\sigma_2^2}{S_2^2/\sigma_1^2}$ tiene una distribución F con $n_1 - 1$ y $n_2 - 1$ grados de libertad.

7) Distribución de la diferencia entre dos medias

d) Varianzas conocidas:

$$\overline{X}_1 - \overline{X}_2 \rightarrow \text{Normal}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right), \quad Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \rightarrow N(0,1)$$

e) Varianzas desconocidas pero iguales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \approx t(n_1 + n_2 - 2), \quad S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

f) Varianzas desconocidas y diferentes ($\sigma_1^2 \neq \sigma_2^2$)

$$T = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \approx t_\upsilon, \quad V = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{\dfrac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$