Capítulo 4 VARIABLES ALEATORIAS

En la investigación econométrica es de interés una clase especial de variables que se denominan variables aleatorias. En principio, se dará un tratamiento general al término variable aleatoria, más adelante, se estudiarán variables aleatorias específicas de los modelos econométricos. Asimismo, se presentan las distribuciones de probabilidad de las variables aleatorias, necesarias para poder realizar las estimaciones y contrastes estadísticos de los parámetros poblacionales.

4.1. Modelos Probabilísticos

En el capítulo anterior quedó establecido el insumo básico del proceso de investigación econométrica: el planteo de una tabla de datos.

Suponiendo que un experimento consista en observar la variable ingreso (X) y otras dos variables cualesquiera (Y, Z) sobre un conjunto de **n** individuos seleccionados al azar de una población de tamaño **N**. El investigador está interesado en conocer la distribución de los ingresos anuales de la población y estudiar si las otras dos variables determinan las variaciones de dichos ingresos, aunque esto, por ahora no será tema de estudio. Para ello plantea una tabla de datos como se muestra en la figura 4.1.

	X	Y	Z
1	x_1	y_1	z_1
2	x_2	y_2	z_2
3	x_3	y_3	z_3
⋮	⋮	⋮	⋮
i	⋮	⋮	⋮
⋮	⋮	⋮	⋮
n	x_n	y_n	z_n
Sumas	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} z_i$
Media aritmética	$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\bar{Y} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\bar{Z} = \frac{\sum_{i=1}^{n} x_i}{n}$

Figura 4.1. Tabla de datos

Para conocer la probabilidad de que un individuo posea ingresos menores a determinada cantidad, el investigador debe transformar la tabla de datos de la siguiente manera:

1) Ordenar el recorrido de la variable aleatoria X de menor a mayor
2) Agrupar los individuos por intervalos de clase; en este caso particular por deciles de ingresos, de tal forma que el primer grupo esté conformado por la cantidad de individuos que posean el 10% de los ingresos totales calculados por $n\bar{X}$.

Por lo tanto, la tabla de datos quedará dispuesta de la siguiente manera:

	Deciles	Punto Medio	Cantidad de individuos	Frecuencia Relativa
Grupo 1	1 decil	x_1	n_1	n_1/n
Grupo 2	2 decil	x_2	n_2	n_2/n
Grupo 3	3 decil	x_3	n_3	n_3/n
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
Grupo 10	10 decil	x_{10}	n_{10}	n_{10}/n
Sumas			n	1

Figura 4.2 Intervalo de clases

Donde ahora

$$x_j;\ \forall j=1,\ldots,10$$

representa el promedio de los ingresos de los individuos que conforman el j-ésimo decil, esto es:

$$x_j=\frac{0{,}10n\bar{X}}{n_j}$$

El *valor probable* de una unidad de observación correspondiente a una variable responde a un *modelo matemático probabilístico* (no determinístico o estocástico); dicho valor es una variable aleatoria. Esto es un principio básico de la Econometría Aplicada; si fuera no estocástico, los instrumentos de la econometría no se aplicarían ya que se especificaría un modelo determinístico y no tendría sentido la corroboración de teorías económicas en espacio y tiempo determinado. Se verá que una parte del modelo econométrico –la representada por variables exógenas – se considerará no estocástica; pero esto hace a la naturaleza misma del modelo probabilístico que se desarrolla en este acápite, ya que una vez ejecutado el experimento el *valor probable* se transforma en un *valor exacto* para una

unidad de observación y será un valor que no cambia si se mantienen las condiciones en que se llevó a cabo el experimento. Como se verá, esto se desprende del concepto de variable aleatoria.

En este manual se tratará un tipo especial de fenómeno. Para investigarlo se formulará un modelo matemático. Al principio es importante distinguir entre el fenómeno observable en sí mismo y el modelo matemático para dicho fenómeno. Evidentemente, no se puede influir sobre lo que se observa; sin embargo, al elegir un modelo, sí se puede aplicar un juicio crítico. Neyman (1954) –University of California Publications in Statistics. Vol I, University of California Press– escribió: "Cada vez que utilizamos las matemáticas con el objeto de estudiar fenómenos observables es indispensable empezar por construir un modelo matemático (determinístico o probabilístico) para esos fenómenos. Necesariamente, este modelo debe simplificar las cosas y permitir la omisión de ciertos detalles. El éxito del modelo depende de si los detalles que se omitieron tienen o no importancia en el desarrollo de los fenómenos estudiados. La solución del problema matemático puede ser correcta y aun así estar muy en desacuerdo con los datos observados, debido sencillamente a que no estaba probada la validez de las suposiciones básicas que se hicieron. Corrientemente, es bastante difícil afirmar con certeza si un modelo matemático es adecuado o no, antes de obtener algunos datos mediante la observación. Para verificar la validez del modelo, debemos deducir un cierto número de consecuencias del mismo y luego comparar con las observaciones esos resultados predichos".

En la naturaleza, hay ejemplos de experimentos para los cuales los modelos determinísticos son apropiados. Así, las leyes gravitacionales describen exactamente lo que sucede a un cuerpo que cae bajo ciertas condiciones. Pero también, las identidades contables en un modelo económico, definen una relación que determina unívocamente la cantidad del primer miembro de la ecuación si se dan las del segundo miembro.

En muchos casos el modelo matemático determinístico descrito anteriormente es suficiente. Sin embargo, hay fenómenos que necesitan un modelo matemático distinto para su investigación. Esos son los llamados modelos no determinísticos o probabilísticos. En este tipo de modelos, las condiciones experimentales determinan el comportamiento probabilístico de los resultados observables (más específicamente, la distribución de probabilidad); mientras que, en los modelos determinísticos se supone que el resultado real (sea numérico o no) está determinado por las condiciones bajo las cuales se efectúa el experimento o procedimiento.

Retomando el ejemplo del ingreso, la distribución de probabilidad será una función P=f(X). Donde P depende de los valores que asume X, teniendo en cuenta que dichos valores ocurren con determinada frecuencia. En este sentido, P es función del recorrido de X; es decir, dado Rx, P asume valores entre 0 y 1, representados en este caso como frecuencias relativas. Así como hay funciones lineales o cuadráticas hay funciones de probabilidad y dentro de ellas la binomial, la poisson y la normal. La figura 4.3 muestra una generalización de la función para la distribución de probabilidad de los deciles de ingresos.

Grupo	Punto medio X	P
1	x_1	P_1
2	x_2	P_2
3	x_3	P_3
⋮	⋮	
i	⋮	
⋮	⋮	
10	X_n	P_n

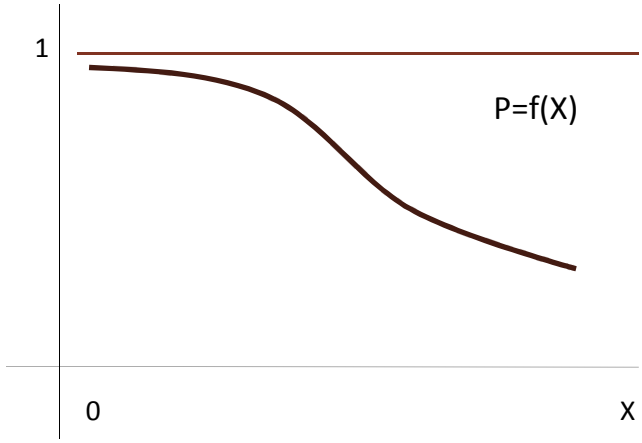


Fig. 4.3. Una generalización de la función de probabilidad

Espacio muestral

Meyer (1974) establece que los *experimentos aleatorios* tienen en común una serie de aspectos que los hacen sencillos de identificar:

a) Es posible repetir cada experimento indefinidamente sin cambiar esencialmente las condiciones
b) Se puede describir el conjunto de todos los resultados posibles del experimento, aunque en general no se pueda indicar cuál será un resultado particular
c) A medida que el experimento se repite, los resultados individuales parecen ocurrir en forma "caprichosa". Sin embargo, como el modelo se repite un gran número de veces, aparece un modelo definido de regularidad. Esta regularidad hace posible la construcción de un modelo matemático preciso con el cual analizar el experimento.

En general, cuando se plantea la tabla de datos como consecuencia del proceso de investigación econométrica (procedimiento experimental), el investigador sabe que con la ocurrencia de un solo fenómeno se pueden calcular "varios" valores numéricos diferentes. Por ejemplo, si se escoge una persona entre un gran grupo de personas (y la elección se hace siguiendo el procedimiento experimental indicado previamente), se puede estar interesado en sexo, peso, ingreso anual, cantidad de hijos, etc. de la persona. En la mayoría de los casos el investigador que siga el procedimiento de investigación econométrica sabe, antes de comenzar el experimento, las características numéricas que le interesan.

El *espacio muestral* se define como el conjunto de todos los resultados posibles del experimento. Este conjunto se lo denomina ***S***.

Observación: En la tabla de datos estarán los resultados posibles para los *elementos* de una población. Estos elementos pueden ser de corte transversal o de tiempo. Cuando se utilizan

elementos de tiempo, los resultados posibles provienen de experimentos repetidos en el tiempo en un espacio determinado.

El resultado de un experimento no necesita ser un número. Por ejemplo, en el experimento de seleccionar una persona entre un gran número de personas hay un resultado posible que no es numérico: el sexo. Aún más, el resultado puede ser un vector o una función. Además, el número de resultados de un espacio muestral puede ser finito o infinito (numerable o no numerable).

Cuando el espacio muestral ***S*** es finito o infinito numerable, *todo* subconjunto se puede considerar como un *suceso* o *evento* (si ***S*** tiene **n** elementos, hay exactamente 2^n subconjuntos o sucesos). En la tabla de datos los valores probables para las unidades de observación (tiempo o individuos) constituirán los sucesos o eventos.

Observación: Se advierte aquí que, aun cuando esto no tenga incidencia en los estudios econométricos, cuando el espacio muestral es infinito no numerable todo subconjunto de ***S*** no es necesariamente un suceso. Asimismo, tanto el espacio muestral, como el conjunto vacío pueden ser considerados sucesos. También, cualquier resultado individual se puede considerar un suceso.

De acuerdo a los fundamentos anteriores, para el estudio econométrico, cada celda de la tabla de datos, , estará representando un suceso o evento. Esto es, el valor probable que asuma para una determinada unidad de información la variable bajo estudio; por ejemplo, si se observa la variable ingreso para un conjunto de individuos, puede interesar el evento de que ocurra para uno de ellos la pertenencia al primer decil.

Cada evento del mundo real se halla relacionado con un conjunto infinito de otros hechos. Toda ciencia estudia solamente cierto número finito de vínculos. De tal modo se establecen las regularidades fundamentales de los eventos estudiados que reflejan las conexiones internas principales inherentes a estos últimos. En principio, es imposible llegar a conocer la infinita diversidad de relaciones existentes con cualquier evento dado.

En cada etapa del conocimiento humano siempre quedan sin estudiar una multitud infinita de vínculos propios a un cierto evento. En consecuencia, cada regularidad puede reflejar solamente un número finito de relaciones fundamentales, debido a lo cual las leyes se cumplen sin precisión, con ciertas desviaciones. Las desviaciones de lo regular, originadas por una infinidad de vínculos no previstos en el evento dado, se llaman *eventos o sucesos aleatorios*. De este modo, la causalidad existe objetivamente en el mundo, porque en principio no es posible revelar todos los nexos accidentales entre el evento estudiado y la multiplicidad infinita de otros sucesos.

“A medida que va desarrollándose la ciencia, dice Pugachev (1973), llegan a ser conocidas nuevas leyes, o sea, las conexiones que tiene el fenómeno estudiado con distintos factores. Por eso, las fronteras entre lo regular o casual no permanecen inalterables sino que cambian a medida que crece el conocimiento humano. Lo que en un período del desarrollo de la ciencia es accidental puede hacerse regular en otro período. Y al contrario, en los

fenómenos considerados como estrictamente regulares en una etapa del desarrollo de la ciencia, a consecuencia de los perfeccionamientos en la técnica del experimento y las exigencias más rigurosas en lo que se refiere a la precisión durante el examen de las dependencias, se descubren desviaciones accidentales de las leyes y surge la necesidad de tenerlas en cuenta" (pp. 11-12).

Si el evento dado se observa una sola vez, no se puede predecir cuál será la desviación accidental de lo regular. Así, por ejemplo, al efectuar una medición sobre el ingreso de cierto residente en una ciudad, para estudiar el ingreso agregado de dicha comunidad, es imposible prever cual será el error de la misma. Sin embargo, si el número de observaciones del evento dado se hace grande, en las mismas desviaciones accidentales se descubren ciertas regularidades que pueden ser estudiadas y utilizadas para determinar la influencia de las desviaciones mencionadas sobre el curso de los eventos por estudiar. Entonces, aparece la posibilidad de investigar eventos aleatorios frecuentes; es decir, observar eventos aleatorios un número ilimitado de veces en las mismas condiciones. La *teoría de la Probabilidad* es, precisamente, la ciencia que estudia las regularidades en los eventos aleatorios frecuentes.

Cierto evento o suceso E puede ocurrir como resultado de un experimento; por ejemplo, al observar determinado individuo, al que se le ha preguntado sobre la variable ingreso ¿resultará un ingreso dentro del primer decil? Ese resultado es, o se lo supone dicotómico; esto significa que existe un evento alternativo al que se designa como $\overline{E}$. Un experimento tiene, generalmente, varios resultados posibles; se denomina x_i a cualquier resultado posible en un experimento.

Cada ciencia se basa sobre algunos hechos experimentales que sirven para formar los conceptos fundamentales de la misma. Estos hechos y conceptos se utilizan para obtener determinadas conclusiones prácticas. En la ciencia económica, las conclusiones se comprueban por la experiencia. La coincidencia, de las conclusiones científicas con los resultados de la experiencia, es el criterio para comprobar una teoría científica.

Enfoque frecuencial de la probabilidad

Es necesario recordar el concepto de probabilidad en forma heurística. La probabilidad de un evento es la frecuencia relativa de su ocurrencia en un gran número de intentos. Esto, en principio, no se contradice con la definición axiomática de la probabilidad que se considera como una función que satisface axiomas que provienen de propiedades elementales de las frecuencias relativas. Más adelante, se asocia el concepto de *distribución de frecuencias* a la distribución del recorrido muestral de las variables aleatorias y el de *distribución de probabilidad* a la distribución del recorrido poblacional de las mismas.

Observación: La teoría de probabilidad es una teoría desarrollada para describir los sucesos aleatorios. Se puede ver la teoría de probabilidad desde dos puntos de vista: sin teoría de medidas o con teoría de medidas. El primer punto de vista es lo que se enseña primero, y entonces se introduce la teoría de medidas. En este capítulo se reseñan los conceptos básicos de la probabilidad. La probabilidad es la característica de un evento del que existen razones para creer que se realizará. Los eventos tienden a ser una frecuencia relativa del número de veces que se realiza el experimento. La probabilidad de aparición de un evento E de un total de M casos posibles igualmente factibles es la razón entre el número de ocurrencias r de dicho suceso y el número total de casos posibles M . La probabilidad es un valor numérico entre 0 y 1. Cuando el evento es imposible se dice que su probabilidad es 0, un evento cierto es aquel que ocurre siempre y su probabilidad es 1. La probabilidad de no ocurrencia de un evento está dada por $q = P(\overline{E}) = 1 - (r/M)$. Simbólicamente el espacio de resultados, que normalmente se denota por S , es el espacio que consiste en todos los resultados que son posibles. Los resultados, que se denotan por x_i, $i = 1, 2, \ldots, n$, son elementos del espacio S .

Según Spiegel (1976) la definición de la probabilidad estimada o empírica está basada en la frecuencia relativa de aparición de un evento E cuando S es muy grande. La probabilidad de un evento es una cantidad, se escribe como $P\{E\}$ y mide con qué frecuencia ocurre algún suceso si se hace algún experimento indefinidamente. La definición anterior es complicada de representar matemáticamente ya que S debiera ser infinito. Otra manera de definir la probabilidad es de forma axiomática, estableciendo las relaciones o propiedades que existen entre los conceptos y operaciones que la componen.

En términos generales y bajo condiciones constantes, si se considera que M es el número de experimentos que se llevan a cabo -el número de veces que la tabla de datos planteada puede ser aplicada para el estudio de un mismo fenómeno económico-, entonces la probabilidad $p(E)$ del evento E se define como

$$p(\mathrm{E}) = \lim_{M \to \infty} \frac{r}{M} \qquad (4.1)$$

donde r es el número de ocurrencias del evento E y M .

Por ejemplo, la probabilidad de que una persona tenga un ingreso perteneciente al segundo decil.

La probabilidad definida en (4.1) proviene evidentemente (y es una extensión) del concepto de frecuencia relativa r/M , donde M es finito.

Aun cuando r/M , con M finito, es inestable en casi todos los fenómenos observables, se requiere que el límite (4.1) exista; aunque para muchos fenómenos puede no existir. La estabilidad final de la frecuencia relativa está explícita en la definición de probabilidad. Por lo tanto, las probabilidades se definen de manera completamente legítima y se usan

provechosamente en aquellos campos en los cuales M puede ser grande y donde, para M grande, la estabilidad de las frecuencias relativas de eventos es un hecho empírico.

No implica necesariamente que se lleven a cabo $M \to \infty$ intentos, sino que infinitos intentos son conceptualmente posibles.

Definición axiomática de la probabilidad

Las frecuencias relativas finitas satisfacen, con pocas excepciones, los axiomas sobre probabilidades. Con respecto a un experimento, sea S un conjunto numerable de puntos, cada uno de los cuales representa un posible resultado de un intento; S es el conjunto de todos los resultados posibles de un intento, esto es, el conjunto de todos los x_i posibles. Sea E cualquier subconjunto de puntos en el conjunto S, p indicará una función que asegura a cada evento o suceso E un número real $p(E)$, llamado la probabilidad de E. En la teoría de la probabilidad, generalmente, se encuentra el siguiente conjunto mínimo de *axiomas*:

1) $p(E) \geq 0$
2) $p(S) = 1$
3) $p(E + F) = p(E) + p(F)$, siendo E y F mutuamente excluyentes.

Para tratar con eventos E y F que no son mutuamente excluyentes, se tratará de expresarlos en términos de eventos que sí lo son, puesto que solamente sobre ellos existe un axioma.

Nótese que al contrario de la definición (4.1), los axiomas no sugieren cómo estimar $p(E)$. Muchas estimaciones numéricas de probabilidades satisfacen estos axiomas; es necesario, entonces, aprender a seleccionar aquellos que son los más adecuados para el problema particular de que se trate.

Los eventos a los que se han asignado probabilidades pertenecen a una clase elemental. Se desea introducir ahora una manera nueva de referirse a algunos de los eventos para los que se pueden asignar probabilidades. Mediante la introducción de una función $X(E)$ se asocia con cada punto E perteneciente a S un número real x_i, esto es, se construye una correspondencia entre los puntos en el espacio S de eventos y los números reales $x_1, x_2, \ldots$ La correspondencia no necesita ser biunívoca; se pueden representar varios puntos de S con números reales x_i. Entonces, el espacio original S puede verse como si estuviera representado en un nuevo espacio de eventos que consiste en puntos sobre la línea real; los eventos particulares E de interés en S, esto es, subconjuntos particulares de puntos de S, se convierten ahora en colecciones particulares de puntos sobre la línea real. Así, por ejemplo, en el experimento de observar la variable ingreso sobre un conjunto de individuos asigna un valor a cada individuo; es decir, mediante la función $X(E)$ se le

asigna al individuo 1 el valor x_1, al individuo 2 el valor x_2, y así siguiendo…, sobre la línea real. Por ejemplo, si se tiene $x_1 = 300, x_2 = 500, \ldots$ puede interpretarse convenientemente como el valor del ingreso para el primero y segundo individuo observado en un experimento, representado en la investigación econométrica en una tabla de datos según lo observado en la figura 4.4.

Cualquier característica cuantitativa del experimento se llama *variable aleatoria*. Por ejemplo, el resultado de $500 en la medición de la variable ingreso.

	X	Y	Z
1	300	y_1	z_1
2	500	y_2	z_2
3	x_3	y_3	z_3
⋮	⋮	⋮	⋮
i	⋮	⋮	⋮
⋮	⋮	⋮	⋮
n	x_n	y_n	z_n

Fig. 4.4. Experimento en la investigación econométrica (tabla de datos)

4.2. Variables aleatorias

Al describir el espacio muestral de un experimento, no se especificó que necesariamente un resultado individual debe ser un número. Aunque esto no debe ser un inconveniente, ya que se podría medir algo y consignarlo como un número. Por ejemplo, el sexo podría ser nominado como 1 femenino y 2 masculino. Lo importante aquí es que, en muchas situaciones experimentales, se desea asignar un número real x a cada uno de los elementos s del espacio muestral S. Esto es, x=X(s) es el valor de una función X del espacio muestral a los números reales. Teniendo esto presente, se puede hacer la siguiente definición formal:

Una *variable aleatoria* se define como el resultado numérico de un experimento aleatorio. Matemáticamente, es una aplicación $X : S \to \Re$ que da un valor numérico, del conjunto de los reales, a cada suceso o evento en el espacio S de los resultados posibles del experimento.

Una variable es un elemento de una fórmula, proposición o algoritmo que puede adquirir o ser sustituido por un valor cualquiera; los valores están definidos dentro de un rango. Mientras que, *variable aleatoria*, es una variable que cuantifica los resultados de un

experimento aleatorio, que puede tomar diferentes valores cada vez que ocurre un suceso; el valor sólo se conocerá determinísticamente una vez acaecido el evento.

Una variable aleatoria tiene una *distribución de probabilidades*. Ésta es un modelo teórico que describe la forma en que varían los resultados de un experimento aleatorio. Es decir, detalla los resultados de un experimento con las probabilidades que se esperarían ver asociadas con cada resultado.

Al igual que las variables cuantitativas, las *variables aleatorias cuantitativas* se clasifican en, *variable aleatoria discreta*, variable que toma un número finito o infinito de valores numerables; *variable aleatoria continua*, variable que toma una cantidad infinita de valores no numerables.

Se denomina *función de probabilidad* a la función que asigna probabilidades a cada uno de los valores de una variable aleatoria *discreta*. Se denomina *función de densidad* a la función que mide concentración de probabilidad alrededor de los valores de una variable aleatoria *continua*. La función *de distribución* es la función que acumula probabilidades asociadas a una variable aleatoria.

Dado una variable aleatoria X se pueden calcular diferentes medidas como la media aritmética, la media geométrica o la media ponderada y también el valor esperado y y la varianza de la distribución de probabilidad de X .

Ejemplo 4.1: Suponga que se lanzan dos monedas al aire, sea X la variable aleatoria que identifica el número de caras obtenidas en el lanzamiento, $S = \{cc, cs, sc, ss\}$ (c identifica una cara, s un sello), $R_X = \{0, 1, 2\}$ (Recorrido de X). Entonces el Recorrido de X , será:

$X : S \to R_X \quad cc \to 2 \quad cs, sc \to 1 \quad ss \to 0$

Freeman (1979) establece que "la mayoría de las variables que nos son conocidas (longitud, peso, ingreso, precio, resistencia, densidad) se manejan cómodamente con representaciones sobre los números reales; otras, como la inteligencia, preferencia, sabor y color, se han cuantificado, y a menudo se puede hacerlo, para manejarse con ese mapeo sobre los números reales. La función $X(S)$ que efectúa este cambio del espacio de eventos anterior al nuevo se llama función aleatoria, o más comúnmente, variable aleatoria; el adjetivo *aleatorio* se emplea para recordarnos que el resultado numérico x_i de un intento o experimento sobre X es cierto y, por tanto, que el valor de la función (y no la función misma) es incierto. Una variable aleatoria es simplemente una función de valores reales definidos en S" (pag. 30 y ss.)

Nótese que los resultados posibles de un experimento son variables aleatorias, pero una vez que se ejecutó el experimento; es decir una vez que, por ejemplo, se seleccionó una

persona y se midió su ingreso anual, se obtiene un valor específico para la variable. Esta distinción entre una variable aleatoria y su valor es muy importante. Este último es un valor no estocástico. No cambia si no cambia el experimento, o lo que es lo mismo, no cambia si no cambian las condiciones bajo las cuales se lleva a cabo el experimento. Es decir, entre las dos interpretaciones posibles que la bibliografía le suele dar al recorrido Rx de una variable aleatoria, en econometría es evidente que como dice Meyer (1973) "el experimento aleatorio termina, de hecho, con la observación de s ∈ S".

4.3 Distribuciones de frecuencia y distribuciones de probabilidad

Una *distribución de frecuencias* representa una ordenación de los datos de tal forma que indique el *número de observaciones*, surgidas de una muestra, en cada clase de la variable. El número de observaciones en cada clase se denomina *frecuencia absoluta;* ésta se distingue de la *frecuencia relativa*, que indica la *proporción de observaciones* en cada clase y no su número.

En una distribución de frecuencias los datos pueden agruparse, como en la figura 4.5, en diferentes clases. Para cada una de las clases de una distribución de frecuencias, los límites nominales de clase inferior y superior indican los valores incluidos dentro de la clase.

Las *fronteras de clase* o *límites exactos* son los puntos específicos de la escala de medición que sirven para separar clases adyacentes cuando se trata de variables continuas. Los límites exactos de clase pueden determinarse identificando los puntos que están a la mitad entre los límites superior e inferior de las clases adyacentes.

El *intervalo de clase* indica el rango de los valores incluidos dentro de una clase y puede ser determinado restando el límite exacto inferior de clase de su límite exacto superior. Los valores de una clase se representan, a menudo, por el *punto medio de clase, el* que puede ser determinado sumando los límites superior e inferior y dividiendo por dos.

Las distribuciones de frecuencia pueden graficarse a través de barras, o *histograma*, en el cual el eje horizontal representa los intervalos de clase y el eje vertical el número de observaciones.

El *polígono de frecuencias* es la línea que une los puntos medios de una distribución de frecuencias. Por su parte la *curva de frecuencias* es un polígono de frecuencias suavizado.

En una población, el concepto correspondiente a la distribución de frecuencias de una muestra se conoce con el nombre de *distribución de probabilidades*.

El 16,67% de *todos* los Estados de Resultados mensuales tuvieron una incidencia de gastos financieros inferior al 5%, esto equivale a afirmar que la probabilidad de seleccionar aleatoriamente un Estado de Resultado con gastos financieros inferiores al 5% es 0,1667.

La figura 4.6 representa una distribución poblacional y la figura 4.5 una distribución muestral. No siempre, obviamente, la muestra representa exactamente a la población como en este caso; pero, con la distribución muestral se puede inferir la poblacional, evaluando estadísticamente el error cometido a partir de la inferencia.

Ejemplo 4.2. Se observa una muestra de 6 estados de resultados de la Empresa "A", considerando en cada uno la incidencia de los gastos financieros sobre el total de la venta. Los resultados aparecen en la figura 4.5.

% gasto financiero (variable)	N° de Estados de Resultados (frec.absoluta)	Proporción de Estados de Resultados (frec.relativa)[1]
menos de 5%	1	0.16672
5% a 9,99%	3	0.50000
10% a 19,99%	2	0.33330
20% o más	0	0.00000
TOTALES	6	1.00000

Figura 4.5. Muestra de los Estados de Resultados de la Empresa “A”

Ejemplo 4.3. La población de *todos* los estados de resultados mensuales de la Empresa "A" en los ejercicios económicos entre los años 1980 y 1991, clasificados según la incidencia de los gastos financieros, se encuentran en la figura 4.6.

% gasto financiero (variable)	% de Estados de Resultados (frec.relativa)
menos de 5%	16.67 %
5% a 9,99%	50.00 %
10% a 19,99%	33.33 %
20% o más	0.00 %
TOTAL	100.00 %

Figura 4.6: Estados de Resultados de la Empresa “A”

Esto define la naturaleza de la *inferencia estadística* que se dedica a efectuar generalizaciones respecto a la población a partir de la información proporcionada por la muestra.

En resumen, se utilizan las muestras para hacer juicios respecto a las poblaciones de las que provienen las muestras. Estos juicios se pueden referir tanto a situaciones del pasado como a estimar lo que ocurrirá en el futuro. En general, sólo interesa alguna característica de la población que se denomina *parámetro*, sobre él serán los juicios basados en el estadístico de la muestra que recibe el nombre de *estimador*.

Ejemplo 4.4. Si el parámetro que se desea estimar es el gasto financiero medio, el estimador muestral que se usa es el promedio muestral del gasto financiero.

Ahora bien, para obtener una distribución muestral deben obtenerse todas las muestras posibles de la población. Puede imaginarse que se van obteniendo una detrás de otra todas las muestras posibles y para cada una se calcula el valor en cuestión (por ejemplo, el gasto financiero). La distribución resultante de este número de muestras se denomina distribución muestral. Una *distribución muestral* es la distribución de probabilidad de un estimador.

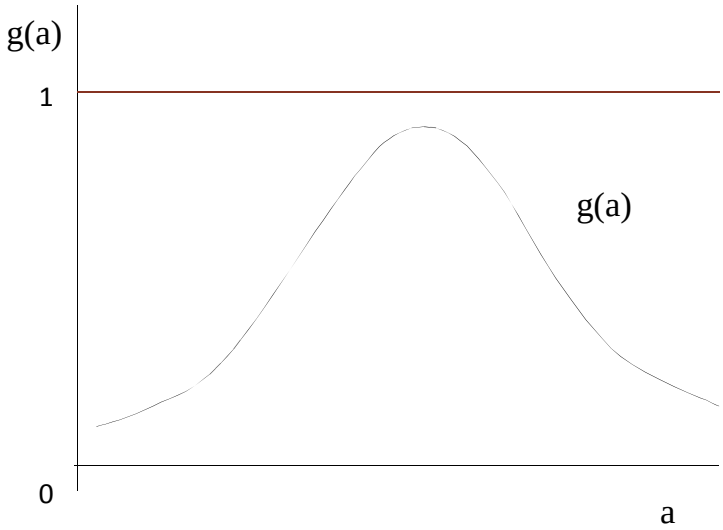


Fig.4.7. Distribución de Probabilidad

De esta forma, si se estudia la población de todos los valores posibles de la variable gasto financiero de los últimos 12 ejercicios económicos de la Empresa "A" (variable X) e interesa estimar el gasto financiero medio (un parámetro α) se usa el estadístico muestral **a**, donde **a** es un estimador de α, y un valor concreto de **a** (obtenido a partir de una muestra determinada) es una estimación de α. Como el gasto financiero de una empresa es una variable continua, la distribución muestral de **a** tendrá, posiblemente, una forma como la que se muestra en la Figura 4.7.

Sin mayor rigurosidad se puede decir que, el concepto de distribución muestral se entiende mejor si se lo considera como una distribución de frecuencias relativas, calculadas a partir de un número grande de observaciones (en este caso, de muestras) del mismo tamaño. En la Figura 4.7, las frecuencias relativas de **a** se miden a lo largo del eje **g(a)**.

Con el análisis anterior se puede ver la importancia que tiene el estudio de las distribuciones muestrales; ya sean experimentales -que son las obtenidas a partir de un número grande o infinito de muestras posibles-, o las teóricas -que están basadas en la teoría de la probabilidad-. Estas últimas, las *distribuciones teóricas*, permiten estudiar poblaciones sin necesidad de realizar repeticiones de experimentos muestrales. Se distinguen según se refieran a variables aleatorias discretas o continuas. Dentro de las continuas, la distribución normal es la más utilizada, ya que responde a la distribución de muchos fenómenos de las ciencias sociales.

4.4. Distribuciones teóricas de probabilidad

La teoría de las Distribuciones Estadísticas es fundamental para el análisis de la información en la toma de decisiones. Es necesario distinguir entre las distribuciones experimentales y las distribuciones teóricas, teniendo en cuenta que estas últimas se basan en la teoría de la probabilidad.

La *función de densidad o función de probabilidad* de una variable aleatoria, es una función a partir de la cual se obtiene la probabilidad de cada valor que toma la variable.

La *función de distribución de probabilidad,* $F(x)$*,* es una función de la probabilidad que representa los resultados que se van obteniendo en un experimento aleatorio.

Para un valor dado x, la probabilidad $p(X \leq x) = F(x)$.

Con las anteriores definiciones se puede realizar una extensión de los axiomas de probabilidad de la siguiente manera:

Para dos números reales cualesquiera a, b tal que $(a < b)$, los eventos $(X \leq a)$ y $p(a \leq x \leq b)$ serán mutuamente excluyentes y su suma es el nuevo evento $(X \leq b)$ por lo que:

$$p(X \leq b) = p(X \leq a) + p(a < X \leq b)$$

$$p(a < X \leq b) = p(X \leq b) - p(X \leq a)$$

Y, finalmente

$$p(a < X \leq b) = F(b) - F(a)$$

Por lo tanto, una vez conocida la función de distribución $F(X)$ para todos los valores de la variable aleatoria X se conocerá completamente la distribución de probabilidad de la variable.

Distribuciones de variable discreta

La *función de probabilidad* de una variable aleatoria discreta X, indicada como $f(X)$, se define como una regla que asigna a cada número real x la probabilidad de que la variable X asuma el valor x. Es decir,

$$f(x) = p(X = x)$$

En cambio, la *función de distribución de probabilidad* de X, indicada como $F(X)$, se define como una regla que asigna a cada número real x la probabilidad de que la variable aleatoria X sea igual o menor que el valor de x. Es decir,

$$F(x) = p(X \leq x) = \sum_{X_i \leq x} f(x_i)$$

Ejemplo 4.5. Se define una variable aleatoria X como las unidades que constituyen la demanda de los productos de la Empresa A durante el año próximo. Se suponen posibles e igualmente probables cuatro niveles de venta: 10, 12, 15 ó 18 unidades. Como las probabilidades de estos cuatro resultados posibles deben sumar 1, la función de probabilidades de X está dada por:

$$F(x) = \begin{cases} P(X = 10) = \dfrac{1}{4} \\ P(X = 12) = \dfrac{1}{4} \\ P(X = 15) = \dfrac{1}{4} \\ P(X = 18) = \dfrac{1}{4} \end{cases} \qquad (1)$$

(1) indica que la probabilidad de que la demanda sea de 10, 12, 15 ó 18 unidades es cada una igual a ¼.

La función de distribución de probabilidades de X, estará dada por:

$$F(x) = \begin{cases} P(X \leq 10) = \dfrac{1}{4} \\ P(X \leq 12) = \dfrac{1}{2} \\ P(X \leq 15) = \dfrac{3}{4} \\ P(X \leq 18) = 1 \end{cases} \qquad (2)$$

(2) dice que hay una probabilidad de ¼ que la demanda sea igual o menor a 10 unidades, una probabilidad de ½ de que la demanda real sea menor o igual a 12 unidades, una probabilidad de ¾ de que la demanda sea menor o igual a 15 unidades y una probabilidad cierta (igual a 1) de que la demanda sea menor o igual a 18 unidades.

La Figura 4.8 representa la función de probabilidad dada por (1) y a la función de distribución de probabilidad dada por (2).

a) Función de Probabilidad b) Función de Distribución de probabilidad

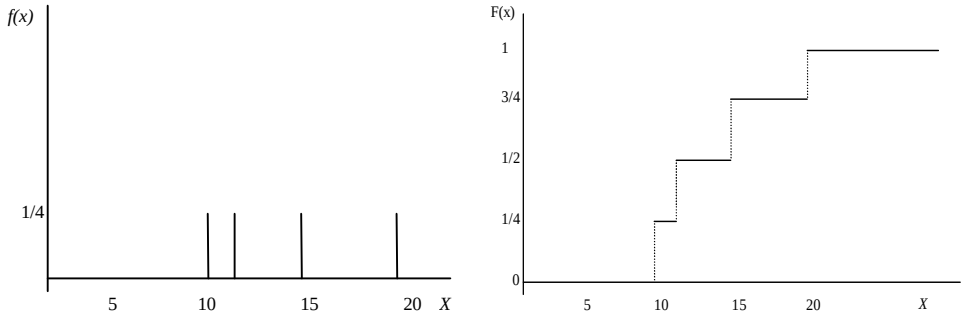


Figura 4.8 Pronóstico de Ventas

La Figura 4.9 muestra la gráfica del caso particular de la Distribución Binomial P(X=k) para k=0,1,2…6.

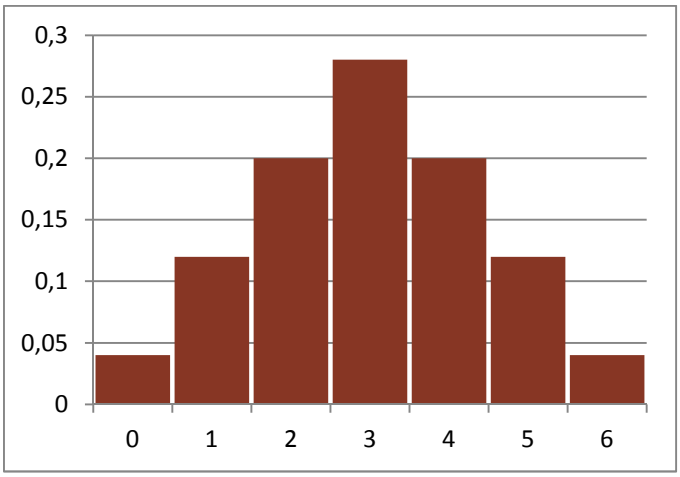


Figura 4.9 Distribución binomial

Distribuciones de variable continua

El análisis anterior se ve modificado en el campo de las variables aleatorias continuas. La *función de densidad* de X es:

$$f(x) = P(x_i \le X \le x_j) = \int_{x_i}^{x_j} f(x)dx \qquad \text{donde, } x_i \le x_j$$

Ejemplo 4.6. Una función de densidad de una variable aleatoria continua es la distribución de probabilidad normal estándar. La función de densidad y la función de distribución de probabilidad de una variable aleatoria X normal estándar son las que muestra la Figura 4.10.

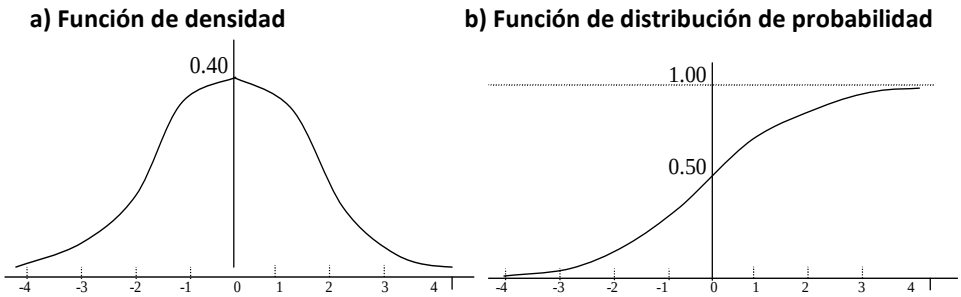


Figura 4.10 Distribución normal estándar

La figura 4.11 muestra la función de densidad para la distribución normal.

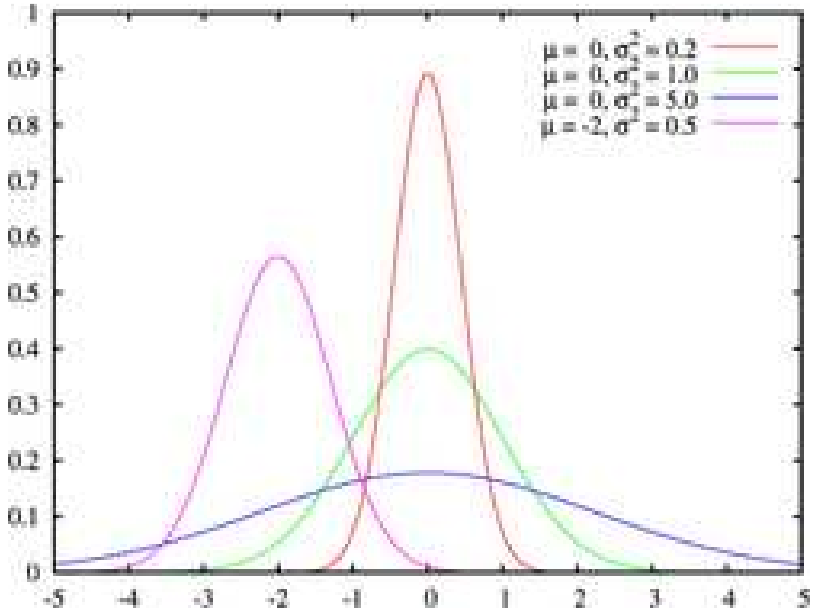


Figura 4.11 Función de densidad para la distribución normal

$f(x)$ es una función donde el área bajo la misma, entre x_i y x_j, es exactamente la probabilidad de que X asuma un valor entre x_i y x_j. De la misma manera, la *función de distribución de probabilidad*, $F(x)$, está dada por la expresión:

$$F(x) = P(X \leq x) = \int_{-\infty}^{X} f(s)ds$$

donde *s* es una variable de integración

Es decir, para determinar la probabilidad acumulativa de que X sea igual o menor que x, se calcula el área bajo la función de densidad, $f(x)$, entre -∞ y x. La probabilidad de que la variable aleatoria continua X sea exactamente igual a cierto valor x es cero.

Esperanza matemática

La *esperanza matemática* de una variable aleatoria es la suma de la probabilidad de cada evento o suceso multiplicada por su valor. *Si todos los sucesos son de igual probabilidad la esperanza es la media aritmética.*

La esperanza matemática o valor esperado de una variable aleatoria discreta *X*, indicado como $E(X)$, se define:

$$E(X) = \sum_{\forall i} x_i f(x_i)$$

donde,

x_i representa cualquier valor posible de *X,* y

$f(x_i)$ es la probabilidad de que $X = x_i$.

Es decir, $E(X)$ es un promedio ponderado de todos los valores posibles de *X*, donde las ponderaciones son las respectivas probabilidades de estos valores.

La varianza de una variable aleatoria *X*, indicada por V(X), se define:

$$V(X) = \sum_{\forall i} (x_i - E(X))^2 f(x_i)$$

donde, todos los términos responden a las definiciones anteriores. Es decir, V(X) es un promedio ponderado de las desviaciones cuadráticas de los valores observados de *X* con respecto al valor esperado de *X*, donde las ponderaciones son las respectivas probabilidades.

La desviación estándar de una variable aleatoria *X*, de suma utilidad práctica, se define como la raíz cuadrada de la V(X).

El valor esperado y la varianza de variables aleatorias con distribución de probabilidad continua, se definen con las respectivas fórmulas como:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

$$V(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx$$

donde, $f(x)$ es la función de densidad de la variable aleatoria *X*.

Algunas distribuciones de probabilidad

No necesariamente todas las variables aleatorias bajo estudio responden a las distribuciones teóricas de probabilidad. Existen las distribuciones experimentales que, una vez obtenidas, pueden o no responder a las formas de las distribuciones teóricas.

En la figura 4.12 y figura 4.13 se presentan algunas distribuciones teóricas de probabilidad, tanto discretas como continuas. Se han incluido en el cuadro las funciones de probabilidad (también llamadas funciones de densidad cuando están asociadas con variables aleatorias continuas) y los principales parámetros (media, varianza) de las distribuciones.

Es importante comentar aquí que, a partir de las distribuciones muestrales se obtienen estimadores de los parámetros poblacionales (como se verá en el próximo capítulo). Estos estimadores, por provenir de una muestra aleatoria constituyen, en sí mismos, variables aleatorias sujetas a distribuciones de probabilidad y a distribuciones acumulativas de probabilidad. Esta es la verdadera naturaleza de la Inferencia Estadística.

Necesidad del uso de probabilidades

En cualquier circunstancia, en el ámbito macro o microeconómico, toda decisión tiene efecto durante un período de tiempo que se extiende hacia el futuro. Esta característica, que es común a todas las decisiones, probablemente se observe con mayor intensidad en las áreas empresariales. Sin embargo, una decisión involucra aspectos del futuro, cualquiera sea la base sobre la que se tome.

Teniendo en cuenta que al evaluar una propuesta se estará mirando hacia el futuro, ésta se traducirá en estimaciones de variables; por ejemplo, costos, ventas, precios, inversiones o impuestos, que estarán sujetas a cierto nivel de incertidumbre. Ante este nivel de incertidumbre en la estimación de variables importantes para la economía, ¿es suficiente trabajar con el valor sospechado, probable o experimental?, o ¿es más conveniente trabajar con la distribución de probabilidad de cada variable?

Hay que tener en cuenta que el riesgo es inseparable en la estimación de cualquier alternativa de decisión. Evidentemente, en el campo de la toma de decisiones, es más importante basarse en los métodos probabilísticos que en los subjetivos.

Distribución de X	Función de Probabilidad (o Función de Densidad)	Parámetros E(X) V(X)
a. Discretas		
Uniforme	$p(x_i) = \frac{1}{n}$	$\sum \frac{x_i}{n}$
Poisson	$P(X = k) = \frac{e^{-\alpha}\alpha^k}{k!}; k = 0,1,\ldots,n,\ldots \quad \alpha > 0$	α α
Geométrica	$P(X = k) = q^{k-1}p; k = 1,2,\ldots$	$\frac{1}{p}$ $\frac{q}{p^2}$
Binomial	$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; k = 0,\ldots,n$	np npq
Binomial Negativa	$b^*(x;k;\theta) = \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k}$ $x = k, k+1, k+2, \ldots$	$\mu = \frac{k}{\theta}$ $\sigma^2 = \frac{k(1-\theta)}{\theta^2}$
Pascal	$P(X = k) = \binom{k-1}{r-1} p^r q^{k-r}; k = r, r+1, \ldots$	$\frac{r}{p}$ $\frac{rq}{p^2}$
Hipergeométrica	$P(X = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}; k = 0,1,2\ldots$	np $npq\frac{N-n}{N-1}$
Multinomial	$P(X_1 = n_1, X_2 = n_2, \ldots X_k = n_k) = \frac{n! p_1^{n_1} \ldots p_k^{n_k}}{n_1! n_2! \ldots n_k!}$	np_i $np_i q_i$ $i = 1,2\ldots k$
Zeta	$P(X = x) = \frac{x^{-s}}{\zeta(s)}$ con $s > 1$ y $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ donde $\zeta(s)$ es la Función Z de Riemann y s es número complejo con parte real >1	Media $\frac{\zeta(s-1)}{\zeta(s)}$ para s>2 Varianza $\frac{\zeta(s)\zeta(s-2)-\zeta(s-1)^2}{\zeta(s)^2}$ para s>3 $x = 1,2, \ldots$

Figura 4.12: Distribuciones Teóricas de Probabilidad de Variable Discreta

Distribución de X	Función de Probabilidad (o Función de Densidad)	Parámetros E(X) V(X)
b. Continuas		
Distribución Uniforme Continua	$f(x)=\begin{cases}\frac{1}{b-a} & a\le x\le b\\ 0 & resto\end{cases}$	$\frac{(a+b)}{2}$ $\frac{(b-a)^2}{12}$
Distribución Pareto	$f(x)=\frac{ab^a}{x^{a+1}}$ $x>b$	$\frac{ab}{a-1}$ $\frac{ab^2}{(a-1)^2(a-2)}$
Distribución beta	$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$ $0<x<1$	Media $\frac{a}{a+b}$ varianza $\frac{ab}{(a+b+1)(a+b)^2}$
Distribución de Cauchy	$f(x)=\frac{1}{s\pi\left[1+\left(\frac{x-t}{s}\right)^2\right]}$ con t y s > 0	En general, no tiene media y varianza
Normal	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}; -\infty<x<\infty<$	μ σ
Exponencial	$f(x)=\alpha e^{-\alpha x}; x>0$	$\frac{1}{\alpha}$ $\frac{1}{\alpha^2}$
Gamma	$f(x)=\frac{\alpha}{\Gamma(r)}(\alpha x)^{r-1}e^{-\alpha x}; x>0$	$\frac{r}{\alpha}$ $\frac{r}{\alpha^2}$

Figura 4.13: Distribuciones Teóricas de Probabilidad de Variable Continua

4.5. Teoría de las pequeñas muestras

Distribución Gamma

La **distribución gamma** servirá de punto de partida para obtener las funciones de densidad de las distribuciones exactas, que comprende a:

Distribución chi cuadrado de Pearson, χ^2
Distribución t de Student, t
Distribución F de Snedecor o Fischer, F

La función gamma del análisis matemático es:

$$\Gamma(\lambda)=\int_0^\infty e^{-x}x^{\lambda-1}dx; \quad \lambda>0 \tag{1}$$

En la Figura 4.14 se observa gráficamente el comportamiento de la función Gamma

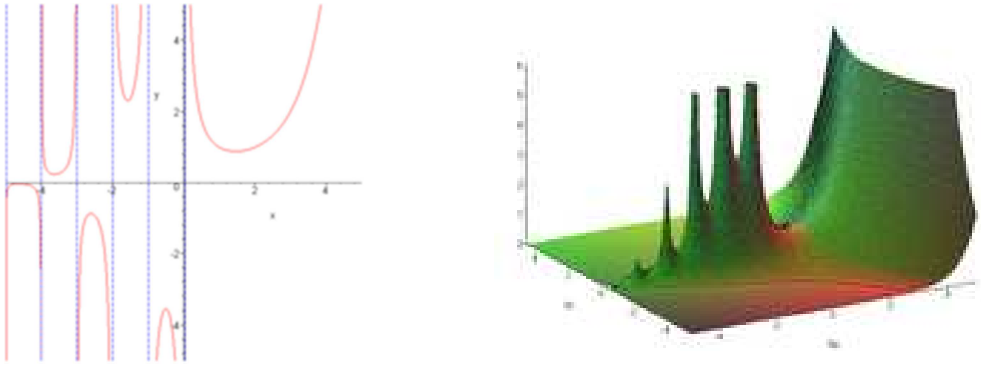


Figura 4.14: Función Gamma

Esta integral se resuelve fácilmente y se obtiene

$$\Gamma(\lambda) = (\lambda - 1)! \tag{2}$$

Se debe tener presente, en los desarrollos posteriores, el valor notable:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{C.D}$$

Hay dos integrales de las que resulta útil conocer su resolución por la aplicación que de ellas se hace más adelante:

1) En la integral:

$$\int_0^{\infty} e^{-\alpha x} x^{\lambda-1} dx; \quad \alpha > 0, \ \lambda > 0 \tag{4}$$

se realiza la siguiente sustitución: $x\alpha = \mu$

de donde se deduce: $x = \frac{\mu}{\alpha}; \quad dx = \frac{du}{\alpha}$

sustituyendo en (4): $\int_0^\infty e^{-\mu}\,\frac{\mu^{\lambda-1}}{\alpha^{\lambda-1}}\frac{du}{\alpha} = \int_0^\infty e^{-\mu}\,\frac{\mu^{\lambda-1}}{\alpha^{\lambda}}du$

si se saca $\frac{1}{\alpha^{\lambda}}$ fuera de la integral y se compara con (1) queda:

$$\frac{1}{\alpha^{\lambda}}\int_0^\infty e^{-\mu}\mu^{\lambda-1}du = \frac{1}{\alpha^{\lambda}}\Gamma(\lambda)$$

o sea que:

$$\int_0^\infty e^{-\alpha x}x^{\lambda-1}dx = \frac{\Gamma(\lambda)}{\alpha^{\lambda}} \qquad (5)$$

2) En la integral

$$\int_0^\infty e^{-\alpha x^2}dx, \quad \alpha > 0$$

sustituyendo $x^2 = \mu$, se tiene

$$x = \mu^{1/2}; \quad dx = \frac{1}{2}\mu^{-\frac{1}{2}}du$$

al sustituir se obtiene:

$$\frac{1}{2}\int_0^\infty e^{-\alpha\mu}\mu^{-\frac{1}{2}}du = \frac{\Gamma\left(\frac{1}{2}\right)}{2\alpha^{\frac{1}{2}}} = \frac{\sqrt{\pi}}{2\alpha^{\frac{1}{2}}} \qquad (6)$$

La **distribución gamma** de la variable aleatoria **X**, con parámetros α y λ, se define mediante la siguiente función de densidad:

$$f(x)=\begin{cases}\dfrac{\alpha^{\lambda}}{\Gamma(\lambda)}e^{-\alpha x}x^{\lambda-1}, & \alpha>0;\lambda>0;x>0\\ 0, & x\le 0\end{cases} \tag{7}$$

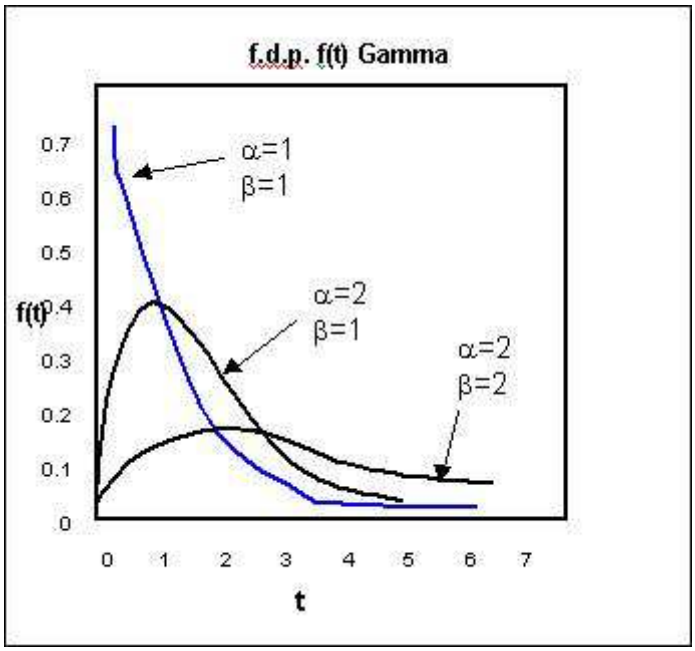


Figura 4.15: Distribución Gamma

Si se integra la función de densidad entre los límites de su recorrido, el resultado debe ser 1. El recorrido de esta función es de 0 a ∞ , si se realiza la integral

$$\int_0^{\infty} f(x)dx=\int_0^{\infty}\frac{\alpha^{\lambda}}{\Gamma(\lambda)}e^{-\alpha x}x^{\lambda-1}dx=\frac{\alpha^{\lambda}}{\Gamma(\lambda)}\int_0^{\infty}e^{-\alpha x}x^{\lambda-1}dx$$

Teniendo en cuenta el resultado de (5)

$$\int_0^{\infty} f(x)dx=\frac{\alpha^{\lambda}}{\Gamma(\lambda)}\frac{\Gamma(\lambda)}{\alpha^{\lambda}}=1$$

Con este resultado se confirma que la expresión hallada en (7) es la función de densidad de la distribución Gamma.

La función generatriz de momentos de esta distribución con función de densidad $f(x)$, es

$$m_x(\theta) = E(e^{\theta x}) = \int_0^\infty e^{\theta x} f(x) dx$$

que, haciendo uso de (7), queda:

$$\frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty e^{-\alpha x} e^{\theta x} x^{\lambda-1} dx = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty e^{-x(\alpha-\theta)} x^{\lambda-1} dx$$

La integral es análoga a (5) pero con $\alpha - \theta$ en lugar de α, luego,

$$m_X(\theta) = \frac{\alpha^\lambda \Gamma(\lambda)}{\Gamma(\lambda)(\alpha-\theta)^\lambda} = \frac{\alpha^\lambda}{(\alpha-\theta)^\lambda} = \frac{(\alpha-\theta)^{-\lambda}}{\alpha^{-\lambda}} = \left(\frac{\alpha-\theta}{\alpha}\right)^{-\lambda} = \left(\frac{\alpha}{\alpha} - \frac{\theta}{\alpha}\right)^{-\lambda}$$

La función generatriz de momentos de una variable X con distribución gamma, en el parámetro θ, es

$$m_X(\theta) = \left(1 - \frac{\theta}{\alpha}\right)^{-\lambda} \tag{8}$$

Derivando sucesivamente la función generatriz se obtienen los momentos de orden 1 y orden 2 (m_1 y m_2), que posibilitan el cálculo de la media y la varianza de la distribución gamma.

$$E(x) = m_1 = \frac{\lambda}{\alpha}$$

$$V(x) = m_2 - m_1^2 = \frac{\lambda}{\alpha^2}$$

Distribución exponencial

En estadística la **distribución exponencial** es una distribución de probabilidad continua con parámetro λ > 0, cuya función de densidad es

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Su función de distribución es

$$F(x) = P(X \leq x) \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

El valor esperado y la varianza de una variable aleatoria X con distribución exponencial son

$E(X) = {}^{1}/_{\lambda}$

$V(X) = {}^{1}/_{\lambda^2}$

Ejemplo 4.7. La distribución exponencial está presente en el tiempo que puede transcurrir para la llegada de un paciente en un servicio de urgencias. También en un proceso de Poisson, donde se repite sucesivamente un experimento a intervalos de tiempo iguales, el tiempo que transcurre entre la ocurrencia de dos sucesos consecutivos sigue un modelo probabilístico exponencial; por ejemplo, el tiempo que transcurre entre que sufrimos dos veces una herida importante.

La función de densidad para λ igual 0.5, 1.0, y 1.5:

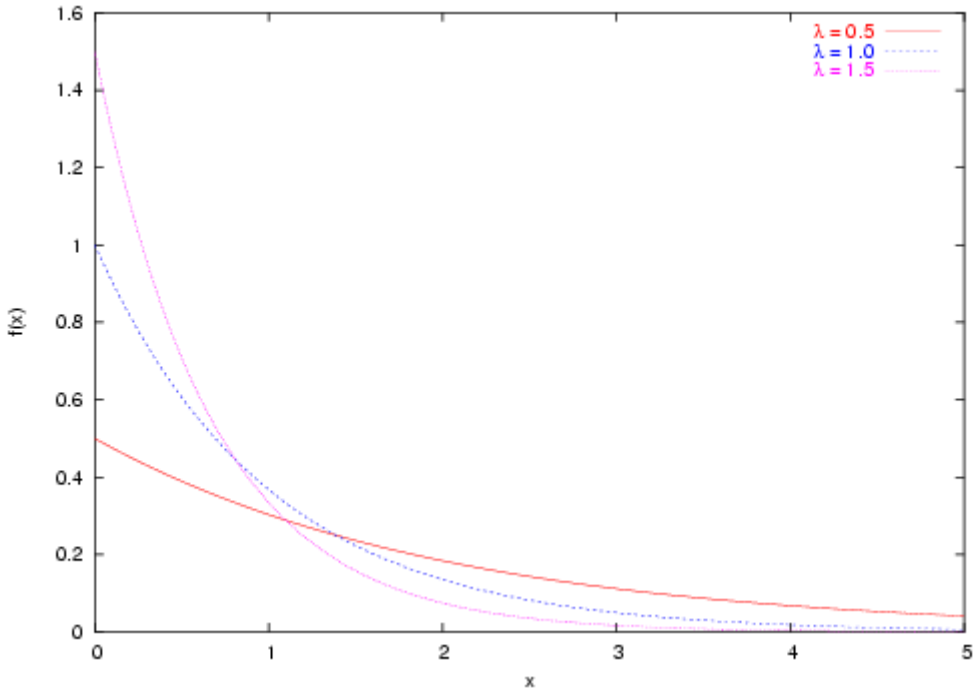


Figura 4.16: Función de densidad exponencial

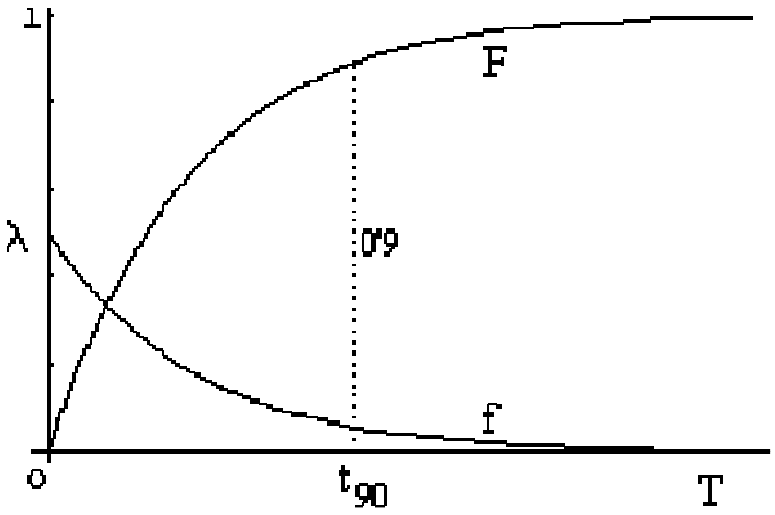


Figura 4.17: Función de densidad (f) y distribución (F) exponencial

---

La suma de *k* variables aleatorias independientes de distribución exponencial con parámetro λ es una variable aleatoria de distribución gamma.

Distribución chi-cuadrado (χ^2) de Pearson

Si $X_1, X_2, \ldots, X_n$, son los datos del recorrido de la variable poblacional X, obtenidos de las **n** unidades de observación de una muestra; el estadístico (variable aleatoria):

$$Y = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2}; \text{ donde } X \sim N(\mu, \sigma) \qquad (9)$$

Se distribuye como una Chi-Cuadrado con **n** grados de libertad, y se representa $Y \sim \chi^2_n$

Al desarrollar el sumatorio, en (9), se tiene que

$$Y = \frac{(X_1 - \mu)^2}{\sigma^2} + \frac{(X_2 - \mu)^2}{\sigma^2} + \ldots + \frac{(X_n - \mu)^2}{\sigma^2} = z_1^2 + z_2^2 + \ldots + z_n^2$$

que conduce a la forma alternativa de escribir (9):

$$Y = \sum_{i=1}^{n} z_i^2 \sim \chi^2_n \qquad \text{donde } z_i \sim N(0,1)$$

La función generatriz de momentos de la variable $Y \sim \chi^2_n$ es:

$$m_Y(\theta) = m_{\sum z_i^2}(\theta) = m_{z_1^2}(\theta)\, m_{z_2^2}(\theta) \ldots m_{z_n^2}(\theta) \Rightarrow m_Y(\theta) = \left(m_{z^2}(\theta)\right)^n \qquad (10)$$

Pero como $z_i \sim N(0,1)$, aplicando lo analizado en la expresión (11) del capítulo anterior,

$$m_{z^2}(\theta) = E\left(e^{\theta z^2}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta z^2} e^{\frac{-z^2}{2}} dz$$

Como se trata de una función par, integramos entre 0 e ∞ y multiplicamos por 2

$$\frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\left(\frac{1}{2}-\theta\right)z^2} dz \qquad (11)$$

(11) es análoga a la integral que da por resultado (6), para $\alpha=(1/2)-\theta$, por lo tanto

$$m_{z^2}(\theta) = \frac{2}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{2\sqrt{\frac{1}{2}-\theta}} = \frac{1}{\sqrt{2}\sqrt{\frac{1}{2}-\theta}} = \frac{1}{\sqrt{1-2\theta}}$$

Entonces, la función generatriz de momentos de z^2 en el parámetro θ es

$$m_{z^2}(\theta) = \frac{1}{\sqrt{1-2\theta}} \qquad (12)$$

Por lo tanto, la función generatriz de momentos de la variable Y, dada en (10), es $m_Y(\theta) = \left[m_{z^2}(\theta)\right]^n$, reemplazando el resultado de (12)

$$m_Y(\theta) = \left[m_{z^2}(\theta)\right]^n = \sqrt{\left(\frac{1}{1-2\theta}\right)^n} \qquad (13)$$

(13) es la función generatriz de momentos de una variable Y con distribución chi cuadrado de n grados de libertad.

La media y la desviación típica de la distribución χ^2 se obtienen derivando la función generatriz de momentos, hallada en (13)

$$E(\chi^2) = \left.\frac{\partial m_Y(\theta)}{\partial(\theta)}\right|_{\theta=0} = n$$

$$V(\chi^2) = \left.\frac{\partial m_Y(\theta)}{\partial^2(\theta)}\right|_{\theta=0} - \left[E(\chi^2)\right]^2 = 2n$$

Comparando (13) con (8) se observa que la función generatriz de momentos tiene una estructura similar a una distribución gamma de:

variable: **y**

parámetros: α = **½**, λ = **n/2**

En consecuencia, teniendo en cuenta (7), la función de densidad de Y es:

$$f(Y) = \frac{e^{-Y/2} Y^{(n/2)-1}}{\sqrt{2^n}\,\Gamma(n/2)} \qquad (14)$$

donde Y se distribuye chi-cuadrado con n grados de libertad.

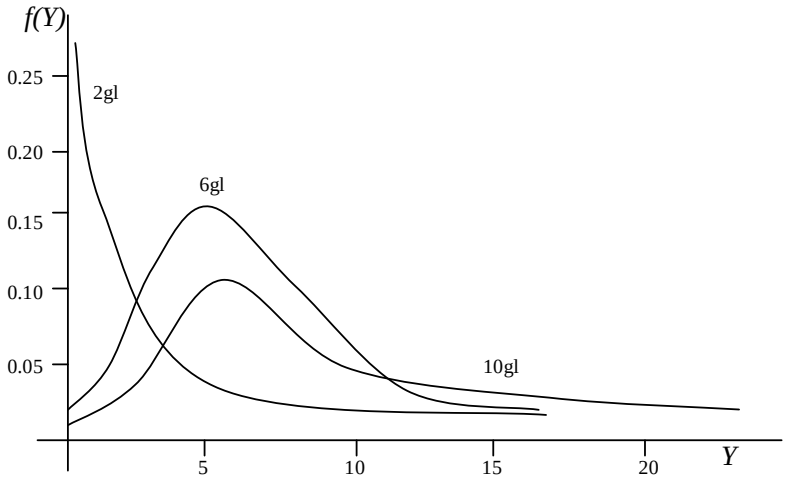


Figura 4.18: Distribución chi Cuadrado

La Figura 4.18 presenta funciones de distribución de chi cuadrado con diferentes grados de libertad

Propiedad reproductiva de χ^2

Dadas dos variables independientes, Y_1 e Y_2, donde ambas se distribuyen χ^2 con n_1 y n_2 grados de libertad, respectivamente; la suma de ellas es otra variable χ^2 con $n_1 + n_2$ grados de libertad.

En efecto,

$$m_{Y_1+Y_2}(\theta) = m_{Y_1}(\theta)\, m_{Y_2}(\theta) = \sqrt{\frac{1}{(1-2\theta)^{n_1}}}\sqrt{\frac{1}{(1-2\theta)^{n_2}}} = \sqrt{\frac{1}{(1-2\theta)^{(n_1+n_2)}}}$$

que según (13) es la función generatriz de momentos de $\chi^2_{n_1+n_2}$

Conclusión. Dada una muestra aleatoria de n unidades de observación sobre la cual se observa la variable poblacional X con distribución normal (μ,σ); si se tipifica se obtiene la variable normal estándar $Z \sim N(0,1)$, con *n* valores aleatorios $Z_1, Z_2, \ldots, Z_n$; si estos valores se elevan al cuadrado y se suman, el estadístico resultante tiene una distribución χ^2 con *n* grados de libertad,

$$Z_1^2 + Z_2^2 + \ldots + Z_n^2 \sim \chi_n^2$$

Cuando los grados de libertad tienden a ∞, $\chi_n^2 \to Z \sim N(0,1)$

Ejemplo 4.8. La variable X tiene distribución chi-cuadrado con 20 grados de libertad:

$$X \sim \chi^2_{20}$$

¿Cuál es la probabilidad de que la variable X asuma valores inferiores o igual a 34.17?

Para hallar esta probabilidad es necesario utilizar una tabla para valores críticos de la chi cuadrado. Por ejemplo, la **Tabla E.4 Valores críticos de** χ^2 publicada por *Mark Berenson y David Levine***.** La tabla tiene los grados de libertad a la izquierda y la probabilidad acumulada desde el valor crítico a infinito en el extremo superior.

La variable tiene 20 grados de libertad, por ende es necesario ubicar la fila correspondiente a 20 grados. Luego debe buscarse sobre la fila el punto 34.17. Una vez hallado debe leerse arriba (en Areas de extremo superior) la probabilidad, en este caso 0.025.

$$P(X \geq 34.17) = 0.025$$

Otra forma de notación es $P(\chi^2 \geq 34.17) = 0.025$

La probabilidad de los menores a 34.17, que es lo que se busca, se halla de la siguiente manera:

$$P(X \leq 34.17) = 1 - P(X \geq 34.17) = 1 - 0.025 = 0.975$$

Distribución F de Snedecor o Fischer

Sea X una variable chi-cuadrado con *m* grados de libertad e Y otra variable chi-cuadrado con *n* grados de libertad, suponiendo que X e Y son independientes, entonces la variable aleatoria

$$F = \frac{X/m}{Y/n} = \frac{X_m^2}{X_n^2} \sim F_{m,n} \qquad (15)$$

se distribuye F de Snedecor con *m* y *n* grados de libertad en el numerador y denominador respectivamente.

De acuerdo al resultado alcanzado en (14), la función de densidad de X es $g(X) = \frac{e^{-\left(\frac{x}{2}\right)} x^{\frac{m}{2}-1}}{2^{\frac{m}{2}}\,\Gamma(m/2)}$, y la función de densidad de Y es $g(Y) = \frac{e^{-\left(\frac{Y}{2}\right)} Y^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\,\Gamma(n/2)}$

Por ende, la función de densidad conjunta de las variables *X* e *Y*, $g(X,Y)$, es:

$$g(X,Y) = g(X)g(Y) = \frac{e^{-\left(\frac{x}{2}\right)} x^{\frac{m}{2}-1}}{2^{\frac{m}{2}}\,\Gamma(m/2)} \frac{e^{-\left(\frac{Y}{2}\right)} Y^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\,\Gamma(n/2)}$$

donde $g(X) \sim \chi_m^2$ y $g(Y) \sim \chi_n^2$

agrupando términos con igual base se obtiene que

$$g(X,Y)=\frac{x^{\frac{m}{2}-1}\,Y^{\frac{n}{2}-1}\,e^{-\left(\frac{X+Y}{2}\right)}}{2^{\frac{m+n}{2}}\,\Gamma(m/2)\,\Gamma(n/2)}$$

de donde la función de distribución de probabilidad conjunta es

$$\iint g(X,Y)dxdy=\iint\frac{x^{\frac{m}{2}-1}\,Y^{\frac{n}{2}-1}\,e^{-\left(\frac{X+Y}{2}\right)}}{2^{\frac{m+n}{2}}\,\Gamma(m/2)\,\Gamma(n/2)}dx\,dy \qquad (16)$$

A partir de (15), se reexpresa *X* en función de *F* e *Y*

$$X=\frac{mYF}{n};\quad dX=\frac{mY}{n}dF\Rightarrow\frac{dX}{dF}=\frac{mY}{n}$$

reemplazando en (16), los valores de X por sus iguales

$$\iint g(F,Y)dFdY=\iint\frac{m^{\frac{m}{2}-1}\,Y^{\frac{m}{2}-1}\,F^{\frac{m}{2}-1}\,Y^{\frac{n}{2}-1}e^{-\left(\frac{mYF}{2n}\right)-\left(\frac{Y}{2}\right)}mY}{2^{\frac{m+n}{2}}\,\Gamma(m/2)\,\Gamma(n/2)\,n^{\frac{m}{2}-1}n}dF\,dY$$

$$=\iint\frac{m^{\frac{m}{2}-1}\,Y^{\frac{m}{2}-1}\,F^{\frac{m}{2}-1}\,Y^{\frac{n}{2}-1}e^{-\left(Y\frac{n+mF}{2n}\right)}mY}{n^{\frac{m}{2}-1}\,2^{\frac{m+n}{2}}\,\Gamma(m/2)\,\Gamma(n/2)\,n}dF\,dY \qquad (17)$$

al integrar $g(F,Y)dFdY$ respecto de Y se obtiene la función de densidad de probabilidad marginal de *F*, esto es

$$\int_0^\infty g(F,Y)dY=h(F)dF$$

Por lo tanto,

$$\int h(F)dF = \int_0^{\infty} \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2)\,\Gamma(n/2)} e^{-Y\frac{n+mF}{2n}} Y^{\frac{n+m}{2}-1} dY \qquad (17)$$

Se observa que:

1) $m^{\frac{m}{2}-1} m = m^{\frac{m}{2}-1+1} = m^{\frac{m}{2}}$

2) $n^{\frac{m}{2}-1} n = n^{\frac{m}{2}}$

3) $Y^{\frac{m}{2}-1} Y^{\frac{n}{2}-1} Y = Y^{\frac{m}{2}-1+\frac{n}{2}-1+1} = Y^{\frac{m+n}{2}-1}$

de donde:

$$h(F)dF = \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2)\,\Gamma(n/2)} \int_0^{\infty} e^{-Y\frac{n+mF}{2n}} Y^{\frac{n+m}{2}-1} dY$$

que es análoga a la integral de la Distribución Gamma hallada en (5)

$$\int_0^{\infty} e^{-\alpha X} X^{\lambda-1} dX = \frac{\Gamma(\lambda)}{\alpha^{\lambda}}; \quad \text{donde: } \alpha = \frac{n+mF}{2n} \text{ y } \lambda = \frac{n+m}{2}$$

Entonces:

$$h(F) = \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2)\,\Gamma(n/2)} \frac{\Gamma\left(\frac{n+m}{2}\right)}{\left(\frac{n+mF}{2n}\right)^{\frac{n+m}{2}}}$$

$$= \frac{m^{\frac{m}{2}} F^{\frac{m}{2}-1}}{n^{\frac{m}{2}} 2^{\frac{m+n}{2}} \Gamma(m/2)\,\Gamma(n/2)} \frac{\Gamma\left(\frac{n+m}{2}\right) 2^{\frac{m+n}{2}} n^{\frac{n+m}{2}}}{(n+mF)^{\frac{n+m}{2}}}$$

$$= \frac{\Gamma\left(\frac{n+m}{2}\right) m^{\frac{m}{2}} F^{\frac{m-2}{2}} n^{-\frac{m}{2}+\frac{n+m}{2}} (n+mF)^{-\left(\frac{n+m}{2}\right)}}{\Gamma(m/2)\,\Gamma(n/2)}$$

$$\Rightarrow h(F) = \frac{\Gamma\left(\frac{n+m}{2}\right) m^{\frac{m}{2}} F^{\frac{m-2}{2}} n^{\frac{n}{2}} (n+mF)^{-\left(\frac{m+n}{2}\right)}}{\Gamma(m/2)\,\Gamma(n/2)} \qquad (18)$$

es la función de densidad de la distribución *F* de Snedecor

Las características de esta distribución son

$$E(F) = \frac{n}{n-2}; \quad n > 2$$

$$V(F) = \frac{n^2(2n+2m-4)}{m(n-2)^2(n-4)}; \quad n > 4$$

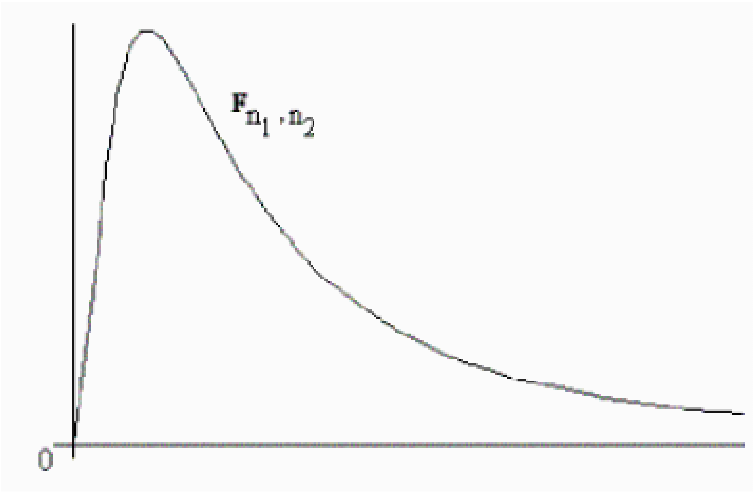


Figura 4.19. Distribución F

Ejemplo 4.9. ¿Cuál es la probabilidad de una variable F, que se distribuye F de Snedecor con 4 grados de libertad en el numerador y 6 grados de libertad en el denominador, para valores igual o mayores a 9,15?.

Para hallar esta probabilidad es necesario utilizar una tabla para valores críticos de la distribución F. Por ejemplo, la **Tabla E.5 Valores críticos de F** publicada por *Mark Berenson y David Levine***.** La tabla tiene los grados de libertad del denominador a la izquierda y los grados de libertad en el numerador arriba. Estos autores presentan una tabla para cada probabilidad en el extremo superior, es decir, la probabilidad acumulada desde el valor crítico a infinito.

La variable tiene 4 grados de libertad en el numerador y 6 grados de libertad en el denominador, por ende es necesario ubicar la celda en la cual, la columna y la fila correspondientes a estos grados de libertad, se crucen. En la Tabla donde se encuentre el punto crítico, allí se hallará la probabilidad.

La probabilidad asociada a los puntos mayores al valor crítico 9.15 es 0.01. Su notación es:

$$P(F_{4,6} \geq 9.15) = 0.01$$

Gomez Villegas (2005) realiza una Aproximación histórica sobre estas distribuciones y dice que "la distribución χ^2 fue introducida por Karl Pearson alrededor de 1900, al estudiar el problema de ajuste entre una distribución poblacional y unos datos muestrales. En 1904, la generalización de este problema al caso en que existan parámetros desconocidos en la distribución, dio lugar a que William Sealy Gosset (1876-1937) introdujera la distribución de Student, así llamada por ser Student el pseudónimo utilizado por Gosset para comunicar los resultados de sus investigaciones con muestra pequeñas, pues la empresa de elaboración de cerveza en la que trabajaba impedía la divulgación de los resultados obtenidos por sus empleados. Snedecor (1881-1974), estadístico estadounidense, introdujo la distribución que lleva su nombre alrededor de 1907. Snedecor trabajó en el laboratorio de estadística del University College fundado por Karl Pearson, junto a Student y Fischer. En agradecimiento a la importante aportación de Fischer a la Inferencia Estadística, Snedecor le dedicó la distribución del cociente de distribuciones χ^2 y la denotó mediante la letra F." (p. 43)

Relaciones entre las distribuciones Normal y χ^2

Teorema: Se tiene $X_1, X_2, \ldots, X_n$ datos provenientes de una muestra aleatoria de una variable X con esperanza μ y varianza σ^2. Se define un estimador de la varianza

$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1} \qquad (19)$$

en donde $\overline{X}$ es el promedio muestral, se tiene que:

(a) $E(S^2) = \sigma^2$

(b) Si X está distribuida normalmente, $\frac{n-1}{\sigma^2}S^2$ tiene una distribución χ^2 con (n-1) grados de libertad

Para demostrar la parte (**a**) del teorema, a la suma de cuadrados de los desvíos respecto de la media, se le resta y suma μ

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

Agrupando convenientemente y desarrollando el cuadrado

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

Distribuyendo el sumatorio

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

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

Multiplicando y dividiendo el segundo término por (-n)

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

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

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

De modo que,

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

Luego, se toma esperanza

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

Aplicando las propiedades de esperanza matemática

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

De modo que

$$E(S^2) = \frac{n\left(\sigma^2 - \sigma_{\overline{X}}^2\right)}{n-1} \qquad (20)$$

Entonces, $E(S^2) = \dfrac{n\sigma^2 - n\dfrac{\sigma^2}{n}}{n-1} = \dfrac{n\sigma^2 - \sigma^2}{n-1} = \dfrac{\sigma^2(n-1)}{n-1} = \sigma^2$

En síntesis

$$E(S^2) = \sigma^2 \qquad (21)$$

con lo que se demuestra la primera parte del Teorema.

El estimador

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n} \qquad (22)$$

es un estimador sesgado de la varianza poblacional; siendo el sesgo igual a $\frac{n-1}{n}$. Para demostrarlo se parte consignando

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

de donde $\frac{n}{n-1}E(\hat{\sigma}^2) = \sigma^2$

Esto puede escribirse reemplazando el estimador por su igual

$$\frac{n}{n-1}E\left[\frac{\sum(X_i - \overline{X})^2}{n}\right] = \sigma^2$$

esto es,

$$E\left(\frac{\sum(X_i - \overline{X})^2}{n-1}\right) = \sigma^2 \qquad (23)$$

Pero $\frac{\sum(X_i - \overline{X})^2}{n-1} = S^2$; es decir,

$$E(S^2) = \sigma^2 \qquad (24)$$

con lo que queda demostrado que el sesgo es $(n-1)/n$. De este modo S^2 es un estimador insesgado de la varianza poblacional y el factor de corrección que permite evitar el sesgo es $n/(n-1)$.

Se demostrará ahora la parte (**b**) del teorema. El enunciado dice que si

$$X \sim N(\mu,\sigma^2)$$

Entonces

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \qquad (25)$$

Si

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma^2} = \frac{(n-1)\dfrac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} \qquad (26)$$

En general, una suma de variables aleatorias independientes con distribución normal, elevadas al cuadrado, se distribuye χ^2 con grados de libertad determinados por la cantidad de variables menos 1.

Ejemplo 4.10. Considerando el caso especial de n=2

$$\sum_{i=1}^{2}(X_i-\overline{X})^2 = (X_1-\overline{X})^2+(X_2-\overline{X})^2$$

$$=\left(X_1-\frac{X_1+X_2}{2}\right)^2+\left(X_2-\frac{X_1+X_2}{2}\right)^2$$

$$=\frac{(2X_1-X_1-X_2)^2}{4}+\frac{(2X_2-X_1-X_2)^2}{4}$$

$$= \frac{(X_1 - X_2)^2 + (X_2 - X_1)^2}{4} = \frac{(X_1 - X_2)^2}{2}$$

Puesto que X_1 y X_2 están independientemente distribuidos $N(\mu, \sigma^2)$ se encuentra que $(X_1 - X_2)$ tiene una distribución $N(0, 2\sigma^2)$, luego

$$\frac{\sum_{i=1}^{2}(X_i - \overline{X})^2}{\sigma^2} = \left(\frac{(X_1 - X_2) - 0}{\sqrt{2}\,\sigma}\right)^2 = \frac{1}{\sigma^2}S^2$$

tiene la distribución χ^2 con 1 grado de libertad.

Distribución T de Student

Sea Z una variable aleatoria normal estándar

$$Z = \frac{x - \mu_x}{\sigma_x}$$

y X^2 una variable aleatoria chi cuadrado con n-1 grados de libertad

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

Se define

$$t = \frac{Z}{\sqrt{X^2 / n-1}} \qquad (27)$$

(27) sigue una distribución t de Student con n-1 grados de libertad.

Reemplazando en (27), las variables Z y X^2

$$t = \frac{\frac{X - \mu_X}{\sigma_X}}{\sqrt{\frac{(n-1)}{\sigma^2}\frac{s^2}{(n-1)}}} = \frac{x - \mu}{s}$$

Se dice que el estadístico *t* se distribuye como una T de Student con $(n-1)$ grados de libertad: $t \sim t_{n-1}$.

La distribución t se caracteriza por

$$E(t) = 0; \quad n > 0$$

$$V(t) = \frac{n}{n-2}; \quad n > 2$$

La distribución *t* es simétrica con respecto a 0 y asintóticamente tiende a una distribución normal tipificada.

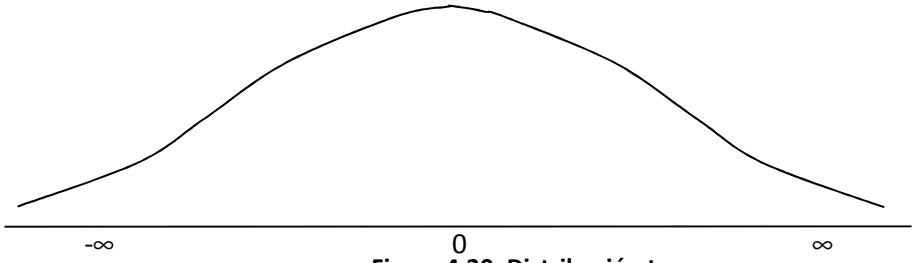


Figura 4.20: Distribución t.

Si se eleva al cuadrado la expresión de t, el resultado se puede escribir como $t^2 = \frac{Z^2/1}{V/v}$ en donde Z^2, al ser el cuadrado de una variable normal tipificada, tiene una distribución χ_1^2.

Así pues, $t^2 = F(1, v)$; es decir, el cuadrado de una variable t con v grados de libertad es una F con (1,v) grados de libertad.

Ejemplo 4.11. En una variable X, que se distribuye t de Student con 10 grados de libertad, ¿cuál es la probabilidad de asumir valores iguales o superiores a 1.3722?

Para hallar esta probabilidad es necesario utilizar una tabla para valores críticos de la distribución t. Por ejemplo, la **Tabla E.3 Valores críticos de t** publicada por *Mark Berenson y David Levine***.** La tabla tiene los grados de libertad a la izquierda y la probabilidad acumulada desde el punto crítico a infinito arriba en el extremo superior.

La variable tiene 10 grados de libertad, debe ubicarse la fila correspondiente a estos grados de libertad y en ellos ubicar el punto crítico 1.3722. Luego leer en la fila correspondiente a Areas de extremo superior la probabilidad desde 1.3722 hasta infinito.

La probabilidad de que la variable asuma valores inferiores a 1.3722 es:

$$P(t_{10} \leq 1.3722) = 1 - P(t_{10} \geq 1.3722) = 1 - 0.10 = 0.90$$

CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

Caso 4.1: Ingreso medio en los hogares de Río Cuarto

En el Cuadro 4.1 se clasifica a los hogares de Río Cuarto de acuerdo al nivel de ingresos declarados a la Encuesta Permanente de Hogares en Mayo de 2003.

Cuadro 4.1 Ingreso en Río Cuarto	
Nivel de ingreso mensual del hogar	Total de hogares (en %)
Hasta 150	4.93
Entre 151 y 300	13.21
Entre 301 y 450	13.41
Entre 451 y 600	13.41
Entre 601 y 750	11.64
Entre 751 y 1000	13.21
Entre 1001 y 1250	7.49
Entre 1251 y 1500	4.14
Entre 1501 y 2000	4.93
Entre 2001 y 3000	3.16
Más de 3000	2.37
Ingreso desconocido	8.10
Total de hogares	100.00

FUENTE: Encuesta Permanente de Hogares. Mayo de 2003

Con esta información, y teniendo en cuenta que la cantidad de hogares relevados fue de 636, se debe calcular:

1) el ingreso medio de los hogares

2) el desvío del ingreso

3) la probabilidad de que un hogar, extraído al azar de la población, tenga

   - ingreso superior a $3000.00
   - ingreso inferior a $750.00
   - ingreso entre $800.00 y $1300.00
   - ingreso inferior a $125.00
   - ingreso inferior a $2400.00

Preguntas

4.1. Explica las diferencias entre

Dato bruto y arreglo ordenado
Distribuciones de frecuencias, distribuciones de frecuencias relativas y distribuciones de porcentajes.
Histogramas, polígonos y ojivas (polígono acumulado).

4.2. Dar la definición de los siguientes términos:

Variable Continua
Parámetros
Variable Discreta
Distribución Muestral
Encuesta
Distribución de Frecuencia
Experimento
Población
Polígono de Frecuencia
Distribución Teórica
Histograma
Variable Categórica
Muestra
Estimador
Distribución de Probabilidad

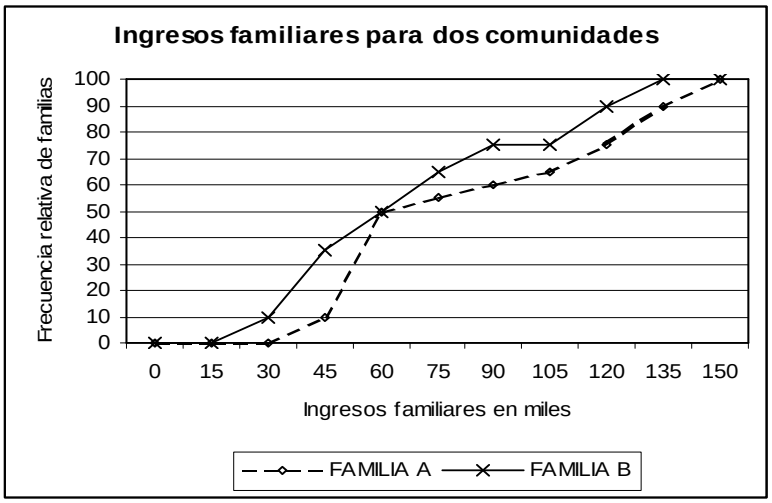
4.3. El cuadro muestra, en intervalos de clase, los importes de alquiler cobrados por 200 departamentos

a. ¿Cuáles son los límites nominales inferior y superior de la primera clase?
b. ¿Cuáles son los límites exactos inferior y superior de la primera clase?
c. El intervalo de clase ¿es igual en todas las clases de la distribución?. ¿Cuál es el tamaño del intervalo?
d. ¿Cuál es el punto medio de la primera clase?
e. ¿Cuáles son los límites exactos inferior y superior de la clase en la que se tabuló la mayor cantidad de rentas de departamentos?
f. Si se reporta una renta de $439,50 identifica los límites nominales inferior y superior de la clase en la que se incluiría esta renta.
g. Construye un Histograma

h. Construye un polígono de frecuencias y una curva de frecuencias.
i. Describe la curva de frecuencias anterior desde el punto de vista de la asimetría.
j. Construye una distribución de frecuencias acumuladas
k. Presenta la distribución de frecuencias acumuladas anterior en forma gráfica mediante una ojiva.

Cuadro 4.2. Departamentos según nivel de renta	
Renta Mensual (en miles de $)	Número de Departamentos
350-379,99	3
380-409,99	8
410-439,99	10
440-469,99	13
470-499,99	33
500-529,99	40
530-559,99	35
560-589,99	30
590-619,99	16
620-649,99	12
Total	200

4.4. La figura contiene los polígonos de frecuencias relativas acumuladas de los ingresos familiares de dos muestras aleatorias (A y B) de 200 familias cada una, trazados en dos comunidades.



Con base en estos datos, contesta cada una de las siguientes preguntas:

a) ¿Cuántas familias en la muestra A, tienen ingresos mayores a $120.000?
b) ¿Cuál es el porcentaje de familias en la muestra A, con ingresos menores de $90.000?
c) ¿Cuál muestra tiene un recorrido más grande de ingresos?
d) ¿Cuántas familias en la muestra B, tienen ingresos entre $90.000 y $105.000?
e) ¿Cuál de las muestras, muestra A o muestra B, tiene más ingresos familiares superiores a $60.000?
f) ¿Qué porcentaje de familias de la muestra A, ganan menos de $60.000?
g) ¿Qué porcentaje de familias de la muestra A, ganan más de $60.000?
h) ¿Qué muestra tiene más ingresos inferiores a $120.000?

4.5. Dados los conjuntos de datos del Cuadro 1-5

a) Forma la distribución de frecuencias y la distribución de porcentajes para cada grupo.
b) Traza los histogramas de frecuencias para cada grupo.
c) Traza los polígonos de porcentajes para cada grupo.
d) Forma las distribuciones de porcentajes acumuladas para cada grupo.
e) Traza las ojivas (polígono de porcentaje acumulada), para cada grupo.
f) Escribe un breve informe, en el cual se compare y contraste los dos grupos.

4.6. ¿Porqué es válida la siguiente igualdad?

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

siendo: $\sum_{i=1}^{n}x_i^2$ la suma del cuadrado de cada observación

$\left(\sum_{i=1}^{n}x_i\right)^2$ el cuadrado de la suma total

Demostrar

Cuadro 4.3. Arreglo ordenados ordenados de promedios de bateo (Muestra de 40 jugadores de cada liga)		
	LIGA AMERICANA	LIGA NACIONAL
	BATEO	BATEO
1	0.184	0.201
2	0.204	0.205
3	0.207	0.214
4	0.209	0.220
5	0.210	0.227
6	0.219	0.233
7	0.220	0.234
8	0.220	0.242
9	0.225	0.242
10	0.227	0.245
11	0.227	0.247
12	0.235	0.248
13	0.240	0.250
14	0.241	0.253
15	0.247	0.253
16	0.248	0.253
17	0.252	0.253
18	0.253	0.261
19	0.255	0.261
20	0.255	0.261
21	0.255	0.263
22	0.268	0.263
23	0.271	0.265
24	0.272	0.274
25	0.278	0.275
26	0.279	0.279
27	0.280	0.287
28	0.282	0.290
29	0.291	0.300
30	0.292	0.306
31	0.294	0.308
32	0.298	0.310
33	0.299	0.319
34	0.301	0.320
35	0.302	0.320
36	0.313	0.338
37	0.316	0.342
38	0.321	0.356
39	0.327	0.359
40	0.348	0.375

Problemas

4.1. Calcular las siguientes probabilidades

$P(\chi^2_{12} \leq 26.2) =$ $P(\chi^2_{12} > 6.3) =$

$P(5.23 \leq \chi^2_{12} \leq 14.8) =$ $P(\chi^2_{12} < X_0) = 0.90$

4.2. Dada una variable G que se distribuye como una t de Student con 15 grados de libertad, calcular

$P(g < 1.341) =$ $P(-2.131 < g < 2.131) =$

$P(g > -1.753) =$ $P(g > 0.691) =$

4.3. Dada una variable aleatoria F se desea saber

$P\left(\frac{1}{3.85} < F_{10,8} < 4.30\right) =$ $P(F_{10,8} > F_0) = 0.95$

$P(3.35 < F_{10,8} < 4.30) =$ $P(F_{10,8} > 4.30) =$

Referencias

- **Berenson, M y Levine, D.** *Estadística Básica En Administración*. México: Prentice Hall, 1996.
- **Dixon, W.J. y Massey, F.J.** *Introduction to Statistical Analysis*. Nueva York: Mc Graw Hill, 1957.
- **Freeman, Harold.** *Introducción a La Inferencia Estadística*. México: Trillas, 1963.
- **Gomez Villegas, Miguel Angel.** *Inferencia Estadística*. España: Ediciones Díaz de Santo, 2005.
- **Hernández Sampieri, R.; Fernández Collado, C. y Baptista Lucio, P.** *Metodología De La Investigación*. México: McGraw Hill, 2010.

- **Kazmier, L y Diaz Mata, A.** *Estaística Aplicada a La Administración Y a La Economía*. México: McGraw Hill, 1993.
- **Kinnear, T.y Taylor, J.** *Investigación De Mercado. Un Enfoque Aplicado*. McGraw Hill, 1993.
- **Mao, J.C.T.** *Análisis Financiero*. Buenos Aires: El Ateneo, 1980.
- **Mendenhall, W. Wackerly, D. Scheaffer, R.** *Estadística Matemática Con Aplicaciones*. México: Grupo Editorial Iberoamérica, 1990.
- **Meyer.** *Probabilidad Y Aplicaciones Estadísticas*. México: Fondo Educativo Interamericano SA, 1973.
- **Pugachev, V.S.** *Introducción a La Teoría De Las Probabilidades*. Moscú: Editorial Mir, 1973.
- **Spiegel, Murray R.** *Teoría Y Problemas De Probabilidad Y Estadística*. México: McGraw Hill, 1976.
- **Tramutola, C.D.** *Modelos Probabilísticos Y Decisiones Financieras*. Capital Federal: E.C.Moderan, 1971.