CAMPUS EXTENS

29 de agosto de 2019

Índice general

1. Estadística descriptiva **6**

1.1. Conceptos básicos 6

1.1.1. Estadística descriptiva 7

1.1.2. Estadística Inferencial 7

1.2. Estadística descriptiva 7

1.2.1. Datos y series estadísticas 7

1.2.2. Clasificación de los datos 8

1.2.3. Descripción de una serie 9

1.2.4. Representación gráfica 9

1.3. Variables unidimensionales 11

1.3.1. Descripción numérica 11

1.3.2. Descripción gráfica 15

1.4. Análisis de las distribuciones 17

1.4.1. Medidas de posición 18

1.4.2. Medidas de dispersión 24

1.4.3. Perfil de una distribución 29

1.4.4. Medidas de simetría 30

1.4.5. Medidas de apuntamiento 33

1.5. Variables multidimensionales 35

1.5.1. Descripción numérica: caso bidimensional 35

1.5.2. Distribuciones marginales 38

1.5.3. Distribuciones condicionadas 40

1.5.4. Momentos bidimensionales 42

1.5.5. Independencia e incorrelación 45

1.6. Asociación, concordancia y correlación 46

1.6.1. Introducción 46

1.6.2. Coeficientes de asociación 47

1.6.3. Correlación lineal 49
1.6.4. Correlación ordinal 51

2. Probabilidad **54**
2.1. Espacio muestral 54
2.1.1. Propiedades del álgebra de sucesos 54
2.2. Definición de probabilidad 55
2.2.1. Probabilidad Condicionada 56
2.3. Fórmula de la probabilidad total 57
2.4. Regla de Bayes 57
2.5. Independencia Estadística 57

3. Variables aleatorias **59**
3.1. Variables aleatorias 59
3.1.1. Definición de variable aleatoria 59
3.1.2. Tipos de variables aleatorias 60
3.2. Distribuciones de probabilidad para v.a. discretas. 61
3.2.1. Función de probabilidad de una variable aleatoria discreta 62
3.2.2. Propiedades de la función de probabilidad 63
3.3. Función de distribución 64
3.4. Momentos de variables aleatorias discretas 66
3.4.1. Esperanza para variables aleatorias discretas 66
3.4.2. Esperanzas de funciones de variables aleatorias discretas 67
3.4.3. Momentos de una v.a. discreta 69
3.4.4. Varianza de una variable aleatoria discreta 70
3.5. Variables aleatorias continuas 72
3.5.1. Distribuciones de probabilidad de v.a. continuas 72
3.5.2. Función de densidad 74
3.6. Momentos para variables aleatorias continuas 77
3.6.1. Esperanza de una v.a. continua 77
3.6.2. Varianza de una v.a. continuas 77
3.7. Transformación de variables aleatorias 78
3.7.1. Transformaciones de v.a. discretas 79
3.7.2. Transformaciones de v.a. continuas 79
3.8. Desigualdad de Chebyshef 80
3.8.1. Desigualdad de Markov 80

4. Distribuciones notables **83**
4.1. Algunas variables aleatorias discretas 83
4.1.1. Bernoulli 83
4.1.2. Binomial: 84
4.1.3. Geométrica 85
4.1.4. Binomial negativa 86
4.1.5. Poisson 86
4.1.6. Distribución hipergeométrica: 88
4.2. Algunas variables aleatorias continuas 89
4.2.1. Distribución uniforme en el intervalo (a,b): 89
4.2.2. Distribución exponencial (exponencial negativa): 90
4.2.3. Distribución normal o Gaussiana 91

5. Variables aleatorias vectoriales **95**
5.1. Variables aleatorias bidimensionales 95
5.1.1. Variables aleatorias conjuntamente discretas 96
5.1.2. Variables aleatorias bidimensionales continuas 98
5.1.3. Distribuciones condicionales. (OPCIONAL) 101
5.2. Valores Esperados 110
5.3. Medida de la variación conjunta 111
5.3.1. Covarianza 111
5.4. Propiedades de las sumas de v.a. 112
5.4.1. Matriz de varianzas-covarianzas, matriz de correlaciones 113
5.5. Distribución normal bivariante 113
5.5.1. Normal bivariante estándar 113
5.5.2. La ley normal multivariante 115
5.6. Teorema del Límite Central (T.L.C.) 115

6. Muestreo Estadístico **120**
6.1. Conceptos básicos 120
6.1.1. Tipos de muestreo 121
6.2. Inferencias 124
6.2.1. Distribución muestral de un estadístico 124
6.2.2. Distribución la media muestral 126
6.2.3. Distribución de una proporción muestral 130
6.2.4. Distribución muestral de la varianza muestral 132

7. Inferencia estadística: estimación de parámetros. 135
7.1. Introducción . . . 135
7.2. Estimadores . . . 135
7.2.1. Estimadores insesgados . . . 136
7.3. Métodos para calcular estimadores . . . 138
7.4. Estimación por intervalos . . . 141
7.4.1. Intervalo de confianza para la media de una población normal: varianza poblacional conocida . . . 141
7.4.2. Intervalo de confianza para la media poblacional: tamaños muestrales grandes . . . 145
7.4.3. Intervalo de confianza para la media de una población normal: varianza poblacional desconocida . . . 146
7.4.4. Intervalos de confianza para una proporción . . . 148
7.4.5. Intervalo de confianza para la varianza de una población normal . . . 149

8. Inferencia estadística: contraste de hipótesis 153
8.1. Tipos de hipótesis . . . 154
8.2. Tipos de Error en un contraste . . . 155
8.3. ¿Inocente o culpable? . . . 157
8.4. Ejemplo de un contraste de hipótesis para la media de una distribución normal: varianza poblacional conocida . . . 158
8.5. Terminología: . . . 159
8.6. Ejemplo tabla de un contraste para la media poblacional de una población normal con varianza poblacional conocida . . . 159
8.7. Método de los seis pasos . . . 159
8.8. Reglas de decisión para contraste de la media de una población normal: varianza poblacional conocida . . . 161
8.9. Contraste para la media: tamaños muestrales grandes . . . 162
8.9.1. Reglas de decisión para el contraste de una media: Tamaños muestrales grandes . . . 164
8.10. Contrastes para la media de una población normal: varianza poblacional desconocida . . . 164
8.10.1. Reglas de decisión para el contraste de una media de una distribución normal: varianza poblacional desconocida . . . 164
8.11. Contraste para la varianza de una distribución normal. . . . 166

8.11.1. Reglas de decisión para el contraste de la varianza de una población normal 166
8.12. Contrastes para la proporción muestral: muestras grandes . . . 168
8.12.1. Reglas de decisión para el contraste de una proporción muestral: tamaño muestral grande 169
8.13. Pruebas de bondad de ajuste 171
8.13.1. Un contraste de bondad de ajuste: distribución totalmente conocida 171
8.13.2. Un contraste de bondad de ajuste: algún parámetro poblacional desconocido 174
8.13.3. Prueba de bondad de ajuste de Kolgomorov-Smirnov (K-S) 175
8.14. Contraste de las medias de dos poblaciones normales o tamaños muestrales grandes 177
8.15. Contraste de dos proporciones 181
8.16. Muestras dependientes 181
8.17. Comparación de dos varianzas 184

Capítulo 1

Estadística descriptiva

Con estas nota de clase iniciamos el curso. El objetivo de este primer capítulo es dar unas nociones básicas de descripción de datos. Tiene que quedar claro que hay muchas otras técnicas básicas y avanzadas de descripción de datos que no veremos. Para la resolución de problemas se tienen que utilizar tanto la resolución tradicional con papel, lápiz y calculadora, como la resolución con hojas de cálculo (Excel,OpenOffice. . .) o paquetes estadísticos (SPSS, R,. . .), o programas de propósito todavía más general (Mathematica, Octave. . .). Algunos de estos paquetes están disponibles en los ordenadores de las aulas de informática de la universidad y de estos algunos otros tienen licencias que os permiten bajarlos por internet de forma gratuita; como el OpenOffice (`www.openoffice.org`)o el paquete estadístico R (`http://www.r-project.org`) en versiones para cualquier sistema operativo usual.

1.1. Conceptos básicos

La estadística es aquella ciencia que tiene por objeto dar métodos tanto para la recopilación, organización y análisis de datos que provienen de un grupo de individuos, así como para la decisión de aceptar o rechazar ciertas afirmaciones o leyes.

Conceptos básicos:

- Población: Conjunto de todos los individuos que tienen en común alguna característica observable y de los que se desea estudiar un determinado fenómeno. Sus características se definen como parámetros.

Tipos de población: finita o infinita.

- Muestra: Es un subconjunto de la población del que se espera represente a la población y en el que se efectúa el estudio del fenómeno. Sus características se definen como estadísticos.

1.1.1. Estadística descriptiva

La Estadística Descriptiva se define como aquella ciencia dedicada a describir las regularidades o características de un conjunto de datos (muestra).

Tareas de la Estadística Descriptiva:

- Organización de los datos numéricos de la muestra mediante tablas y representaciones gráficas.

- Análisis de los datos obtenidos mediante la obtención de índices representativos de la muestra como son las medidas de tendencia central y de dispersión.

1.1.2. Estadística Inferencial

La Estadística Descriptiva basa su estudio sobre las muestras. Ver si éstas son representativas de la población es tarea de la estadística inferencial.

La misión principal de la Estadística Inferencial es extraer conclusiones de las características de la población mediante una muestra representativa de la misma.

1.2. Estadística descriptiva

1.2.1. Datos y series estadísticas

El análisis estadístico parte siempre de un conjunto de datos. Dado un conjunto de objetos cualesquiera (individuos, países, municipios, etc...), la observación de una determinada característica o medida de ésta (cualidad o atributo) da lugar a un dato estadístico.

Ejemplos:

- Población: Países del mundo.

Característica a estudiar: Producto Interior Bruto (P.I.B.). Los datos estadísticos serán los valores del P.I.B. de los países en cuestión.

- Población: Estudiantes de segundo curso de Informática.

  Característica a estudiar: Altura. Los datos estadísticos serán los valores de la altura en cm. para cada estudiante.

1.2.2. Clasificación de los datos

Una clasificación elemental de los datos estadísticos es la siguiente, dividida en tres criterios:

- Tipo de dato
  - Cualitativos o de atributos: cuando la comparación entre ellos sólo puede ser de igualdad o desigualdad.

    Por ejemplo: color de los ojos, afiliación política, lugar de residencia, etc,...
  - Ordinales: cuando los datos no son numéricos y la comparación entre ellos establece un orden.

    Por ejemplo: estado de ánimo (valores posibles: depresivo, normal y eufórico), estudios (valores posibles: ninguno, primarios, secundarios, superiores), etc...
  - Cuantitativas: cuando los datos son numéricos. Entre los datos cuantitativos podemos señalar dos tipos más:
    - Discretos: cuando entre dos posibles valores no hay otro. Por ejemplo: número de hijos de una familia, número de letras de una palabra en un texto, etc,...
    - Continuas: cuando entre dos posibles valores, siempre podemos encontrar otro valor posible. Por ejemplo: altura, intereses de una cuenta bancaria, etc,...
- Dimensión
  - Unidimensionales: si sólo se considera una única característica.

    Ejemplos: altura, edad, etc,...

- Multidimensionales: si se consideran conjuntamente varias características.

  Ejemplos: edad y altura, altura y peso, edad, altura y sexo, etc,...

- Tiempo

  - Atemporales: cuando los datos no están referidos, o no se considera, el momento de tiempo en el que fueron obtenidos.

    Ejemplos: color de los ojos de cierto conjunto de individuos, peso de los estudiantes que han asistido a la clase de hoy, etc,...

  - Temporales o series cronológicas: en caso contrario.

    Ejemplos: P.I.B. anual de España durante el periodo 1980 hasta 2004, número de turistas llegados al aeropuerto de Palma el mes de agosto durante los años 1970 al 2004, etc,...

1.2.3. Descripción de una serie

Una vez realizada la recogida de datos, se ha de hacer una representación numérica y descriptiva de éstos que se adecue de la mejor manera posible al estudio que se desea realizar.

Cuando los datos son atemporales y unidimensionales, es habitual presentarlos en forma de distribución de frecuencias asociando a cada modalidad o valor las veces que se repite (frecuencias absolutas).

En el caso en que los datos sean bidimensionales y atemporales se suele hacer una tabla de frecuencias de denominada también como tabla de contingencia.

En el caso en que los datos sean series cronológicas, se representan como una función matemática en el tiempo, es a decir, una serie de valores (t, Y_t) donde el primer elemento es el tiempo y el segundo valor es el dato en ese tiempo.

1.2.4. Representación gráfica

Una vez descrita la serie estadística en forma de tabla, el paso siguiente es hacer una representación gráfica de la misma porque lo interesante es observar de golpe el aspecto general de los datos.

Veamos con unos cuantos ejemplos en los que esquemáticamente veremos las representaciones gráficas más habituales.

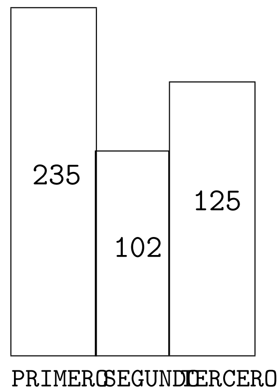


Figura 1.1: Alumnos de los distintos cursos de Informática

- Diagramas de barras. Como ejemplo, ver en la figura 1.1 la representación gráfica de la cantidad de alumnos que hay en distintos cursos de informática.
- Gráficos de sectores: Es un gráfico circular dividido en sectores donde cada sector representa el tanto por ciento de individuos que pertenecen a una determinada modalidad.
- Pictogramas: Son representaciones gráficas que guardan relación con el objeto de estudio estadístico.

Diagramas causa-efecto

Se utilizan en las empresas e industrias, para representar los factores que influyen en un fenómeno.

Por ejemplo, consideremos el diagrama de la figura 1.2:

En la figura anterior tenemos un problema en el que inciden las materias primas (MP, 2 diferentes), los métodos de elaboración (ME, 3 diferentes), la temperatura (TE, 15 grados a 20, o bien de 20 a 30) y los turnos de trabajo (TU, 2 diferentes).

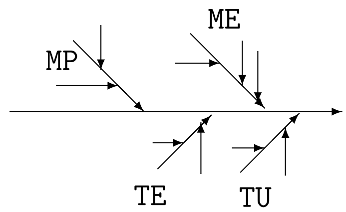


Figura 1.2: Diagrama causa-efecto

1.3. Variables unidimensionales

1.3.1. Descripción numérica

La representación ordenada de las observaciones de una muestra se hace mediante una tabla numérica, en la que aparecen los valores de la variable y sus frecuencias absolutas; número de veces que aparece cada dato en la muestra. Además la tabla se puede completar con las frecuencias absolutas acumuladas.

Mas concretamente, sean $x_1, \ldots, x_n$ las observaciones de una muestra, de tamaño n. Supongamos que los distintos valores que aparecen en las muestra son $X_1, \ldots, X_J$ y que, si es posible están ordenados de menor a mayor:

$$X_1 < X_2, < \ldots, X_J.$$

Denotaremos por n_1 las veces que aparece el valor X_1 en la muestra, n_2 las veces que aparece el valor X_2, ..., y n_J las veces que aparece el valor X_J. Las frecuencias absolutas serán, por lo tanto los valores: $n_j,\ j = 1, \ldots, J$.

Es evidente que se verifica la siguiente relación:

$$n = \sum_{j=1}^{J} n_j$$

Las frecuencias relativas se definen como el cociente entre las absolutas y el tamaño de la muestra: $f_j = \frac{n_j}{N}$. La frecuencia relativa de X_j es el tanto por uno de veces que aparece en la muestra. En ocasiones se utilizan los tantos por cien, por mil,..., pero en la práctica para el cálculo es más cómodo utilizar tantos por uno.

Cuando los datos pueden ser ordenados, se define la frecuencia absoluta acumulada N_j del valor X_j como el número de observaciones que son menores o iguales a X_j. Se verifica la siguiente relación:

$$N_j = \sum_{k=1}^{j} n_k.$$

La frecuencia relativa acumulada F_j del valor X_j es el cociente entre N_j y n, que corresponde a la suma de las frecuencias relativas de los datos anteriores a X_j. Así podemos escribir

$$F_j = \sum_{k=}^{j} f_k = \frac{N_j}{N}.$$

Todos los resultados anteriores se pueden presentar en forma de tabla, como por ejemplo la que sigue:

X_j	n_j	N_j	f_j	F_j
X_1	n_1	N_1	f_1	F_1
X_2	n_2	N_2	f_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_J	n_J	N_J=n	f_J	$F_J = 1$
Suma $\sum$	n		1	

Cuando los datos son continuos y de una precisión elevada (por ejemplo el tiempo en segundos, con una precisión hasta milisegundos, de una transmisión) o discretos con un número elevado de posibles valores (por ejemplo tamaño en bits de ficheros en un HD), se corre el riesgo de que las frecuencias resuman escasamente la muestra, es decir que las frecuencias absolutas de cada valor sean 1 o a lo más 2. En ambos casos se suele recurrir al conteo de datos por grupos o intervalos de valores a los que se denomina clases; es lo que se llama recuento de datos agrupados.

Consideremos el caso del peso en kilogramos de una persona. Cuando decimos yo peso 60 kilos ¿qué estoy diciendo en realidad? o si consideramos la edad y digo que tengo 21 años ¿qué estoy diciendo en realidad?

En la variable continua peso en kilos tenemos que los valores se calculan hasta las unidades, si estamos haciendo una medida en forma correcta decir que pesamos 60 Kg. debería ser equivalente a decir que pesamos $60 \pm 0{,}5$ Kg, es decir nuestro instrumento de medida debería medir así; pues el error cometido será la mitad de la precisión del instrumento de medida. Lo mismo sucede con la edad; cometemos menos error si decimos que tenemos 21 años cuando tengamos $21 \pm 0{,}5$ años [1]. Así que 60 Kg. corresponde al intervalo $(59{,}5, 60{,}5)$ este tipo de extremos recibe el nombre de límites reales y puede abarcar más de un tipo de dato, por ejemplo el intervalo $(59, 5, 70, 5)$. Por contra tenemos los a veces llamados límites aparentes así podríamos definir el agrupamiento de 60 a 70 Kg.

De forma más general tenemos que si las observaciones vienen dadas con una precisión de una cifra decimal, los extremos reales de los intervalos serán de la forma #.#5, donde el símbolo # simboliza un dígito para la parte decimal y uno o varios para la entera.

Por ejemplo si nos dan los datos:

$$1{,}3, 3{,}6, 4{,}7, 4{,}9, 1{,}2, 0{,}6,$$

unos posibles intervalos de agrupamiento con límites reales y de amplitud 1 son:

$$\begin{gathered}[0{,}55, 1{,}55),\\ [1{,}55, 2{,}55),\\ [2{,}55, 3{,}55),\\ [3{,}55, 4{,}55),\\ [4{,}55, 5{,}55).\end{gathered}$$

Si los valores vienen dados con dos cifras decimales de precisión, los extremos de los intervalos serían de la forma #.##5.. Por ejemplo, si los datos son:

$$0{,}23, 1{,}26, 3{,}54, 5{,}76, 8{,}76, 3{,}67,$$

unos posibles intervalos con límites reales de amplitud 2 son:

$$\begin{gathered}[0{,}225, 2{,}225),\\ [2{,}225, 4{,}225),\\ [4{,}225, 6{,}225),\\ [6{,}225, 8{,}225),\\ [8{,}225, 10{,}225).\end{gathered}$$

[1] Es evidente que las personas no hacemos esto y que decimos que tenemos 21 años hasta el día anterior a nuestro aniversario, con lo cual durante la mitad del año, de cada año de nuestra vida, estamos cometiendo un error superior a medio año.

Para escoger el primer extremo se suele calcular el mínimo de la muestra y se toma como valor mínimo el extremo inferior del límite real de ese valor.

En el primer ejemplo, el valor mínimo era 0,6; por lo tanto el primer extremo es $0{,}6-0{,}05=0{,}55$. En el segundo ejemplo, el valor mínimo es 0,23; por lo tanto, el primer extremo será $0{,}23-0{,}005=0{,}225$.

Los otros extremos se obtienen sumando una amplitud, de la misma precisión que los datos, desde el valor mínimo.

Como receta general, que no es de obligado cumplimiento, a la hora de agrupar se recomienda:

i) Decidir el número de clases a considerar. Este número no debe ser inferior a 5 y como máximo entre 15 y 20. Se pueden utilizar las siguientes heurísticas, si J es el número de clases tomar $J\geq\sqrt{n}$ (para tamaños muestrales inferiores a 150) o también $2^J\geq n$.

ii) Seleccionar los límites de clase que definen los intervalos, de forma que, si es posible, todos tengan la misma amplitud, salvo quizás los extremos.

iii) Intentar no dejar clases con frecuencias muy bajas, para evitar esto se pueden unir estas clases a una de sus adyacentes.

A cada clase o agrupamiento se le asigna ahora un valor representativo que recibe el nombre de marca de clase. Se suele tomar, salvo que se diga lo contrario, como marca de clase el punto medio de un intervalo; que se obtiene dividiendo por dos la amplitud del mismo.

En el primer ejemplo las marcas de clase son:

$[0{,}55, 1{,}55)$	1,05
$[1{,}55, 2{,}55)$	2,05
$[2{,}55, 3{,}55)$	3,05
$[3{,}55, 4{,}55)$	4,05
$[4{,}55, 5{,}55)$	5,05

mientras que para el segundo son estas:

$[0{,}225, 2{,}225)$	1,225
$[2{,}225, 4{,}225)$	3,225
$[4{,}225, 6{,}225)$	5,225
$[6{,}225, 8{,}225)$	7,225
$[8{,}225, 10{,}225)$	9,225

La tabla final es:

intervalos	(Marca de clase) X_j	n_j	N_j	f_j	F_j
$[L_1, L_2)$	X_1	n_1	N_1	f_1	F_1
$[L_2, L_3)$	X_2	n_2	N_2	f_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[L_J, L_{J+1})$	X_I	n_I	N_I	f_I	F_I
Suma $\sum$		n		1	

Ejemplo 1 *Consideremos las puntuaciones de* 50 *aspirantes a un puesto de trabajo:*

8	11	11	8	9	10	16	6	12	19
13	6	9	13	15	9	12	16	8	7
14	11	15	6	14	14	17	11	6	9
10	19	12	11	12	6	15	16	16	12
13	12	12	8	17	13	7	12	14	12

La tabla de frecuencias agrupadas con límites reales y amplitud fija de los intervalos 3 *es:*

intervalos	X_j	n_j	N_j	f_j	F_j
[5,5, 8,5)	*7*	*11*	*11*	*0.22*	*0.22*
[8,5, 11,5)	*10*	*11*	*22*	*0.22*	*0.44*
[11,5, 14,5)	*13*	*17*	*39*	*0.34*	*0.78*
[14,5, 17,5)	*16*	*9*	*48*	*0.18*	*0.96*
[17,5, 20,5)	*19*	*2*	*50*	*0.04*	*1.00*

1.3.2. Descripción gráfica

La representación gráfica de los datos cuantitativos discretos se hace mediante diagramas de barras.

El gráfico 1.3 nos muestra el diagrama de barras de las frecuencias absolutas y absolutas acumuladas para variables discretas. Las frecuencias absolutas n_i son las alturas de las barras con base el punto X_i. Las frecuencias absolutas acumuladas N_i son también las alturas de las barras con base el punto X_i.

La descripción gráfica de los datos continuos (agrupados) se hace mediante histogramas. En la figura 1.4 tenemos un ejemplo de histograma. En

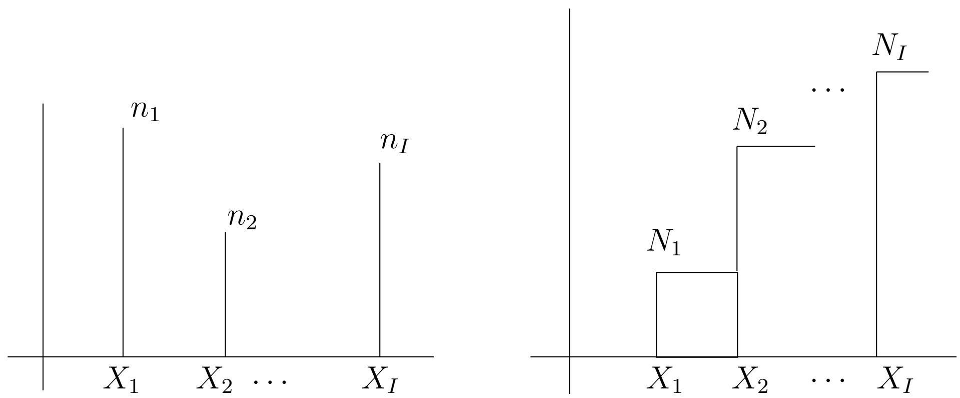


Figura 1.3: Frecuencias absolutas. Variables discretas

este caso es el histograma de las frecuencias absolutas a la izquierda y a la derecha tenemos el gráfico de las frecuencias absolutas acumuladas. Las frecuencias absolutas n_j del gráfico de la izquierda representen las areas de los rectángulos de la base $L_{j+1} - L_j$ (amplitud del intervalo de clase) mientras que las frecuencias absolutas acumuladas N_j del gráfico de la derecha representen las alturas de los rectángulos de base $L_{i+1} - L_i$. La curva que une los pares ordenados (X_j, h_j) recibe el nombre se llama polígono de frecuencias absolutas (léase igual para relativas), mientras que el polígono de frecuencias absolutas acumuladas (de forma similar para relativas) es el formado por los puntos $(L_1, 0), (L_2, N_1), \ldots, (L_{J+1}, N_J)$.

Ejemplo 2 *Consideremos las puntuaciones de los* 50 *aspirantes del ejemplo* 1. *Tomamos intervalos de amplitud* 3. *El histograma de frecuencias absolutas con el correspondientes polígono de frecuencias acumuladas se muestra en la figura* 1.5.

Notemos que las alturas de los rectángulos se calculan teniendo en cuenta que la amplitud de los intervalos es 3*:*

$$h_1 = \frac{n_1}{3} = \frac{11}{3} = 3{,}6666, \qquad h_2 = \frac{n_2}{3} = \frac{11}{3} = 3{,}666,$$

$$h_3 = \frac{n_3}{3} = \frac{17}{3} = 5{,}6666, \qquad h_4 = \frac{n_4}{3} = \frac{9}{3} = 3,$$

$$h_5 = \frac{n_5}{3} = \frac{2}{3} = 0{,}666.$$

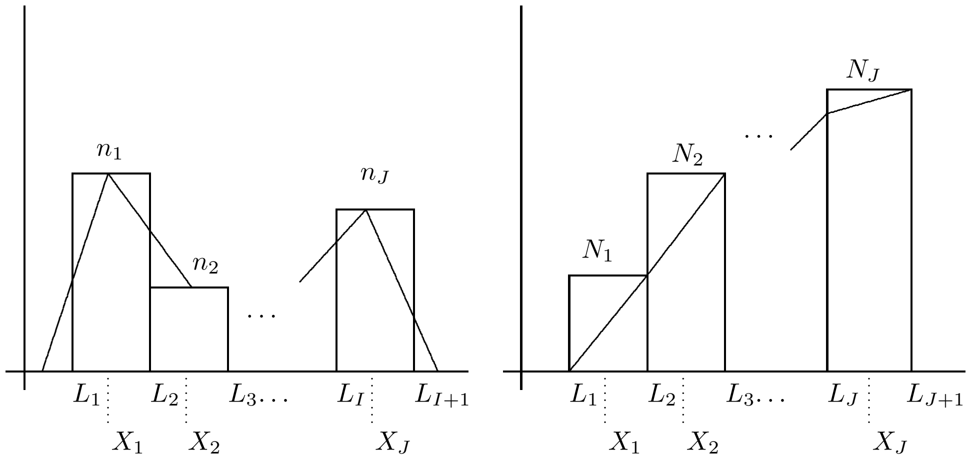


Figura 1.4: Frecuencias absolutas. Variables continuas

1.4. Análisis de las distribuciones

El análisis de las distribuciones de una variable o dato estadístico consiste en reducir los datos estadísticos a unas pocas medidas o índices, que reciben el nombre de estadísticos, que nos permitan una interpretación de las regularidades de todo el colectivo.

Tenemos los siguientes tipos de medidas:

- Medidas de posición: Intentan representar toda la distribución. Las más importantes son la media aritmética, la moda y la mediana.

- Medidas de dispersión: Intentan señalar la dispersión o separación del conjunto de datos respecto a las medidas de posición adoptadas. Las más importantes son la varianza, la desviación típica, el coeficiente de variación y los recorridos.

- Medidas de simetría y apuntamiento: Estudian si el polígono de frecuencias relativas es simétrico y lo *estirado* que está (apuntamiento). Se suele comparar este polígono con la curva de frecuencias de una distribución ideal llamada normal o campana de Gauss.

- Otras como las medidas de concentración; que no veremos. Por ejemplo el índice de Gini.

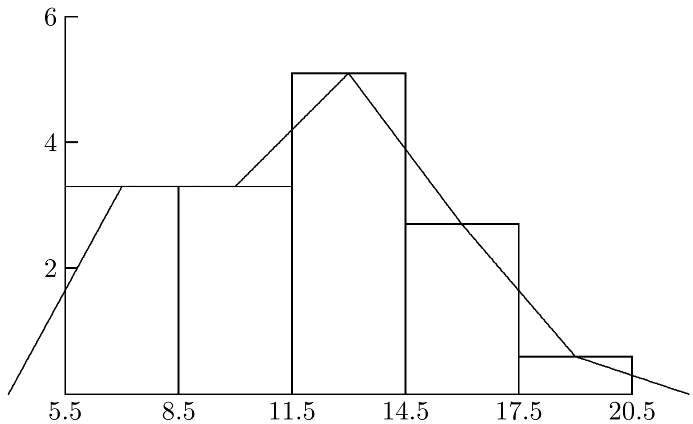


Figura 1.5: Histograma de frecuencias absolutas (ejemplo 1)

Nota importante: Algunas de estas medidas sólo se pueden calcular cuando tenga sentido operar con los datos, es decir, si estos son cantidades o al menos órdenes. Si tengo que una variable que responde al deporte que practica una persona de determinad población, aunque la variable esté codificada a valores enteros, no tiene sentido hacer la media aritmética. En lo que sigue dejaremos al lector que decida, siempre de forma razonada, que estadísticos no son aplicables a estas variables.

1.4.1. Medidas de posición

Veremos aquí las más conocidas medidas de posición.

Media aritmética

La media aritmética es la medida de tendencia central más utilizada, simboliza el valor central de toda la distribución. Su fórmula general es[2]:

$$overlinex = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{j=1}^{J} n_j X_j}{n} = \sum_{j=1}^{J} f_j X_j.$$

[2]Como se ve damos dos fórmulas una para datos no agrupados y otra para datos posiblemente agrupados, en lo que sigue no especificaremos cuales son las fórmulas para datos agrupados o no.

Para el caso de distribuciones de variables discretas, los X_j son los posibles valores de la variable mientras que en el caso continuo, son las marcas de clase de los intervalos.

Una de las propiedades fundamentales de la media es que si hacemos una transformación lineal de los datos digamos $Y = aX + b$[3] donde X son los valores de la variable, la relación entre la media aritmética de Y y la de X es:

$$\overline{y} = a\overline{x} + b.$$

Medias armónica y geométrica

Las medias armónica y geométrica no son de gran utilidad salvo en problemas concretos. Se calculan de la siguiente forma:

$$M_h = \frac{n}{\sum\limits_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\sum\limits_{j=1}^{I} \frac{n_j}{X_j}}, \quad M_g = \sqrt[n]{\prod_{i=1}^{n} x_j} = \sqrt[n]{\prod_{j=1}^{I} X_j^{n_j}}.$$

Estas medias tienen restricciones sobre los datos, no pueden tener datos nulos, y en general se utilizan para datos positivos.

Ejemplo 3 *Consideremos los siguientes datos:*

10	5	2	7	9	5	7	6	5	9
12	2	6	6	9	12	6	6	6	4
9	7	12	11						

La media aritmética de los datos anteriores sin agrupar en intervalos es:

$$\overline{x} = \frac{10 + 5 + 2 + \cdots + 12 + 11}{24} = \frac{173}{24} = 7{,}20833$$

Si los agrupamos en intervalos de amplitud 3*, la media será (hacemos*

[3]Multiplicar por una constante positiva se suel denominar cambio de escala. SUmar una constante a una varible recibe el nombre de cambio de origen. Así podemos decir que la media arimética queda igual de afectada por los cambios de escala y origen en los datos.

primero la correspondiente tabla de frecuencias)

intervalos	X_j	n_j	n_jX_j
[1,5, 4,5)	3	3	9
[4,5, 7,5)	6	12	72
[7,5, 10,5)	9	5	45
[10,5, 13,5)	12	4	48
Suma		24	174

$$\overline{x} = \frac{174}{24} = 7{,}25$$

Notemos que los valores difieren ya que el agrupamiento provoca una pérdida de información.

Media general de orden m

Definimos la media general $M_{(m)}$ de orden m como:

$$M_{(m)} = \left(\frac{\sum_{i=1}^{n} n_i x_i^m}{n} \right)^{\frac{1}{m}} = \left(\frac{\sum_{j=1}^{J} n_j X_j^m}{n} \right)^{\frac{1}{m}}$$

Se cumple que:

$$M_{(-1)} = M_h;\; M_{(0)} = M_g;\; M_{(1)} = \overline{x}$$

Además se cumple que $M_{(m)}$ es una función creciente en m y por lo tanto :

$$M_h \leq M_g \leq \overline{x}.$$

Mediana y percentiles

La mediana es aquel valor que, cuando consideremos todos los valores de la muestra ordenados, ocupa el lugar central. Es decir, quedan la misma cantidad de valores a su izquierda que a su derecha.

Supongamos que los datos ordenados son $x_1, x_2, \ldots x_n$, la **mediana** vale:

$$\begin{array}{rl} x_{\frac{n+1}{2}}, & \text{si } n \text{ es impar,} \\ \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}, & \text{si } n \text{ es par.} \end{array}$$

Por ejemplo, la mediana de los datos $1, 2, 5, 8, 8, 9, 11$ vale 8 ya que 8 es el que ocupa el lugar central y la mediana de $2, 3, 3, 4, 5, 6$ vale $\frac{3+4}{2} = 3{,}5$.

La manera anterior de calcular la mediana no es práctica en el caso en que haya muchos datos es muy costoso ordenarlos (orden n^2 o $n\log(n)$). Veamos alguna manera de cálculo aproximado de la mediana a partir de la tabla de distribución de frecuencias.

Necesitaremos las columnas de frecuencias absolutas y la de frecuencias absolutas acumuladas:

intervalos	X_j	n_j	N_j
$[L_1, L_2)$	X_1	n_1	N_1
$[L_2, L_3)$	X_2	n_2	N_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[L_I, L_{I+1})$	X_I	n_I	N_I
$\sum$		n	

Llamaremos intervalo crítico para la mediana al primer intervalo en el que su frecuencia absoluta acumulada supere o iguale a $\frac{n}{2}$. Denotemos por $[L_c, L_{c+1})$ el intervalo crítico. Sea N_{c-1} la frecuencia absoluta acumulada del intervalo anterior al crítico. En el caso en que el interval crítico sea el primero, $N_{c-1} = 0$. Sea n_c la frecuencia absoluta del intervalo crítico. Sea $A_c = L_{c+1} - L_c$ la amplitud del intervalo crítico. Calcularemos la **mediana** mediante:

$$M = L_c + A\frac{\left(\frac{n}{2} - N_{c-1}\right)}{n_c}.$$

La justificación de la fórmula anterior es la siguiente: si representásemos las frecuencias absolutas acumuladas entre los extremos de los intervalos, la mediana seria la antiimagen de $\frac{n}{2}$ en el intervalo crítico haciendo una interpolación por rectas (ver figura 1.6).

Los percentiles son una generalización de la mediana. La mediana es el percentil 50 ya que deja el 50 % de las observaciones a su izquierda.

En general el **percentil** P es aquel valor que deja el P % de las observaciones a su izquierda. El cálculo, dada la distribución de frecuencias es semejante al cálculo de la mediana.

Definimos el intervalo crítico en este caso como el primer intervalo del que su frecuencia absoluta acumulada supera o iguala a $\frac{n \cdot P}{100}$.

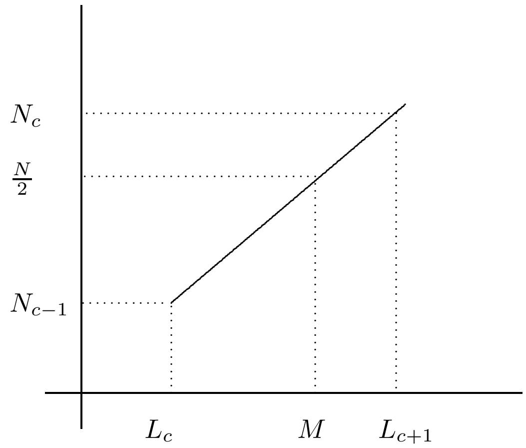


Figura 1.6: Interpretación geométrica de la Mediana

Sean entonces, $[L_c, L_{c+1})$ el intervalo crítico, N_{c-1} la frecuencia absoluta acumulada del intervalo anterior al crítico y n_c la frecuencia absoluta del intervalo crítico. Si denomamos por A_c a la amplitud del intervalo crítico, la fórmula para calcular el **percentil** P es:

$$M_p = L_c + A_c \frac{\left(\frac{n \cdot P}{100} - N_{c-1}\right)}{n_c}.$$

Ejemplo 4 *Calculemos la mediana, sin agrupar, de los siguientes datos:*

$$14, 15, 16, 18, 18, 18, 18, 19, 20, 20, 22.$$

El tamaño de la muestra es $n = 11$ *observaciones y ya están ordenadas. El lugar central, es el que ocupa el sexto puesto, el valor que ocupa este lugar es el* 18*, por lo tanto, la mediana es* 18*.*

En la siguiente muestra tenemos un número par de datos:

$$24, 25, 26, 26, 27, 27, 27, 29.$$

El tamaño muestral es $n = 8$ *observaciones que ya están ordenadas. El lugar central estará entre el cuarto y el quinto puesto. Los datos que ocupan estos lugares son el* 26 *y el* 27*. Por lo tanto la mediana vale*

$$M = \frac{26 + 27}{2} = 26{,}5.$$

Ejemplo 5 *Consideremos la siguiente distribución de frecuencias:*

intervalos	X_j	n_j	N_j
$[1{,}5, 4{,}5)$	3	3	3
$[4{,}5, 7{,}5)$	6	12	15
$[7{,}5, 10{,}5)$	9	5	20
$[10{,}5, 13{,}5)$	12	4	24

Tenemos que $n = 24$ *y que* $\frac{n}{2} = 12$*. El intervalo crítico es:* $[4{,}5, 7{,}5)$ *La mediana valdrá entonces:*

$$M = 4{,}5 + 3\frac{(12-3)}{12} = 6{,}75.$$

Percentil 25*:* $25\,\% \Rightarrow \frac{n\cdot P}{100} = 6$*. Intervalo crítico:* $[4{,}5, 7{,}5)$*.*

$$M_{25} = 4{,}5 + 3\frac{(6-3)}{12} = 5{,}25$$

Percentil 75*:* $75\,\% \Rightarrow \frac{n\cdot P}{100} = 18$*. Intervalo crítico:* $[7{,}5, 10{,}5)$*.*

$$M_{75} = 7{,}5 + 3\frac{(18-15)}{5} = 9{,}3$$

En general se habla de cuantiles para denominar a todos estos estadísticos. Los cuartiles que dividen a la población en cuartos son llamados cuartiles, así el primer cuartil Q_1 deja a su izquierda el 25 % de las observaciones, el segundo cuartil Q_2 es la mediana y el tercer cuartil Q_3 deja a su izquierda el 75 % de las observaciones. También se habla de los deciles que son los estadísticos que dividen a la población en décimas partes.

Moda

La moda de una muestra es un valor que tenga la frecuencia absoluta más grande. En consecuencia la moda no tiene por qué ser única puede haber más de un valor con frecuencia absoluta máxima. Si una distribución tiene una sola moda diremos que es unimodal, si dos bimodal, . . . La presencia de dos modas puede indicar la existencia de dos poblaciones diferenciadas en la muestra (por ejemplo el peso según sexo).

En el caso en que tengamos una tabla de distribuciones, para encontrar la moda, hemos de localizar el intervalo o intervalos con frecuencia absoluta más alta.

Sean $[L_j, L_{j+1})$ los extremos del intervalo con frecuencia absoluta máxima.

Para calcular la moda podemos utilizar la siguiente fórmula, en la que suponemos que todos los intervalos tienen la misma amplitud A (en caso contrario se utilizan otras aproximaciones):

$$M_o = L_j + A\frac{n_{j+1}}{(n_{j-1} + n_{j+1})}.$$

Donde:

- A: amplitud de los intervalos
- n_{j-1}: frecuencia absoluta del intervalo anterior al de frecuencia máxima.
- n_{j+1}: frecuencia absoluta del intervalo posterior al de frecuencia máxima.

Ejemplo 6 *Consideremos la siguiente distribución de frecuencias:*

intervalos	X_j	n_j	N_j
$[1{,}5, 4{,}5)$	3	3	3
$[4{,}5, 7{,}5)$	6	12	15
$[7{,}5, 10{,}5)$	9	5	20
$[10{,}5, 13{,}5)$	12	4	24

El intervalo con la frecuencia absoluta mas alta es el $[4{,}5, 7{,}5)$*. Por lo tanto, la moda vale:*

$$M_0 = 4{,}5 + 3\frac{5}{(3+5)} = 6{,}375.$$

1.4.2. Medidas de dispersión

Una vez estudiadas las medidas de posición, vamos a estudiar algunos estadísticos que miden lo separadas que están las observaciones entre sí.

Algunas medidas de dispersión respecto a la media aritmética son la varianza, la desviación típica, la desviación media respecto de la media y el coeficiente de variación.

Las medidas de dispersión respecto a la a la mediana es la desviación media respecto de la mediana.

Otras medidas de dispersión son el recorrido, el rango , el recorrido intercuartílico, el rango intercuartílico.

Varianza y desviación típica o estándar

La varianza y la desviación típica nos indican si los datos están muy dispersos respecto de la media aritmética $\overline{x}$.

La fórmula del cálculo de la varianza es:

$$s^2 = \frac{1}{n}\sum_{j=1}^{J} n_j (X_j - \overline{x})^2 = \frac{1}{n}\sum_{j=1}^{J} n_j X_j^2 - \overline{x}^2.$$

o bien para datos sin agrupar

$$s^2 = \frac{1}{n}\sum_{j=1}^{n} (x_i - \overline{x})^2 = \frac{1}{n}\sum_{j=1}^{n} x_i^2 - \overline{x}^2.$$

La segunda expresión es más útil que la primera de cara al cálculo de la varianza.

La propiedad fundamental de la varianza es que minimiza las desviaciones al cuadrado respecto a cualquier punto X_0. Es decir:

$$\min_{X_0} \frac{1}{n}\sum_{j=1}^{J} n_j (X_j - X_0)^2 = s^2.$$

La desviación típica o estándar es la raíz cuadrada positiva de la varianza:

$$s = \sqrt{\frac{1}{n}\sum_{j=1}^{J} n_j X_j^2 - \overline{x}^2}.$$

Por motivos que veremos en temas posteriores existe otra fórmula para el cálculo de la varianza de una muestra a la que en ocasiones se le denomina cuasivarianza o también se le llama varianza muestral [4]:

$$\tilde{s}^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{j=1}^{n} n_j (X_j - \overline{X})^2.$$

Notemos que la cuasivarianza es una pequeña corrección de la varianza, en lugar de dividir por el tamaño muestral se divide por el tamaño muestral

[4] Quizá algunos de vosotros descubra aquí el motivo por el que las calculadoras llevan dos teclas s_n^2 o σ_n^2 y s_{n-1} o σ_{n-1}.

menos 1. Para muestras grandes la corrección puede resultar insignificante pero para muestras pequeñas es necesaria.

Cuando la variable X se vea afectada por un cambio lineal: $Y = aX + b$, la varianza de Y cumple la siguiente relación:

$$s_Y^2 = a^2 s_X^2.$$

De aquí deducimos que la varianza es independiente respecto a cambios de origen y que queda afectada por el cuadrado de los cambios de escala.

Para las desviaciones típicas tendremos:

$$s_Y = |a| s_X.$$

Ejemplo 7 *Consideremos la siguiente distribución de frecuencias*

intervalos	X_j	n_j	n_jX_j
[9,5, 29,5)	19.5	38	741.0
[29,5, 49,5)	39.5	18	711.0
[49,5, 69,5)	59.5	31	1844.5
[69,5, 89,5)	79.5	20	1590.0
Sumas		107	4886.5

Vamos a calcular la varianza

Primero calculamos la media:

$$\overline{x} = \frac{4886{,}5}{107} = 45{,}6682$$

Para calcular la varianza hemos de añadir dos columnas a la tabla anterior:

X_j	X_j^2	$n_jX_j^2$
19.5	380.25	14449.50
39.5	1560.25	28084.50
59.5	3540.25	109747.75
79.5	6320.25	126405.00
Suma		278686.75

La varianza y la desviación típica valen:

$$s_X^2 = \frac{278686{,}75}{107} - 45{,}6682^2 = 518{,}962$$

$$s_X = \sqrt{518{,}962} = 22{,}7807$$

Coeficiente de variación

El coeficiente de variación se define como el cociente entre la desviación típica y la media aritmética, se utiliza para variables en las que la media represente a la magnitud de los datos (por ejemplo si todos son positivos y la distribución es unimodal) :

$$CV = \frac{s}{\overline{x}}.$$

El coeficiente de variación es independiente del cambio de escala. Más concretamente, si hacemos el cambio lineal de la variable X: $Y = aX$, con $a > 0$, el coeficiente de variación de la variable Y es el mismo que el de la variable X:

$$CV_Y = CV_X.$$

El coeficiente de variación será útil para comparar la dispersión de distribuciones medidas en diferentes escalas.

Ejemplo 8 *Consideremos la siguiente distribución de frecuencias:*

intervalos	X_j	n_j	$n_j X_j$
$[9{,}5, 29{,}5)$	19.5	38	741.0
$[29{,}5, 49{,}5)$	39.5	18	711.0
$[49{,}5, 69{,}5)$	59.5	31	1844.5
$[69{,}5, 89{,}5)$	79.5	20	1590.0
Sumas		107	4886.5

La media y la desviación típica son:

$$\overline{x} = 45{,}6682, \quad s_X = 22{,}7807$$

Por lo tanto el coeficiente de variación es:

$$CV = \frac{s}{\overline{x}} = \frac{22{,}6807}{45{,}6682} = 0{,}4988$$

Desviación media

La desviación media es un índice de dispersión respecto a la mediana o a la media. Queda definido por:

$$D_M = \frac{1}{n}\sum_{j=1}^{J} n_j |X_j - M|.$$

donde M es la mediana o la media aritmética.

La propiedad fundamental de la desviación media respecto a la mediana es que minimiza las desviaciones en valor absoluto respecto de un punto cualquiera X_0. Es decir:

$$\min_{X_0} \frac{1}{n}\sum_{j=1}^{J} n_j |X_j - X_0| = D_M.$$

Recorrido

Otra medida de dispersión es el recorrido. Se define como la diferencia entre el valor máximo y mínimo de los valores observados.

Ejemplo 9 *Consideremos la siguiente distribución de frecuencias:*

intervalos	X_j	n_j	n_jX_j
[0,5, 15,5)	8	4	32
[15,5, 30,5)	23	4	92
[30,5, 45,5)	38	2	76
Sumas		10	200

Vamos a calcular la desviación media:

Calculamos la media

$$\overline{x} = \frac{200}{10} = 20$$

Añadimos dos columnas más a la tabla de frecuencias:

X_j	$\lvert X_j - \overline{x}\rvert$	$n_j\lvert X_j - \overline{x}\rvert$
8	*12*	*48*
23	*3*	*12*
38	*18*	*36*
Sumas		*96*

La desviación es:

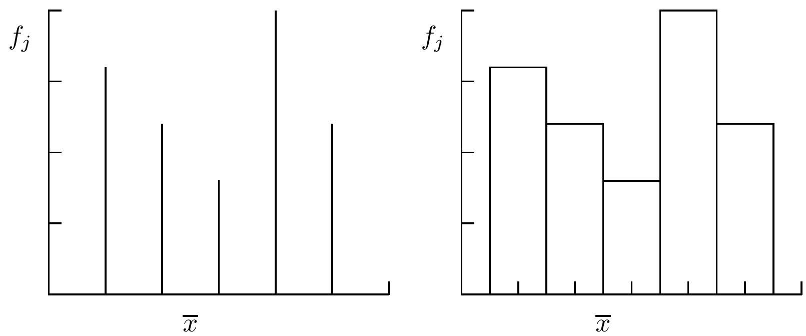


Figura 1.7: Diagrama de barras e histograma de las frecuencias relativas

$$D_M = \frac{96}{10} = 9{,}6$$

También se utiliza el recorrido intercuartílico que es $Q_3 - Q_1$; la diferencia entre el tercer y primer cuartil. También se pueden calcular recorridos con deciles, percentiles y cuantiles en general.

1.4.3. Perfil de una distribución

El perfil de una distribución viene determinado por alguno de sus polígonos de frecuencias. Es mejor utilizar las frecuencias relativas ya que no dependen del tamaño de la muestra. La idea es encontrar la curva a donde tiende el polígono de frecuencias cuando la muestra se hace grande, que en definitiva sería la curva de frecuencias de toda la población.

Una curva continua en forma de campana llamada curva de Gauss [5] puede servir como un modelo matemático ideal para comparar el perfil de cualquier distribución. Esta curva corresponde a la gráfica de la función:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

donde μ se aproxima por $\overline{x}$ y σ por s. Su representación gráfica es la de la figura 1.8, gaussiana o campana de gauss, para el caso (estándar) en el que $\mu = 0$ y $\sigma = 1$.

Las propiedades más importantes de la curva normal son:

[5] Es una buena broma pedir a un amigo el cálculo de la primitiva de la curva de Gauss.

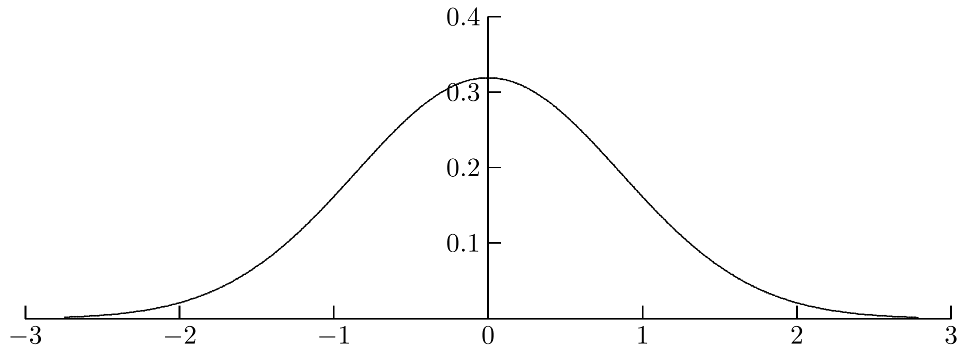


Figura 1.8: Curva normal o campana de Gauss

a) Está definida para cualquier real y es siempre positiva.

b) El área comprendida entre la curva y el eje de abcisas vale siempre 1 para cualquier valor de μ y $\sigma > 0$.

c) Es simétrica respecto a la recta vertical $X = \mu$ y en este punto tiene un máximo absoluto que vale $\frac{1}{\sqrt{2\pi}\sigma}$.

d) Tiene dos puntos de inflexión en $x = \mu \pm \sigma$.

e) El eje de abcisas es una asíntota de la curva.

Las medidas de simetría y apuntamiento se suelen referir a la correspondiente distribución normal; aquella en la que los parámetros se estiman por $\mu = \overline{x}$ y $\sigma = s$. (o por la cuasivarianza).

Se entiende, entonces, que la distribución normal es simétrica y es perfecta respecto al apuntamiento. Es decir, que no es ni apuntada ni chata.

1.4.4. Medidas de simetría

Para ver si una distribución es simétrica o asimétrica por la derecha o por la izquierda se toma como índice de simetría:

$$g_1 = \frac{m_3}{s^3},$$

donde m_3 es el momento central de tercer orden y se calcula de la siguiente forma:

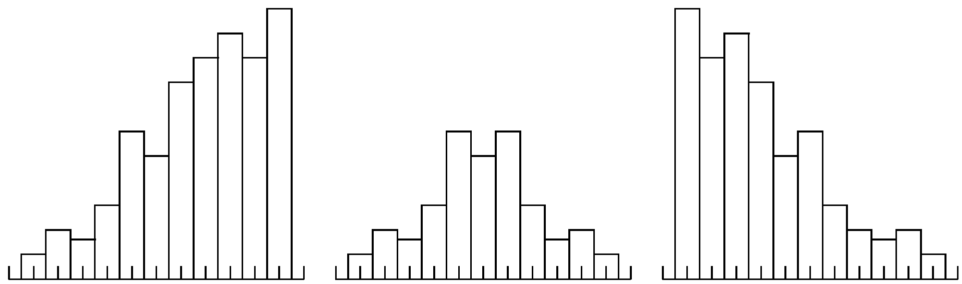


Figura 1.9: Histogramas

$$m_3 = \frac{1}{n}\sum_{j=1}^{J} n_j (X_j - \overline{x})^3,$$

y s es la desviación típica. Tenemos, entonces que:

- Si $g_1 > 0$, la distribución es asimétrica por la derecha o asimetría positiva.

- Si $g_1 = 0$, la distribución es simétrica o el índice no decide.

- Si $g_1 < 0$, la distribución es asimétrica por la izquierda o asimetría negativa.

Ejemplo 10 *Consideremos la siguiente distribución de frecuencias:*

intervalos	X_j	n_j	n_jX_j	$n_jX_j^2$
[14,5, 19,5)	17	4	68	1156
[19,5, 24,5)	22	6	132	2904
[24,5, 29,5)	27	8	216	5832
[29,5, 34,5)	32	11	352	11264
[34,5, 39,5)	37	35	1295	47915
[39,5, 44,5)	42	100	4200	176400
[44,5, 49,5)	47	218	10246	481562
Sumas		382	16509	727033

La media y la varianza valen:

$$\overline{x} = \frac{16509}{382} = 43{,}2173, \quad s_X^2 = \frac{727033}{382} - \left(\frac{16509}{382}\right)^2 = 35{,}49$$

Calculemos el coeficiente de asimetría g_1. Para hacerlo, hemos de añadir una columna más a la tabla anterior:

X_j	n_j	$n_j(X_j - \overline{x})^3$
17	4	-72081.33
22	6	-57308.66
27	8	-34121.16
32	11	-15525.84
37	35	-8411.41
42	100	-180.37
47	218	11799.67
Sumas	382	-175829.09

El momento de tercer orden vale:

$$m_3 = \frac{-175829{,}09}{382} = -460{,}285$$

A continuación calculamos el índice de asimetría:

$$g_1 = \frac{m_3}{s^3} = \frac{-460{,}285}{\left(\sqrt{35{,}49}\right)^3} = -2{,}18$$

Por lo tanto podemos decir que se trata de una distribución asimétrica por la izquierda o negativa

El índice de simetría es independiente de cambios lineales de la forma $Y = aX + b$, con $a > 0$, es decir:

$$g_1(X) = g_1(Y).$$

En otras palabras, el índice de simetría no queda afectado por cambios de origen, ni por cambios de escala positivos, mientras que para cambios de escala negativos cambia el signo de la simetría (ejercicio).

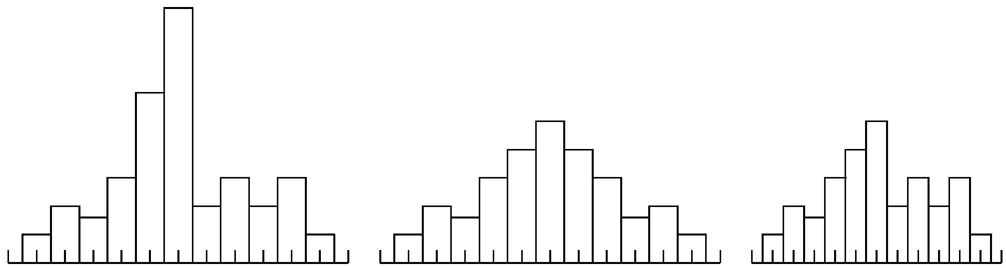


Figura 1.10: Histogramas de los tres tipos de apuntamiento

1.4.5. Medidas de apuntamiento

Las medidas de apuntamiento nos miden si el perfil de una distribución muestral está muy apuntado o no en comparación con un perfil ideal, como por ejemplo el de la campana de gauss asociada. Para estudiar el apuntamiento se utiliza un índice basado en el momento de cuarto orden, que recibe el nombre de coeficiente de apuntamiento o curtosis[6]:

$$g_2 = \frac{m_4}{s^4} - 3,$$

donde m_4 es el llamado momento central de cuarto orden y se calcula de la siguiente forma:

$$m_4 = \frac{1}{n}\sum_{j=1}^{J} n_j (X_j - \overline{x})^4,$$

y s es la desviación típica. Tenemos, pues que:

- Si $g_2 > 0$, la distribución es puntiaguda o leptocúrtica.
- Si $g_2 = 0$, la distribución es similar a la normal o mesocúrtica.
- Si $g_2 < 0$, la distribución es achatada o platicúrtica.

[6] En inglés *kurtosis*.

Ejemplo 11 *Consideremos la siguiente distribución de frecuencias:*

intervalos	X_j	n_j	n_jX_j	$n_jX_j^2$
[14,5, 19,5)	17	4	68	1156
[19,5, 24,5)	22	6	132	2904
[24,5, 29,5)	27	8	216	5832
[29,5, 34,5)	32	11	352	11264
[34,5, 39,5)	37	35	1295	47915
[39,5, 44,5)	42	100	4200	176400
[44,5, 49,5)	47	218	10246	481562
Sumas		382	16509	727033

La media y la varianza valen:

$$\overline{x} = \frac{16509}{382} = 43{,}22, \quad s_X^2 = \frac{727033}{382} - \left(\frac{16509}{382}\right)^2 = 35{,}49$$

El coeficiente de apuntamiento g_2. Hemos de añadir una columna a la tabla:

intervalos	X_j	n_j	$n_j(X_j - \overline{x})^4$
[14,5, 19,5)	17	4	1889776.11
[19,5, 24,5)	22	6	1215933.65
[24,5, 29,5)	27	8	553352.38
[29,5, 34,5)	32	11	174157.64
[34,5, 39,5)	37	35	52296.07
[39,5, 44,5)	42	100	219.56
[44,5, 49,5)	47	218	44634.88
Sumas		382	3930370.29

El momento de cuarto orden vale:

$$m_4 = \frac{3930370{,}29}{382} = 10288{,}93$$

A continuación, calculamos el índice de apuntamiento

$$g_2 = \frac{m_4}{s^4} - 3 = \frac{10288{,}93}{35{,}49^2} - 3 = 5{,}17$$

Por lo tanto se trata de una distribución puntiaguda o leptocúrtica.

El índice de apuntamiento es independiente respecto cambios lineales de la forma $Y = aX + b$, es decir:

$$g_2(X) = g_2(Y).$$

El índice g_2 no queda afectado por cambios de origen ni de escala.

1.5. Variables multidimensionales

Hasta ahora sólo hemos estudiado una variable, es evidente que en la realidad interesa el comportamiento conjunto de dos o más variables. En cualquier disciplina técnica o científica, economía, ciencias de la computación, bioinformática, telecomunicaciones,... son muy utilizados los conceptos de asociación, independencia y otros, entre dos o más variables. Para introducirlos estudiaremos el caso más sencillo; el de las variables estadísticas bidimensionales. En lo que respecta a esta sección cada individuo de la población tiene asociado más de un valor o cualidad observada. Per ejemplo peso y altura de un grupo de personas, peso y sexo, altura y nivel de estudios, ... Por ejemplo si estudiamos el peso (p) y la altura (h) de una población una muestra genérica de tamaño n tendría el siguiente aspecto:

$$(p_1, h_1), (p_2, h_2), \ldots, (p_n, h_n),$$

donde (p_i, h_i) es el peso y la estatura correspondientes a la observación i-ésima.

Otro ejemplo sería el estudio de la relación entre los turistas llegados a nuestra isla y el año de llegada. Los datos serían:

$$(t_1, n_1), (t_2, n_2), \ldots, (t_N, n_N),$$

donde t_i es el año i-ésimo y n_i = número de turistas llegados ese año.

1.5.1. Descripción numérica: caso bidimensional

Supongamos que tenemos (X, Y) un par de variables que se pueden medir conjuntamente en un individuo de la población que se desea estudiar.

Supondremos que las variables son discretas.

Sean $\{X_1, X_2, \ldots, X_I\}$ los valores posibles de X y $\{Y_1, Y_2, \ldots, Y_J\}$ los de Y.

El conjunto de valores que puede tomar la variable conjunta (X, Y) son:

$$\{(X_1, Y_1), \ldots, (X_1, Y_J), (X_2, Y_1) \ldots, (X_2, Y_J), \ldots, (X_I, Y_1), \ldots, (X_I, Y_J)\}.$$

Sean n_{ij} la frecuencia absoluta correspondiente al valor (X_i, Y_j), o sea, es el nombre de individuos de la muestra que tienen la variable X igual a X_i y la variable Y igual a Y_j.

Toda esta información se puede resumir en la siguiente tabla de frecuencias absolutas o tabla de contingencia:

X/Y	Y_1	Y_2	$\ldots$	Y_j	$\ldots$	Y_J	$n_{i\bullet}$
X_1	n_{11}	n_{12}	$\ldots$	n_{1j}	$\ldots$	n_{1J}	$n_{1\bullet}$
X_2	n_{21}	n_{22}	$\ldots$	n_{2j}	$\ldots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_i	n_{i1}	n_{i2}	$\ldots$	n_{ij}	$\ldots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_I	n_{I1}	n_{I2}	$\ldots$	n_{Ij}	$\ldots$	n_{IJ}	$n_{I\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	$\ldots$	$n_{\bullet j}$	$\ldots$	$n_{\bullet J}$	$N = n_{\bullet\bullet}$

En la tabla anterior, los valores de $n_{i\bullet}$ representan el número de individuos con $X = X_i$, $n_{\bullet j}$ el nombre de individuos con $Y = Y_j$ y n es el nombre total de individuos.

Ejemplo 12 *Consideremos la siguiente muestra de tamaño* 12 *de dos características conjuntas; la edad y peso de unas personas:*

$$\begin{array}{cccc} (20,75) & (20,75) & (30,75) & (40,85) \\ (30,65) & (20,75) & (40,85) & (30,65) \\ (20,65) & (40,75) & (30,65) & (20,75) \end{array}$$

La variable X es "edad" y toma los valores $\{20, 30, 40\}$ y la variable Y es "peso" y toma los valores $\{65, 75, 85\}$.

La tabla de frecuencias será:

X/Y	65	75	85	
20	1	4	0	5
30	3	1	0	4
40	0	1	2	3
	4	6	2	12

En el caso en que las variables X y Y sean agrupadas es hace una tabla semejante al caso discreto. En este caso, los datos de las dos variables X e Y se agrupan en intervalos de clase. La tabla de valores queda como se muestra a continuación:

X/Y	intervalos	$[L'_0, L'_1)\cdots[L'_{j-1}, L'_j)\cdots[L'_{J-1}, L'_J)$					
intervalos	M. Clase	c'_1	$\cdots$	c'_j	$\cdots$	c'_J	$n_{i\bullet}$
$[L_0, L_1)$	c_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1J}	$n_{1\bullet}$
$[L_1, L_2)$	c_2	n_{21}	$\cdots$	n_{2j}	$\cdots$	n_{2J}	$n_{2\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[L_{i-1}, L_i)$	c_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{iJ}	$n_{i\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$[L_{I-1}, L_I)$	c_I	n_{I1}	$\cdots$	n_{Ij}	$\cdots$	n_{IJ}	$n_{I\bullet}$
	$n_{\bullet j}$	$n_{\bullet 1}$	$\cdots$	$n_{\bullet j}$	$\cdots$	$n_{\bullet J}$	n

En la tabla anterior las c_i son las marcas de clase correspondientes a los intervalos de la variable X y las c'_j son las marcas de clase correspondientes a los intervalos de la variable Y.

Ejemplo 13 *Consideremos la siguiente tabla que nos da el peso y la estatura de* 15 *individuos:*

Individuo	X=peso	Y=estatura
1	65	1.6
2	62	1.6
3	71	1.6
4	72	1.7
5	75	1.8
6	80	1.6
7	74	1.6
8	77	1.7
9	81	1.8
10	90	1.8
11	89	1.7
12	83	1.8
13	82	1.8
14	81	1.7
15	71	1.7

Tomamos intervalos de amplitud 10 *para la variable* X =*peso. Así los intervalos para* X *empiezan en el límite real del mínimo peso* 62*:*

$$[61{,}5, 71{,}5),\ [71{,}5, 81{,}5),\ [81{,}5, 91{,}5).$$

Tomamos intervalos de amplitud 0,1 *para la variable* Y =*talla. Los intervalos para* Y *empiezan en el límite real de la mínimo altura* 1,6*.*

$$[1{,}55, 1{,}65),\ [1{,}65, 1{,}75),\ [1{,}75, 1{,}85).$$

La tabla de frecuencias agrupadas conjunta será:

X/Y	intervalos	$[1{,}55, 1{,}65)$	$[1{,}65, 1{,}75)$	$[1{,}75, 1{,}85)$	
intervalos	M. Clase	1.6	1.7	1.8	$n_{i\bullet}$
$[61{,}5, 71{,}5)$	66.5	3	1	0	4
$[71{,}5, 81{,}5)$	76.5	2	3	2	7
$[81{,}5, 91{,}5)$	86.5	0	1	3	4
	$n_{\bullet j}$	5	5	5	15

1.5.2. Distribuciones marginales

Consideremos una distribución conjunta de las variables (X, Y) donde X toma valores

$$\{X_1, X_2, \ldots, X_I\},$$

mientras que Y toma los valores

$$\{Y_1, Y_2, \ldots, Y_J\},$$

con la correspondiente tabla de frecuencias conjunta n_{ij}.

A la distribución unidimensional de la variable X la llamaremos distribución marginal de X y es la que toma los valores:

$$\{X_1, X_2, \ldots, X_I\},$$

y para la que la frecuencia absoluta correspondiente a X_i es:

$$n_{i\bullet} = \sum_{j=1}^{J} n_{ij},$$

es decir, la frecuencia absoluta del valor X_i es el número total de individuos que tienen la variable $X = X_i$.

De la misma forma, la distribución marginal de$Y$ es aquella variable unidimensional que toma los valores

$$\{Y_1, Y_2, \ldots, Y_J\},$$

y para la que la frecuencia absoluta correspondiente al valor Y_j vale:

$$n_{\bullet j} = \sum_{i=1}^{I} n_{ij},$$

o sea, el número total de individuos observados que tienen la variable $Y = Y_j$.

Las tablas de frecuencias correspondientes a las distribuciones marginales son :

Distribución marginal de la variable X

X_i	$n_{i\bullet}$
X_1	$n_{1\bullet}$
X_2	$n_{2\bullet}$
$\vdots$	$\vdots$
X_i	$n_{i\bullet}$
$\vdots$	$\vdots$
X_I	$n_{I\bullet}$
	n

Distribución marginal de la variable Y

Y_i	$n_{\bullet j}$
Y_1	$n_{\bullet 1}$
Y_2	$n_{\bullet 1}$
$\vdots$	$\vdots$
Y_i	$n_{\bullet j}$
$\vdots$	$\vdots$
Y_I	$n_{\bullet J}$
	n

Ejemplo 14 *Consideremos una distribución conjunta (X, Y) no agrupada con tabla de frecuencias:*

X/Y	65	75	85	
20	1	4	0	5
30	3	1	0	4
40	0	1	2	3
	4	6	2	12

Las distribuciones marginales de X e Y son:

Distribución marginal de X		Distribución marginal de Y	
X_i	$n_{i\bullet}$	Y_i	$n_{\bullet j}$
20	5	65	4
30	4	75	6
40	3	85	2
	12		12

Ejemplo 15 *Consideremos una distribución conjunta (X, Y) en este caso de valores agrupados con tabla de frecuencias:*

X/Y	intervalos	[1,55, 1,65)	[1,65, 1,75)	[1,75, 1,85)	
intervalos	M. Clase	1.6	1.7	1.8	$n_{i\bullet}$
[61,5, 71,5)	66.5	3	1	0	4
[71,5, 81,5)	76.5	2	3	2	7
[81,5, 91,5)	86.5	0	1	3	4
	$n_{\bullet j}$	5	5	5	15

Les distribuciones marginales de X e Y son:

Distribución marginal de X			Distribución marginal de Y		
Intervalo	X_i	$n_{i\bullet}$	Intervalo	Y_i	$n_{\bullet j}$
[61,5, 71,5)	66.5	4	[1,55, 1,65)	1.6	5
[71,5, 81,5)	76.5	7	[1,65, 1,75)	1.7	5
[81,5, 91,5)	86.5	4	[1,75, 1,85)	1.8	5
		15			15

1.5.3. Distribuciones condicionadas

Consideremos una distribución conjunta de variables (X, Y) donde X toma valores

$$\{X_1, X_2, \ldots, X_I\},$$

e Y toma valores

$$\{Y_1, Y_2, \ldots, Y_J\},$$

con la correspondiente tabla de frecuencias conjunta n_{ij}.

Consideremos un valor concreto de la variable Y, Y_j. Definimos la distribución condicionada de X respecte al valor Y_j de Y y lo denotaremos por $X/Y = Y_j$ como aquella distribución unidimensional que toma los mismo valores que X, es decir,

$$\{X_1, X_2, \ldots, X_I\},$$

y tal que la frecuencia absoluta del valor X_i (a la que denotaremos por $n_{i/j}$) se define como el número de individuos observados que tienen $X = X_i$ e $Y = Y_j$.

De la misma manera, podemos considerar un valor concreto de la variable X, X_i. Definimos distribución condicionada de Y respecto del valor X_i y la denotaremos por $Y/X = X_i$ como aquella distribución unidimensional que toma los mismo valores que Y,

$$\{Y_1, Y_2, \ldots, Y_J\},$$

y tal que la frecuencia absoluta del valor Y_j (a la que denotaremos por $n_{j/i}$) se define como el número de individuos observados que tienen la $Y = Y_j$ y la $X = X_i$.

Observemos que existen tantas distribuciones condicionadas $X/Y = Y_j$ como valores distintos toma Y y que existen tantas condicionales $Y/X = X_i$ como valores distintos toma X.

Ejemplo 16 *Consideremos una distribución conjunta (X, Y) sin agrupar, con tabla de frecuencias:*

X/Y	65	75	85	
20	1	4	0	5
30	3	1	0	4
40	0	1	2	3
	4	6	2	12

Fijemos $Y = 75$. La tabla de frecuencias de la distribución $X/Y = 75$ es:

$X_i/Y = 75$	$n_{i/75}$
20	4
30	1
40	1
	6

Fijemos por ejemplo $X = 30$. *La tabla de frecuencias de la distribución* $Y/X = 30$ *es:*

$Y_j/X = 30$	$n_{j/30}$
65	3
75	1
85	0
	4

Ejemplo 17 *Consideremos una distribución conjunta* (X, Y) *caso agrupado con tabla de contingencia:*

X/Y	intervalos	$[1{,}55, 1{,}65)$	$[1{,}65, 1{,}75)$	$[1{,}75, 1{,}85)$	
intervalos	M. Clase	1.6	1.7	1.8	$n_{i\bullet}$
$[61{,}5, 71{,}5)$	66.5	3	1	0	4
$[71{,}5, 81{,}5)$	76.5	2	3	2	7
$[81{,}5, 91{,}5)$	86.5	0	1	3	4
	$n_{\bullet j}$	5	5	5	15

Fijemos por ejemplo $Y = 1{,}6$. *La tabla de frecuencias de* $X/Y = 1{,}6$ *es:*

Intervalo	X_i	$n_{i/1,6}$
$[61{,}5, 71{,}5)$	66.5	3
$[71{,}5, 81{,}5)$	76.5	2
$[81{,}5, 91{,}5)$	86.5	0
		5

Fijemos por ejemplo $X = 86{,}5$. *La tabla de frecuencias de* $Y/X = 86{,}5$ *es:*

Intervalo	Y_j	$n_{j/86,5}$
$[1{,}55, 1{,}65)$	1.6	0
$[1{,}65, 1{,}75)$	1.7	1
$[1{,}75, 1{,}85)$	1.8	3
		4

1.5.4. Momentos bidimensionales

Consideremos una distribución conjunta de las variables (X, Y) donde X toma valores

$$\{X_1, X_2, \ldots, X_I\},$$

mientras que Y toma los valores

$$\{Y_1, Y_2, \ldots, Y_J\},$$

con la correspondiente tabla de frecuencias conjunta n_{ij}.

Vamos a estudiar los momentos bidimensionales de primer y segundo orden.

Los momentos de primer orden son la media de X ($\overline{x}$) y la media de Y ($\overline{y}$). Se calculan de la siguiente forma:

$$\overline{x} = \frac{\sum_{i=1}^{I} n_{i\bullet} X_i}{n}, \quad \overline{y} = \frac{\sum_{j=1}^{J} n_{\bullet j} Y_j}{n}.$$

Los momentos de segundo orden son la varianza de X (s_X^2), la varianza de Y (s_Y^2) y la covarianza de X e Y (s_{XY}).

Las fórmulas respectivas son:

$$s_X^2 = \frac{1}{n}\sum_{i=1}^{I} n_{i\bullet}\left(X_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{I} n_{i\bullet} X_i^2 - \overline{x}^2,$$

$$s_Y^2 = \frac{1}{n}\sum_{j=1}^{J} n_{\bullet j}\left(Y_j - \overline{y}\right)^2 = \frac{1}{n}\sum_{j=1}^{J} n_{\bullet j} Y_j^2 - \overline{y}^2,$$

$$s_{XY} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}(X_i - \overline{x})(Y_j - \overline{y})}{n} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} X_i Y_j}{n} - \overline{x} \cdot \overline{y}.$$

Ejemplo 18 *Consideremos las siguientes distribución de las variables* (X, Y)*:*

X/Y	65	75	85	
20	1	4	0	5
30	3	1	0	4
40	0	1	2	3
	4	6	2	12

Los momentos de primer orden son:

$$\overline{x} = \frac{n_{1\bullet}X_1 + n_{2\bullet}X_2 + n_{3\bullet}X_3}{n} = \frac{5 \cdot 20 + 4 \cdot 30 + 3 \cdot 40}{12} = 28{,}333,$$

$$\overline{y} = \frac{n_{\bullet 1}Y_1 + n_{\bullet 2}Y_2 + n_{\bullet 3}Y_3}{n} = \frac{4 \cdot 65 + 6 \cdot 75 + 2 \cdot 85}{12} = 73{,}333$$

Los momentos de segundo orden son:

$$
\begin{aligned}
s_X^2 &= \tfrac{n_{1\bullet}X_1^2+n_{2\bullet}X_2^2+n_{3\bullet}X_3^2}{n} - \overline{x}^2 = \tfrac{5\cdot 20^2+4\cdot 20^2+3\cdot 40^2}{12} - 28{,}333^2 = 63{,}888,\\
s_Y^2 &= \tfrac{n_{\bullet 1}Y_1^2+n_{\bullet 2}Y_2^2+n_{\bullet 3}Y_3^2}{n} - \overline{y}^2 = \tfrac{4\cdot 65^2+6\cdot 75^2+2\cdot 85^2}{12} - 73{,}333^2 = 47{,}222,\\
s_{XY} &= \tfrac{1}{n}\left(n_{11}X_1Y_1 + n_{12}X_1Y_2 + n_{13}X_1Y_3 + n_{21}X_2Y_1 + n_{22}X_2Y_2\right.\\
&\quad \left.+n_{23}X_2Y_3 + n_{31}X_3Y_1 + n_{32}X_3Y_2 + n_{33}X_3Y_3\right) - \overline{x}\cdot\overline{y}\\
&= \tfrac{1}{12}\left(1\cdot 20\cdot 65 + 4\cdot 20\cdot 75 + 0\cdot 20\cdot 85 + 3\cdot 30\cdot 65 + 1\cdot 30\cdot 75 +\right.\\
&\quad \left. 0\cdot 30\cdot 85 + 0\cdot 40\cdot 65 + 1\cdot 40\cdot 75 + 2\cdot 40\cdot 85\right) - 28{,}333\cdot 73{,}333\\
&= 22{,}222
\end{aligned}
$$

Ejemplo 19 *Consideremos una distribución conjunta (X, Y) (datos agrupados) con tabla de frecuencias:*

X/Y	intervalos	$[1{,}55, 1{,}65)$	$[1{,}65, 1{,}75)$	$[1{,}75, 1{,}85)$	
intervalos	M. Clase	1.6	1.7	1.8	$n_{i\bullet}$
$[61{,}5, 71{,}5)$	66.5	3	1	0	4
$[71{,}5, 81{,}5)$	76.5	2	3	2	7
$[81{,}5, 91{,}5)$	86.5	0	1	3	4
	$n_{\bullet j}$	5	5	5	15

Los momentos de primer orden son:

$$\overline{x} = 76{,}5, \quad \overline{y} = 1{,}7$$

Los momentos de segundo orden son:

$$s_X^2 = 53{,}333, \quad s_Y^2 = 0{,}006, \quad s_{XY} = 0{,}4$$

1.5.5. Independencia e incorrelación

Vamos a introducir dos conceptos nuevos: el de independencia y el de incorrelación.

El concepto de independencia formaliza la idea conocer el valor de la variable X no aporta información alguna sobre el valor de Y y viceversa.

Dada una variable bidimensional (X, Y) con tabla de frecuencias conjunta n_{ij}, diremos que X e Y son independientes si:

$$\frac{n_{ij}}{n} = \frac{n_{i\bullet}}{n}\frac{n_{\bullet j}}{n}, \text{ para todo } i \in \{1, 2, \ldots, I\} \text{y para todo } j \in \{1, 2, \ldots, J\}.$$

En el caso en que la relación anterior falle para un i y un j diremos que las dos variable no son independientes.

Ejemplo 20 *En este ejemplo las variables X e Y no son independientes:*

X/Y	65	75	85	
20	1	4	0	5
30	3	1	0	4
40	0	1	2	3
	4	6	2	12

ya que por ejemplo

$$\frac{n_{11}}{n} \neq \frac{n_{1\bullet}}{n}\frac{n_{\bullet 1}}{n}, \quad \frac{1}{12} \neq \frac{5}{12} \cdot \frac{4}{12}.$$

En cambio en este otro caso, si son independientes:

X/Y	65	75	85	
20	3	2	1	6
30	6	4	2	12
40	6	4	2	12
	15	10	5	30

Dejamos al lector la comprobación como ejercicio.

El concepto de incorrelación formaliza la idea de relación lineal en el sentido de que las variables crecen de forma lineal conjuntamente (relación directa) o bien si una crece, la otra decrece (relación inversa). Dada una variable bidimensional (X, Y) con tabla de frecuencias conjunta n_{ij}, diremos que X e Y son incorreladas si su covarianza $s_{XY} = 0$.

La relación que existe entre los dos conceptos introducidos, el de independencia y el de incorrelación viene dada por la siguiente propiedad:

Teorema 21 *Si las variables X e Y son independientes entonces son incorreladas.*

El recíproco del teorema anterior no es cierto en general. Podemos decir que independencia implica incorrelación pero lo contrario no.

$$\textbf{independencia} \Rightarrow \textbf{ incorrelación}$$

Demostración del teorema:
Si X es independiente de Y tenemos que

$$\frac{n_{ij}}{n} = \frac{n_{i\bullet}}{n}\frac{n_{\bullet j}}{n}.$$

Por lo tanto:

$$\begin{aligned} s_{XY} &= \tfrac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}\left(x_i-\overline{x}\right)\left(y_j-\overline{y}\right) \\ &= \sum_{i=1}^{I}\left(x_i-\overline{x}\right)\tfrac{n_{i\bullet}}{n}\sum_{j=1}^{J}\left(y_j-\overline{y}\right)\tfrac{n_{\bullet j}}{n} \\ &= 0\cdot 0 = 0, \end{aligned}$$

teniendo en cuenta que si X es una variable unidimensional con valores $\{x_1, x_2, \ldots, x_I\}$, con las correspondientes frecuencias absolutas $\{n_1, n_2, \ldots, n_I\}$, tenemos:

$$\sum_{i=1}^{I} n_i\left(x_i-\overline{x}\right) = \sum_{i=1}^{I} n_i x_i - n\overline{x} = 0.$$

1.6. Asociación, concordancia y correlación

1.6.1. Introducción

En esta sección estudiaremos si existe algún tipo de relación entre dos variables X e Y.

Hasta ahora sabemos cuando dos variables son independientes o no. En caso de que no se sean independientes, nos interesará medir el grado de dependencia que tienen, es decir, si son "muy dependientes o no".

Para medir la dependencia utilizaremos una serie de coeficientes como son el coeficiente de contingencia de Pearson y en el caso de variables con sólo dos valores el coeficiente de contingencia de Yule.

1.6.2. Coeficientes de asociación

Coeficiente de contingencia o de correlación de Pearson

Dada una variable bidimensional (X, Y) con tabla de frecuencias conjunta n_{ij}, definimos el coeficiente de correlación de Pearson como:

$$C_P = \sqrt{\frac{\chi^2}{n + \chi^2}},$$

donde χ^2 es el llamado estadístico de Pearson, que se define como:

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(n_{ij} - \frac{n_{i\bullet} n_{\bullet j}}{n}\right)^2}{\frac{n_{i\bullet} n_{\bullet j}}{n}}.$$

Las n_{ij} son las conocidas como frecuencias empíricas u observadas y las expresiones $\frac{n_{i\bullet} n_{\bullet j}}{n}$ se las conocen como frecuencias teóricas; o sea, las frecuencias conjuntas que tendrían las variables (X, Y) si fueran independientes.

Por lo tanto, cuando más cerca estén las frecuencias empíricas de las teóricas, más pequeño será el estadístico de Pearson χ^2 y el coeficiente de contingencia de Pearson C_P.

Ejemplo 22 *Consideremos la siguiente distribución conjunta:*

X/Y	65	75	85	
20	1	4	0	5
30	3	1	0	4
40	0	1	2	3
	4	6	2	12

La tabla anterior es la de frecuencias empíricas. A continuación construimos la tabla de frecuencias teóricas:

X/Y	65	75	85	
20	1.66	2.5	0.83	5
30	1.33	2	0.66	4
40	1	1.5	0.5	3
	4	6	2	12

Como podemos observar, las frecuencias teóricas no coinciden con las empíricas. Por lo tanto, deducimos que las dos variables X e Y no son independientes.

Vamos a calcular ahora el coeficiente de contingencia de Pearson C_P.

En primer lugar hemos de calcular el estadístico χ^2:

$$\begin{aligned}\chi^2 = & \tfrac{(1-1{,}66)^2}{1{,}66} + \tfrac{(4-2{,}5)^2}{2{,}5} + \tfrac{(0-0{,}83)^2}{0{,}83} + \tfrac{(3-1{,}33)^2}{1{,}33} + \tfrac{(1-2)^2}{2} + \\ & \tfrac{(0-0{,}66)^2}{0{,}66} + \tfrac{(0-1)^2}{1} + \tfrac{(1-1{,}5)^2}{1{,}5} + \tfrac{(2-0{,}5)^2}{0{,}5} \\ = & \ 10{,}916\end{aligned}$$

Por último calculamos el coeficiente de contingencia de Pearson C_P:

$$C_P = \sqrt{\frac{10{,}916}{12 + 10{,}916}} = 0{,}690$$

Propiedades C_P:

1) El valor de C_P es mayor o igual que 0 y menor que 1. En el caso en que X e Y sean independientes, las frecuencias empíricas y teóricas coinciden y $C_P = 0$.

2) Cuanto más dependientes son las variables X e Y, C_P se aproxima más a 1.

Por lo tanto si C_P es pequeño podemos decir que el grado de dependencia es bajo , mientras que si C_P aumenta es alto.

Para el caso más trivial en el que tengamos una tabla 2×2, es decir $I = J = 2$, o sea cuando las variables sólo toman dos valores cada una, se utiliza otro coeficiente, el coeficiente de contingencia de Yule. Se define así:

$$C_\gamma = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}}.$$

Recordamos al lector que la tabla de frecuencias tendrá el siguiente aspecto:

X/Y	Y_1	Y_2	
X_1	n_{11}	n_{12}	$n_{1\bullet}$
X_2	n_{21}	n_{22}	$n_{2\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	n

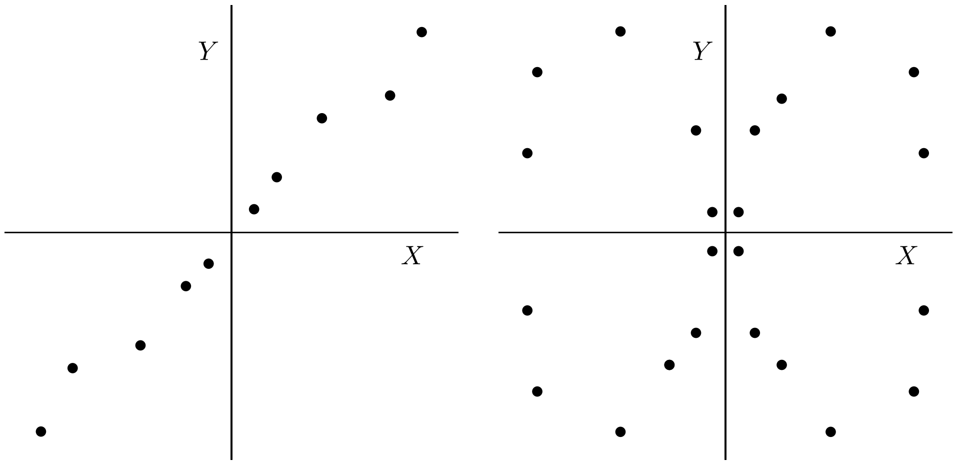


Figura 1.11: Nube de puntos con relación lineal y sin relación lineal

Ejemplo 23 *Calculemos el coeficiente de contingencia de Yule de la siguiente distribución conjunta:*

X/Y	Y_1	Y_2	
X_1	2	8	10
X_2	3	7	10
	5	15	20

$$C_\gamma = \frac{2 \cdot 7 - 3 \cdot 8}{2 \cdot 7 + 3 \cdot 8} = -0{,}263158$$

El coeficiente de contingencia de Yule siempre está entre -1y1. En caso de independencia entre las variables, se tiene que $C_\gamma = 0$.

1.6.3. Correlación lineal

El problema que nos planteamos en este punto es saber si existe una relación lineal entre X e Y, es decir, si existen dos valores numéricos a y b tales que:

$$Y \approx a + bX.$$

La relación lineal no tiene por que ser perfecta. Lo que nos interesa es medir esa relación lineal. En el gráfico de la izquierda de la figura 1.11 se vislumbra una relación lineal mayor que en el de la derecha ya que podemos encontrar un a recta que aproxime mejor Y en función de X.

Vamos a introducir un coeficiente que mide la relación lineal entre dos variables. Este coeficiente es el coeficiente de correlación lineal de Pearson r_{XY} y se define de la manera siguiente

$$r_{XY} = \frac{s_{XY}}{\sqrt{s_X^2 s_Y^2}} = \frac{s_{XY}}{s_X s_Y}.$$

Propiedades de r_{XY}:

1) $-1 \leq r_{XY} \leq 1$

2) $r_{XY} = r_{YX}$

3) Interpretación de r_{XY}:

- Si $r_{XY} > 0$ y a medida que se aproxima a 1, aumenta la relación lineal positiva entre las dos variables X e Y; lo que quiere decir que si X crece, la variable Y también y si X decrece, Y también, Obsérvese la parte izquierda de la figura 1.11 como ejemplo de este caso.
- Si $r_{XY} < 0$ y su valor está muy cerca de -1 quiere decir que hay una buena relación lineal negativa entre las dos variables X e Y; lo que significa que si la variable X crece, la variable Y decrece o viceversa. Como ejemplo ver el gráfico de la derecha de la figura 1.12.
- Si $r_{XY} = 0$ o es pequeño, quiere decir que no hay ningún tipo de relación lineal entre las variables X e Y.
- Si $r_{XY} = \pm 1$, hay relación lineal exacta entre X e Y, o sea, existen dos números reales a y b tales que $Y = a + bX$.

Ejemplo 24 *Consideremos la siguiente distribución conjunta de la variable* (X, Y)*: (ver ejemplo 18)*

Los valores de s_X^2, s_Y^2 *y de* s_{XY} *son:*

$$s_X^2 = 63{,}888, \quad s_Y^2 = 47{,}222, \quad s_{XY} = 22{,}222$$

El coeficiente de correlación lineal vale:

$$r_{XY} = \frac{22{,}222}{\sqrt{63{,}888 \cdot 47{,}222}} = 0{,}405$$

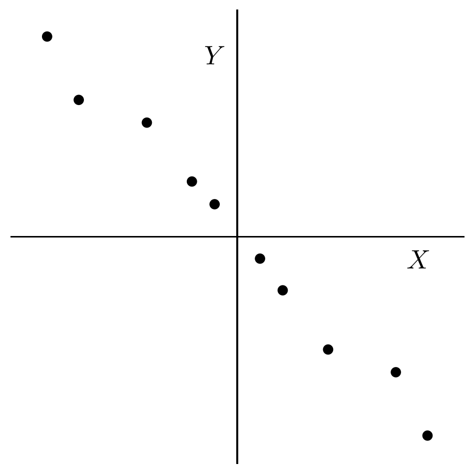


Figura 1.12: Nube de puntos con relación lineal negativa

1.6.4. Correlación ordinal

Vamos a estudiar ahora la relación que existe entre dos ordenaciones dadas por una muestra de datos bidimensionales.

Los estadísticos que miden este tipo de relaciones reciben el nombre de coeficiente de correlación ordinal y nos darán medidas de la similitud de las dos ordenaciones a lo que se suele llamar concordancia.

Más concretamente, consideremos un conjunto de individuos y los ordenamos según dos criterios. Tendremos así dos ordenaciones de los individuos. Estas ordenaciones las podemos disponer como si se tratara de una estadística bidimensional, donde la primera componente de la observación de un individuo correspondería al número de orden del primer criterio de ordenación y la segunda componente al otro.

Por ejemplo consideremos las observaciones en 5 humanos de su peso X en Kg. y estatura Y en metros:

Individuo i	(X_i, Y_i)	Orden X	Orden Y
Individuo 1	$(80, 1{,}75)$	3	2
Individuo 2	$(75, 1{,}92)$	2	4
Individuo 3	$(85, 1{,}67)$	4	1
Individuo 4	$(66, 1{,}80)$	1	3
Individuo 5	$(90, 2{,}00)$	5	5

Si ordenamos los individuos en orden ascendente (de menor a mayor) según el peso quedan así:

Rango Peso	1	2	3	4	5
Individuo	4	2	1	3	5

mientras que si los ordenamos en orden ascendente según su altura:

Rango	1	2	3	4	5
Individuo	3	1	4	2	5

Tenemos así dos ordenaciones de números ordinales enteros que reciben el nombre de rangos[7]. En general podemos escribir:

$$\begin{array}{ll} \text{para } X \rightarrow & \{r_{x_1}, r_{x_2}, r_{x_3}, \ldots, r_{x_n}\}, \\ \text{para } Y \rightarrow & \{r_{y_1}, r_{y_2}, r_{y_3}, \ldots, r_{y_n}\}, \end{array}$$

donde los valores de r_{x_i} y r_{y_i} dan el lugar que ocupa el valor x_i o el y_i en cada una de las muestras ordenadas. Estos valores están comprendidos entre 1 y n, luego son dos permutaciones de orden n.

Las diferencias entre las ordenaciones son:

$$d_i = r_{x_i} - r_{y_i}\,; i = 1, 2, \ldots, n.$$

El coeficiente de correlación ordinal o por rangos de Spearman queda definido por:

$$r_S = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n\,(n^2 - 1)}.$$

De hecho, r_S no es más que el coeficiente de correlación lineal introducido en la sección anterior aplicado a los rangos.

Propiedades de r_S:

- Si $r_S = 1$, las dos ordenaciones coinciden; o sea, $r_{x_i} = r_{y_i}$ para cualquier i entre 1 y n.

- Si $r_S = -1$, la ordenación de Y es exactamente la opuesta a la de X, es decir , $r_{x_i} = r_{y_{n-i+1}}$ para cualquier i entre 1 y n.

[7] El cálculo de rangos se complica en el caso de empates, es decir cuado hay valores repetidos enlas series de datos. En estos casos se puede romper el empate de varias maneras.

- El coeficiente r_S esta siempre comprendido entre -1 y 1. Si $r_S > 0$, podemos decir que las dos ordenaciones son del mismo sentido y si $r_S < 0$, las dos ordenaciones son de sentidos opuestos.

Ejemplo 25 *Consideremos la muestra anterior de pesos y estaturas de* 5 *individuos:*

Individuo i	(X_i, Y_i)	r_{X_i}	r_{Y_i}	d_i^2
Individuo 1	$(80, 1{,}75)$	3	2	1
Individuo 2	$(75, 1{,}92)$	2	4	4
Individuo 3	$(85, 1{,}67)$	4	1	9
Individuo 4	$(66, 1{,}80)$	1	3	4
Individuo 5	$(90, 2{,}00)$	5	5	0
Σ				18

Luego tenemos que :

$$r_s = 1 - \frac{6 \cdot 18}{5 \cdot (25-1)} = 1 - \frac{108}{120} = 0{,}1$$

Capítulo 2

Probabilidad

2.1. Espacio muestral

Llamaremos espacio muestral (e.m.) asociado a un experimento aleatorio al conjunto de todos sus posibles resultados. Normalmente lo denotaremos por Ω.

A los elementos de Ω les denominaremos sucesos elementales.

Llamaremos suceso del e.m. Ω a cualquier subconjunto de Ω. El suceso Ω recibe el nombre de suceso seguro o cierto y $\emptyset$ es el suceso vacío o imposible.

Dado un e.m. Ω llamaremos **σ-álgebra de sucesos** de Ω a todo subconjunto $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ tal que

a) $\Omega \in \mathcal{F}$

b) $A \in \mathcal{F} \Longrightarrow \overline{A} \in \mathcal{F}$

c) $A_1, A_2, \ldots \in \mathcal{F} \Longrightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{F}$

Diremos que un suceso $A \in \mathcal{F}$ ha ocurrido si el resultado del experimento es ω y $\omega \in A$.

2.1.1. Propiedades del álgebra de sucesos

- $\emptyset \in \mathcal{F}$
- $A_1, A_2, \ldots \in \mathcal{F} \Longrightarrow \cap_{n=1}^{\infty} A_n \in \mathcal{F}$
- $A_1, A_2, \ldots, A_n \in \mathcal{F} \Longrightarrow \cup_{k=1}^{n} A_n \in \mathcal{F}$

- $A_1, A_2, \dots, A_n \in \mathcal{F} \Longrightarrow \cap_{k=1}^{n} A_n \in \mathcal{F}$
- $A, B \in \mathcal{F} \Longrightarrow A - B \in \mathcal{F}$

Ejemplo 26 *a)* $\mathcal{P}(\Omega)$ *es la* σ*-álgebra más grande sobre* Ω

b) $\{\emptyset, \Omega\}$ *es la mínima* σ*-álgebra sobre* Ω

c) $\{\emptyset, E, \overline{E}, \Omega\}$ *es la menor* σ*-álgebra que contiene al suceso* E

d) Si $\Omega = I\!R$ *llamaremos* σ*-álgebra de Borel* $\mathcal{B}(I\!R)$ *a la menor* σ *álgebra que contiene a todos los intervalos* $]a, b[$ *con* $a, b \in I\!R$*.*

2.2. Definición de probabilidad

- Definición como frecuencia relativa (Von Mises, 1883-1953)

  Sea $N_A(n)$ el número de veces que ha ocurrido el suceso A en n repeticiones del mismo experimento aleatorio. Entonces la frecuencia relativa de A es

  $$f_A(n) = \frac{N_A(n)}{n}$$

  y definimos

  $$P(A) = \lim_{n \to \infty} f_A(n)$$

- Definición Clásica (Laplace, 1812) Sea $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ un e.m. finito. Supongamos que los sucesos elementales sean equiprobables, es decir,
  $P(\{\omega_i\}) = \frac{1}{n}$ para todo $i = 1, \dots, n$
  Entonces definimos la probabilidad del suceso A como:
  $P(A) = \frac{card(A)}{n} = \frac{\text{c. favorables al suc.}A}{\text{casos posibles}}$

• Definición Axiomática (Kolgomorov, 1933) Diremos medida de probabilidad o simplemente probabilidad sobre una σ-álgebra $\mathcal{F}$ de un e.m. Ω a toda aplicación

$$P : \mathcal{F} \to [0, 1]$$

tal que:
a) $P(\Omega) = 1$
b) Si $A_1, A_2, \ldots \in \mathcal{F}$ y $A_i \cap A_j = \emptyset$ para $i, j = 1, 2, \ldots$ con $i \neq j$, entonces

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

A $(\Omega, \mathcal{F}, P)$ se le denomina espacio de probabilidad.
Consecuencias

- $P(\emptyset) = 0$
- Si $A_1, \ldots, A_n \in \mathcal{F}$ y $A_i \cap A_j = \emptyset$ para todo $i, j = 1, \ldots n$ con $i \neq j$, entonces
  $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
- $P(\overline{A}) = 1 - P(A)$
- $P(A - B) = P(A) - P(A \cap B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Si $A \subseteq B$ entonces $P(A) \leq P(B)$
- $P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$

2.2.1. Probabilidad Condicionada

Definición 27 *Si B es un suceso no nulo, es decir $P(B) > 0$, definimos la probabilidad condicional del suceso A al ocurrir el suceso B como*

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Propiedad (Bayes)
$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B/A)P(A)}{P(B)}$

Regla del producto

- $P(A \cap B) = P(A)P(B/A) = P(B)P(A/B)$
- $P(A_1 \cap A_2 \cap \ldots \cap A_n) =$
  $P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \ldots P(A_n/A_1 \cap \ldots \cap A_{n-1})$

2.3. Fórmula de la probabilidad total

Sean $A_1, \ldots, A_n$ sucesos tales que
• $\cup_{i=1}^n A_i = \Omega$
• $A_i \cap A_j = \emptyset$ para todo $i, j = 1, \ldots n \quad i \neq j$ Una partición así se denomina sistema completo de sucesos.

Entonces si A es otro suceso

$$P(A) = \sum_{i=1}^n P(A \cap A_i) = \sum_{i=1}^n P(A_i)P(A/A_i)$$

2.4. Regla de Bayes

Sean $A_1, \ldots, A_n$ un sistema completo de sucesos, y sea B otro suceso

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B/A_i)}{\sum_{k=1}^n P(A_k)P(B/A_k)}$$

Propiedad
Si B es un suceso no nulo, podemos definir

$$\begin{array}{rccc} P(\bullet/B): & \mathcal{F} & \rightarrow & [0,1] \\ & A & \mapsto & P(A/B) \end{array}$$

Entonces $P(\bullet/B)$ es una medida de probabilidad y por lo tanto cumple todas las propiedades de las medidas de probabilidad.

2.5. Independencia Estadística

a) Dos sucesos A y B son estadísticamente independientes si $P(A \cap B) = P(A)P(B)$.

b) Diremos que los sucesos $\{A_i\}_{i\in I}$ son estadísticamente independientes si

$$P(\cap_{i\in J} A_i) = \prod_{i\in J} P(A_i) \text{ para todo } J \subseteq I \text{finito}$$

c) Diremos que los sucesos $\{A_i\}_{i\in I}$ son estadísticamente independientes dos a dos si

$P(A_i \cap A_j) = P(A_i)P(A_j)$ para todo
$i, j \in I\,, i \neq j$

Propiedad Si los sucesos $\{A_i\}_{i\in I}$ son estadísticamente independientes entonces son independientes dos a dos. El recíproco no es cierto en general.

Propiedad

- A y B son independientes sii $P(A/B) = P(A)$
- Las siguientes afirmaciones son equivalentes
  1. A y B son independientes
  2. $\overline{A}$ y B son independientes.
  3. $\overline{A}$ y $\overline{B}$ son independientes.
- si A, B, C, D son independientes entonces $A\cup B$ y $C \cup D$ son independientes. También son independientes $A \cup B$, C y D.

Capítulo 3

Variables aleatorias

Hasta ahora nuestros sucesos han sido de varios tipos: $\{C,+\}$ en la moneda, nombres de periódicos, ángulos en una ruleta, número de veces que sale cara en el lanzamiento de una moneda etc....

Necesitamos estandarizar de alguna manera todos estos sucesos. Una solución es asignar a cada suceso un cierto conjunto de números reales, es decir, convertir todos los sucesos en *sucesos de números reales* para trabajar con ellos de forma unificada. Para conseguirlo utilizaremos unas funciones que transformen los elementos del espacio muestral en números; esta funciones son las variables aleatorias.

3.1. Variables aleatorias

Comenzaremos dando una definición práctica de variable aleatoria.

3.1.1. Definición de variable aleatoria

Definición 28 *De manera informal*[1] *una variable aleatoria (v.a.) es una aplicación que toma valores numéricos determinados por el resultado de un experimento aleatorio*

$$X : \Omega \to I\!R$$

Notación:

[1]Formalmente dado un espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y el espacio probabilizable $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ una aplicación $X : \Omega \to \mathbb{R}$ es una v.a. si y sólo si $X^{-1}(B) \in \mathcal{F}$ para todo $B \in \mathcal{B}(\mathbb{R})$.

- Normalmente representaremos las v.a. por letras mayúsculas $X, Y, Z \ldots$
- Los valores que "*toman*" las v.a. los representaremos por letras minúsculas (las mismas en principio) $x, y, z \ldots$

Ejemplo 29 *Lanzar un dado de seis caras entonces*

$$\Omega = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$$

y una v.a

$$X : \Omega \to I\!R$$

sobre este espacio queda definida por $X(⚀) = 1, X(⚁) = 2, X(⚂) = 3, X(⚃) = 4, X(⚄) = 5, X(⚅) = 6$. *Ahora el suceso* $A = \{⚁, ⚃, ⚅\}$, *es decir "salir número par", es equivalente a* $\{X = 2, X = 4, X = 6\}$ *y el suceso* $B = \{⚀, ⚁, ⚂\}$, *es decir "salir un número inferior o igual a* 3.^{es} *en términos de la v.a.* $\{X = 1, X = 2, X = 3\}$ *o también* $\{X \leq 3\}$.

Ejemplo 30 *Consideremos el experimento lanzar una anilla al cuello de una botella. Si acertamos a ensartar la anilla en la botella el resultado del experimento es* éxito *y* fracaso *en caso contrario. El espacio muestral asociado a este experimento será* $\Omega = \{\textit{éxito, fracaso}\}$. *Construyamos la siguiente variable aleatoria:*

$$X : \{\textit{éxito, fracaso}\} \to I\!R$$

definida por $X(\textit{éxito}) = 1$ *y* $X(\textit{fracaso}) = 0$.

3.1.2. Tipos de variables aleatorias

Hay dos tipos fundamentales de variables aleatorias, las discretas y las continuas. Damos a continuación una definición informal[2] de estos tipos.

2

- La distinción entre variables aleatorias discretas y continuas es teóricamente más sofisticada. En la práctica serán discretas aquellas variables a las que podamos asignar probabilidades a los sucesos elementales y queden así descritas, siendo continuas las "restantes". En los problemas que trataremos en este curso esto será casi siempre así.
- Hay variables aleatorias mixtas que, no por ser menos frecuentes, tienen menos importancia. Por ejemplo el tiempo que está ocupado un procesador atendiendo los

Definición 31 *a) Una variable aleatoria es discreta si sólo puede tomar una cantidad numerable de valores con probabilidad positiva.*

b) La variables aleatorias continuas toman valores en intervalos.

Ejemplo 32 *Son variables aleatorias discretas:*

- *Número de artículos defectuosos en un cargamento.*
- *Número de clientes que llegan a una ventanilla de un banco en una hora.*
- *Número de errores detectados en las cuentas de una compañía.*
- *Número de reclamaciones de una póliza de un seguro médico.*

Son variables aleatorias continuas:

- *Renta anual de una familia.*
- *Cantidad de petróleo importado por un país*
- *Variación del precio de las acciones de una compañía de telecomunicaciones.*
- *Porcentaje de impurezas en un lote de productos químicos.*

3.2. Distribuciones de probabilidad para v.a. discretas.

Pasamos ahora a describir el comportamiento de la v.a. Para ello utilizaremos distintas funciones que nos darán algunas probabilidades de la variable aleatoria y que intentan emular a las frecuencias relativas y relativas acumuladas que vimos para muestras. En el caso discreto estas funciones son la de probabilidad, y la de distribución.

---

trabajos que llegan en una hora, si llega al menos un trabajo nos dará una v.a. continua pero si no llega ningún trabajo el tiempo que está ocupado es cero en ese punto, que posiblemente tenga probabilidad mayor que cero, la v.a. tiene un comportamiento discreto. Otro ejemplo más artificial es el siguiente juego: lanzar un dado al aire y si sale numero par anotar el resultado mientras que si sale impar lanzar un dardo a una diana y anotar la distancia al centro.

3.2.1. Función de probabilidad de una variable aleatoria discreta

En el caso discreto la función de probabilidad es la que nos da las probabilidades de los sucesos elementales de la v.a.

Definición 33 *La* función de probabilidad[3] *de una variable aleatoria discreta X a la que denotaremos por $P_X(x)$ está definida por*[4] $P_X(x) = P(X = x)$ *es decir la probabilidad de que X tome el valor x. Si X no asume ese valor entonces $P_X(x) = 0$. El conjunto $D_X = \{x \in I\!R \mid P_X(x) > 0\}$ recibe el nombre de soporte o dominio de la v.a. En el caso discreto lo más habitual es que $X(\Omega) = D_X$.*

Ejemplo 34 *En un dado de seis caras que se lanza una vez, en esta ocasión representaremos los sucesos elementales por el número de puntos de la cara obtenida, tenemos que $\Omega = \{1, 2, 3, 4, 5, 6\}$ y $X : \Omega \to I\!R$ viene definida por $X(i) = i$ para $i = 1, 2, 3, 4, 5, 6$. Supongamos que el dado está bien balanceado. Entonces*

$$P_X(1) = P_X(2) = P_X(3) = P_X(4) = P_X(5) = P_X(6) = \frac{1}{6}$$

Concretamente:

$$P_X(x) = \begin{cases} \frac{1}{6} & si\ x = 1, 2, 3, 4, 5, 6 \\ 0 & en\ otro\ caso \end{cases}$$

Ejemplo 35 *Sea X la v.a. asociada al lanzamiento de una moneda. $\Omega = \{c, +\}$:*

$$X(\omega) = \begin{cases} 1 & si\ \omega = c \\ 0 & si\ \omega = + \end{cases}$$

entonces:

$$P_X(x) = P(X = x) = \begin{cases} \frac{1}{2} & si\ x = 1, 0 \\ 0 & en\ otro\ caso \end{cases}$$

[3]También llamada función de cuantía o de masa de probabilidad, inglés " *probability mass function*"por lo que se abrevia con el acrónimo *pmf*, recordar esto cuando busquéis comandos en paquetes estadísticos

[4]Concretamente $P(X = x) = P(X^{-1}(x))$.

Ejemplo 36 *Tenemos una urna con tres bolas rojas,una negra y dos blancas. Realizamos una extracción y observamos el color de la bola entonces un espacio muestral es* $\Omega = \{roja, blanca, negra\}$ *una variable aleatoria asociada al experimento es:*

$$X(\omega) = \begin{cases} 1 & si\ \omega = roja \\ 2 & si\ \omega = negra \\ 3 & si\ \omega = blanca \end{cases}$$

entonces

$$P_X(x) = \begin{cases} \frac{3}{6} & si\ x = 1 \\ \frac{1}{6} & si\ x = 2 \\ \frac{2}{6} & si\ x = 3 \\ 0 & en\ otro\ caso \end{cases}$$

3.2.2. Propiedades de la función de probabilidad

Sea X una variable aleatoria discreta $X : \Omega :\to \mathbb{R}$ entonces los valores que *asume* la variable aleatoria X son $X(\Omega)$. Su función de probabilidad P_X verifica las siguientes propiedades:

1) $0 \leq P_X(x) \leq 1$ para todo $x \in \mathbb{R}$

2) $\sum_{x\in X(\Omega)} P_X(x) = 1$

Ejemplo 37 *Lanzamos al aire tres veces, de forma independiente, una moneda perfecta. El espacio muestral de este experimento es* $\Omega = \{ccc, cc+, c+c, +cc, c++, +c+, ++c, +++\}$ *(expresados en orden de aparición). Este espacio tiene todos los sucesos elementales equiprobables. Consideremos la variable aleatoria asociada a este experimento* $X =$ *número de caras en los tres lanzamientos. Entonces*

$$\begin{array}{l} P(X = 0) = P(\{+++\}) = \frac{1}{8} \\ P(X = 1) = P(\{c++, +c+, ++c\}) = \frac{3}{8} \\ P(X = 2) = P(\{cc+, c+c, +cc\}) = \frac{3}{8} \\ P(X = 3) = P(\{ccc\}) = \frac{1}{8} \end{array}$$

La función de probabilidad de X *es:*

$$P_X(x) = \begin{cases} \frac{1}{8} & si\ x = 0,3 \\ \frac{3}{8} & si\ x = 1,2 \\ 0 & en\ otro\ caso \end{cases}$$

3.3. Función de distribución

Otra función que nos será útil es la que da las probabilidades de los sucesos del tipo $\{X \leq x\}$, por lo que también recibe el nombre de distribución de frecuencias acumuladas.

Definición 38 *La función de probabilidad acumulada*[5] X *(de cualquier tipo; discreta o continua)* $F_X(x)$ *representa la probabilidad de que* X *no tome un valor superior a* x *es decir*[6]*:*

$$F_X(x) = P(X \leq x)$$

Proposición 39 *Sea* F_X *la función de distribución de una v.a.* X *entonces:*

a) $0 \leq F_X(x) \leq 1$.

b) La función F_X *es no decreciente.*

c) La función F_X *es continua por la derecha.*

d) Se cumple que $\lim\limits_{x\to\infty} F_X(x) = 1$; $\lim_{x\to-\infty} F_X(x) = 0$.

e) Toda función F *verificando las propiedades anteriores es función de distribución de alguna v.a.* X.

f) $P(X > x) = 1 - F_X(x)$

g) Dados $a, b \in \mathbb{R}$ *con* $a < b$

$$P(a < X \leq b) = F_X(b) - F_X(a)$$

En la última propiedad anterior no se pueden cambiar en general las desigualdades a estrictas o no estrictas, veamos que propiedades tenemos cuando se cambian estas desigualdades.

[5] También llamada función de distribución de probabilidad o simplemente función de distribución de una v.a., y en inglés *cumulative distribution function* por lo que se abrevia con el acrónimo *cdf*, que es el que se suele utilizar en los paquetes estadísticos

[6] Más concretamente

$$F_X(x) = P(X \leq x) = P(X^{-1}((-\infty, x])$$

Proposición 40 *Sea F_X una función de distribución de la v.a. X y denotamos por $F_X(x_0^-) = \lim_{x\to x_0^-} F_X(x)$, entonces.*

a) $P(X = x) = F_X(x) - F_X(x^-)$

b) $P(a < X < b) = F_X(b^-) - F_X(a)$

c) $P(a \leq X < b) = F_X(b^-) - F_X(a^-)$

d) $P(X < a) = F_X(a^-)$

e) $P(a \leq X \leq b) = F_X(b) - F_X(a^-)$

f) $P(X \geq a) = 1 - F_X(a^-)$

La siguiente afirmación es consecuencia inmediata del primer ítem de la proposición anterior.

Corolario 41 *Si F_X es continua en x se tiene que $P(X = x) = 0$.*

Proposición 42 *Sea X una variable aleatoria discreta que toma valores en $X(\Omega)$ y que tiene por función de probabilidad $P_X(x)$ entonces su función de distribución $F_X(x_0)$ es*

$$F_X(x_0) = \sum_{x\leq x_0} P_X(x)$$

donde $\sum_{x\leq x_0}$ indica que sumamos todos los $x \in X(\Omega)$ tales que $x \leq x_0$

Demostración:

$$F_X(x_0) = P(X \leq x_0) = P\left(\bigcup_{x\leq x_0; x\in X(\Omega)} \{x\}\right) = \sum_{x\leq x_0} P(X = x) = \sum_{x\leq x_0} P_X(x).$$

Ejemplo 43 *En el experimento del dado se tiene que:*

$$P_X(x) = \begin{cases} \frac{1}{6} & \text{si } x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{en el resto de casos} \end{cases},$$

por lo tanto

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & si\ x < 1 \\ \frac{1}{6} & si\ 1 \leq x < 2 \\ \frac{2}{6} & si\ 2 \leq x < 3 \\ \frac{3}{6} & si\ 3 \leq x < 4 \\ \frac{4}{6} & si\ 4 \leq x < 5 \\ \frac{5}{6} & si\ 5 \leq x < 6 \\ 1 & si\ 6 \leq x \end{cases}$$

Calculemos más detalladamente algún valor de F_X, por ejemplo:

$F_X(3{,}5) = P(X \leq 3{,}5) = P(\{X=1\} \cup \{X=2\} \cup \{X=3\}) = P(\{X = 1\}) + P(\{X=2\}) + P(\{X=3\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$,

o de otra forma

$$F_X(3{,}5) = \sum_{x \leq 3{,}5} P_X(x) = \sum_{x=1}^{3} P(X=x) = \sum_{x=1}^{3} \frac{1}{6} = 3 \cdot \frac{1}{6} = \frac{1}{2}.$$

3.4. Momentos de variables aleatorias discretas

Al igual que el capítulo de estadística descriptiva utilizamos distintas medidas para resumir los valores centrales y para medir la dispersión de una muestra, podemos definir las correspondiente medidas para variables aleatorias. A estas medidas se les suele añadir el adjetivo poblacionales mientras que a las que provienen de la muestra se las adjetiva como muestrales.

3.4.1. Esperanza para variables aleatorias discretas

Una vez hemos descrito el comportamiento de las probabilidades de una variable aleatoria discreta mediante las funciones de probabilidad y de distribución, profundizaremos un poco más en la descripción de una v.a. Vamos a buscar un valor que resuma toda la variable. Este valor es el que "*esperamos*" que se resuma la v.a. o esperamos que las realizaciones de la v.a. queden cerca de él.

Definición 44 *El valor esperado $E(X)$ de una v.a. discreta X, se define* *como*

$$E(X) = \sum_{x \in X(\Omega)} x P_X(x)$$

En ocasiones se le domina media (poblacional) y muy frecuentemente se la denota $\mu_X = E(X)$ *o simplemente* $\mu = E(X)$.

Relación de la esperanza para variables aleatorias discretas con la media aritmética

Supongamos que lanzamos un dado n veces y obtenemos unas frecuencias absolutas n_i para el resultado i con $i = 1, \ldots, 6$. Sea X la v.a. que nos representa el valor de una tirada del dado.

Calculemos la media aritmética (o media muestral) de los datos

$$\overline{x} = \frac{1n_1 + 2n_2 + 3n_3 + 4n_4 + 5n_5 + 6n_6}{n} = \sum_{x=1}^{6} x \frac{n_x}{n}$$

Si $n \to \infty$ entonces $lim_{n\to\infty} \frac{n_x}{n} = P_X(x)$ y por lo tanto

$$E(X) = lim_{x\to\infty} \sum_{x=1}^{6} x \frac{n_x}{n}$$

Entonces el valor esperado en una v.a. discreta puede entenderse como el valor promedio que tomaría una v.a. en un número grande de repeticiones.

Ejemplo 45 *Sea* $X =$ *número de errores de imprenta en una página de un libro, y resulta que*

$$P(X = 0) = 0{,}42, \qquad P(X = 1) = 0{,}4, \quad P(X = 2) = 0{,}18$$

entonces $E(X) = 0 \cdot 0{,}42 + 1 \cdot 0{,}4 + 2 \cdot 0{,}18 = 0{,}76$

Elegida una página del libro con igual probabilidad esperamos encontrar 0,76 *errores.*

3.4.2. Esperanzas de funciones de variables aleatorias discretas

Supongamos que en el ejemplo anterior el autor nos paga 2 euros por cada página que encontremos con 1 error y 3 euros por cada página con dos errores (y nada por las páginas correctas) ¿Cuánto esperamos cobrar si analizamos una página?

$$0 \cdot 0{,}42 + 2 \cdot 0{,}4 + 3 \cdot 0{,}18 = 1{,}34$$

Definición 46 *Sea X una v.a. discreta con función de probabilidad P_X y de distribución F_X. Entonces el valor esperado de $g(x)$ es :*

$$E(g(X)) = \sum_x g(x) P_X(x)$$

La demostración de las siguientes propiedades se deja como ejercicio.

Proposición 47

- *$E(k) = k$. para cualquier constante k.*
- *Si $a \leq X \leq b$ entonces $a \leq E(X) \leq b$.*
- *Si X es una v.a. discreta que toma valores enteros no negativos entonces $E(X) = \sum_{x=0}^{+\infty}(1 - F_X(x))$*

Ejemplo 48 *Supongamos que estamos sentados delante de nuestro ordenador con un amigo y le decimos que en dos minutos podemos programar una paleta para poner colores a unos gráficos. Queremos la que paleta tenga dos botones con las opciones color rojo y color azul. Como hemos programado a gran velocidad resulta que el programa tiene un error; cada vez que se abre la paleta los colores se colocan al azar (con igual probabilidad) en cada botón, así que no sabemos en que color hemos de pinchar. Además, como nos sobraron* 15 *segundos para hacer el programa y pensando en la comodidad del usuario, la paleta se cierra después de haber seleccionado un color y hay que volverla a abrir de nuevo. La pregunta es ¿cuál es el valor esperado del número de veces que hemos pinchar el botón de color azul antes de obtener este color?*

Llamemos X al número de veces que pinchamos en el botón azul (y nos sale rojo) hasta obtener el primer azul. La variable X toma valores en los enteros no negativos. Su función de probabilidad queda determinada por

$$P_X(x) = P(X = x) = P(\overbrace{rojo, rojo, \ldots, rojo}^{x\,veces}, azul) = \left(\frac{1}{2}\right)^{x+1}$$

Por lo tanto [7]

$$E(X)=\sum_{x=0}^{+\infty} xP(X=x)=\sum_{x=0}^{+\infty} x\left(\frac{1}{2}\right)^{x+1}=\left(\frac{1}{2}\right)^{2}\sum_{x=1}^{+\infty} x\left(\frac{1}{2}\right)^{x-1}=\left(\frac{1}{2}\right)^{2}\frac{1}{(1-\frac{1}{2})^{2}}=1$$

Por otra parte, calculemos su función de distribución

$$F_X(x)=P(X\leq x)=\sum_{k=0}^{x} P(X=k)=\sum_{k=0}^{x}(\frac{1}{2})^{k+1}=\frac{\frac{1}{2}-\frac{1}{2}^{x+1}\frac{1}{2}}{1-\frac{1}{2}}=1-(\frac{1}{2})^{x+1}$$

Ahora como la variable toma valores enteros positivos, podemos calcular su valor esperado de esta otra manera
$E(X)=\sum_{x=0}^{+\infty}(1-F_X(x))=\sum_{x=0}^{+\infty}(\frac{1}{2})^{x+1}=\frac{1}{2}\frac{1}{1-\frac{1}{2}}=1.$

Queda como ejercicio que calculéis el valor esperado de la variable $Y=$ *número de intentos para conseguir el color azul.*

3.4.3. Momentos de una v.a. discreta

Se llama momento de orden r respecto al punto C a $E((X-C)^m)$. Cuando $C=0$ los momentos reciben el nombre de momentos respecto al origen y

[7] Recordemos conceptos básicos de las series geométricas. Una progresión geométrica es de la forma $r^0,r^1,\ldots,r^n,\ldots$, el valor r recibe el nombre de razón de la progresión geométrica. La serie geométrica es la suma de todos los valores de la progresión geométrica $\sum_{k=0}^{+\infty} r^k$. Se cumplen las siguientes propiedades:

- Las sumas parciales desde el término n_0 al n de una progresión geométrica son $\sum_{k=n_0}^{n} r^k=\frac{r^{n_0}-r^n r}{1-r}$
- Si $|r|<1$ la serie geométrica es convergente y $\sum_{k=0}^{+\infty} r^k=\frac{1}{1-r}$. En el caso en que se comience en n_0 se tiene que $\sum_{k=n_0}^{+\infty} r^k=\frac{r^{n_0}}{1-r}$.
- Si $|r|<1$ también son convergentes las derivadas, respecto de r, de la serie geométrica y convergen a la derivada correspondiente. Así tenemos que

$$\left(\sum_{k=0}^{+\infty} r^k\right)'=\sum_{k=1}^{+\infty} kr^{k-1}=\left(\frac{1}{1-r}\right)'=\frac{1}{(1-r)^2}$$

$$\left(\sum_{k=0}^{+\infty} r^k\right)''=\sum_{k=2}^{+\infty} k(k-1)r^{k-2}=\left(\frac{1}{1-r}\right)''=\frac{2}{(1-r)^3}$$

cuando $C = E(X)$ reciben el nombre de centrales o respecto de la media, así la esperanza es el momento de orden 1 respecto al origen. Estos momentos son la versión poblacional de los momentos que vimos en el capítulo de estadística descriptiva, recibiendo estos último el nombre de momentos muestrales.

3.4.4. Varianza de una variable aleatoria discreta

Ya tenemos descrito el comportamiento aleatorio de una v.a. discreta mediante P_X y F_X, también tenemos un valor central; el valor esperado $E(X)$. Nos queda encontrar una medida de lo lejos que están los datos del valor central $E(X)$ una de estas medidas es la varianza de X.

Definición 49 *Sea X una v.a. Llamaremos varianza de X a*

$$Var(X) = E((X - E(X))^2)$$

Por lo tanto la varianza es el momento central de orden 2.

De forma frecuente se utiliza la notación $\sigma_X^2 = Var(X)$. A la raíz cuadrada positiva de la varianza $\sigma_X = \sqrt{Var(X)}$ se la denomina desviación típica o estándar de X.

En el caso discreto:

$$\sigma_X^2 = Var(X) = E((X - E(X))^2) = \sum_x (x - E(X))^2 P_X(x)$$

Proposición 50 *Sea X una v.a.*

$$Var(X) = E(X^2) - (E(X))^2 = \sum_x x^2 P_X(X) - (E(X))^2$$

Demostración:

$$Var(X) = \sum_x (x - E(X))^2 P_X(x) = \sum_x (x^2 - 2xE(X) + (E(X)^2) P_X(x) =$$

$$\sum_x x^2 P_X(x) - E(X) \sum_x 2x P_X(x) + (E(X)^2) \sum_x P_X(x) =$$

$$E(X^2) - 2E(X)E(X) + (E(X))^2 = E(X^2) - (E(X))^2.$$

Ejemplo 51 *Vamos a calcular en el ejemplo anterior la varianza del número de errores. Recordemos que:*

$$P(X=0)=0{,}42, \quad P(X=1)=0{,}4, \quad P(X=2)=0{,}18$$

y $E(X)=0{,}76$

Entonces:

$$Var(X)=E(X^2)-(E(X))^2=E(X^2)-(0{,}76)^2$$

Ahora

$$E(X^2)=0^2(0{,}41)+1^2(0{,}4)+2^2(0{,}18)=0{,}4+0{,}72=1{,}12$$

y por lo tanto

$$Var(X)=E(X^2)-(0{,}76)^2=1{,}12-0{,}5776=0{,}542$$

y

$$\sqrt{Var(X)}=\sqrt{0{,}542}$$

En resumen $\sigma_X^2=0{,}542$ *y* $\sigma_X=\sqrt{0{,}542}$

Proposición 52

- $Var(X)\geq 0$
- $Var(cte)=E(cte^2)-(E(cte))^2=cte^2-cte^2=0$
- *El mínimo de* $E((X-C)^2)$ *se alcanza cuando* $C=E(X)$ *y es* $Var(X)$*. Esta propiedad es una de las que hace útil a la varianza como medida de dispersión.*

Demostración:(ejercicio)

Esperanza y varianza de transformaciones lineales

Un cambio lineal o transformación lineal de una v.a. X es otra v.a. $Y=a+bX$ donde $a,b\in\mathbb{R}$. Sea X una v.a. con $E(X)=\mu_X$ y $Var(X)=\sigma_X^2$ y $a,b\in\mathbb{R}$. Entonces si $Y=a+bX$:

- $E(Y)=E(a+bX)=a+bE(X)=a+b\mu_X$.
- $Var(Y)=Var(a+bX)=b^2Var(X)=b^2\sigma_X^2$

- $\sigma_Y = \sqrt{Var(Y)} = \sqrt{b^2 Var(X)} = |b|\sigma_X$

Demostración:

$$E(Y) = E(a+bX) = \sum_x (a+bX)P_X(x) = a\sum_x P_X(x) + b\sum_x xP_X(x) = a+bE(X) = a+b\mu_X$$

Las demostración de las demás propiedades queda como ejercicio.

3.5. Variables aleatorias continuas

Como ya hemos dicho las variables aleatorias continuas toman valores en intervalos. Lo más habitual es que estas variables tengan función de distribución continua y derivable (salvo a los más en una cantidad finita o numerable de puntos). En lo que sigue supondremos que la función de distribución de variables aleatorias continuas cumplen estas propiedades.

3.5.1. Distribuciones de probabilidad de v.a. continuas

Notemos que si X es una v.a. con función de distribución continua se tiene que $P(X = x_0) = F_X(x_0) - F(x_0^-) = 0$. Por lo que no tiene sentido definir ‘’*función de probabilidad*”. Más en general tendremos que $P(X < x_0) = P(X \leq x_0)$. Por otra parte podemos utilizar una regla parecida del cociente entre casos favorables y casos posibles de Laplace pero en este caso el conteo se hace por la “medida” de los casos posibles partida por la “medida” de los casos favorables. Veamos un ejemplo de v.a. continua, que ampliaremos en el tema siguiente, en el que se utilizan todos estos conceptos.

Ejemplo 53 ***Distribución uniforme en el intervalo unidad.***

Supongamos que lanzamos un dardo a una diana de radio 1, de forma que sea “equiprobable” cualquier distancia al centro[8]*. Consideremos la v.a. continua X =distancia al centro.*

$$F_X(x) = \begin{cases} 0 & si\ x \leq 0 \\ x & si\ 0 < x < 1 \\ 1 & si\ x \geq 1 \end{cases}$$

Ya que

[8]!Cuidado! esto no es equivalente a que cualquier punto de la diana sea “*equiprobable*”.

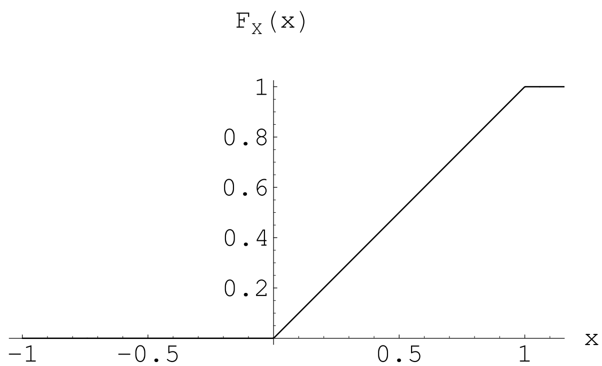


Figura 3.1: Función de distribución de una v.a. uniforme en el intervalo unidad.

- *C.F. ''longitud favorable''* $x - 0$
- *C.P. ''longitud posible''* $1 - 0$
- *Luego* $P(X \leq x) = \frac{x-0}{1-0} = x$

Por ejemplo

$$F_x(0{,}75) = 0{,}75$$

En las variables continuas los sucesos del tipo $\{X \leq x\}$ y $\{X < x\}$ tendrán la misma probabilidad, y otros tipos de sucesos similares también, algunas de estas propiedades se explicitan en la siguiente proposición.

Proposición 54 *Dada una v.a. continua* X *se tiene que:*

a) $P(X \leq b) = P(X < b)$

b) $P(X < b) = P(X < a) + P(a < X < b)$

c) $P(a < X < b) = P(X < b) - P(X < a)$

Demostración:

b) $\{X < a\} \cap \{a < X < b\} = \emptyset$ $\{X < a\} \cup \{a < X < b\} = \{X < a\}$ entonces

$$P(X \leq b) = P(\{X < a\} \cup \{a < X < b\}) = P(X < a) + P(a < X < b)$$

a) $P(X < b) = P(X < b) + P(X = b) = P(X < b)$

c) Ídem que b) aplicando a).

Las propiedades anteriores y combinaciones de ellas se pueden escribir utilizando la función de distribución de X:

Proposición 55 *Dada una variable aleatoria continua se tiene que:*

a) $F_X(b) = F_X(a) + P(a < X < b)$

b) $P(a < X < b) = F_X(b) - F_X(a)$

c) $P(a \leq X \leq b) = F_X(b) - F_X(a)$

Demostración: ejercicio.

Ejemplo 56 *En los dardos:*

$$P(0{,}25 < X < 0{,}3) = F_X(0{,}3) - F_X(0{,}25) =$$

$$= 0{,}3 - 0{,}25 = 0{,}05$$

3.5.2. Función de densidad

Una función $f : \mathbb{R} \to \mathbb{R}$ es una función de densidad sobre $\mathbb{R}$ si cumple que

a) $f_X(x) \geq 0$ para todo $x \in \mathbb{R}$.

b) f es continua salvo a lo más en una cantidad finita de puntos sobre cada intervalo acotado de $I\!R$, es decir es integrable Riemman.

c) $\int\limits_{-\infty}^{+\infty} f_X(x)dx = 1$.

Definición 57 *Sea X una v.a. con función de distribución F_X. Sea $f : I\!R \to R$ una función de densidad tal que*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt. \text{ para todo } x \in I\!R.$$

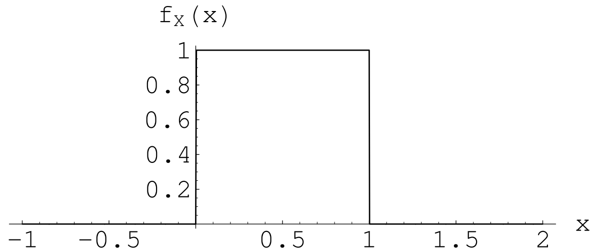


Figura 3.2: Función de densidad de una v.a. uniforme en el intervalo $(0, 1)$

Entonces X es una variable aleatoria continua y f_X es la densidad de la v.a. X.

El conjunto $D_X = \{x \in \mathbb{R} | f_x(x) > 0\}$ recibe el nombre de soporte o dominio de la variable aleatoria continua y se interpreta su conjunto de resultados posibles.

Ejemplo 58 *En nuestra diana la función f es una densidad*

$$f_X(x) = \begin{cases} 0 & si\ x \leq 0 \\ 1 & si\ 0 < x < 1 \\ 0 & si\ 1 \leq x \end{cases}$$

que es la densidad de X, en efecto:

- *Si $x \leq 0$ entonces $\int_{-\infty}^{x} f_X(t)dt = 0$.*
- *Si $0 \leq x \leq 1$ entonces $\int_{-\infty}^{x} f_X(t)dt = \int_0^x 1dt = x$.*
- *Si $x \geq 1$ entonces $\int_{-\infty}^{x} f_X(t)dt = \int_0^1 1dt = 1$.*

Luego $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$ para todo $x \in \mathbb{R}$.

La función de densidad nos permite calcular diversas probabilidades.

Proposición 59 *Sea X una v.a. continua con función de distribución F_X y de densidad f_X, entonces*

1. $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f_X(x)dx$.

2. *Si A es un recinto adecuado*[9] *de* $\mathbb{R}$ *entonces* $P(X \in A) = \int_A f(x)dx = \int_{A\cap D_X} f(x)dx$.

La siguiente proposición fija algunas propiedades de la función de densidad para v.a. continuas y nos da un método de cálculo; la densidad es la derivada de la función de distribución.

Proposición 60 *Sea X una v.a. continua con función de distribución F_X y de densidad f_X, entonces:*

a) F_X *es continua.*

b) Si f_x es continua en un punto x, F_X es derivable en ese punto y $F'_X(x) = f_X(x)$.

c) $P(X = x) = 0$ *para todo* $x \in \mathbb{R}$.

Ejemplo 61 *Sea X = tiempo de ejecución de un proceso. Se supone que X sigue una distribución uniforme en dos unidades de tiempo, si tarda más el proceso se cancela. Entonces*

$$F_X(x) = P(X \leq x) = \frac{CF}{CP} = \frac{x}{2}$$

Luego su función de distribución es:

$$F_X(x) = \begin{cases} 0 & si\ x \leq 0 \\ \frac{x}{2} & si\ 0 < x < 2 \\ 1 & si\ 2 \leq x \end{cases}$$

mientras que su función de densidad es:

$$f_X(x) = F'_X(x) = \begin{cases} 0 & si\ x \leq 0 \\ \frac{1}{2} & si\ 0 < x \leq 2 \\ 0 & si\ 2 \leq x \end{cases}$$

Efectivamente

- $f_X(x) \geq 0$, *y tiene un conjunto finito de discontinuidades.*
- $F_X(x) = \int_{-\infty}^{x} f_X(t)dt.$ *para todo* $x \in \mathbb{R}$ *(ejercicio, resolverlo gráficamente.)*

[9] Un boreliano

- $\int\limits_{-\infty}^{+\infty} f_X(x)dx = \int\limits_{0}^{2} \frac{1}{2}dx = \frac{x}{2}\mid_0^2 == \frac{2}{2} - \frac{0}{2} = 1.$

Ejercicio: Calcular la probabilidad de que uno de nuestros procesos tarde más de una unidad de tiempo en ser procesado. Calcular también la probabilidad de que dure entre 0,5 y 1,5 unidades de tiempo.

3.6. Momentos para variables aleatorias continuas

Los mismos comentarios y definiciones que se dieron en la sección correspondiente del tema de estadística descriptiva son aplicables aquí. Así que sólo daremos las definiciones, la forma de cálculo y algunos ejemplos.

3.6.1. Esperanza de una v.a. continua

Sea X una v.a. continua con función de densidad $f_X(x)$ entonces:

- su esperanza es : $E(X) = \int\limits_{-\infty}^{+\infty} x f_X(x)dx.$
- Si $g(x)$ es una función de la variable X entonces

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x)dx.$$

3.6.2. Varianza de una v.a. continuas

- $Var(X) = \sigma_X^2 = E((X - \mu_X)^2) = \int\limits_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x)dx.$
- A $\sigma_X = +\sqrt{\sigma_X^2}$ se le denomina desviación típica o estándar de X.

Propiedades

- $\sigma_X^2 \geq 0$
- $Var(cte) = E(cte^2) - (E(cte))^2 = cte^2 - cte^2 = 0$

- $Var(x) = E(X^2) - \mu_X^2 = \int_{-\infty}^{+\infty} x^2 f_X(x)dx - \mu_X^2$.
- El mínimo de $E((X-C)^2)$ se alcanza cuando $C = E(X)$ y es $Var(X)$.

Ejemplos Calcular μ_X y σ_X^2 en el dardo.
Resultado $\mu_X = \frac{1}{2}$, $E(X^2) = \frac{1}{3}$, $Var(X) = \frac{1}{12}$.

Esperanza y varianza de trasformaciones lineales

Sea X una v.a. continua con $E(X) = \mu_X$ y $Var(X) = \sigma_X^2$ sea $Y = a + bX$, donde $a, b \in \mathbb{R}$, es una nueva v.a. continua obtenida mediante una transformación lineal de X. Se verifican las mismas propiedades que en el caso discreto:

- $E(Y) = E(a + bX) = a + bE(X)$
- $Var(Y) = Var(a + bX) = b^2 Var(X)$
- $\sigma_Y = |b|\sigma_X$
- $Z = \frac{X-\mu_X}{\sigma_X}$ es una transformación lineal de X de forma que

$$E(Z) = 0 \text{ y } Var(Z) = 1$$

Ejemplo En una empresa de venta de vinos por internet, sea $X =$ número de litros de vino del país vendidos en un año. Supongamos que sabemos que $E(X) = 10000$ y que $Var(X) = 100$ Supongamos que los gastos fijos de distribución son 50000 y el beneficio por litro es de 10 pts por botella. Definimos $T = 10X - 50000$ que será el beneficio después de gastos entonces:

$$E(T) = 10E(X) - 50000 = 50000$$

y

$$Var(T) = 10^2 VAR(X) = 10000$$

3.7. Transformación de variables aleatorias

Muchas variables aleatorias son funciones de otras v.a. En lo que sigue resumiremos diversas técnicas para dada una v.a. X y una transformación $Y = h(X)$ encontrar F_Y a partir de F_X.

3.7.1. Transformaciones de v.a. discretas

Proposición 62 *Sea X una v.a. discreta con $X(\Omega) = \{x_1, x_2, \ldots, x_n, ..\}$ y sea $h : \mathbb{R} \to \mathbb{R}$ una aplicación. Entonces $Y = h(X)$ es también una v.a. discreta. Además si P_X y F_X son las funciones de probabilidad y de distribución de X entonces*

a) $P_Y(y) = \displaystyle\sum_{x_i | h(x_i) = y} P_X(x_i).$

b) $F_Y(y) = \displaystyle\sum_{x_i | h(x_i) \leq y} P_X(x_i).$

3.7.2. Transformaciones de v.a. continuas

Desafortunadamente este caso no es tan sencillo como el anterior, pues la transformación de una v.a. continua puede ser continua, discreta, mixta ...

Proposición 63 *Sea X una v.a. continua cuya función de densidad es f_X. Sea $h : \mathbb{R} \to \mathbb{R}$ una aplicación estrictamente monótona y derivable, tal que $h'(x) \neq 0$ para todo $x \in \mathbb{R}$. Sea $Y = h(X)$ la transformación de X por h. Entonces Y es una v.a. continua con función de densidad*

$$f_Y(y) = \left. \frac{f_X(x)}{|h'(x)|} \right|_{x = h^{-1}(y)}$$

Proposición 64 *Sea X una v.a. continua cuya función de densidad es f_X. Si $h : \mathbb{R} \to \mathbb{R}$ es una aplicación, no necesariamente monótona, pero sí derivable con derivada no nula, y si la ecuación $h(x) = y$ tiene un número finito de soluciones $x_1, x_2, .., x_n$ entonces:*

$$f_Y(y) = \sum_{k=1}^{n} \left. \frac{f_X(x)}{|h'(x)|} \right|_{x = x_k}$$

Método general

Cuando no podamos aplicar las propiedades anteriores intentaremos calcular primero la función de distribución de la transformación y luego su densidad.

Notemos que en general si $Y = g(X)$ es una v.a. transformación de la v.a. X entonces

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

Por ejemplo si g es estrictamente creciente y cont.

$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

y si g es estrictamente decreciente y cont.

$$F_Y(y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

3.8. Desigualdad de Chebyshef

Veremos en esta sección distintas desigualdades que acotan determinadas probabilidades de una variable aleatoria. Estas desigualdades sirven en algunos casos para acotar probabilidades de determinados sucesos, también son interesantes desde el punto de vista teórico y por ejemplo para justificar que la varianza es una mediada de la dispersión de los datos

3.8.1. Desigualdad de Markov

Proposición 65 *Sea X una v.a. positiva con $E(X)$ finita. Entonces $P(X \geq a) \leq \frac{E(X)}{a}$ para todo $a > 0$.*

Demostración:

Si X es continua y sólo toma valores positivos

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{0}^{+\infty} x f_X(x) dx = \int_{0}^{a} x f_X(x) dx + \int_{a}^{+\infty} x f_X(x) dx \geq \int_{a}^{+\infty} x f_X(x) dx \geq a \int_{a}^{+\infty} f_X(x) dx = aP(X \geq a)$$

de donde se sigue que $P(X \geq a) \leq \frac{E(X)}{a}$

Corolario 66 *Sea X una v.a. con $E(X)$ finita entonces $P(|X| \geq a) \leq \frac{E(|X|)}{a}$. Para todo $a > 0$*

Desigualdad de Chebyshef

Proposición 67 *Sea X una v.a.con $E(X) = \mu$ y $Var(X) = \sigma^2$ entonces*

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

para todo $a > 0$

Demostración:

Apliquemos la consecuencia de la desigualdad de Markov a la v.a. no negativa $Y^2 = (X - \mu)^2$ entones

$P(Y^2 \geq a^2) \leq \frac{E(Y^2)}{a^2} = \frac{E((X-\mu)^2)}{a^2} = \frac{Var(X)}{a^2} = \frac{\sigma^2}{a^2}$

Por otra parte

$$P(Y^2 \geq a^2) = P(|Y| \geq a) = P(|X - \mu| \geq a)$$

hecho que, junto con la desigualdad anterior, demuestra el resultado.

Observación: Supongamos que X es una v.a. con $Var(X) = 0$ entonces aplicando la desigualdad anterior

$P(|X - E(X)| \geq a) = 0$ para todo $a > 0$ lo que implica que $P(X = E(X)) = 1$ luego la probabilidad de que X sea constantemente $E(X)$ es 1. Lo que nos confirma la utilidad de la varianza es una medida de la dispersión de los datos.

Ejemplo 68 *Se sabe que el tiempo de respuesta medio y la desviación típica de un sistema multiusuario son 15 y 3 u.t. respectivamente. Entonces:*

$P(|X - 15| \geq 5) \leq \frac{9}{25} = 0{,}36.$

Si substituimos a por $a\sigma$ en la desigualdad de Chebyshef. Entonces nos queda:

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

Que es otra manera de expresar la desigualdad de Chebyshef.

La desigualdad de Chebyshef también se puede escribir de al menos dos maneras más:

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

$P(\mu - a \cdot \sigma \leq X \leq \mu + a \cdot \sigma)$

La varianza como medida de dispersión

Tomando la segunda expresión que hemos visto para la desigualdad de Chebyshef para distintos valores de $a > 0$ tenemos la siguiente tabla.

a	$P(\|X - E(X)\| \geq a\sigma)$
1	≤ 1
2	$\leq 0{,}25$
3	$\leq 0{,}111$
4	$\leq 0{,}0025$

Lo que se interpreta, por ejemplo para $a = 2$, como que dada una v.a. X con cualquier distribución que tenga $E(X)$ y $Var(X)$ finitos la probabilidad de que un valor se aleje de la media μ más de 2 desviaciones típicas es menor o igual que 0,25. Es decir sólo el 25 % de los valores estarán alejados de la media más de 2σ !sea cual sea la distribución de la v.a.!

Capítulo 4

Distribuciones notables

En este tema estudiaremos diversos tipos de experimentos que son muy frecuentes y algunas de las variables aleatorias asociadas a ellos. Estas variables reciben distintos nombres que aplicaremos sin distinción al tipo de población del experimento a la variable o a su función de probabilidad, densidad o distribución.

4.1. Algunas variables aleatorias discretas

Vamos a ver distintas v.a. discretas que se presentan con frecuencia ya que están relacionadas con situaciones muy comunes como el número de caras en varios lanzamiento de una moneda, el número de veces que una maquina funciona hasta que se estropea, el numero de clientes en una cola,...

4.1.1. Bernoulli

Consideremos un experimento con dos resultados posibles éxito (E) y fracaso (F). Sea $\Omega = \{E, F\}$ el e.m. asociado al experimento . De forma que sabemos que $P(E) = p$ y $P(F) = 1 - p = q$ con $0 < p < 1$. Consideremos la aplicación $X : \Omega = \{E, F\} \to \mathbb{R}$ definida por $X(E) = 1$, $X(F) = 0$ entonces su función de probabilidad es

$$P_X(x) = \begin{cases} q & \text{si } x = 0 \\ p & \text{si } x = 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Bajo estas condiciones diremos que X sigue una distribución de probabilidad Bernoulli de parámetro p y lo denotaremos por $X \equiv Ber(p)$ o $X \equiv B(1,p)$. A los experimentos de este tipo (éxito/fracaso) se les denomina experimentos Bernoulli.

Resumen v.a con distribución Bernoulli, $Ber(p)$

Valores admisibles.	$P_X(x) = P(X = x) =$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = \{0,1\}$	$\left\{\begin{array}{ll} q & \text{si } x = 0 \\ p & \text{si } x = 1 \\ 0 & \text{en otro caso} \end{array}\right.$	$\left\{\begin{array}{ll} 0 & \text{si } x < 0 \\ q & \text{si } 0 \leq x < 1 \\ 1 & \text{si } 1 \leq x \end{array}\right.$	p	pq

4.1.2. Binomial:

Supongamos que repetimos n veces de forma independiente un experimento Bernoulli de parámetro p. Entonces Ω estará formado por cadenas de E's y F's de longitud n.

Sea $X : \Omega \to \mathbb{R}$ definida por $X(\omega)$ =número de éxitos en ω. Entonces

$$P_X(x) = \left\{\begin{array}{ll} \binom{n}{x} p^x (1-p)^{n-x} \text{ si } x = 0, 1, \ldots, n & \\ 0 & \text{en otro caso} \end{array}\right. .$$

Entonces diremos que la v.a. sigue una ley de probabilidad binomial con parámetros n y p y lo denotaremos por $X \equiv B(n,p)$. (Nota: $B(1,p) = Ber(p)$).

Su función de distribución no tienen una forma general, por ello esta tabulada en campus extens disponéis de unas tablas de esta distribución para distintos valores de n y p. Hoy por hoy cualquier paquete estadístico, hoja de cálculo,... dispone de funciones para el cálculo de estas probabilidades, así que el uso de las tablas queda algo anticuado. Veremos algunas formas de aproximar las probabilidades de una Binomial, en este capítulo y en el siguiente, mediante otras variables con distribución Poisson o con distribución normal.

Resumen v.a con distribución binomial $B(n,p)$

Valores admisibles.	$P_X(x)=P(X=x)=$	$F_X(x)=$ $P(X\leq X)=$	$E(X)$	$Var(X)$
$D_X=\{0,1,\ldots,n\}$	$\begin{cases} \binom{n}{x}p^x q^{n-x} & \text{si } x=0,1,\ldots,n \\ 0 & \text{en otro caso} \end{cases}$	Tabulada	np	npq

4.1.3. Geométrica

Consideremos la experiencia consistente en repetir un experimento Bernoulli, de parámetro p, de forma independiente hasta obtener el primer éxito.

Sea X la v.a. que cuenta el número de intentos necesarios para obtener el primer éxito. Entonces

$$P_X(x)=P(X=x)=\begin{cases} q^{x-1}p & \text{si } x=1,2,\ldots \\ 0 & \text{en otro caso} \end{cases}.$$

Una v.a. de este tipo diremos que sigue una distribución geométrica de parámetro p y lo denotaremos por $Ge(p)$ especificando que los valores admisibles de X son $D_X=\{1,2,\ldots\}$.

Propiedad de la carencia de memoria

Sea X una v.a. discreta con valores admisibles $D_X=\{1,2,\ldots\}$. X sigue una ley $Ge(p)$ si y sólo si $P(X\geq k+j/X>j)=P(X\geq k)$ para todo $k,j=1,2,3\ldots$ y $P(X=1)=p$

La variable geométrica que cuenta el número de fracasos

Supongamos que sólo estamos interesados en el número de fracasos antes de obtener el primer éxito. Sea Y= número de fracasos antes del primer éxito, entonces Y toma valores en $\{0,1,2,\ldots\}$ y su función de probabilidad es:

$$P_Y(y)=\begin{cases} q^k p & \text{si } k=0,1,2,\ldots \\ 0 & \text{en otro caso} \end{cases}.$$

Notemos que $Y=X-1$

Resumen v.a con distribución geométrica $Ge(p)$

X = número de intentos para conseguir el primer éxito:				
Valores admisibles.	$P_X(x) = P(X = x) =$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = \{1, 2, \ldots\}$	$\begin{cases} q^{x-1}p & \text{si } x = 1, 2, \ldots \\ 0 & \text{en otro caso} \end{cases}$	$\begin{cases} 0 & \text{si } x < 1 \\ 1 - q^k & \text{si } \begin{cases} k \leq x < k+1 \\ \text{para } k = 1, 2, \ldots \end{cases} \end{cases}$	$\frac{1}{p}$	$\frac{q}{p^2}$

Y = número de fracasos para conseguir el primer éxito.				
Valores admisibles.	$P_Y(y) = P(Y = y) =$	$F_Y(y) = P(Y \leq y) =$	$E(X)$	$Var(X)$
$D_X = \{0, 1, \ldots\}$	$\begin{cases} q^k p & \text{si } k = 0, 1, 2, \ldots \\ 0 & \text{en otro caso} \end{cases}$	$\begin{cases} 0 & \text{si } y < 0 \\ 1 - q^{k+1} & \text{si } \begin{cases} k \leq y < k+1 \\ \text{para } k = 0, 1, 2, \ldots \end{cases} \end{cases}.$	$\frac{q}{p}$	$\frac{q}{p^2}$

4.1.4. Binomial negativa

Bajo las mismas condiciones que en el caso anterior repetimos el experimento hasta obtener el r-ésimo éxito. Sea X la v.a que cuenta el número de repeticiones del experimento hasta el r-ésimo éxito. Entonces

$$P_X(x) = P(X = x) = \begin{cases} \binom{x-1}{r-1} (q)^{x-r} p^r & \text{si } x = r, r+1, \ldots \\ 0 & \text{en otro caso} \end{cases}.$$

Una v.a. con este tipo de distribución recibe el nombre de binomial negativa y la denotaremos por $BN(p, r)$. Notemos que $BN(p, 1) = Ge(p)$.

Resumen v.a con distribución Binomial Negativa $BN(r, p)$

Valores admisibles.	$P_X(x) = P(X = x) =$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = \{r, 1, \ldots\}$	$\left(\begin{array}{c} x-1 \\ r-1 \end{array} \right) q^{x-r} p^r$ si $x = r, r+1, \ldots$	Hay que sumarla.	$\frac{r}{p}$	$\frac{rq}{p^2}$

4.1.5. Poisson

Formalmente diremos que una v.a. discreta X con $X(\Omega) = I\!N$ tiene distribución de Poisson con parámetro $\lambda > 0$, y lo denotaremos por $Po(\lambda)$ si su función de probabilidad es:

$$P_X(x) = P(X = x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{si } x = 0, 1, \ldots \\ 0 & \text{en otro caso} \end{cases}.$$

Efectivamente la función anterior es una función de probabilidad (ejercicio[1])

La distribución Poisson como "límite" de una binomial.

La distribución Poisson aparece en el conteo de determinados eventos que se producen en un intervalo de tiempo o en el espacio.

Supongamos que nuestra variable de interés es X= número de eventos en el intervalo de tiempo $(0, t]$ (por ejemplo el número de llamadas a una centralita telefónica) y que sabemos que se cumplen las siguientes condiciones:

a) El número promedio de eventos en el intervalo $(0, t]$ es $\lambda > 0$

b) Es posible dividir el intervalo de tiempo en un gran número de subintervalos (denotemos por n al número de intervalos) de forma que:

   1) La probabilidad de que se produzcan dos o más eventos en un subintervalo es despreciable.
   2) La ocurrencia de eventos en un intervalo es independiente de los demás.
   3) La probabilidad de que un evento ocurra en un subintervalo es $p = \lambda/n$

Bajo estas condiciones podemos considerar que el número de eventos en el intervalo $(0, t]$ será el número de "éxitos" en n repeticiones independientes de un proceso Bernoulli de parámetro p

Entonces si $n \to \infty$ y pn se mantiene igual a λ resulta que la función de probabilidad de X se puede poner como[2]:

$$f_X(k) = \lim_{n\to\infty} \binom{n}{k} p^k q^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

Propiedad Si tenemos un experimento *tipo* Poisson con λ igual al promedio de eventos en una unidad de tiempo (u.t.) entonces si t es una cantidad de tiempo en u.t., la v.a. X_t=numero de eventos en el intervalo $(0, t]$ es una $Po(\lambda \cdot t)$.

[1] Recordemos que el desarrollo de Taylor de la exponencial es $e^\lambda = \sum_{x=0}^{+\infty} \frac{\lambda^x}{x!}$.

[2] La demostración de esta convergencia la podéis encontrar, por ejemplo, en *Probability and Random Processes for Electrical Engineering. Alberto-Leon Garcia 2 Ed. Addison Wesley 1994*, páginas 106 a 109

Resumen v.a con distribución Poisson $Po(\lambda)$

Valores admisibles.	$P_X(x) = P(X = x) =$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = \{0, 1, \ldots\}$	$\begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{si } x = 0, 1, \ldots \\ 0 & \text{en otro caso} \end{cases}$	Tabulada.	λ	λ

Aproximación de la distribución binomial por la Poisson:

Bajo el punto de vista anterior y si p es pequeño y n suficientemente grande (existen distintos criterios por ejemplo $n > 20$ ó 30 y $p \leq 0{,}1$) podemos aproximar una $B(n, p)$ por una $Po(np)$

4.1.6. Distribución hipergeométrica:

Es la que modeliza el número de bolas blancas extraídas de una urna sin reposición. Sea una urna que contiene N bolas de las que N_1 son blancas y las restantes N_2 no. Obviamente $N = N_1 + N_2$. Extraemos n bolas de la urna sin reemplazarlas. Sea X la v.a. que cuenta el número de bolas blancas extraídas. Entonces

$$P_X(x) = \begin{cases} \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} & \text{si } \max\{0, n - N_2\} \leq x \leq \min\{n, N_1\} \text{ con } x \in \mathbb{N} \\ 0 & \text{en otro caso} \end{cases}$$

Una v.a. hipergeométrica con los parámetros[3] anteriores la denotaremos por $H(N_1, N_2, n)$.

Resumen v.a con distribución hipergeométrica $H(N_1, N_2, n)$.

Valores admisibles.	$P_X(x) = P(X = x) =$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = \{x \in \mathbb{N} \mid \max\{0, n - N_2\} \leq x$ $x \leq \min\{n, N_1\}\}$	$\begin{cases} \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}} & \text{si } x \in D_X \\ 0 & \text{en otro caso} \end{cases}$	No tiene expresión.	$\frac{nN_1}{N}$	$n\frac{N_1}{N}\left(1 - \frac{N_1}{N}\right)\frac{N-n}{N-1}$

[3] En ocasiones se parametriza una v.a. hipergeométrica mediante N =número total de bolas, n=número de extracciones y p = probabilidad de una bola blanca. Así podríamos $H(N, n, p)$ donde $p = \frac{N_1}{N}$, $N = N_1 + N_2$.

4.2. Algunas variables aleatorias continuas

Al igual que en el caso discretos veremos distintos tipos de v.a. continuas que son utilizadas de forma muy frecuente.

4.2.1. Distribución uniforme en el intervalo (a,b):

Una v.a. continua X diremos que tiene una distribución uniforme sobre el intervalo real (a, b) ,$(a < b)$, si su función de densidad es

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b \\ 0 & \text{en cualquier otro caso} \end{cases}$$

(como ejercicio comprobar que el área comprendida entre f_X y la horizontal vale 1.)

Entonces su función de distribución es

$$F_X(x) = \begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a < x < b \\ 1 & \text{si } b \leq x \end{cases}$$

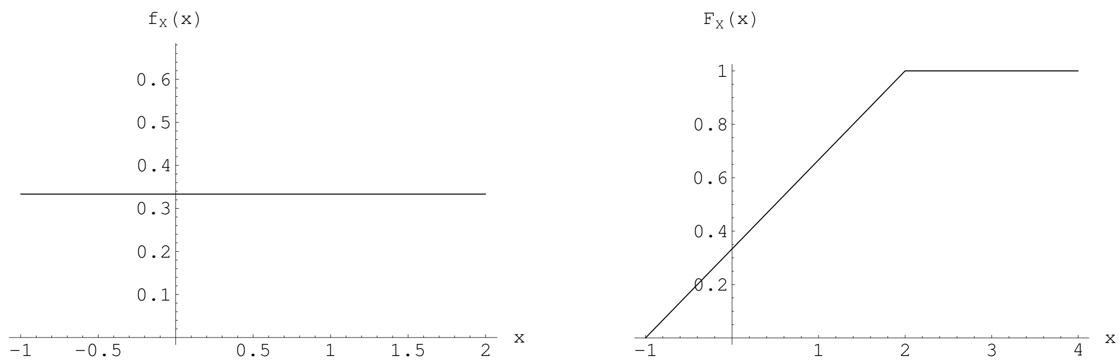
Efectivamente:

- Si $x \leq a$ entonces $F_X(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} 0dt = 1 \mid_{-\infty}^{x} = 1 - 1 = 0$
- Si $a < x < b$ entonces $F_X(x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{a} 0dt + \int_{-\infty}^{x} \frac{1}{b-a}dt = 1 \mid_{-\infty}^{x} + \frac{t}{b-a} \mid_{a}^{x} = (1-1) + \frac{x}{b-a} - \frac{t}{b-a} = \frac{x-a}{b-a}$
- Por último si $x \geq b$ entonces $F_X(x) \int_{-\infty}^{x} f(t)dt = 1$ (ejercicio).

Si X es una v.a. uniforme en el intervalo (a, b) escribiremos $X \equiv U(a, b)$.

Esperanza y varianza para $X \equiv U(a, b)$

$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx = \int_{-\infty}^{+\infty} x\frac{1}{b-a}dx = \frac{x^2}{2(b-a)} \mid_a^b = \frac{b+a}{2}$
$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)dx = \int_{-\infty}^{+\infty} x^2\frac{1}{b-a}dx = \frac{x^3}{3(b-a)} \mid_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{b^2+ab+a^2}{3}$
$Var(X) = E(X^2) - (E(X))^2 = \frac{b^2+ab+a^2}{3} - (\frac{b+a}{2})^2 = \frac{(b-a)^2}{12}$



a) b)

Figura 4.1: Gráficas de la función de densidad (a) y de la función de distribución (b) de una v.a. $U(-1, 2)$.

Cambio lineal v.a. uniforme

Si X sigue una distribución $U(a, b)$ entonces $Z = \frac{x-a}{b-a}$ sigue una distribución $U(0, 1)$.

En general si d y e son dos constantes reales $T = d \cdot X + e$ sigue una ley $U(d \cdot a + e, d \cdot b + e)$ si $d > 0$, cuando d sea negativo T sigue una ley $U(d \cdot b + e, d \cdot a + e)$. Las demostración se dejan como ejercicios.

Resumen v.a con distribución uniforme, $U(a, b)$

Valores admisibles.	$f_X(x)$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = (a, b)$	$\begin{cases} \frac{1}{b-a} & \text{si } a < x < b \\ 0 & \text{en cualquier otro caso} \end{cases}$	$\begin{cases} 0 & \text{si } x \leq a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq \\ 1 & \text{si } b \leq x \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

4.2.2. Distribución exponencial (exponencial negativa):

Supongamos que tenemos un proceso Poisson con parámetro λ en una unidad de tiempo.

Sea t una cantidad de tiempo en u.t. entonces $N_t =$ número de eventos en el intervalo de tiempo $(0, t]$ es una $Po(\lambda \cdot t)$. Consideremos la v.a. T =tiempo transcurrido entre dos eventos Poisson consecutivos.

Sea $t > 0$, entonces

$$P(T > t) = P(\text{Cero eventos en el intervalo}(0,t]) = P(N_t = 0) = \frac{(\lambda t)^0}{0!}e^{-\lambda t} = e^{-\lambda t}.$$

Tomando complementarios, la función de distribución de T será

$$F_T(t) = P(T \leq t) = \begin{cases} 0 & \text{si } t \leq 0 \\ 1 - P(T > t) = 1 - e^{-\lambda t} & \text{si } t > 0 \end{cases}$$

Entonces

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{si } t > 0 \\ 0 & \text{si } t \leq 0 \end{cases}$$

La exponencial se denota por $Exp(\lambda)$

Propiedad de la falta de memoria

Sea X una v.a. $Exp(\lambda)$ entonces

$$P(X > s + t / X > s) = P(X > t) \text{ para todo } s, t \in \mathbb{R}$$

Toda v.a. absolutamente continua, que tome valores positivos y que verifique la propiedad de la falta de memoria es una v.a. exponencial.

Resumen v.a con distribución exponencial, $Exp(\lambda)$

Sea $X \equiv Exp(\lambda)$.

Valores admisibles.	$f_X(x)$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = (0, +\infty)$	$\begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$	$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{si } x \leq 0 \\ 1 - e^{-\lambda x} & \text{si } x > 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

4.2.3. Distribución normal o Gaussiana

Diremos que una v.a. X sigue una ley normal de parámetros μ y σ^2 y lo denotaremos por $N(\mu, \sigma^2)$ si tiene por función de densidad

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \text{ para todo } x \in \mathbb{R}$$

La gráfica de esta función es la conocida campana de Gauss.

La v.a. normal con $\mu = 0$ y $\sigma = 1$ recibe el nombre de normal estándar y se suele denotar por la letra Z.

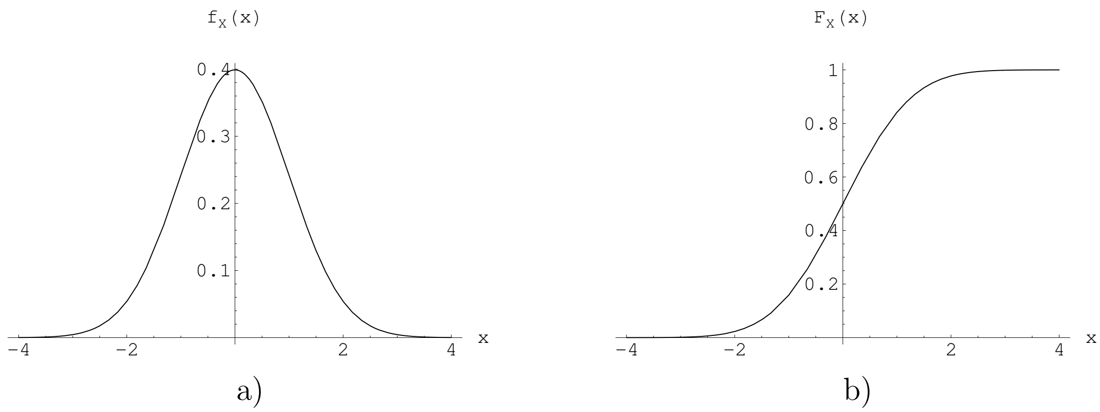


Figura 4.2: Gráficas de la función de densidad (a) y de la función de distribución (b) de una v.a. $N(0,1)$.

Su función de distribución es, como sabemos :

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

Que no tiene ninguna expresión algebraica "decente". Es por esta razón, y por comodidad, que esta función está tabulada.

Cuando una variable tiene distribución normal con parámetros μ, σ la denotamos por $X \equiv N(\mu, \sigma^2)$

Resumen v.a con distribución normal, $N(\mu, \sigma^2)$

Valores admisibles.	$f_X(x)$	$F_X(x) = P(X \leq X) =$	$E(X)$	$Var(X)$
$D_X = \mathbb{R}$	$= \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ para todo $x \in \mathbb{R}$	Tabulada la $N(0,1)$	μ	σ^2

Propiedades de la distribución normal.

La función de densidad de la distribución normal tiene las siguientes propiedades:

a) f es continua

b) $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$ (propiedad de todas las densidades).

c) $f(\mu + x) = f(\mu - x)$ y $F(x + \mu) = 1 - F(\mu - x)$ para todo $x \in \mathcal{R}$

d) $\lim\limits_{x \to +\infty} f(x) = \lim\limits_{x \to -\infty} f(x) = 0$ es decir tiene asíntota horizontal a derecha e izquierda.

e) f es estrictamente creciente si $x < \mu$ y decreciente si $x > \mu$.

f) Alcanza el máximo en $x = \mu$ y en este punto vale $f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$

g) Tiene dos puntos de inflexión en $x = \mu + \sigma$ y en $x = \mu - \sigma$.

Transformaciones lineales de variables aleatorias normales

Proposición 69 *Sea $X \equiv N(\mu, \sigma^2)$ entonces la variable $Y = aX + b$ con $a \neq 0, b \in \mathcal{R}$ tiene distribución $N(a\mu + b, a^2\sigma^2)$*

En particular si $X \equiv N(\mu, \sigma^2)$, tomando $a = \frac{1}{\sigma}$ y $b = \frac{-\mu}{\sigma}$ la v.a. $Z = \frac{X-\mu}{\sigma}$ se distribuye $N(0, 1)$.

Esta propiedad es muy importante, ya que utilizándola sólo necesitaremos tabular la $N(0, 1)$. A la función de distribución de una $Z \equiv N(0, 1)$ la llamaremos F_Z y a una normal $N(0, 1)$ se le denomina normal estándar. Por lo tanto si $F_X(x) = F_Z(\frac{x-\mu}{\sigma})$.

La propiedad siguiente se desprende de las propiedades generales de una normal y nos será muy útil en los cálculos de probabilidades de una normal.

Propiedad Si $Z \equiv N((0, 1)$ entonces $F_Z(x) = 1 - F_Z(-x)$.

Ejemplo 70 *Sea $Z \equiv N(0, 1)$ Calcular :*

a) Dado $\delta > 0$, $P(-\delta \leq Z \leq \delta) = F_Z(\delta) - F_Z(-\delta) = F_Z(3) - (1 - F_Z(\delta)) = 2F_Z(\delta) - 1$

b) $P(-4 \leq Z \leq 4) = F_Z(4) - F_Z(-4) = 2F_Z(4) - 1$

c) $P(-2 \leq Z \leq 2) = F_Z(2) - F_Z(-2) = 2F_Z(2) - 1$

d) $P(Z \leq -2) = F_Z(-2) = 1 - F_Z(2)$

e) $P(Z \leq 2) = F_Z(2)$

f) $P(Z \geq 2) = 1 - P(Z < 2) = 1 - F_Z(2)$

g) $P(Z > 2) = 1 - P(Z \leq 2) = 1 - F_Z(2)$

h) $P(Z = 2) = 0$

i) $P(Z \geq -2) = 1 - P(Z < -2) = 1 - F_Z(-2) = 1 - (1 - F_Z(2)) = F_Z(2)$.

Resumiendo podemos utilizar las siguientes propiedades, $X \equiv N(\mu, \sigma)$

- Z es su variable tipificada, es decir, $Z = \frac{X-\mu}{\sigma} \equiv N(0,1)$ entonces:

$$P(X \leq x) = P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) = F_Z(\frac{x-\mu}{\sigma})$$

- Cuando tengamos un intervalo

$$P(a < X < b) = P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}) =$$

$$= P(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}) = F_Z(\frac{b-\mu}{\sigma}) - F_Z(\frac{a-\mu}{\sigma})$$

- Si $\delta > 0$ $P(\mu - \delta \leq X \leq \mu + \delta) = 2F_Z(\frac{\delta}{\sigma}) - 1$

Ejemplo 71 *Sea X una normal com media 2 y varianza 4, entonces*

a) $P(1 < X < 2) = P(\frac{1-2}{2} < \frac{X-2}{2} < \frac{2-2}{2}) = P(\frac{-1}{2} < Z < 0) = F_Z(0) - F_Z(-0{,}5) = \frac{1}{2} - 1 + F_Z(0{,}5)$.

b) $P(X > 3) = P(\frac{X-2}{2} > \frac{3-2}{2}) = P(Z > 0{,}5) = 1 - F_Z(0{,}5)$.

Capítulo 5

Variables aleatorias vectoriales

En este tema estudiaremos brevemente el comportamiento aleatorio conjunto de varias variables. También veremos uno de los teoremas más utilizados en estadística el Teorema del Límite central.

5.1. Variables aleatorias bidimensionales

Vamos a exponer las nociones básicas de variables aleatorias bidimensionales.

Definición 72 *Dadas X, Y dos variables aleatorias, continuas o discretas, llamaremos función de distribución o de probabilidad acumulada conjunta a:*

$$F_{XY}(x, y) = P(X \leq x, Y \leq y).$$

La funciones de distribución de las variables X e Y, F_X y F_Y reciben el nombre de distribuciones marginales. Esta dewfinición se generaliza de forma natural a n variables.

Independencia

Definición 73 *Sean X, Y dos variables aleatorias con función de distribución conjunta F_{XY}. Diremos que X, Y son independientes cuando*

$$F_{XY}(x, y) = F_X(x) F_Y(y).$$

Esta propiedad se generaliza a n variables. Es decir n variables aleatorias son independientes si su función de distribución conjunta es el producto de las distribuciones de cada una de las variables.

5.1.1. Variables aleatorias conjuntamente discretas

Definición 74 *Sean X, Y dos v.a. discretas. Llamaremos función de probabilidad conjunta de X , Y a*

$$P_{X,Y}(x,y) = P(X = x, Y = y)$$

donde $P(X = x, Y = y) = P(\{X = x\} \cap \{Y = y\})$.

Llamaremos función de probabilidad marginal de X y de Y a $P_X(x)$ y a $P_Y(y)$ respectivamente.

Proposición 75 *Sean X, Y dos v.a. discretas con función de probabilidad conjunta P_{XY}, entonces:*

a) $0 \leq P_{X,Y}(x,y) \leq 1$.

b) $\sum_x \sum_y P_{XY}(x,y) = 1$

c) $P_X(x) = \sum_y P_{XY}(x,y)$, $P_Y(y) = \sum_x P_{XY}(x,y)$.

Proposición 76 *Sean X, Y dos v.a. discretas con función de probabilidad conjunta P_{XY} y función de distribución F_{XY}. Entonces*

$$F_{XY}(x,y) = \sum_{x_i \leq x} \sum_{y_i \leq y} P(X = x_i, Y = y_i).$$

Proposición 77 *Dadas dos v.a. X , Y discretas con función de probabilidad conjunta P_{XY}. Se tiene que X e Y son independientes si y sólo si $P_{XY}(x,y) = P_X(x)P_Y(y)$ para todo x, y. Esta propiedad se generaliza a n variables.*

Ejemplo 78 *Supongamos que X e Y son las v.a. que nos dan el valor del lanzamiento (de forma independiente) de dos dados bien balanceados. Es fácil deducir que*

$$P_{XY}(i,j) = \begin{cases} \frac{1}{36} & si\ i,j = 1,2,3,4,5,6 \\ 0 & en\ el\ resto\ de\ casos \end{cases}$$

Normalmente se disponen estos resultados en forma de tabla:

		X						
		1	2	3	4	5	6	P_Y
Y	1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
	P_X	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

En esta tabla se pueden comprobar fácilmente las propiedades anteriores. Además debido a la "independencia" en el lanzamiento de los dados se cumple que $P_{XY}(x,y) = P_X(x)P_Y(y)$ *las variables son independientes y podemos "recuperar la función de probabilidad conjunta conociendo las marginales".*

Pero supongamos que los dados están bien balanceados pero que no son "independientes" como por ejemplo en el siguiente caso:

		X						
		1	2	3	4	5	6	P_Y
Y	1	$\frac{2}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{6}$
	2	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{6}$
	3	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{6}$
	4	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{6}$
	5	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{1}{42}$	$\frac{1}{6}$
	6	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{1}{42}$	$\frac{2}{42}$	$\frac{1}{6}$
	P_X	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Aquí $P_{XY}(x,y) \neq P_X(x)P_Y(y)$. *En este caso sabiendo las marginales no podemos saber la distribución conjunta, las variables no son independientes.*

Ejemplo 79 *Sea* $N =$ *número de bytes de un mensaje, supongamos que* N *sigue una distribución* $Ge(1-p)$. N *toma valores en* $\mathbb{N}$. *y* $P(N=n) = (1-p)p^n$. *Supongamos que los mensajes se distribuyen en paquetes con un máximo de* M *bytes cada uno. Sea* $C =$*número de paquetes completos que ocupa un mensaje y sea* $R = N - CM$ *el número de bytes "sobrantes".*

a) *Vamos a determinar la distribución conjunta de* C *y* R *es decir del vector aleatorio bidimensional* (C,R).

Si el mensaje tiene N bytes entonces sea $N = MC + R$ la división entera de N por M, sabemos que $0 \leq R \leq M-1$ es decir $R \in \{0, 1, 2, .., M-1\}$ y por lo tanto si $c \in I\!N$ y $r \in \{0, 1, 2, .., M-1\}$ entonces:

$$P(C = c, R = r) = P(N = cM + r) = (1-p)p^{cM+r}.$$

Luego $P_{CR}(c, r) = (1-p)p^{cM+r}$ con $c \in N$ y $0 \leq r \leq M-1$ siendo cero en el resto de casos.

b) *Calculemos las funciones de probabilidad marginales.*

$$P_C(c) = P(C = c) = \sum_{r=0}^{M-1} P_{CR}(c, r) = \sum_{r=0}^{M-1} (1-p)p^{cM+r} = (1-p)p^{cM} \sum_{r=0}^{M-1} p^r =$$

$$(1-p)p^{cM} \frac{1 - p^{M-1}p}{1-p} = (1-p^M)(p^M)^c c = 0, 1, 2, \ldots$$

Luego C es una $Ge(1-p^M)$.

De forma similar $P_R(r) = \frac{1-p}{1-p^M} p^r$ para $r = 0, 1, \ldots, M-1$

c) *¿Son las variables C y R independientes? Lo serán si el producto de las marginales es igual a la conjunta? Esto sucederá si el cociente C es independiente del resto R como sucesos. Efectivamente:*

$P_C(c)P_R(r) = (1-p^M)(p^M)^c \frac{1-p}{1-p^M} p^r = (1-p)p^{cM+r} = P_{CR}(c, r)$ y por lo tanto C y R son independientes.

5.1.2. Variables aleatorias bidimensionales continuas

Definición 80 *Una función $f : I\!R^2 \to R$ es una función de densidad bidimensional si es no negativa, es integrable en R^2 y $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$*

Definición 81 *Dadas dos v.a. continuas X e Y diremos que son conjuntamente continuas si existe una función de densidad $f_{XY} : I\!R^2 \to I\!R$ tal que*

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v) dv du$$

Definición 82 *Dadas dos v.a. X e Y conjuntamente continuas las funciones de densidad y de distribución de X e Y reciben el nombre de densidades y distribuciones marginales.*

Proposición 83 *Se verifican las siguientes propiedades:*

a) $P((X,Y) \in A) = \int\int_A f_{XY}(x,y)dxdy$

b) [1] $f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$

c) $f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dy$; $f_Y(y) = \int_{-\infty}^{+\infty} f_{XY}(x,y)dx$.

Ejemplo 84 *Distribución uniforme en un recinto. Sea R un "recinto" adecuado del plano. Sean (X,Y) las coordenadas resultantes de elegir al azar un punto de ese recinto y sea A= área del recinto R. Entonces diremos que (X,Y) sigue una ley uniforme en ese recinto si su función de densidad es*

$$f_{XY}(x,y) = \begin{cases} \frac{1}{A} & \text{si } (x,y) \in R \\ 0 & \text{en cualquier otro caso} \end{cases}$$

En particular si $R = (a,b)^2$ entonces (X,Y) siguen una ley uniforme en rectángulo $(a,b)^2$ si su densidad es

$$f_{XY}(x,y) = \begin{cases} \frac{1}{(b-a)^2} & \text{si } a < x, y < b \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Es fácil comprobar que las marginales de esta última distribución conjunta son uniformes en (a,b).

Ejercicio; Construir la densidad de (X,Y) uniforme sobre los recintos $R = (a,b) \times (c,d)$ y $R = \{(x,y)|x^2+y^2 \leq r^2\}$ con $r > 0$

Ejemplo 85 *Consideremos una diana cuadrada $(0,1) \times (0,1)$. Sean (X,Y) las coordenadas de un impacto "al azar en esta diana".*

Al ser al azar la probabilidad de un suceso será igual al cociente

$$\frac{\text{área total favorable}}{\text{área total posible}}$$

entonces si $0 < x < 1$ y $0 < y < 1$

[1]Esta propiedad se verifica salvo en un conjunto de "medida nula", es decir de probabilidad 0. Al igual que en el caso discreto no siempre es posible obtener la distribución o la densidad conjunta desde las marginales.

$P(X \leq x, Y \leq y) = xy$

Pero si $0 < x < 1$ *y* $1 \leq y$ *entonces*

$P(X \leq x, Y \leq y) = P(X \leq x, Y \leq 1) = P(X \leq x) = x$

Análogamente si $1 \leq x$ $0 < y < 1$ *tenemos que*

$P(X \leq x, Y \leq y) = y$

En resumen completando los casos que faltan resulta que la función de distribución conjunta es:

$$F_{XY}(x,y) = \begin{cases} 0 & \text{si } x < 0 \text{ ó } y < 0 \\ xy & \text{si } 0 \leq x \leq 1 \text{ y } 0 \leq y \leq 1 \\ x & \text{si } 0 \leq x \leq 1 \text{ y } y > 1 \\ y & \text{si } 0 \leq y \leq 1 \text{ y } x > 1 \\ 1 & \text{si } x > 1 \text{ y } y > 1 \end{cases}$$

Entonces la densidad [2] *de* F_{XY} *es*

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) = \begin{cases} 1 & \text{si } 0 \leq x \leq 1 \text{ y } 0 \leq y \leq 1 \\ 0 & \text{en otro caso} \end{cases}$$

Ahora es fácil comprobar que

$$F_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\,du\,dv$$

También es fácil ver que en este caso el producto de las densidades (distribuciones) marginales es igual a la densidad (distribución) conjuntas.

Ejemplo 86 *Encontrar el valor de c para que*

$$f_{XY}(x,y) = \begin{cases} ce^{-x}e^{-y} & \text{si } 0 \leq y \leq x \\ 0 & \text{en otro caso} \end{cases}$$

Sea función de densidad.

Al igual que en el caso unidimensional es suficiente con que la función sea integrable, no negativa e integre 1 en $\mathbb{R}^2$*. Las dos primeras condiciones son evidentes (si* $c > 0$*), veamos la última:*

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dx\,dy = 1 \text{ y como } \int_{0}^{+\infty} \int_{0}^{x} e^{-x}e^{-y}\,dy\,dx = \frac{1}{2}$$

entonces $c = 2$*. Luego* $f_{XY}(x,y) = \begin{cases} 2e^{-x}e^{-y} & \text{si } 0 \leq y \leq x \\ 0 & \text{en otro caso} \end{cases}$

Las densidades marginales serán:

[2] Notemos que $\frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$ no existe en los bordes del cuadrado que forman un conjunto de medida nula.

$$f_X(x) = \int_0^{+\infty} f_{XY}(x,y)dy = \int_0^{x} 2e^{-x}e^{-y}dy = 2e^{-x}(1-e^{-x})$$

si $x \geq 0$ y cero en cualquier otro caso.

De forma análoga:$f_Y(y) = 2e^{-2y}$ si $0 \leq y$, siendo cero en cualquier otro caso.

b) Calculemos ahora $P(X+Y \leq 1)$. Sea A el conjunto de puntos del plano tales que $X+Y \leq 1$ entonces:
$P(X+Y \leq 1) = \int\int_A f_{XY}(u,v)dudv =$
$= \int_{y=0}^{\frac{1}{2}} \int_{x=y}^{1-y} 2e^{-x}e^{-y}dxdy = 1 - 2e^{-1}$

5.1.3. Distribuciones condicionales. (OPCIONAL)

Sea (X,Y) una v.a. bidimensional discreta. Estudiemos la v.a. $X/Y = y$.

Definición 87 *Sea (X,Y) una v.a. bidimensional discreta con función de probabilidad conjunta $P_{XY}(x,y)$. Sea y tal que $P(Y=y) > 0$. Definimos la función de probabilidad de X condicionada a que $Y = y$ como*

$$P_X(x|y) = P(X = x/Y = y)$$

para todo $x \in I\!R$.

$X/Y = y$ resulta ser una v.a. discreta que tiene por función de probabilidad $P_X(x|y)$

Análogamente dado x tal que $P(X=x) > 0$ definimos $Y/X = x$ que tiene por función de probabilidad $P_Y(y|x) = P(Y = y/X = x)$ para todo $y \in I\!R$

Proposición 88 *Sean X e Y dos v.a. discretas. Enunciamos las siguientes propiedades para $X/Y = y$, se enuncian de forma similar para $Y/X = x$*

a) $\sum_x P_X(x|y) = 1$. Efectivamente

$$\sum_x P_X(x|y) = \sum_x P(X = x/Y = y) =$$

$$= \sum_x \frac{P(X=x, Y=y)}{P(Y=y)} = \sum_x \frac{P_{XY}(x,y)}{P_Y(y)} =$$

$$= \frac{1}{P_Y(y)} \sum_x P_{XY}(x,y) = \frac{P_Y(y)}{P_Y(y)} = 1$$

b) $P_X(x|y) = \frac{P_{XY}(x,y)}{P_Y(y)} = \frac{P_{XY}(x,y)}{\sum_x P_{XY}(x,y)}$

Luego es suficiente conocer la distribución conjunta para calcular la condicional.

c) $F_X(x_0|y) = P(X \leq x_0 / Y = y) = \sum_{x \leq x_0} P_X(x|y)$

d) $P_{XY}(x,y) = P_Y(y) P_X(x|y)$. *Luego podemos calcular la función de probabilidad conjunta a partir de la marginal y la condicional.*

e) *X es independiente de Y sii $P_X(x|y) = P_X(x)$ para todo x, y.*

Ejemplo 89 *Un técnico tiene que revisar un sistema de dispositivos. No los revisa todos cada vez sino que escoge el número a revisar de forma equiprobable entre 5 y 8. Sea $X=$ número de dispositivos que revisa el técnico. Sabiendo que $X = x$ dispositivos el tiempo en segundos que el técnico tarda en revisarlos es una v.a. Y (discreta) con función de probabilidad*

$P_Y(y|x) = \frac{10-x}{x}\left(\frac{x}{10}\right)^y$ *para $y = 1, 2, 3\ldots$ y cero en el resto de casos.*

Entonces

$P_{XY}(x,y) = P_X(x) P_Y(y|x) = \frac{1}{4}\frac{10-x}{x}\left(\frac{x}{10}\right)^y$

para $y = 1, 2, 3, \ldots; x = 5, 6, 7, 8$

Calculemos la marginal de Y desde la conjunta

$$P_Y(y) = \sum_{x=5}^{8} P_{XY}(x,y) = \sum_{x=5}^{8} \frac{1}{4}\frac{10-x}{x}\left(\frac{x}{10}\right)^y = \frac{1}{4 \cdot 10^y} \sum_{x=5}^{8} (10-x) x^{y-1}$$

También podemos calcular la marginal de X:

$$P_X(x) = \sum_{1 \leq y} f_{XY}(x,y) = \sum_{1 \leq y} \frac{1}{4}\frac{10-x}{x}\left(\frac{x}{10}\right)^y =$$

$$\frac{10-x}{4x} \sum_{1 \leq y} \left(\frac{x}{10}\right)^y = \frac{10-x}{4x}\left(\frac{1}{1-\frac{x}{10}} - 1\right) =$$

$$= \frac{10-x}{4x}\frac{10-(10-x)}{10-x} = \frac{1}{4}$$

Variables condicionadas a un suceso

Tambien podemos condicionar una variable a un suceso.

Definición 90 ***Definición*** *Sea (X,Y) una v.a. bidimensional discreta y sea B un conjunto del plano tal que $P((X,Y)\in B)>0$. Definimos la función de probabilidad de X condicionada al suceso $\{(x,y)\in B\}$ como*
$P_X(x|B)=P(X=x/(X,Y)\in B)$

Ejemplo 91 *Consideremos un canal de comunicación sea (X,Y) una v.a. donde X=símbolo emitido, Y=símbolo recibido. La función de probabilidad conjunta P_{XY} viene determinada por la siguiente tabla:*

		Y			
		1	2	3	P_X
X	1	0.1	0.2	0.1	0.4
	2	0.2	0.1	0	0.3
	3	0.1	0.2	0	0.3
	P_Y	0.4	0.5	0.1	1

Consideremos el suceso "error"
$E=\{(1,2),(1,3),(2,1),(2,3),(3,1),(3,2)\}$.
Que también se puede expresar como $X\neq Y$. Entonces $P((X,Y)\in E)=1-P((X,Y)\notin E)=1-0{,}1-0{,}1-0=0{,}8$

Queremos calcular la función de probabilidad de Y condicionado a error.
$P_Y(y|X\neq Y)=P(Y=y/(X,Y)\in E)=$
$\frac{P(Y=y\cap\{(X,Y)\in E\}}{P((X,Y)\in E)}=\frac{P((X,y)\in\{(x,y)|x\neq y\})}{0{,}8}$
Entonces
$P_Y(1|X\neq Y)=\frac{P((X,1)\in\{(x,1)|x\neq 1\})}{0{,}8}=$
$=\frac{P((X,1)\in\{(2,1),(3,1)\})}{0{,}8}=\frac{3}{8}$
$P_Y(2|X\neq Y)=\frac{P((X,2)\in\{(x,2)|x\neq 2\})}{0{,}8}=$
$=\frac{P((X,2)\in\{(1,2),(3,2)\})}{0{,}8}=\frac{4}{8}$
$P_Y(3|X\neq Y)=\frac{P((X,3)\in\{(x,3)|x\neq 3\})}{0{,}8}=$
$=\frac{P((X,3)\in\{(1,3),(2,3)\})}{0{,}8}=\frac{1}{8}$

Nota: En este caso las funciones de distribución condicionadas se pueden calcular así

$$F_X(x_0|(x,y)\in B)=P(X\leq x_0/(x,y)\in B)=\sum_{x\leq x_0}P_X(x|(x,y)\in B)$$

Definición 92 *Sea (X, Y) una v.a. bidimensional conjuntamente continua y sea y un valor "posible" de Y, es decir, $f_Y(y) > 0$. Definimos la función de distribución de X condicionado a que $Y = y$, la variable $X/Y = y$,como*

$F_X(x|y) = \lim_{h\to 0} P(X \leq x/y \leq Y \leq y + h)$.

Llamaremos función de densidad condicionada de X a que $Y = y$ a la función de densidad de la variable $X/Y = y$ y la denotaremos por $f_X(x/y)$.

Justifiquemos la definición. En primer lugar es claro que no podemos proceder como en el caso discreto pues al ser la v.a. continua $P(Y = y) = 0$.

Por otra parte

$$P(X \leq x/y \leq Y \leq y + h) =$$

$$= \frac{P(\{X \leq x\} \cap \{y \leq Y \leq y + h\})}{P(y \leq Y \leq y + h)}$$

entonces $F_X(x|y) = \frac{\lim_{h\to 0} P(\{X\leq x\}\cap\{y\leq Y\leq y+h\})}{\lim_{h\to 0} P(y\leq Y\leq y+h)}$

Al ser las variables continuas $\lim\limits_{h\to 0} P(\{X \leq x\} \cap \{y \leq Y \leq y + h\}) = = P(X \leq x, Y = y) = 0$

$\lim\limits_{h\to 0} P(y \leq Y \leq y + h) = P(Y = y) = 0$

Luego el cociente está indeterminado. Pero este límite existe pues:

$F_X(x|y) = \lim_{h\to 0} \frac{P(\{X\leq x\}\cap\{y\leq Y\leq y+h\})}{P(y\leq Y\leq y+h)} =$

$= \lim_{h\to 0} \frac{F_{XY}(x,y+h)-F_{XY}(x,y)}{F_Y(y+h)-F_Y(y)} =$

$= \lim_{h\to 0} \frac{\frac{F_{XY}(x,y+h)-F_{XY}(x,y)}{h}}{\frac{F_Y(y+h)-F_Y(y)}{h}} = \frac{\frac{\partial F_{XY}(x,y)}{\partial y}}{f_Y(y)}$

Lo que demuestra la existencia de la distribución condicionada y la siguiente propiedad, que no servirá para el cálculo de las distribuciones condicionales de variables aleatorias conjuntamente continuas.

Proposición 93 *Con las notaciones y condiciones anteriores se tiene que:*

a) $F_X(x|y_0) = \left[\frac{\frac{\partial F_{XY}(x,y)}{\partial y}}{f_Y(y)}\right]_{y=y_0}$

Además $\frac{\partial}{\partial x} F_X(x|y) = f_X(x|y)$ y por lo tanto $F_X(x|y)$ es una auténtica función de distribución.

b) $f_X(x/y) = \frac{f_{XY}(x,y)}{f_y y}$.

Efectivamente:

$\frac{\partial}{\partial x} F_X(x|y) = \frac{\frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)}{f_Y(y)} = \frac{f(x,y)}{f_Y(y)} = f_X(x|y)$

Además como $\frac{f(x,y)}{f_Y(y)} > 0$ deducimos que $F_X(x|y)$ es creciente.

Por otra parte $F_X(+\infty|y) = \frac{\frac{\partial}{\partial y} F_{XY}(+\infty,y)}{f_Y(y)} = \frac{\frac{d}{dy} F_Y(y)}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$

De forma similar $F_X(-\infty|y) = 0$

Estos últimos resultados junto a que $F_X(x|y)$ es creciente en x implican que $0 \leq F_X(x|y) \leq 1$.

Por último al ser $F_X(x|y)$ derivable es continua.

Observaciones

a) Algunos resultados anteriores son similares a los de v.a. discretas cambiando las funciones de probabilidad por las densidades.

b) $f_{XY}(x,y) = f_Y(y) f_X(x|y)$ si $f_Y(y) > 0$

c) Si queremos calcular $F_X(x|(x,y) \in B)$ donde B es una región (adecuada) del plano y $P((x,y) \in B) > 0$ podemos definir

   $F_X(x|(x,y) \in B) = P(X \leq x/(x,y) \in B)$

   Una vez calculada la distribución podemos derivar para obtener la densidad.

d) Todo lo dicho (cambiando X por Y) es válido para $Y/X = x$, $F_Y(y|x)$ y $f_Y(y/x)$.

Proposición 94 *Sean X e Y dos v.a. conjuntamente continuas con función de densidad conjunta f_{XY} entonces X e Y son independientes si y sólo si $f_{XY}(x,y) = f_X(x) f_Y(y)$ para todo x, y.*

Ejemplo 95 *Sea X= demanda en tiempo de un servicio; $X > 0$. Sea Y=coste aleatorio del servicio; $10 \leq Y \leq 15$*

Supongamos que (X,Y) es una v.a. continua con densidad

$f_{XY}(x,y) = \begin{cases} \frac{1}{5} y e^{-xy} & \text{si } x > 0, 10 \leq y \leq 15 \\ 0 & \text{en otro caso} \end{cases}$

Queremos calcular la probabilidad de que la demanda supere 100 cuando el coste Y es 12. Calculemos en primer lugar la densidad condicionada:

$f_X(x|12) = \frac{f_{XY}(x,12)}{f_Y(12)}$ *Necesitamos la densidad de Y*

$f_Y(y) = \int_0^{+\infty} \frac{1}{5} y e^{-xy} dx =$

$= -\frac{1}{5} e^{-xy} \Big|_0^{+\infty} = \frac{1}{5}$ *si* $10 \leq y \leq 15$ *y cero en el resto de casos.*

Entonces $f_X(x|12) = \begin{cases} 12e^{-12x} & si\ x > 0 \\ 0 & en\ otro\ caso \end{cases}$

Por lo tanto la probabilidad pedida es

$P(X > 100/Y = 12) = 1 - F_X(100|12) = 1 - \int_0^{100} 12e^{-12x} dx = 1 - [(-e^{-12x})]_0^{100} = 1 + e^{-1200} - 1 = e^{-1200}$

Calculemos en general $F_{XY}(x, y)$

Si $x < 0$ *ó* $y < 10$ *es fácil ver que* $F_{XY}(x, y) = 0$.

Si $0 \leq x$ *y* $10 \leq y \leq 15$

$$F_{XY}(x, y) = \int_{10}^{y} \int_0^x \frac{1}{5} v e^{-uv} du dv =$$

$$= -\frac{1}{5x}\left(e^{-10x} - e^{-xy}\right) + \frac{1}{5}(y - 10)$$

Si $0 \leq x$; $y \geq 15$

$F_{XY}(x, y) = 1 - \frac{1}{5x}\left(e^{-10x} - e^{-15x}\right)$

En resumen:

$F_{XY}(x, y) =$

$$\begin{cases} 0 & si\ y < 10,\ ó\ x < 0 \\ -\frac{1}{5x}\left(e^{-10x} - e^{-xy}\right) + \frac{1}{5}(y-10) & si\ x < 0, 10 < y < 15 \\ 1 - \frac{1}{5x}\left(e^{-10x} - e^{-15x}\right) & si\ 0 \leq x, y \geq 15 \end{cases}$$

Ahora para comprobar si el resultado es correcto calculamos la parcial segunda respecto de x y de y

$f_{XY}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x, y) =$

$$= \begin{cases} 0 & si\ y < 10,\ ó\ (y > 15, x < 0) \\ \frac{1}{5} y e^{-xy} & si\ x < 0, 10 < y < 15 \\ 0 & si\ 0 \leq x, y \geq 15 \end{cases}$$

Ejemplo 96 *Consideremos la v.a.* (X, Y) *con función de densidad conjunta*

$f_{XY}(x, y) = \begin{cases} \frac{1}{24}(x + y) & si\ 0 < x < 4, 0 < y < 2 \\ 0 & en\ otro\ caso \end{cases}$

Si $0 < x < 4$ $f_X(x) =$

$= \int_0^2 \frac{1}{24}(x + y) dy = \left[\frac{1}{24}(xy + \frac{y^2}{2})\right]_{y=0}^{2} =$

$= \frac{x+1}{12}$

De forma análoga calculamos $f_Y(y)$. *En resumen tenemos que:*

$f_X(x) = \begin{cases} \frac{1}{12}(x + 1) & si\ 0 < x < 4 \\ 0 & en\ otro\ caso \end{cases}$

$f_Y(y) = \begin{cases} \frac{1}{6}(y+2) & si\ 0 < y < 2 \\ 0 & en\ otro\ caso \end{cases}$

Vamos a calcular $f_Y(y|3)$, $F_Y(y|3)$,*la probabilidad de que* Y *esté entre 1 y 2 cuando* $X = 3$, $F_{XY}(x, y)$ *y* $F_X(x|X > 2Y)$.

$f_Y(y|3) = \frac{f_{XY}(3,y)}{f_X(3)} = \frac{\frac{1}{24}(3+y)}{\frac{3+1}{12}} = \frac{3+y}{8}$

si $0 < y < 2$ *y vale cero en el resto de casos.*

$F_Y(y|3) = \int_0^y \frac{3+v}{8} dv = \left[\frac{1}{8}(3v + \frac{v^2}{2})\right]_0^y = \frac{1}{8}(3y + \frac{y^2}{2})$

si $0 < y < 2$ *valiendo* 0 *si* $y \leq 0$ *y* 1 *si* $y \geq 2$.

$P(1 \leq Y \leq 2/X = 3) = F_Y(2|3) - F_Y(1|3) = \frac{9}{16}$

Para calcular F_{XY} *basta integrar en los siguientes casos* f_{XY}, *obteniéndose que*

$$F_{XY}(x, y) = \begin{cases} 0 & si\ x \leq 0\ ó\ y \leq 0 \\ \frac{1}{48}xy(x+y) & si\ {\scriptstyle 0<x<4,0<y<2} \\ \frac{1}{24}(x^2+2x) & si\ 0 < x < 4, 2 \leq y \\ \frac{1}{12}(y^2+4y) & si\ 4 \leq x, 0 < y < 2 \\ 1 & si\ 4 \leq x, 2 \leq y \end{cases}$$

Ahora podemos comprobar si la distribución condicional calculada es correcta

$F_Y(y|3) = \frac{\frac{\partial}{\partial x} F_{XY}(x,y)\big|_{x=3}}{f_X(3)} =$

$$= \begin{cases} 0 & si\ y \leq 0 \\ \frac{3}{48}(6y + y^2) & si\ 0 < y < 2 \\ \frac{\frac{1}{24}(2\cdot 3+2)}{\frac{1}{3}} = 1 & si\ y \geq 2 \end{cases}$$

Simplificando obtenemos la misma función.

Por último:

$F_X(x|X > 2Y) = P(X \leq x/X > 2y) =$

$= \frac{P(X \leq x, X > 2Y)}{P(X > 2Y)}$

Ayudándose de algún gráfico (o de forma analítica) obtenemos que

$P(X > 2Y) = \int_0^4 \int_0^{\frac{u}{2}} \frac{1}{24}(u+v)dvdu = \frac{5}{9}$

$P(X \leq x, X > 2Y) = \int_0^x \int_0^{\frac{u}{2}} \frac{1}{24}(u+v)dvdu = \int_0^x \frac{5v^2}{192} du = \frac{5x^3}{576}$

Por lo tanto

$$F_X(x|X > 2Y) = \begin{cases} 0 & si\ x \leq 0 \\ \frac{x^3}{64} & si\ 0 < x < 4 \\ 1 & si\ x \geq 4 \end{cases}$$

Derivando la función anterior obtenemos la densidad condicional

$$f_X(x|X > 2Y) = \begin{cases} \frac{3x^2}{64} & 0 < x < 4 \\ 0 & \textit{en otro caso} \end{cases}$$

Observaciones

a) Si A es un suceso entonces $P(Y \in A/X = x) = \int_A f_Y(y|x)dy$.

b) Si X, Y son independientes entonces

- $F_Y(y|x) = P(Y \leq y) = F_Y(y)$
- $f_Y(y|x) = \frac{d}{dy}F_Y(y|x) = f_Y(y)$

Ejemplo 97 *Sean X, Y dos v.a. continuas tal que su función de densidad sea*

$$f_{XY}(x,y) = \begin{cases} \frac{1}{2} & si \; -a < x < a, -a < y < a \\ 0 & \textit{en el resto de casos} \end{cases}$$

a) Encontrar el valor de a.

b) Encontrar las densidades marginales de X e Y.

c) Encontrar F_{XY}.

d) Calcular $E(X)$ y $E(Y)$.

a) Se puede hacer gráficamente.
De forma analítica tenemos:
$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = \int_{-a}^{a}\int_{-a}^{a}\frac{1}{2}dxdy = \int_{-a}^{a}[\frac{x}{2}]_{-a}^{a}dy = \int_{-a}^{a}dy = [ay]_{-a}^{a} = aa - a(-a) = 2a^2$
entonces $1 = 2a^2$ de donde $a = \frac{\sqrt{2}}{2}$, luego :

$$f_{XY}(x,y) = \begin{cases} \frac{1}{2} & \text{si } -\frac{\sqrt{2}}{2} < x < \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} < y < \frac{\sqrt{2}}{2} \\ 0 & \text{en el resto de casos} \end{cases}$$

b) Sea x tal que $-\frac{\sqrt{2}}{2} < x < \frac{\sqrt{2}}{2}$ entonces
$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_{-\frac{\sqrt{2}}{2}}^{\frac{\sqrt{2}}{2}}\frac{1}{2}dy = \frac{\sqrt{2}}{2}$.
Si $x \notin (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ entonces $f_{XY}(x,y) = 0$ para cualquier y, por lo tanto $f_X(x) = 0$
En definitiva:

$$f_X(x) = \begin{cases} \frac{\sqrt{2}}{2} & \text{si } -\frac{\sqrt{2}}{2} < x < \frac{\sqrt{2}}{2} \\ 0 & \text{en el resto de casos} \end{cases}$$

De forma análoga:
$f_Y(y) = \begin{cases} \frac{\sqrt{2}}{2} & \text{si } -\frac{\sqrt{2}}{2} < y < \frac{\sqrt{2}}{2} \\ 0 & \text{en el resto de casos} \end{cases}$
Notemos que: $f_{XY}(x,y) = f_X(x) f_Y(y)$ luego X e Y son independientes.
Si $-\frac{\sqrt{2}}{2} < y < \frac{\sqrt{2}}{2}$ entonces $F_X(x) = \int_{-\frac{\sqrt{2}}{2}}^{x} \frac{\sqrt{2}}{2} dt = [\frac{\sqrt{2}}{2}]_{-\frac{\sqrt{2}}{2}}^{x} = \frac{\sqrt{2}x-1}{2}$

Luego la función de distribución es: $F_X(x) = \begin{cases} 0 & \text{si } x \leq -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}x-1}{2} & \text{si } -\frac{\sqrt{2}}{2} < y < \frac{\sqrt{2}}{2} \\ 1 & \text{en el resto de casos} \end{cases}$

Análogamente:
$F_Y(y) = \begin{cases} 0 & \text{si } y \leq -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}y-1}{2} & \text{si } -\frac{\sqrt{2}}{2} < y < \frac{\sqrt{2}}{2} \\ 1 & \text{en el resto de casos} \end{cases}$
Como X e Y son independientes entonces:
$F_{XY}(x,y) = F_X(x) F_Y(y) =$

$$\begin{cases} 0 & \text{si } \begin{cases} x \leq -\frac{\sqrt{2}}{2} \text{ o} \\ y \leq -\frac{\sqrt{2}}{2} \end{cases} \\ \frac{1}{4}(\sqrt{2}x-1)(\sqrt{2}y-1) & \text{si } \begin{cases} -\frac{\sqrt{2}}{2} < x < \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} < y < \frac{\sqrt{2}}{2} \end{cases} \\ \frac{1}{2}(\sqrt{2}x-1)1 & \text{si } \begin{cases} -\frac{\sqrt{2}}{2} < x < \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \leq y \end{cases} \\ 1\frac{1}{2}(\sqrt{2}y-1) & \text{si } \begin{cases} -\frac{\sqrt{2}}{2} < y < \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \leq x \end{cases} \\ 1 & \text{otros casos, } \begin{cases} \text{es decir} \\ \frac{\sqrt{2}}{2} \leq x \\ \frac{\sqrt{2}}{2} \leq y \end{cases} \end{cases}$$

d)

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\frac{\sqrt{2}}{2}}^{\frac{\sqrt{2}}{2}} x \frac{\sqrt{2}}{2} dx =$$

$$\left[\frac{\sqrt{2}}{2} \frac{x^2}{2} \right]_{-\frac{\sqrt{2}}{2}}^{\frac{\sqrt{2}}{2}} = 0$$

De forma análoga $E(Y) = 0$.

5.2. Valores Esperados

Sea $g(X,Y)$ es una función $g : \mathcal{R}^2 \rightarrow \mathcal{R}$ y sean X, Y dos v.a. entonces:

a) Si las variables son conjuntamente discretas:

$$E(g(X,Y)) = \sum_x \sum_y g(x,y) P_{XY}(x,y) dxdy.$$

b) Si las variables son conjuntamente continuas

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y) dxdy.$$

En general si $g(X_1, \ldots, X_n)$ es una función de n variables aleatorias $g : \mathcal{R}^n \rightarrow \mathcal{R}$ entonces:

En el caso discreto tenemos que

$$E(g(X_1, \ldots, X_n)) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} g(x_1, \ldots, x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

mientras que en el continuo

$$E(g(X_1, \ldots, X_n)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, .., x_n) f_{X_1,..,X_n}(x_1, .., x_n) dx_1..d_{x_n}$$

.

Ejemplo 98 *Consideremos las v.a. X, Y con función de densidad*

$$f_{XY}(x,y) = \begin{cases} 1 & si\ 0 < x < 1, 0 < y < 1 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

Consideremos la función $g(X,Y) = X + Y$ entonces
$E[g(X,Y)] = E(X+Y) =$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y) dxdy =$$

$$\int_0^1 \int_0^1 (x+y) 1 dxdy =$$

$$\int_0^1 \left[\frac{x^2}{2} + yx\right]_0^1 dy = \int_0^1 (\frac{1}{2} + y) dy =$$

$$\left[\frac{1}{2} y + \frac{y^2}{2} x\right]_0^1 = 1$$

Vector de esperanzas

Al igual que en el caso discreto si $X_1, \ldots, X_n$ son n v.a. (continuas) su vector de medias o de esperanzas es el formado por las epsperanzas de cada una de las variables aleatorias:

$$\begin{pmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

5.3. Medida de la variación conjunta

En esta sección estudiaremos como medir la variación conjunta de dos variables.

5.3.1. Covarianza

Sean X, Y dos v.a. definimos la covarianza de X con Y

$$Cov(X,Y) = E((X - E(X))(Y - E(Y))$$

y la correlación es

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}.$$

La covariaza se suele denotar $Cov(X,Y) = \sigma_{XY}$. Diremos que las v.a. X e Y son incorreladas si $Cov(X,Y) = 0$ o lo que es lo mismo si su correlación es cero. Al igual que vimos en estadística descriptiva la covarianza y la correlación miden el grado de dependencia lineal entre dos variables.

Propiedades

- $Cov(X,Y) = Cov(Y,X)$ y $\rho_{XY} = \rho_{YX}$
- $Cov(X,Y) = E(XY) - E(X)E(Y)$.
- $Cov(X,-Y) = -Cov(X,Y)$.
- Si X e Y son independientes entonces $Cov(X,Y) = 0$ (son incorreladas). El recíproco no es siempre cierto.

- $-1 \leq \rho_{XY} \leq 1$.
- $Cov(X,X) = Var(X)$, $\rho_{XX} = 1$

Ejemplo 99 $f_{XY}(x,y) = \begin{cases} 1 & si\ 0 < x < 1, 0 < y < 1 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$

entonces $E(XY) = \int_0^1 \int_0^1 xydxdy = \int_0^1 \left[\frac{x^2}{2}y\right]_0^1 dy =$

$\int_0^1 \frac{y}{2}dy = \left[\frac{y^2}{4}\right]_0^1 = \frac{1}{4}$

Como ejercicio demostrar que $E(X) = \frac{1}{2}$ *y* $E(Y) = \frac{1}{2}$ *luego* $Cov(X,Y) = \frac{1}{4} - \frac{1}{2}\frac{1}{2} = 0$.

5.4. Propiedades de las sumas de v.a.

Sean $X_1, X_2, \ldots, X_n$ v.a. entonces:

I) $E(X_1 + \ldots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$.

II) Si $Cov(X_i, Y_j) = 0\ \forall\ i,j\ i \neq j$, es decir si son incorreladas dos a dos(cosa que sucede si son independientes) entonces $Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n)$

III) En general $Var(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^n \sum_{j=1}^n Cov(X_i, X_j) = \sum_{i=1}^n Var(X_i) + 2\sum_{i=1}^n \sum_{i<j}^n Cov(X_i, X_j)$

Observaciones
De las propiedades anteriores se deducen fácilmente las siguientes:

- $E(X - Y) = E(X) - E(Y)$.
- $Var(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$. Si además son incorreladas $Var(X + Y) = Var(X) + Var(Y)$.
- $Var(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$.
- X e Y con incorreladas si y sólo si $E(XY) = E(X)E(Y)$.

5.4.1. Matriz de varianzas-covarianzas, matriz de correlaciones

Dadas dos v.a. X, Y distribuidas conjuntamente se definen:

Matriz de varianzas-covarianzas $\begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$

Matriz de correlaciones $\begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix}$

Estas matrices, simétricas, resumen la variación conjunta de ambas variables. Para el caso de n la variables la definición se generaliza de forma natural obteniéndose matrices simétricas de orden n.

5.5. Distribución normal bivariante

También se le denomina gaussiana bivariante. Sean X e Y dos v.a. continuas con $E(X) = \mu_X$, $E(Y) = \mu_Y$, $Var(X) = \sigma_X^2$ y $Var(Y) = \sigma_Y^2$ y correlación ρ_{XY}. Diremos que X, Y son conjuntamente gaussianas si tienen por densidad conjunta:

Simplificando:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}} \quad e^{-(x^2+y^2-2\rho_{XY}xy)/(2(1-\rho_{XY}^2))}$$

En forma matricial si $M = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}$

entonces

$f_{XY}(x,y) = \frac{1}{|M|^{\frac{1}{2}}2\pi} e^{-\frac{1}{2}(x-\mu_X, y-\mu_Y)M^{-1}\begin{pmatrix} x-\mu_X \\ y-\mu_Y \end{pmatrix}}$

Donde efectivamente $det(M) = \sigma_X^2\sigma_Y^2 - \sigma_{XY}^2$ y entonces $det(M)^{\frac{1}{2}}2\pi = 2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}$

5.5.1. Normal bivariante estándar

La normal bivariante estándar (Z_1, Z_2) es aquella que tiene por vestor de medias el cero y por matriz de covarianzas la matriz identidad, es decir $\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ y $M = \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{YX} & \sigma_Y^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

entonces su densidad es

$$f_{Z_1Z_2}(z_1,z_2)=\frac{1}{2\pi}e^{-\frac{1}{2}(z_1^2+z_2^2)}.$$

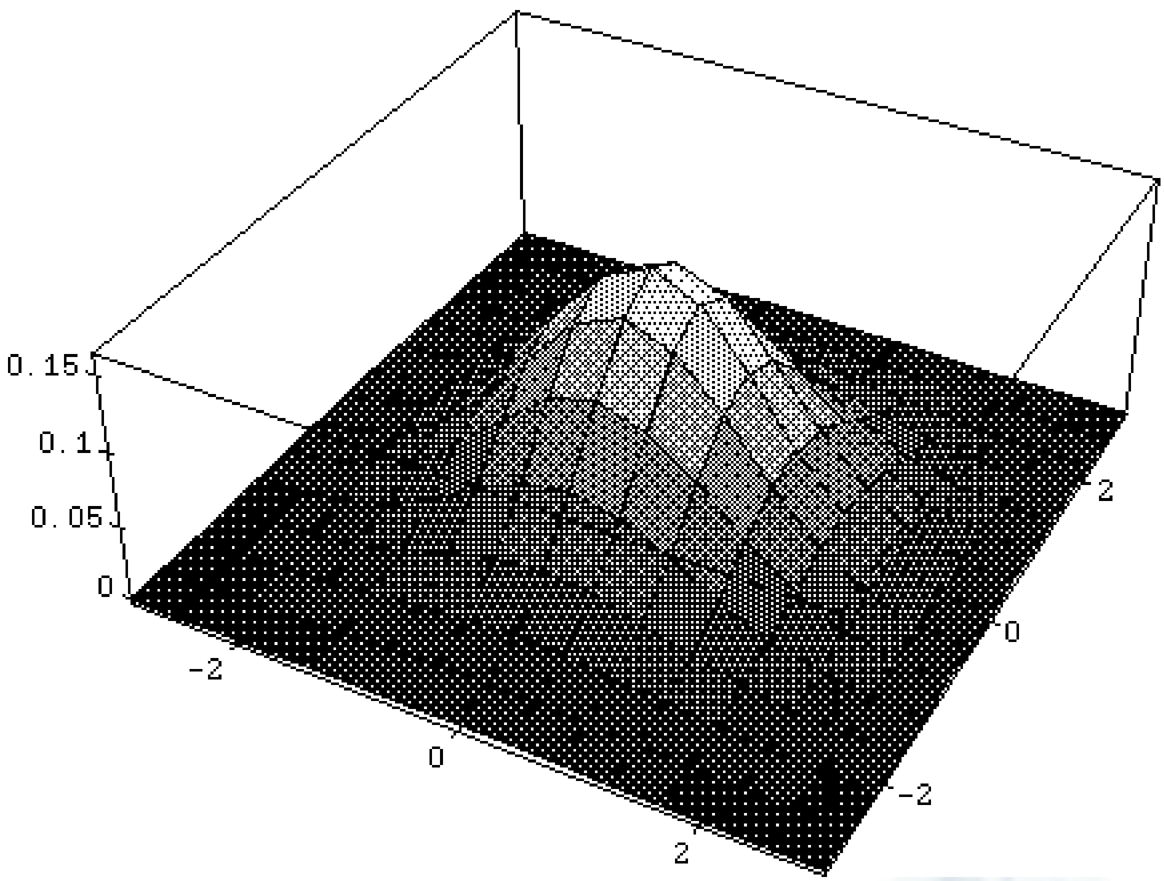


Figura 5.1: Gráfica de la función de densidad normal estándar

Proposición 100 *Si X e Y son conjuntamente gaussianas entonces:*

a) *$X+Y$ sigue una distribución normal de media $E(X+Y)=\mu_X+\mu_Y$ y varianza $Var(X+Y)=\sigma_X^2+\sigma_Y^2+2\sigma_{XY}$. De hecho cualquier combinación lineal no trivial de X e Y es normal.*

b) *$Cov(X,Y)=0$ si y sólo si X e Y son independientes.*

c) *sus marginales son normales.*

Proposición 101 *Si X es una v.a. normal e Y es otra variable normal y son independientes entonces X e Y son conjuntamente normales.*

Proposición 102 *Sean X e Y dos v.a. conjuntamente normales entonces:*

a) *$X+Y$ sigue una distribución normal de media $E(X+Y)=\mu_X+mu_Y$ y varianza $Var(X+Y)=\sigma_X^2+\sigma_Y^2+2\sigma_{XY}$.*

b) *$Cov(x,y)=0$ si y sólo si X e Y son independientes.*

5.5.2. La ley normal multivariante

No la definimos aquí, es una generalización de la normal bidimensional y al igual que ella su función de distribución queda definida por el vector de medias y la matriz de varianzas-covarianzas.

Proposición 103 *a)* *Sean $X_1,X_2,\ldots,X_n$ n v.a. que sigan una ley conjunta normal entonces cualquier combinación lineal no trivial de $X_1,\ldots X_n$ sigue una ley normal unidimensional*

b) *Si $X_1,X_2,\ldots,X_n$ son v.a. normales univariantes e independientes entonces son conjuntamente normales.*

c) *Como consecuencia de los resultados anteriores si $X_1,X_2,\ldots,X_n$ son v.a. normales univariantes e independientes entonces cualquier combinación lineal no trivial de ellas es normal univariante.*

d) *Las marginales de una normal multivariante son normales.*

Estas propiedades nos serán de utilidad en la siguiente sección.

5.6. Teorema del Límite Central (T.L.C.)

Sean $X_1,X_2,\ldots,X_n$ v.a. independientes e idénticamente distribuidas (iid.) con $E(X_i)=\mu$ y $Var(X_i)=\sigma^2$ para $i=1,2,\ldots,n$. Sea $X=X_1+X_2+\ldots+X_n$ entonces se tiene que: $E(X)=n\mu$ y $Var(x)=n\sigma^2$

Además podemos “tipificar” X de forma que :

$$Z=\frac{X-E(X)}{\sqrt{Var(X)}}=\frac{X-n\mu}{\sqrt{n\sigma^2}}$$

Operando obtenemos que Z se puede poner como:

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

donde
$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{X}{n}$$

Se deja como ejercicio la demostración de las siguientes propiedades:

Proposición 104 *Con las notaciones y condiciones anteriores:*

a) $E(\sum_{i=1}^{n} X_i) = n\mu$ *y* $Var(\sum_{i=1}^{n} X_i) = n\sigma^2$.

b) $E(\overline{X}) = \mu$ *y* $Var(\overline{X}) = \frac{\sigma^2}{n}$.

c) $E(Z) = 0$ *y* $Var(Z) = 1$.

Notemos que si la distribución común a las variables es normal entonces Z es una combinación lineal de normales y por lo tanto conocemos su distribución. La pregunta es ¿qué pasa cuando la distribución común no es normal sino otra cualquiera? La respuesta nos la da el siguiente Teorema llamado del Límite Central.

Teorema del Límite Central

Teorema 105 *Con las notaciones anteriores y cuando n tiende a ∞ entonces*

$$Z = \frac{X - n\mu}{\sqrt{n\sigma^2}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

tiende a tener una distribución $N(0,1)$. Dicho de otra forma:
dado $x \in \mathbb{R}$ se tiene que

$$\lim_{n\to\infty} P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Nota: La importancia de este teorema es tremenda pues nos dice que la distribución de $\overline{X}$ en cualquier v.a. con cualquier distribución se aproxima, de alguna forma, cuando n crece a una distribución normal.

En particular podemos utilizar este resultado para aproximar las distribuciones que de alguna manera sumen v.a. independientes con la misma distribución como es el caso de la binomial y como es el espíritu de la Poisson. Pero tendremos un pequeño problema que estaremos aproximando una v.a. discreta por una continua, por lo que tendremos que realizar pequeñas correcciones (se suelen llamar correcciones de continuidad de Fisher).

Aproximación de la distribución binomial y la Poisson por la normal

Aproximación de una Binomial por una distribución normal

Veamos como utilizando el T.L.C. podemos aproximar por una distribución normal algunas distribuciones binomiales.

Sea X una v.a. con distribución $B(n,p)$ entonces $E(X)=np$ y $Var(X)=npq$. También sabemos que $X=X_1+\cdots+X_n$ donde cada X_i es una Bernouilli de parámetro p e independientes entre si, entonces:

$Z=\frac{X-E(X)}{\sqrt{Var(X)}}=\frac{X-np}{\sqrt{np(1-p)}}$

Estamos en condiciones de aplicar el T.L.C. La aproximación se realiza de la siguiente manera:

- $P(X=k)\approx P\left(\frac{k-0{,}5-np}{\sqrt{npq}}\leq Z\leq\frac{k+0{,}5-np}{\sqrt{npq}}\right)$

Mediante un razonamiento similar :

$$\bullet\ P(X\leq k)\approx P\left(Z\leq\frac{k+0{,}5-np}{\sqrt{npq}}\right)$$

y

- $P(a\leq X\leq b)\approx$

$\approx P\left(\frac{a-0{,}5-np}{\sqrt{npq}}\leq Z\leq\frac{b+0{,}5-np}{\sqrt{npq}}\right)$

Donde, en ambos casos Z es la normal estándar.

Corrección de continuidad El motivo de sumar o restar 0.5 en las aproximaciones es corregir el efecto que tienen aproximar una v.a. discreta por una continua. Esta operación recibe el nombre de corrección de continuidad de Fisher. Gráficamente el área que hacemos corresponder a la probabilidad de cada valor entero k en una binomial corresponde a la comprendida entre la curva normal y el segmento centrado en k de amplitud 1.

Ejemplo 106 *X=número de caras en 100 lanzamientos de una moneda. $P(cara)=\frac{1}{2}$. Calcular:*

a) $P(40 \leq X \leq 49)$

b) $P(X = 37)$

c) $P(X \leq 50)$

Solución:

$E(X) = 50 = \mu_X$, $Var(X) = 25$ *y* $\sigma_X = 5$

$Z = \frac{X-50}{5}$ *por el T.L.C. se aproxima a normal estándar*

Entonces

$P(40 \leq X \leq 49) \approx P(\frac{40-0{,}5-50}{5} \leq Z \leq \frac{49+0{,}5-50}{5}) = P(-\frac{10{,}5}{5} \leq Z \leq -\frac{0{,}5}{5}) = F_Z(-\frac{0{,}5}{5}) - F_Z(-\frac{10{,}5}{5}) = 1 - F_Z(\frac{0{,}5}{5}) - 1 + F_Z(\frac{10{,}5}{5}) = F_Z(\frac{10{,}5}{5}) - F_Z(\frac{0{,}5}{5}) = F_Z(2{,}1) - F_Z(0{,}1) = 0{,}9821 - 0{,}5398 = 0{,}4423$

La probabilidad exacta da 0,442605*. !!La aproximación es bastante buena!!*

b) $P(X = 37) = P(37 \leq X \leq 37) \approx P(\frac{37-0{,}5-50}{5} \leq Z \leq \frac{37+0{,}5-50}{5}) = P(-\frac{13{,}5}{5} \leq Z \leq -\frac{12{,}5}{5}) = F_Z(\frac{13{,}5}{5}) + F_Z(\frac{12{,}5}{5}) = F_Z(2{,}7) - F_Z(2{,}5) = 0{,}9965 - 0{,}9938 = 0{,}0027$

La probabilidad exacta da 0,0026979*. !!La aproximación es bastante buena!!*

c) $P(X \leq 50) \approx P(Z \leq \frac{50+0{,}5-50}{5}) = P(Z \leq 0{,}5) = F_Z(0{,}1) = 0{,}5398$

La probabilidad exacta calculada con un programa adecuado da 0,539795 *!!La aproximación es bastante buena!!*

Aproximación de una Poisson por una distribución normal

De forma similar, y aplicando también la corrección de continuidad podemos aproximar la probabilidad de una v.a. Poisson por una normal.

Si $X \equiv Po(\lambda)$ y λ es grande, entonces podemos utilizar el TLC o también esta otra aproximación:

- $$P(X = k) \approx P\left(\frac{k - 0{,}5 - \lambda}{\sqrt{\lambda}} \leq Z \leq \frac{k + 0{,}5 - \lambda}{\sqrt{\lambda}}\right)$$

Mediante un razonamiento similar :

- $$P(X \leq k) \approx P\left(Z \leq \frac{k + 0{,}5 - \lambda}{\sqrt{\lambda}}\right)$$

y

- $$P(a \leq X \leq b) \approx P\left(\frac{a - 0{,}5 - \lambda}{\sqrt{\lambda}} \leq Z \leq \frac{b + 0{,}5 - \lambda}{\sqrt{\lambda}}\right)$$

Ejemplo 107 *Sea X =número de trabajos que llegan a un centro de cálculo en un lapso de 60 minutos.*

Supongamos que X siga una ley Poisson y que el número medio de trabajos que llegan por minuto sea 0,2. Entonces $E(X) = 0{,}2 \cdot 60 = 12$ por lo tanto X es una $Po(12)$ es decir $\lambda = 12$ y por lo tanto $\mu_X = 12$ y $\sigma_X^2 = 12$.

Si queremos calcular

$P(Y \leq 10) \approx P(Z \leq \frac{10+0{,}5-12}{\sqrt{12}}) = P(Z \leq -0{,}4330127)$

$F_Z(-0{,}4330127) \approx 1 - F_Z(0{,}43) = 1 - 0{,}6664 = 0{,}3336$

La probabilidad exacta[3] *da 0,347229. La aproximación es buena.*

[3] Con R es ppois(10,12)

Capítulo 6

Muestreo Estadístico

En esta tema sentaremos las bases del muestreo estadístico y estudiaremos las distribuciones de algunos estadísticos muestrales; como la media aritmética, las proporciones y la varianza.

6.1. Conceptos básicos

Aunque en el tema dedicado a la estadística descriptiva ya vimos algunos de los conceptos básicos sobre muestras, no está de más que los repasemos y ampliemos:

Población: Conjunto de individuos con una característica observable común.

Muestra: Subconjunto de la población del que se espera que la represente.

El objetivo de la **estadística inferencial** es obtener información sobre el conjunto de la población a partir de un subconjunto representativo de ella llamado muestra.

En la práctica lo más común es conocer sólo una parte de la población y lo que queremos es averiguar por ejemplo qué esperanza o qué varianza o ... tiene determinada población.

Inferir información de una muestra es contestar preguntas sobre el total de la población a partir del estudio de una muestra representativa de la misma.

Pasos en un estudio con muestreo

a) ¿Qué información se necesita?

b) ¿Cuál es la información relevante? ¿Se dispone de acceso a todos los individuos de la población?

c) ¿Cómo seleccionamos los individuos de la muestra?

d) ¿Qué método emplearemos para obtener la información de los individuos de la muestra?

e) ¿Qué herramientas utilizaremos para hacer inferencias?

f) ¿Qué conclusiones podemos obtener?

g) Si las conclusiones son fiables y suficientes redactar informe, en caso contrario volver a empezar.

6.1.1. Tipos de muestreo

El objetivo de las técnica de muestreo es encontrar métodos para seleccionar muestras representativas de la población.

Las técnicas básicas son: el muestreo aleatorio simple, el muestreo aleatorio estratificado, el muestreo sistemático y el muestreo polietápico. Cada una de estas técnicas proporciona una muestra representativa de la población

- **Muestreo aleatorio probabilístico** El muestreo aleatorio consiste en seleccionar muestra de la población con igual probabilidad. Esto quiere decir que cualquier conjunto de individuos tiene la misma probabilidad de ser seleccionado.

  Pensemos en una urna con 100 bolas de colores. Hay dos maneras de obtener una muestra de 10 bolas. Una podría ser sacar una bola de la urna, observar su color y devolverla a la urna. Es decir vamos obteniendo individuos y los volvemos a poner en la urna. Este tipo de muestreo recibe el nombre de **muestreo con reposición** o **muestreo aleatorio simple**. Otra forma sería repetir la experiencia anterior pero no devolver las bolas a la urna. En este caso también sucede que cualquier selección de los 10 individuos es equiprobable. En este caso se habla de **muestreo aleatorio sin reemplazo**.

Notemos que el muestreo aleatorio con reemplazo o simple y el muestreo aleatorio sin reemplazo serán prácticamente el mismo cuando el tamaño de la población sea muy grande en relación a la muestra y la probabilidad de que dos individuos se repitan sea muy pequeña. De todas maneras si la población es pequeña se suelen aplicar estadísticos corregidos por el efecto de población finita.

- **Muestreo Aleatorio Estratificado**

  En el caso en que la población esté dividida en grupos o estratos y que estos sean de interéas para la variable del estudio se toman muestras donde cada grupo esté representado en función de su tamaño. Por ejemplo los estratos podrían ser las grupos de edad, o en las Islas Baleares por islas en proporción a su número de habitantes,, en una provicia por municipios también en función de su número de habitantes, por estratos de nivel educativo .....

  En estos casos Se determina el tamaño de la muestra en cada estrato y luego se toma una muestra aleatoria simple en ese bloque.

- **Muestreo por conglomerados**

  El proceso de obtener una muestra aleatoria es caro. Por ejemplo si el estudio se realiza por ejemplo sobre conjuntos de personas. Pesemos que queremos saber los hábitos de seguridad vial que tienen los estudiantes de primaria de Baleares. Para ello, previo permiso de la autoridad responsable queremos seleccionar una muestra representativa de los escolares de Baleares. Lo lógico es que una vez los encuestadores estén en un centro pregunten a varios alumnos. La selección de los colegios debe ser al azar.

  Lo mismo para una encuesta en una ciudad, seleccionaremos un edificio e intentaremos encuesta a todos sus habitantes. La selección de los bloques debe ser al azar.

- **Muestreo Polietápico** En este caso se selecciona en etapas sucesivas grupos de la población. Por ejemplo encuesta a escolares de ESO de Mallorca, primero seleccionan al azar los centros y luego las clases y luego los individuos o se hace un muestreo por bloque de una clase.

-

- **Muestreo no probabilístico** Cuando nos tenemos que confomar con la información disponible o la obtenida voluntariamente
- Combinaciones de las técnicas anteriores y otro tipos de técnicas dan lugar a nuevo tipos de muestreo estadístico.

En cada tipo de muestro se emplean técnicas para estimar los resultados de interés: proporciones, medias , varianzas...

En cada uno de los tipos de muestreo las técnicas de estimación puede ser diferentes.

En este módulo sólo se estudian técnicas de estimación para el caso de muestreo aleatorio simple.

Muestreo aleatorio simple

Estudiemos un poco más en detalle el caso del muestreo aleatorio simple con o sin reposición. La idea es que queremos seleccionar una muestra de tamaño n (es decir formada por n individuos) de una población de tamaño N. Obtendremos una muestra aleatoria simple cuando todas las muestras posibles de n individuos tengan la misma probabilidad de ser elegidas.

El tener una m.a.s de una población junto con un tamaño muestral adecuado nos asegurará la suficiente representatividad de la muestra.

Hagamos algunas observaciones:

- El proceso mismo del muestreo aleatorio simple es complejo.
- Una forma sencilla es numerar, si es posible a todos los individuos de la población y sortearlos eligiendo números como si se tratase de una lotería; por ejemplo con una tabla de números aleatorios[1] o con un generador de números aleatorios; por ejemplo el que llevan nuestra hoja de cálculo.
- En ocasiones esto es impracticable o caro:
  a) Población mundial de seres humanos.
  b) Población de llamadas a una centralita telefónica.

[1]En realidad los números aleatorios generados por diversos tipos de algoritmos son pseudoalatorios; son números que superan determinados test de aleatoriedad

c) Población de votantes en las próximas elecciones locales y autonómicas.

- En algunos de estos casos será luego impracticable localizar a los individuos seleccionados y convencerlos de que respondan, muchos no querrán.

6.2. Inferencias

Nuestro interés es estudiar la distribución de probabilidad de la muestra o de alguna función de la muestra y de esta inferir resultados de la distribución de probabilidad de la población.

Estadísticos y distribuciones muestrales

Supongamos que tenemos una muestra aleatoria simple de una población y deseamos obtener información sobre la media o la varianza poblacionales. Estas inferencias las basaremos en un estadístico, que estudiaremos en más profundidad en los temas siguientes y que no es más que una función que depende de la muestra. p e: media aritmética, proporción muestral…

6.2.1. Distribución muestral de un estadístico

La **distribución muestral o distribución en el muestreo** de un estadístico es la distribución de probabilidad de los valores que puede tomar el estadístico en todas las posibles muestras, es decir la distribución de la variable aleatoria que define el estadístico.

Veamos un ejemplo:

Ejemplo 108 *Supongamos que queremos estimar cuál es número medio de municiones defectuosas una determinada marca por cada caja de 10 municiones cada una. Para ello tomamos una muestra aleatoria simple de cuatro cajas X_1, X_2 X_3, X_4, las probamos y obtenemos los siguientes resultados:*

primera caja	:	*1*	*defectuoso*
segunda caja	:	*2*	*defectuosos*
tercera caja	:	*0*	*defectuoso*
cuarta caja	:	*1*	*defectuosos*

Definimos el estadístico media aritmética de munición defectuosa como:

$$\overline{X} = T(X_1, X_2, X_3, X_4) = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

En este caso $\overline{X} = 1$.

Supongamos que tomamos repetidas muestras de tamaño 4 y los resultados son:

M. 1	M. 2	M. 3	M. 4	M. 5	M. 6	M. 7	M. 8	M. 9	M. 10
0	1	3	0	0	1	0	0	0	1
1	1	1	0	1	1	1	0	0	2
0	1	2	1	0	0	1	2	0	1
1	1	2	2	1	3	0	0	1	1

M. 11	M. 12	M. 13	M. 14	M. 15	M. 16	M. 17	M. 18	M. 19	M. 20
0	0	1	2	0	2	1	2	1	1
1	0	1	0	1	1	2	0	0	1
1	0	2	0	1	1	0	1	1	0
3	3	1	0	0	2	1	0	1	1

Las medias aritméticas de cada muestra son:

0.50	1.00	2.00	0.75	0.50
1.25	0.50	0.50	0.25	1.25
1.25	0.75	1.25	0.50	0.50
1.50	1.00	0.75	0.75	0.75

Entonces:

$$P_{\overline{X}}(0{,}25)) = P(\overline{X} = 0{,}25) = \frac{1}{20} = 0{,}05$$

$$P_{\overline{X}}(0{,}50)) = P(\overline{X} = 0{,}50) = \frac{6}{20} = 0{,}30$$

$$P_{\overline{X}}(0{,}75)) = P(\overline{X} = 0{,}75) = \frac{5}{20} = 0{,}25$$

$$P_{\overline{X}}(1)) = P(\overline{X} = 1) = \frac{2}{2} = 0{,}10$$

$$P_{\overline{X}}(1{,}25)) = P(\overline{X} = 1{,}25) = \frac{4}{20} = 0{,}20$$

$$P_{\overline{X}}(1{,}50)) = P(\overline{X} = 1{,}5) = \frac{1}{20} = 0{,}05$$

$$P_{\overline{X}}(2)) = P(\overline{X} = 2) = \frac{1}{20} = 0{,}05$$

Esta sería una aproximación a la distribución muestral del estadístico $\overline{X}$ a partir de los datos de varias muestras.

6.2.2. Distribución la media muestral

La distribución del estadístico puede seguir un modelo preestablecido si se cumplen varias condiciones. Por ejemplo, supongamos que hemos tomado una muestra aleatoria simple de n observaciones de una v.a. X en una población de media μ_X y desviación típica σ_X.

Representemos por $X_1, X_2, \ldots, X_n$ los elementos de n observaciones independientes que forman una muestra aleatoria simple de ésta población. Cada una de las observaciones de la población son así mismo variables aleatorias con la misma esperanza y varianza que la población.

Llamaremos **media aritmética o media muestral** de la muestra $X_1, \ldots, X_n$ a

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

Resaltemos que:

a) $E(\overline{X}) = \frac{1}{n}E(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}(\mu_X + \mu_X + \cdots + \mu_X) = \mu_X$

b) El valor esperando de la media aritmética de la muestra es la media poblacional. Entonces el estadístico media muestral *estima* la media poblacional. Dicho de otra forma la esperanza de la distribución muestral de la media aritmética es la media poblacional.

Pero que el valor esperado sea μ_X no quiere decir que $\overline{X}$ sea exactamente μ_X. Estudiemos la varianza de $\overline{X}$. Como $X_1, \ldots, X_n$ son independientes tenemos que:

a) $Var(\overline{X}) = \frac{1}{n^2}Var(X_1 + X_2 + \cdots + X_n) = \frac{1}{n^2}(Var(X_1) + Var(X_2) + \cdots + Var(X_n)) = \frac{1}{n^2}n\sigma_X^2 = \frac{1}{n}\sigma_X^2$

b) Luego si n es suficientemente grande (o cuando $n \to \infty$) la varianza tenderá a estar muy próxima a cero.

Ejemplo 109 *No siempre tendremos independencia entre $X_1, \ldots, X_n$. Por ejemplo supongamos que queremos averiguar cuántos votos afirmativos hay en una urna con 10 votos. Tenemos dos opciones para realizar la muestra aleatoria simple:*

a) Tomar un voto al azar anotar su resultado y devolverlo a la urna, repetir el proceso 3 veces más. En este caso es un muestreo con reemplazamiento.

b) Tomar sucesivamente 4 votos de la urna sin reemplazarlos. En este caso es un muestreo reemplazamiento.

En ambos casos la muestra obtenida es una muestra aleatoria pues todos los subconjuntos de individuos tienen igual probabilidad de ser elegidos.

Pero en el primer caso tenemos independencia entre cada una de las observaciones mientras que en el segundo esto no es así.

En la práctica se elige casi siempre el muestreo consistente en observar n individuos distintos. Además si n es pequeño con respecto al tamaño de la población N podemos suponer que las variables son prácticamente independientes. Si no, tenemos que corregir la varianza multiplicándola por lo que se llama **factor de población finita** y tendremos que

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{1}{n}\sigma^2_X \frac{N-n}{N-1}$$

Frecuentemente utilizaremos la expresión tipificada de la media muestral:

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}$$

Además para tamaños muestrales grandes aplicando el Teorema del Límite Central la distribución de Z es una normal estándar. Este resultado es muy importante pues **sea cual sea la distribución de X la distribución de $\overline{X}$ será conocida si n es suficientemente grande**.

Distribución muestral de $\overline{X}$

Sea X la variable aleatoria de interés que queremos observar en una cierta población. Supongamos que la esperanza pobalacional es $E(X) = \mu_X$ y su varianza $Var(X) = \sigma_X^2$. Sea $X_1, \ldots, X_n$ una muestra aleatoria simple de dicha población:

Entonces se cumplen las propiedades siguientes:

a) $\mu_{\overline{X}} = E(\overline{X}) = \mu_X$; el valor esperado de la media es la media poblacional.

b) La varianza y la desviación típica de $\overline{X}$ se pueden obtener con las siguientes fórmulas $\sigma_{\overline{X}}^2 = \frac{1}{n}\sigma_X^2$ y la desviación típica de $\overline{X}$ es $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}$ que también recibe el nombre de error estándar de $\overline{X}$.

c) En el caso en que el tamaño muestral n no sea *pequeño* en relación al tamaño de la población entonces tenemos que aplicar el factor de corrección de población finita en el cálculo del error estándar de $\overline{X}$:

$$\sigma_{\overline{X}}^2 = \frac{1}{n}\sigma_X^2 \frac{N-n}{N-1}$$

y el error estándar será $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

d) Si la distribución de la población (X) es normal entonces la variable aleatoria:

$$Z = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}}$$

es una normal estándar. O lo que es lo mismo $\overline{X}$ es una normal con media μ_X y desviación típica $\sigma_{\overline{X}}$

e) Si la distribución de la población no es normal pero el tamaño muestral es suficientemente grande entones por el T.L.C. la distribución de Z también se aproxima a una normal estándar y por lo tanto $\overline{X}$ se aproxima a una normal con media μ_X y desviación típica $\sigma_{\overline{X}}$

Ejemplo 110 *El precio medio por m^2 de venta de casas nuevas durante el último año en una determinada ciudad fue de 115000 pts. La desviación típica de la población fue de 25000 pts. Se toma una muestra aleatoria de 100 casas nuevas de esta ciudad.*

a) ¿Cuál es la probabilidad de que la media muestral de los precios de venta sea menor que 110000 pts?

b) ¿Cuál es la probabilidad de que la media muestral de los precios de venta esté entre 113000 pts y 117000 pts?

c) ¿Cuál es la probabilidad de que la media muestral de los precios de venta esté entre 114000 pts y 116000 pts?

d) Sin hacer cálculos, razonar en cuál de los siguientes rangos resulta más probable que se encuentre la media muestral de los precios de venta:

113000 pts.-	*115000 pts.*
114000 pts.-	*116000 pts.*
115000 pts.-	*117000 pts.*
116000 pts.-	*118000 pts.*

Supongamos que el número de casas de la ciudad sea muy grande en relación al tamaño muestral $n = 100$. Entones si X es la v.a. precio de una casa de la ciudad el enunciado nos dice que $\mu_X = E(X) = 115000.$ y $\sigma_X = 25000$. Tomamos una muestra aleatoria simple $X_1, \ldots, X_{100}$ de precios entonces F $\mu_{\overline{X}} = \mu_X = 115000$ y $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{25000}{\sqrt{100}} = 2500$

Además $Z = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} = \frac{\overline{X} - 115000}{2500}$ sigue aproximadamente una distribución normal estándar.

Solución:

a) $P(\overline{X} \leq 110000) =$
$P(Z \leq \frac{110000-115000}{2500}) = P(Z \leq -2) = F_Z(-2) = 1 - F_Z(2) = 1 - 0{,}9772 = 0{,}0228$

b) $P(113000 \leq \overline{X} \leq 117000) =$
$P(\frac{113000-115000}{2500} \leq Z \leq \frac{117000-115000}{2500}) = F_Z(0{,}8) - F_Z(-0{,}8) = 2F_Z(0{,}8) - 1 = 2(0{,}7881) - 1 = 0{,}5762$

c) $P(114000 \leq \overline{X} \leq 116000) =$
$P(\frac{114000-115000}{2500} \leq Z \leq \frac{116000-115000}{2500}) = F_Z(0{,}4) - F_Z(-0{,}4) = 2F_Z(0{,}8) - 1 = 2(0{,}6554) - 1 = 0{,}3108$

d) La media aritmética de los precios $\overline{X}$ sigue aproximadamente una distribución normal entonces gráficamente el intervalo de mayor probabilidad será el que mayor área cubra bajo la curva normal (centrada en 115000) y ese intervalo es 114000 pts.-116000 pts.

6.2.3. Distribución de una proporción muestral

La proporción muestral de un evento en una población vendrá generalmente asociada a una variable binomial (si la población es pequeña será Hipergeométrica).

Por ejemplo si tomamos una muestra de tamaño n, determinar el porcentaje de votos que recibirá el Partido P.X. en las próximas elecciones es lo mismo que determinar el parámetro p de $X = \sum_i^n X_i$ número de votantes de P.X. en la muestra de tamaño n, que es $B(n,p)$ y donde cada X_i es una $Ber(p)$ independiente de forma que $X_i = 1$ si el iésimo individuo y cero en caso contrario, así que la proporción muetral es la media aritmética de observaciones $Ber(p)$.

¿Será realmente binomial? Notemos que en la muestra no preguntaremos dos veces al mismo individuo, luego las observaciones no son exactamente independientes, pero si el tamaño de la población es grande respecto a la muestra podemos considerarlas así, ya que la probabilidad de repuesta afirmativa no cambia (es despreciable el cambio).

Definición 111 *Sea X el número de éxitos en una muestra binomial de n observaciones, con probabilidad de éxito p. Entonces la proporción de éxitos en la muestra es:*

$\hat{p}_X = \frac{X}{n}$, *y se denomina proporción muestral.*

Distribución en el muestreo de $\hat{p}_X$

Sea $\hat{p}_X$ la proporción de éxitos en una muestra aleatoria de n observaciones. Entonces:

a) $E(\hat{p}_X) = E(\frac{X}{n}) = \frac{E(X)}{n} = \frac{np}{n} = p$

b) La distribución muestral de $\hat{p}_X$ tiene varianza $\sigma^2_{\hat{p}_X} = Var(\frac{X}{n}) = \frac{Var(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$ y por lo tanto su desviación típica es $\sigma_{\hat{p}_X} = \sqrt{\frac{p(1-p)}{n}}$ que recibe también el nombre de error estándar de la proporción muestral

c) Si n es pequeño en relación al tamaño de la población N tenemos que aplicar el factor de corrección de población finita y entonces el error estándar de $\hat{p}_X$ es

$$\sigma_{\hat{p}_X} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}.$$

d) Si el tamaño muestral es grande (por ejemplo $n > 30$ o mejor $n > 40$) entonces

$$Z = \frac{\hat{p}_X - p}{\sigma_{\hat{p}_X}},$$

y que se distribuye aproximadamente como una normal estándar o lo que es lo mismo $\hat{p}_X$ se distribuye aproximadamente como una normal con esperanza p_X y varianza $\sigma_{\hat{p}_X}$.

e) Cuando no se verifiquen las condiciones de aproximación utilizaremos la distribución t de Student que veremos el el siguiente tema.

Observación Notemos que si n crece el error estándar disminuye y entonces $\hat{p}$ estará más cerca del valor real p.

Ejemplo 112 *El dueño de una tienda de discos ha comprobado que el 20 % de los clientes que entran en su tienda realizan una compra. Cierta mañana entraron en esa tienda 180 personas, que pueden ser consideradas como una muestra aleatoria de todos sus clientes.*

a) ¿Cuál será la media de la proporción muestral de clientes que realizaron alguna compra?

b) ¿Cuál es la varianza de la proporción muestral?

c) ¿Cuál es el error estándar de la proporción muestral?

d) ¿Cuál es la probabilidad de que la proporción muestral sea mayor que 0.15?

Solución: *El tamaño de la muestra es pequeño en relación al número total de clientes. Tenemos que $p = 0{,}2$ (probabilidad de éxito en la venta). Sea $X =$ número de clientes que compran entre los 180, entonces:*

a) $E(\hat{p}_X) = p = 0{,}2$

b) $\sigma^2_{\hat{p}_X} = \frac{p(1-p)}{n} = \frac{0{,}2(1-0{,}2)}{180} = 0{,}0009$

c) $\sigma_{\hat{p}_X} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{0{,}0009} = 0{,}03$

d) *Como* n *es grande entonces* $Z = \frac{\hat{p}_X - p}{\sigma_{\hat{p}_X}} = \frac{\hat{p}_X - 0{,}2}{0{,}03}$ *sigue aproximadamente una distribución normal estándar, entonces:*

$P(\hat{p}_X > 0{,}15) = 1 - P(\hat{p}_X \leq 0{,}15) = 1 - P(Z \leq \frac{0{,}15-0{,}2}{0{,}03}) = 1 - F_Z(-1{,}67) = F_Z(1{,}67) = 0{,}9525$

6.2.4. Distribución muestral de la varianza muestral

Definición 113 *Sea* $X_1, \ldots, X_n$ *una muestra aleatoria simple de una población* (X) *con* $E(X) = \mu_X$ *y* $Var(X) = \sigma_X^2$. *Llamaremos* <u>varianza muestral</u> *a :*

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

$\tilde{S}_X = +\sqrt{S_X^2}$ *recibe el nombre de desviación típica muestral.*

Denotaremos por $S_X^2 = \frac{n-1}{n}\tilde{S}_X^2$ *y* $S_X = +\sqrt{S_X^2}$.

Proposición 114

1. $S_X^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n} = \left(\frac{\sum_{i=1}^n X_i^2}{n} - \overline{X}^2\right)$
2. $E(S_X^2) = \frac{n-1}{n}\sigma_X^2$
3. $\tilde{S}_X^2 = \frac{n}{n-1}\left(\frac{\sum_{i=1}^n X_i^2}{n} - \overline{X}^2\right)$
4. $E(\tilde{S}_X^2) = \sigma_X^2$

Distribución en el muestreo de $\tilde{S}_X^2$

Con las notaciones anteriores tenemos que:

a) $E(\tilde{S}_X^2) = \sigma_X^2$

b) Si la distribución de la población es normal entonces la variable $\frac{(n-1)\tilde{S}_X^2}{\sigma_X^2}$ se distribuye según una ley χ_{n-1}^2

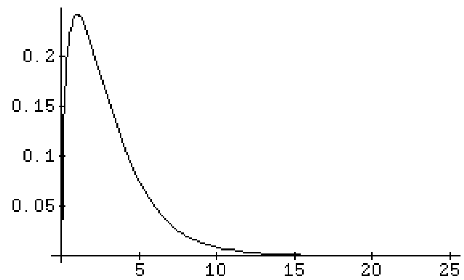


Figura 6.1: Gráfica de la función de densidad de una χ^2

La distribución χ_n^2 (chi-cuadrado con n g.l.)

Si $X_1, X_2, \ldots, X_n$ son n v.a. independientes y $X_i \equiv N(0,1)$ entonces:

$$X = X_1^2 + X_2^2 + \ldots + X_n^2$$

es una v.a. que diremos que se distribuye chi-cuadrado con n grados de libertad y lo notaremos por χ_n^2

La función de densidad de una χ_n^2 es :

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}$$

con $x \geq 0$ y $\Gamma(n/2) = \int_0^{+\infty} u^{(n/2)-1} e^{-u} du$ la llamada función gamma.

Su función de distribución se puede calcular pero para nuestra comodidad está tabulada.

Ejemplo 115 *Las rentabilidades mensuales de cierto tipo de acciones son independientes unas de otras, y siguen una distribución normal con desviación típica 1.7. Se toma una muestra de 12 meses.*

a) Hallar la probabilidad de que la desviación típica muestral sea menor que 2.5.

b) Hallar la probabilidad de que la desviación típica muestral sea mayor que 1.

Solución Sea X= rentabilidad de las acciones. Sabemos que $\sigma_X^2 = (1{,}7)^2$ además como la distribución de la población es normal y $n = 12$ tenemos que $\frac{(n-1)\tilde{S}_X^2}{\sigma_X^2}$ sigue una distribución χ_{11}^2.

a) $P(\tilde{S}_X < 2{,}5) = P(\tilde{S}_X^2 < (2{,}5)^2) = P(\frac{(12-1)\tilde{S}_X^2}{(1{,}7)^2} < \frac{(12-1)(2{,}5)^2}{(1{,}7)^2}) = P(\chi_{11}^2 < 23{,}7889) \approx P(\chi_{11}^2 < 24{,}725) = 0{,}99.$

b) $P(\tilde{S}_X > 1) = P(\tilde{S}_X^2 > 1) = P(\frac{(12-1)\tilde{S}_X^2}{1{,}7^2} > \frac{(12-1)1}{1{,}7^2}) = P(\chi_{11}^2 > 3{,}80623) =\approx 1 - P(\chi_{11}^2 < 3{,}816) = 1 - 0{,}025 = 0{,}975$

Capítulo 7

Inferencia estadística: estimación de parámetros.

7.1. Introducción

En este tema estudiaremos como aproximar distintos parámetros poblacionales a partir de una m.a.s. formada por observaciones independientes de una población, en los que sigue cuando digamos m.a.s. entenderemos que es una muestra aleatoria formada por observaciones independientes.

Normalmente el parámetro (por ejemplo μ, $\sigma\ldots$) tendrá distribución conocida o la aproximaremos por el T.L.C.

7.2. Estimadores

Definición 116 ***Estadístico****: Sean $X_1, \ldots, X_n$ n v.a. iid que forman una m.a.s. de una población. Un estadístico es una función de una de una muestra.*

Podemos decir que un estadístico una variable aleatoria que es función de la muestra.

Definición 117 ***Estimador puntual****: Un* estimador puntual *de un parámetro θ es un estadístico que da como resultado un único valor del que se espera que se aproxime a θ.*

Una realización del estimador $T(x_1, \ldots, x_n) = \hat{\theta}$ *en una muestra se llama* estimación puntual de parámetro*.*

Ejemplo 118 *Dada una m.a.s. $X_1,\ldots,X_n$ y una realización de la misma $x_1,\ldots,x_n$ los principales estimadores de los parámetros poblacionales que hemos visto son:*

Parámetro Poblacional	*Estimador(θ)*	*Estimación($\hat{\theta}$)*
μ_X	$\overline{X}=\frac{\sum_{i=1}^n X_i}{n}$	$\overline{x}=\frac{\sum_{i=1}^n x_i}{n}$
σ_X	$\tilde{S}_X=\frac{\sum_{i=1}^n (X_i-\overline{X})^2}{n-1}$	$\tilde{s}_X=\frac{\sum_{i=1}^n (x_i-\overline{x})^2}{n-1}$
p	$\hat{p}_X=\frac{\sum_1^n X_i}{n}$	$\frac{\sum_1^n x_i}{n}$

Ejemplo 119 *Consideremos una m.a.s. X_1,X_2,X_3,X_4,X_5 del lanzamiento de un dado ($n=5$).*

Una realización de esta muestra es $x_1=2, x_2=3, x_3=3, x_4=5, x_5=6$.

Sabemos que, si el dado es perfecto, $\mu=3{,}5$; el estadístico de esta muestra es

$$\overline{X}=\frac{X_1+X_2+X_3+X_4+X_5}{5}$$

y una estimación es

$$\overline{x}=\frac{x_1+x_2+x_3+x_4+x_5}{5}=\frac{2+3+3+5+6}{5}=\frac{19}{5}=3{,}8$$

Si queremos estimar la proporción de veces que sale 3 es $p_3=\frac{1}{6}$ el estadístico es

$$\hat{p}_3=\frac{\textit{frec. de 3 en la muestra}}{5}$$

y una realización será $\frac{2}{5}$.

7.2.1. Estimadores insesgados

Vamos a ver en esta sección algunas propiedades de los estimadores. La más inmediata es pedirles que a medida que se aumente el tamaño muestral se aproximen más al verdadero valor del parámetro.

Definición 120 ***Estimador insesgado*** *Sea $\hat{\theta}$ un estimador de un parámetro poblacional θ. Diremos que $\hat{\theta}$ es insesgado si $E(\hat{\theta})=\theta$.*

Es este caso la estimación puntual se dice que es insesgada.

Ejemplo 121 *En el ejemplo anterior para cualquier muestra de tamaño n, $X_1, \ldots, X_n$, tenemos que $E(\overline{X}) = \mu_X$ por lo tanto $\overline{X}$ es un estimador insesgado de μ_X.*

Proposición 122 *Dada una m.a.s. La media, varianza y proporción muestrales son estimadores insesgados de sus correspondientes parámetros poblacionales.*

Definición 123 ***Sesgo***: *Sea $\hat{\theta}$ un estimador puntual de un parámetro poblacional θ, llamaremos sesgo de $\hat{\theta}$ a:*

$$Sesgo(\hat{\theta}) = E(\hat{\theta}) - \theta$$

Observación Evidentemente un estimador es insesgado si y sólo si tiene sesgo cero.

Una propiedad buena para un estimador es la carencia de sesgo. Pero podría suceder que tuviera una gran variabilidad, entonces aunque su valor central sea el verdadero valor del parámetro que se estima una realización del estadístico podría estar lejos del verdadero valor del parámetro. Parece pues interesante emplear aquellos estimadores que tengan varianza más pequeña.

Definición 124 ***Eficiencia***: *Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores de un parámetro poblacional θ obtenidos de la misma muestra.*

a) Diremos que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

b) Definimos la eficiencia relativa de $\hat{\theta}_2$ respecto de $\hat{\theta}_1$ como

$Eff.rel = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}$

de forma que si $Eff.rel < 1$ entonces $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$

Ejemplo 125 *Sea $x_1, \ldots, x_n$ la realización ordenada de menor a mayor de una muestra de tamaño n. Se define la mediana muestral como*

$$Me = \begin{cases} x_{\frac{n+1}{2}} & si\ n\ es\ impar \\ \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & si\ n\ es\ par \end{cases}$$

Como vimos en problemas la mediana es también un valor de tendencia central, pero ¿es un buen estimador de μ?

Se puede demostrar que si la población tiene distribución normal con media μ *y varianza* σ_X^2 *entonces* $E(Me) = \mu$ *y* $Var(Me) = \frac{\pi}{2}\frac{\sigma_X^2}{n} \approx \frac{1{,}57\sigma_X^2}{n}$

entonces $Eff.rel = \frac{Var(Me)}{Var(\overline{x})} = 1{,}57$

Luego si la muestra es de una población normal $\overline{X}$ *es más eficiente (un 57 % más de varianza) que la Mediana.*

Definición 126 ***Estimador más eficiente:***

Diremos que un estimador insesgado $\hat{\theta}$ *del parámetro* θ *es el estimador más eficiente si no existe ningún otro estimador insesgado que tenga menor varianza que él (también se le denomina estimador insesgado de varianza mínima).*

Ejemplo 127 [1]

- *Si la población es normal la media muestral es el estimador insesgado más eficiente de la media poblacional.*
- *Si la población es normal la varianza muestral es el estimador insesgado más eficiente de la varianza poblacional.*
- *Si la población es binomial la proporción muestral es el estimador insesgado más eficiente de la proporción poblacional.*

7.3. Métodos para calcular estimadores

Existen diversos métodos para el cálculo de estimadores:

- Método de los momentos. Momento central de orden r
  $m_r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^r}{n}$
- El de menor error cuadrático medio
  $E((\hat{\theta} - \theta)^2)$
- Convergencia en probabilidad
  $P(|\hat{\theta}_n - \theta| < \epsilon) \to 1$

[1] Más concretamente estos estimadores son del tipo UMVUE del acrónimo inglés “Uniformly Minimum Variance Unbiased Estimator”: Estimadores insesgados uniformemente de mínima varianza”.

- Estimadores máximo verosímiles.

En esta sección veremos sólo este último método.

Estimadores máximo verosímiles.

Definición 128 ***Función de verosimilitud***
Sea X una v.a. tal que su distribución (densidad o función de probabilidad) depende de un parámetro desconocido λ (En el caso discreto $P_X(x;\lambda)$ y en el continuo $f_X(x;\lambda)$). Sea $X_1,\ldots X_n$ una m.a.s. de X (es decir son n v.a. iid como X) y sean $x_1,x_2,\ldots,x_n$ una realización de la muestra. Entonces la función de verosimilitud de la muestra es:

a) En el caso discreto $L(\lambda)=P_X(x_1;\lambda)\cdots P_X(x_n;\lambda)$

b) En el caso continuo $L(\lambda)=f_X(x_1;\lambda)\cdots f_X(x_n;\lambda)$

Definición 129 *Dada una función de verosimilitud $L(\lambda)$ de una muestra, sea $\hat{\lambda}=g(x_1,\ldots,x_n)$ el punto donde se alcanza en máximo de $L(\lambda)$ para la realización de la muestra $x_1,\ldots,x_n$, es decir $L(\hat{\lambda})=\max\limits_{\lambda} L(\lambda)$. Entonces definimos el estimador de máxima verosimilitud de λ como el valor:*

$$\hat{\Lambda}=g(X_1,\ldots,X_n)$$

En ocasiones es conveniente trabajar con el logaritmo de la función de verosimilitud ya que el máximo de $\log(L(\lambda))$ y $L(\lambda)$ es el mismo y suele ser más fácil de maximizar.

Ejemplo 130 *Sea $X_1,\ldots X_n$ una muestra con observaciones independientes, de una población Bernouilli, por ejemplo se pregunta a 100 personas si votarán al partido P.X. en las próximas elecciones y se anota un 1 si lo votan y cero en cualquier otro caso. Sea $\hat{\theta}=T(X_1,\ldots,X_n)$ un estimador cualquiera. Sea p la proporción poblacional de personas que votarán a P.X. Entonces*

$$P(X_i=1)=p \text{ y } P(X_i=0)=1-p=q,$$

o lo que es lo mismo

$$P(X=x_i)=p^{x_i}q^{1-x_i} \text{ si } x_i=0,1$$

Como las observaciones son independientes. la función de verosimilitud es:

$$L(p) = P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 = x_1,\ldots,X_n = x_n) = P(X_1 = x_1)\cdots P(X_n = x_n) =$$

$$p^{x_1}q^{1-x_1}\cdots p^{x_n}q^{1-x_n} = p^{\sum_{i=1}^n x_i}q^{\sum_{i=1}^n(1-x_i)} = p^{\sum_{i=1}^n x_i}q^{n-\sum_{i=1}^n x_i}$$

entonces el valor de p que hace máxima esta probabilidad es el más verosímil o el de máxima verosimilitud de esta muestra.
El problema se reduce a estudiar qué valor de p maximiza

$$p^{\sum_{i=1}^n x_i}q^{n-\sum_{i=1}^n x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$$

Tomando logaritmos

$$\log\left(p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}\right) = \sum_{i=1}^n x_i \log p + (n - \sum_{i=1}^n x_i)\log(1-p)$$

Derivando respecto de p

$$(\sum_{i=1}^n x_i)\frac{1}{p} - (n - \sum_{i=1}^n x_i)\frac{1}{1-p} = 0$$

Despejando

$$(1-p)\sum_{i=1}^n x_i - p(n - \sum_{i=1}^n x_i) = 0$$

por lo tanto

$$p = \frac{\sum_{i=1}^n x_i}{n}$$

luego el estimador máximo verosímil de p es la proporción muestral, que es el que maximiza la función de verosimilitud $L(p)$.

De modo similar se puede definir los estimadores máximo verosímiles cuando el número de parámetros no conocidos de la distribución son más de uno.

7.4. Estimación por intervalos

Una estimación por intervalos de un parámetro poblacional es una regla para determinar un rango o un intervalo donde, con cierta probabilidad, se encuentre el verdadero valor del parámetro. La estimación correspondiente se llama estimación por intervalo. Más formalmente:

Definición 131 *Sea θ un parámetro, el intervalo (A, B) es un intervalo de confianza del $(1-\alpha)\dot{1}00\,\%$ para el parámetro θ si*

$$P(A < \theta < B) = 1 - \alpha.$$

El valor $1-\alpha$ recibe el nombre de nivel de confianza, α es la "cola"*de probabilidad sobrante que normalmente se reparte por igual ($\alpha/2$) a cada lado del intervalo. Es muy frecuente que el nivel de confianza se dé en tanto por ciento.*

En lo que sigue daremos distintas maneras de calcular o aproximar intervalos de confianza para distintos parámetros.

7.4.1. Intervalo de confianza para la media de una población normal: varianza poblacional conocida

Sea $X_1, \ldots, X_n$ una m.a.s. de una v.a. X con distribución normal y $Var(X) = \sigma^2$ conocida.

Encontremos un intervalo de confianza al *nivel de confianza* del $90\,\%$ para la media poblacional μ.

Sabemos por el tema anterior que bajo estas condiciones la variable $Z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ sigue una distribución normal estándar pues es una trasformación lineal de una combinación lineal de variables normales e independientes..

Ejemplo 132 *Comencemos calculando un intervalo centrado en 0 para esta Z que tenga probabilidad $0{,}975$.*

$$0{,}975 = P(-\delta < Z < \delta) = F_Z(\delta) - F_Z(-\delta) = 2F_Z(\delta) - 1$$

Entonces

$$F_Z(\delta) = \frac{1{,}975}{2} = 0{,}9875$$

mirando en las tablas de la distribución normal estándar, entonces $F_Z(2{,}24) = 0{,}9875$ *y por lo tanto* $\delta = 2{,}24$

Luego $P(-2{,}24 < Z < 2{,}24) = 0{,}975$

En resumen, hemos obtenido lo siguiente

$$0{,}975 = P(-2{,}24 < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < 2{,}24) =$$

$$P(\overline{X} - 2{,}24\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2{,}24\frac{\sigma}{\sqrt{n}})$$

Hemos encontrado un intervalo de confianza para μ*, y además la probabilidad de que* μ *se encuentre en el intervalo*

$\left(\overline{X} - 2{,}24\frac{\sigma}{\sqrt{n}}, \overline{X} + 2{,}24\frac{\sigma}{\sqrt{n}}\right)$

es 0.975; luego es un intervalo de confianza con nivel de confianza 97.5 %

Ejemplo 133 *Supongamos que tenemos una muestra con* $n = 16$ *de una v.a. normal de forma que* $\overline{x} = 20$*, y la desviación típica poblacional es conocida* $\sigma = 4$*. Entonces un intervalo de confianza al 97.5 % para* μ *será:*

$$\left(20 - \frac{(2{,}24)4}{\sqrt{16}}, 20 + \frac{(2{,}24)4}{\sqrt{16}}\right)$$

La probabilidad con que el verdadero valor del parámetro μ *se encuentra en el intervalo* $(17{,}76, 22{,}24)$ *es* $0{,}975$*, o lo que es lo mismo:*

$P(17{,}76 < \mu < 22{,}24) = 0{,}975$

Interpretación: *En el 97.5 % de la muestras de tamaño 16 el verdadero valor del parámetro* μ *se encontrará dentro del intervalo correspondiente.*

En general si tenemos una m.a.s. $X_1, \ldots, X_n$ de una población normal (representado por la v.a. X) con distribución normal de media μ y varianza conocida σ^2 el intervalo de confianza para μ al nivel de confianza $(1-\alpha)\cdot 100\,\%$ será:

$$1-\alpha = P(z_{\alpha/2} < Z < z_{1-\alpha/2}) = P(z_{\alpha/2} < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\alpha/2}) =$$

$$P(z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}) = P(\overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}})$$

Resumen: Intervalo de confianza para μ: σ^2 conocida.

Condiciones:

a) Población Normal con media μ y varianza σ^2 conocida

b) Muestra aleatoria de tamaño n

Entonces el intervalo de confianza del $100(1-\alpha)\,\%$ para μ es:

$$\left(\overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

donde $z_{\frac{\alpha}{2}}$ es el cuantil $\frac{\alpha}{2}$, es decir $P(Z \leq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$, cuando Z tiene distribución normal estándar, mientras que $z_{1-\frac{\alpha}{2}}$ es el cuantil $1-\frac{\alpha}{2}$, es decir $P(Z \leq z_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$, cuando Z tiene distribución normal estándar. Notemos que $z_{\frac{\alpha}{2}} = -z_{1-\frac{\alpha}{2}}$

Ejemplo 134 *Para discutir la conveniencia de aumentar sus instalaciones una empresa desea estimar la demanda que espera recibir. Para ello, selecciona al azar a diez de sus clientes, observando el número de unidades demandadas en el último año por éstos se distribuye de la forma siguiente:*

Núm. Unidades	*Núm. Clientes*	*Unidades× Clientes*
1.000	*1*	*1.000*
1.002	*2*	*2.004*
1.004	*1*	*1.004*
1.006	*2*	*2.012*
1.008	*1*	*1.008*
1.010	*2*	*2.020*
1.012	*1*	*1.012*
Total	*10*	*10.06*

Supongamos que la demanda sigue una distribución normal con varianza poblacional conocida $\sigma^2 = 16$ y que se espera que en el futuro siga comportándose como en el periodo anterior, calcular un intervalo de confianza al 90 % para la media de la demanda futura.

Solución: Tenemos las siguientes condiciones:

- Población de demandas normal varianza $\sigma^2 = 16$ conocida

- Muestra aleatoria de tamaño $n = 10$

Podemos entonces aplicar la formula anterior para $1-\alpha = 0{,}9$, de donde $\alpha = 0{,}1$, entonces $\frac{\alpha}{2} = 0{,}05$ y $1 - \frac{\alpha}{2} = 0{,}95$

Calculemos la media aritmética de las observaciones

$$\overline{x} = \frac{10{,}06}{10} = 1{,}006,$$

entonces el intervalo es

$$\left(1{,}006 + z_{0{,}05}\frac{4}{\sqrt{10}}, 1{,}006 + z_{1-0{,}05}\frac{4}{\sqrt{10}}\right).$$

Mirando en las tablas de la normal $P(Z \leq 1{,}65) = 0{,}9505 \approx 0{,}95$ entonces $z_{0{,}95} = 1{,}65$, y $z_{0{,}05} = -1{,}65$ sustituyendo tenemos que

$z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 1{,}65\frac{4}{\sqrt{10}} = 2{,}0871$ $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = -1{,}65\frac{4}{\sqrt{10}} = 2{,}0871$

por lo que el intervalo de confianza del 90 % para la media de la demanda es :

$$(1{,}006 - 2{,}0871, 1{,}006 + 2{,}0871) = (-1{,}081, 3{,}093)$$

Lo que quiere decir que en el 90 % de la ocasiones en que tomemos una muestra de tamaño 10 la demanda media está comprendida entre $-1{,}081$ y 3,093. Como se ve hay un abuso, en este caso, de la suposición de normalidad en la distribución de la demanda.

Amplitud del intervalo de confianza

Como de todos es conocido la amplitud (longitud) de un intervalo es la diferencia entre sus extremos superior e inferior. En el caso anterior la amplitud A será

$A = \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} - \left(\overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

El *error* máximo, al nivel $(1-\alpha)$, que cometemos al estimar μ por $\overline{X}$ será la mitad de la amplitud del intervalo de confianza $z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

Si queremos calcular el tamaño n de la muestra para asegurarnos que el intervalo de confianza para μ al nivel $(1-\alpha)$ tiene amplitud prefijada A (o un error $\frac{A}{2}$) se puede despejar así:

$n = \left(2z_{1-\frac{\alpha}{2}}\frac{\sigma}{A}\right)^2$

Observaciones:

- El intervalo está centrado en $\overline{X}$.
- Para n y $1-\alpha$ fijos si la varianza poblacional aumenta entonces A aumenta.
- Para una varianza poblacional conocida y $1-\alpha$ fijos si n aumenta entonces A disminuye.
- Para una varianza poblacional conocida y n fijos si $1-\alpha$ aumenta entonces A aumenta.

7.4.2. Intervalo de confianza para la media poblacional: tamaños muestrales grandes

Condiciones:

- Población con media μ y varianza σ^2 conocida o si no se estima por $\tilde{S}^2$
- Muestra aleatoria de tamaño n grande (criterio $n \geq 30$)

Entonces el intervalo de confianza del $100(1-\alpha)\,\%$ para μ es:

$$\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\tilde{S}}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}}\frac{\tilde{S}}{\sqrt{n}}\right)$$

En caso de que σ sea conocida pondremos σ en lugar de $\tilde{S}$

Ejemplo 135 *Se tomó una muestra de 147 expertos en investigación de mercados y se les pidió que calificasen en una escala de 1 (totalmente en desacuerdo) a 10 (totalmente de acuerdo) la siguiente afirmación: "A veces utilizo técnicas de investigación que garantizan la obtención de los resultados que mi cliente o jefe desea". La calificación media de la muestra fue* 6,06 *y la desviación típica muestral fue 1.43. Se pide calcular un intervalo de confianza al 90 % para la media de las puntuaciones.*

Solución: El enunciado no nos asegura que la población sea normal pero como el tamaño de la población es grande podemos aplicar el resultado anterior.

Tenemos $n = 147$, $\tilde{S} = 1{,}43$, $1-\alpha = 0{,}9$ entonces $\frac{\alpha}{2} = 0{,}05$ y por lo tanto $z_{1-0{,}05} \approx 1{,}65$

El intervalo para la media poblacional de las puntuaciones al nivel de confianza del 90 % es
$\left(6{,}06-1{,}65\frac{1{,}43}{\sqrt{147}}, 6{,}06+1{,}65\frac{1{,}43}{\sqrt{147}}\right)=$
$(5{,}8654, 6{,}2546)$

Distribución t de Student

Si queremos calcular un intervalo de confianza para μ en una población normal con varianza poblacional desconocida necesitamos una nueva distribución: la t de Student.

Dada una muestra de n observaciones con media muestral $\overline{X}$ y desviación típica muestral $\tilde{S}_X$ procedente de una población normal con media μ la variable aleatoria:

$$t=\frac{\overline{X}-\mu}{\frac{\tilde{S}_X}{\sqrt{n}}}$$

sigue una distribución t de Student con $n-1$ grados de libertad.

Proposición 136 *La distribución t de Student es similar a la normal si el número de grados de libertad es grande. Su función de densidad es simétrica respecto al origen como la de la normal estándar. Es decir si t_ν es una v.a. que sigue la distribución t de Student con ν g.l. entonces:*

$$P(t_\nu \leq -t)=1-P(t_\nu \leq t)$$

Notación: Sea t_ν una v.a. que sigue una distribución t de Student con ν g.l. Denotaremos por $t_{\nu,\alpha}$ al valor para el que se verifica que:

$$P(t_\nu \leq t_{\nu,\alpha})=\alpha.$$

Luego $t_{\nu,\alpha}$ es el α cuantil de una t de Student con ν g.l. y $t_{\nu,\alpha}=-t_{\nu,1-\alpha}$.

7.4.3. Intervalo de confianza para la media de una población normal: varianza poblacional desconocida

Condiciones:

- Muestra aleatoria de n observaciones independientes
- Población normal varianza desconocida

Entonces si $\overline{X}$ y $\tilde{S}_X$ son respectivamente la media y la desviación típica muestrales un intervalo de confianza al nivel $(1-\alpha)100\,\%$ para la media de la población μ es:

$$\left(\overline{X}+t_{n-1,\frac{\alpha}{2}}\frac{\tilde{S}_X}{\sqrt{n}},\overline{X}+t_{n-1,1-\frac{\alpha}{2}}\frac{\tilde{S}_X}{\sqrt{n}}\right)$$

Siendo $t_{n-1,\frac{\alpha}{2}}$ y $t_{n-1,\frac{\alpha}{2}}$ los cuantiles de una v.a. t_{n-1} con distribución t de Student con n-1 g.l., respectivamente.

Ejercicio Demostrar que la probabilidad con que μ se encuentra en el intervalo anterior es $1-\alpha$

Ejemplo 137 *Un fabricante de cartuchos de tinta para impresoras afirma en su publicidad que sus cartuchos imprimirán un promedio de 500 páginas*; donde el asterisco remite a una nota a pie de página donde afirma que: “ `Datos técnicos: Muestra mensual de tamaño` $n=25$ `población supuesta normal nivel de confianza del 90%`”.*

Una organización de consumidores desea comprobar estas afirmaciones y toma también una muestra al azar de tamaño $n=25$ obteniendo como media $\overline{x}=518$ páginas y una desviación estándar $\tilde{S}_X=40$. Comprobar que con esta muestra la media poblacional que afirma el fabricante cae dentro del intervalo de confianza del 90 %

Solución: El problema se reduce a calcular, bajo las condiciones que afirma el fabricante el intervalo de confianza para μ con $\alpha=0{,}1$.

Mirando en las tablas de la t de Student para $n-1=24$ g.l. tenemos que $t_{n-1},1-\frac{\alpha}{2}=t_{24,1-0{,}05}=1{,}71$

El intervalo para la media al 90 % es

$$\left(518-1{,}71\frac{40}{\sqrt{25}},518+1{,}71\frac{40}{\sqrt{25}}\right)=(504{,}32,531{,}68)\,.$$

Es este caso la afirmación del fabricante queda contradicha por la muestra pues 500 cae fuera del intervalo. En cualquier caso se equivoca a favor del consumidor.

7.4.4. Intervalos de confianza para una proporción

El procedimiento es similar al caso de las medias. Comencemos con un ejemplo.

Ejemplo 138 *En una muestra aleatoria de 500 familias que poseen televisores en una ciudad se encontró que 340 se habían suscrito al canal TEVE. Encontrar un intervalo de confianza del 95 % para la proporción actual de familias de esta ciudad que están suscritas a TEVE.*

Tenemos una población binomial donde los éxitos son las familias que tienen contrato con TEVE. Sea X el número de familias contratadas con TEVE entre una muestra aleatoria de tamaño n. Entonces X sigue una distribución binomial con n repeticiones y probabilidad de éxito p (proporción poblacional de familias contratadas a TEVE). Si llamamos $\hat{p}_X = \frac{X}{n}$ a la proporción muestral, sabemos que $Z = \frac{\hat{p}_X - p}{\sqrt{\frac{p(1-p)}{n}}}$ sigue aproximadamente una distribución normal estándar.

Pero como es evidente no conocemos p así que no tenemos más remedio que aproximar el denominador

$$\sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$$

Si la muestra es grande $Z = \frac{\hat{p}_X - p}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}}$ seguirá siendo aproximadamente normal estándar.

Intervalos de confianza para la proporción poblacional:(muestras grandes)

Condiciones:

- Una muestra aleatoria de tamaño n grande.
- Población Bernouilli con proporción de éxitos p (desconocida)

Bajo estas condiciones y si $\hat{p}_X$ es la proporción de éxitos en la muestra, un intervalo de confianza al nivel $(1-\alpha)100\,\%$ es

$$\left(\hat{p}_X - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}, \hat{p}_X + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}\right)$$

Criterio: los intervalos de confianza anteriores son fiables si $n \geq 40$.
Observaciones

- El intervalo de confianza anterior está centrado en la proporción muestral.
- Cuando n crece se reduce la amplitud del intervalo de confianza.
- La amplitud del intervalo de confianza es $A = 2z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$
- De la fórmula anterior no podemos determinar el tamaño de la muestral sin conocer $\hat{p}_X$ así que nos podremos en el caso peor:

  El máximo de
  $\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n}}$
  se alcanza en $\hat{p}_X = 0{,}5$ y en este caso
  $\sqrt{\frac{0{,}5(1-0{,}5)}{n}}$ por lo tanto en el peor de los casos[2]
  $n = \frac{0{,}25 z^2_{1-\frac{\alpha}{2}}}{(A/2)^2}$.

7.4.5. Intervalo de confianza para la varianza de una población normal

Recordemos que si tenemos una población normal con varianza σ^2 y una muestra aleatoria de tamaño n de esta población con varianza muestral $\tilde{S}^2_X$ entonces el estadístico

$$\chi^2_{n-1} = \frac{(n-1)S^2_X}{\sigma^2}$$

sigue una distribución χ^2 con $n-1$ g.l.

Notación Si χ^2_ν es una v.a. que tiene distribución χ^2 con ν g.l. denotaremos por $\chi^2_{\nu,\alpha}$ al valor que verifica:

$$P(\chi^2_\nu \leq \chi^2_{\nu,\alpha}) = \alpha$$

[2]Por esto en las especificaciones o detalles técnicos de las encuestas se suele leer, por ejemplo: "Universo población Balear mayor de 18 años. Encuesta telefónica, selección aleatoria, de tamaño mil, error en las proporciones $\pm 3\,\%$ con una confianza del $95\,\%$ supuesto que $p = q = \frac{1}{2}$"

es decir el cuantil $\frac{\alpha}{2}$ de una v.a. con distribución χ^2_ν. Estos valores están tabulados para distintos g.l. en la tabla de la distribución χ^2.

Ejemplo 139 *Sea χ^2_{10} una v.a. que tiene distribución χ^2 con 10 g.l. Entonces $\chi^2_{10,0{,}995} = 25{,}19$ y $\chi^2_{10,0{,}005} = 2{,}16$, es decir*

$$P(\chi^2_{10} \leq 25{,}19) = 0{,}995 \text{ y } P(\chi^2_{10} \leq 2{,}16) = 0{,}005$$

Además tendremos que

$$P(2{,}16 \leq \chi^2_{10} \leq 25{,}19) = P(\chi^2_{10} \leq 25{,}19) - P(\chi^2_{10} \leq 2{,}16) = (1-0{,}005)-(1-0{,}995) = 0{,}995-0{,}005 =$$

En general dado α entre 0 y 1 tendremos que

$$1-\alpha = P(\chi^2_{\nu,\frac{\alpha}{2}} \leq \chi^2_\nu \leq \chi^2_{\nu,1-\frac{\alpha}{2}})$$

Si tenemos una muestra de tamaño n de una población normal con desviación típica muestral $\tilde{S}^2_X$, dado un nivel de confianza $1-\alpha$ tendremos que $\chi^2_{n-1} = \frac{(n-1)\tilde{S}^2_X}{\sigma^2}$ y entonces:

$$1-\alpha = P(\chi^2_{n-1,\frac{\alpha}{2}} \leq \chi^2_{n-1} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}) =$$

$$P(\chi^2_{n-1,\frac{\alpha}{2}} \leq \frac{(n-1)S^2_X}{\sigma^2} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}) = P(\frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,\frac{\alpha}{2}}})$$

Luego, bajo estas condiciones, un intervalo de confianza para la varianza poblacional del $(1-\alpha)100\,\%$ es

$$\left(\frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\tilde{S}^2_X}{\chi^2_{n-1,\frac{\alpha}{2}}}\right).$$

Intervalo de confianza para la varianza de una población normal

Condiciones

- Población normal
- Muestra aleatoria de tamaño n con varianza muestral S^2_X

Entonces un intervalo de confianza del $(1-\alpha)100\,\%$ es

$$\left(\frac{(n-1)\tilde{S}_X^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)\tilde{S}_X^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right)$$

Donde $\chi^2_{n-1,\frac{\alpha}{2}}$ es el valor que verifica

$$P(\chi^2_{n-1} < \chi^2_{n-1,\frac{\alpha}{2}}) = \frac{\alpha}{2}$$

y

$$\chi^2_{n-1,1-\frac{\alpha}{2}}$$

es el valor tal que

$$P(\chi^2_{n-1} \leq \chi^2_{n-1,1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$$

donde χ^2_{n-1} es una v.a. que sigue una distribución χ^2 con $n-1$ g.l.

Observación El intervalo de confianza para σ^2 no está centrado en $\tilde{S}_X^2$.

Ejemplo 140 *Una cadena de hoteles tiene una* Línea 900 *para recibir reservas telefónicas. Un índice de la calidad del servicio es el tiempo de espera, el tiempo que transcurre desde que el teléfono suena por primera vez hasta que el operador responde. El estándar de la cadena es que el tiempo promedio de espera no debe ser superior a 30 segundos además se supone que la distribución del tiempo de espera será aproximadamente normal. La cadena tiene inspectores que visitan los distintos hoteles y verifican todos los aspectos del servicio. Estas personas realizan cada semana 30 llamadas para hacer reservas y anotan, entre otros indicadores el tiempo de espera en cad una de ellas. En una semana los tiempos de espera en segundos son:*

12, 13, 13, 14, 14, 14, 15, 15, 16, 17, 17, 18, 18, 19, 19, 25, 25, 26, 27, 30, 33, 34, 35, 40, 40, 51, 51, 58, 59, 83

Calcular un intervalo de confianza para la varianza y la desviación poblacionales al nivel $95\,\%$.

Solución: *Sea X el tiempo de espera. Haciendo los cálculos tenemos que (redondeando al segundo decimal):*

$\overline{X} = 28{,}37$ *y* $\tilde{s}_X = 17{,}37$

Como $1-\alpha = 0{,}95$ *tenemos que* $\frac{\alpha}{2} = 0{,}025$*, entonces mirando en las tablas de la* χ^2 *(y redondeando también al segundo decimal)*

$$\chi^2_{n-1,\frac{\alpha}{2}} = \chi^2_{29,0{,}975} = 45{,}72 \text{ y } \chi_{n-1}, 1 - \frac{\alpha}{2}^2 = \chi^2_{29,0{,}025} = 16{,}05.$$

Por lo tanto un intervalo de confianza del 95 % para σ^2 *es*

$$\left(\frac{(30-1)(17{,}37)^2}{45{,}72}, \frac{(30-1)(17{,}37)^2}{16{,}05} \right) = (191{,}38, 545{,}16)$$

Es decir $P(191{,}28 \leq \sigma^2 \leq 545{,}16) = 0{,}95$ *y operando tenemos que*

$$P(\sqrt{191{,}28} \leq \sigma \leq \sqrt{545{,}16}) = 0{,}95,$$

luego un intervalo de confianza del 95 % para σ *es*

$$(13{,}83, 23{,}35)\,.$$

Capítulo 8

Inferencia estadística: contraste de hipótesis

En los temas anteriores hemos visto como puede estimarse un parámetro a partir de los datos contenidos en una muestra. Puede encontrarse una estimación puntual o bien una estimación por intervalo. Sin embargo muchos problemas de economía y administración requieren tomar una *decisión* es decir se debe aceptar o rechazar alguna afirmación sobre, por ejemplo, el valor de un parámetro.

Esta afirmación recibe el nombre de *hipótesis* y el método estadístico de toma de decisión sobre la hipótesis recibe el nombre de prueba (o contraste) de hipótesis.

éste es uno de los aspectos más útiles de la inferencia estadística puesto que muchos problemas de toma de decisiones pueden plantearse en términos de contraste de hipótesis.

Ejemplo 141 *a) Un fabricante de bombillas afirma que la duración media de sus productos es de 1000 horas. Para verificar esta hipótesis se toma una muestra aleatoria y se infiere el resultado a la población general.*

b) Una distribuidora recibe una partida de productos. El encargado tiene orden de aceptar los envíos que contengan menos de un 5 % de piezas defectuosas. La decisión del encargado se podría basar en una muestra aleatoria de la partida.

c) Estamos interesados en comparar dos métodos de enseñanza de idiomas. Para ello se toman dos muestras de alumnos que han seguido cada método *de enseñanza y se les evalúa con el mismo examen. Tenemos que decidir cuál de los dos métodos es mejor a la vista de estas muestras.*

d) Un experto de una determinada compañía de tarjetas de crédito desea saber si las nuevas comisiones serán aceptadas en igual proporción por los pequeños y grandes comercios. Para ello realiza una encuesta de opinión a pequeños comerciantes y a grandes superficies y de ella tiene que inferir la conclusión.

Definición 142 *Una hipótesis estadística es una afirmación que se realiza sobre los parámetros de una o más poblaciones*

*Las hipótesis estadística se contrastan una contra otra. Habitualmente las denominaremos **Hipótesis nula** H_0 e **hipótesis alternativa** H_1.*

Ejemplo 143 *Un fabricante de sobrasada asegura en su etiqueta que sus piezas pesan 200 Kg. Un fabricante de la competencia sospecha que el peso es inferior al que figura en la etiqueta para ello toma una muestra aleatoria de sobrasadas y las pesa. El contraste de interés será:*

$$\left\{ \begin{array}{l} H_0 : \mu = 200 \\ H_1 : \mu < 200 \end{array} \right.$$

Donde μ será el contenido medio en gramos de la población de sobrasadas.

Al competidor sólo le interesa contrastar $\mu = 200$ contra $\mu < 200$, pues sólo quiere decidir si el peso es inferior al declarado.

Pero si es el encargado del control de la producción le interesará contrastar

$$\left\{ \begin{array}{l} H_0 : \mu = 200 \\ H_1 : \mu \neq 200 \end{array} \right.$$

Pues no puede engañar al consumidor pero tampoco quiere darle más peso gratis.

8.1. Tipos de hipótesis

- $H : \theta = \theta_0$ hipótesis simple (en caso contrario compuesta).
- $H : \theta > \theta_0$ hipótesis unilateral.
- $H : \theta \neq \theta_0$ hipótesis bilateral.

Resumiendo: Un contraste de hipótesis consiste en plantear una hipótesis nula y una alternativa

$$\begin{cases} H_0 : \text{hipótesis nula} \\ H_1 : \text{hipótesis alternativa} \end{cases}$$

y generar un regla de decisión para aceptar la hipótesis nula o rechazarla en favor de la alternativa a partir de la información contenida en una muestra.

Ejemplo 144 *Supongamos que queremos decidir si una moneda está bien balanceada. Para ello lanzamos la moneda 100 veces obteniéndose X caras.*

Sea p la probabilidad de cara en esta moneda, queremos contrastar:

$$\begin{cases} H_0 : p = 0{,}5 \\ H_1 : p \neq 0{,}5 \end{cases}$$

Una regla podría ser aceptar H_0 contra H_1 si X no es muy distinto de 50 por ejemplo si $48 \leq X \leq 52$.

En lo que sigue definiremos los elementos necesarios para estudiar que reglas (regiones) de rechazo son las más adecuadas para distintos tipos de contrastes.

8.2. Tipos de Error en un contraste

Cuando realizamos un contraste de hipótesis pueden darse las situaciones que detallamos en la tabla siguiente:

Decisión	Estados de la naturaleza	
	H_0 cierta	H_0 falsa
Aceptar H_0	Dec. correcta Prob=$1-\alpha$	Error tipo II Prob=β
Rechazar H_0	Error tipo I Prob=α	Dec. correcta Prob $=1-\beta$

- La probabilidad de Error Tipo I es

$$P(\text{Error Tipo I}) = P(\text{Rechazar} H_0 / H_0 \text{ cierta}) = \alpha$$

y recibe el nombre de *nivel de significación* del contraste.

- La probabilidad de Error Tipo II es

$$P(\text{Error Tipo II}) = P(\text{Aceptar}H_0/H_0\text{falsa}) = \beta$$

  el valor $1-\beta$ recibe el nombre de *potencia* del contraste.

En ocasiones daremos los niveles de significación y la potencia en tantos por cien, así un nivel de significación del 5 % implica que $\alpha = 0{,}05$ Lo ideal es encontrar aquella regla de rechazo de H_0 que tenga menor probabilidad de Error Tipo I α y también tenga menor probabilidad de Error Tipo II β o lo que es lo mismo mayor potencia $1-\beta$.

Lo que sucede es que si modificamos la regla de rechazo para que disminuya α entonces aumentamos β. Buscaremos reglas de decisión que para un α fijo nos den un β lo más pequeño posible. Lo que se hace normalmente es fijar α y esto nos da la región crítica y luego, si es posible, controlar el tamaño de la muestra n para obtener la mayor potencia y por lo tanto el menor Error de Tipo II al menor coste.

En resumen: Si el investigador fija un nivel de significación obtiene una regla de decisión que fija un Error de Tipo II.

Terminología

Resumamos los conceptos vistos hasta ahora:

- Hipótesis nula H_0: Es la hipótesis que se desea aceptar si no hay prueba de que es falsa.

- Hipótesis Alternativa H_1: Es la hipótesis frente a la que se contrasta la hipótesis nula y que se acepta si se rechaza la nula.

- Hipótesis simple: Es la que especifica un sólo valor para el parámetro a contrastar.

- Hipótesis compuesta: Es la que especifica un rango de valores para el parámetro a contrastar.

- Alternativa unilateral: Es una H_1 compuesta formada por un semiintervalo es decir $\theta > \theta_0$ o $\theta < \theta_0$.

  Alternativa bilateral: Es aquella H_1 compuesta que es el complementario de una H_0 simple.

- Decisión de un contraste de hipótesis: puede ser aceptar o rechazar la hipótesis nula lo que se hace en función de una regla de decisión que recoge la información de una muestra.

- Error de Tipo I: Se comete cuando se rechaza H_0 siendo cierta. Su probabilidad se denota por α.

  Error Tipo II: Se comete cuando se acepta una H_0 falsa. Su probabilidad se denota por β.

- Nivel de significación α: Es la probabilidad de cometer un Error Tipo I, es decir , $\alpha = P(\text{Error Tipo I}) = P(\text{Rechazar} H_0/H_0 \text{ cierta})$

- Potencia de un contraste: Es la probabilidad de rechazar una hipótesis nula que es falsa. Entonces la potencia es

$$P(\text{Rechazar} H_0/H_0 \text{ es falsa}) = 1-P(\text{Aceptar} H_0/H_0 \text{ es falsa}) = 1-P(\text{Error Tipo II}) = 1-\beta$$

8.3. ¿Inocente o culpable?

La decisión de aceptar o rechazar una hipótesis nula se asemeja al concepto de declarar a un acusado en juicio inocente o culpable.

El acusado es la hipótesis nula H_0, las pruebas son los elementos de la muestra. Si el jurado no encuentra suficientes las pruebas tiene que declarar inocente al acusado (Aceptar H_0), sólo en el caso en que las pruebas sean lo suficientemente incriminatorias condenará al culpable y se aceptará la hipótesis nula. El jurado siempre corre el riesgo de declarar inocente a un culpable cometiendo un Error de Tipo I o condenar a un inocente cometiendo un Error de Tipo II.

Desde este punto de vista es más conveniente controlar el Error de Tipo I pues es mejor declarar inocente a un culpable que culpable a un inocente.

Ejercicio Construir de forma similar al ejemplo anterior una similitud entre las pruebas de hipótesis y un combate de boxeo por un título mundial entre un aspirante y el actual poseedor del título. En caso de empate a puntos el título queda en poder del campeón.

8.4. Ejemplo de un contraste de hipótesis para la media de una distribución normal: varianza poblacional conocida

En lo que sigue, comenzando por esta sección, daremos distintos contrastes de hipótesis para la media de una población.

Para contrastar las hipótesis dispondremos de una m.a.s. de n observaciones $X_1, \ldots, X_n$. En este caso procedentes de una distribución normal con media μ y varianza σ^2. Supondremos que la varianza es conocida.

Consideremos el contraste:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Como es natural la regla de rechazo se basará en observar si la media aritmética $\overline{X}$ es suficientemente mayor que valor μ_0. Si es así rechazaremos la hipótesis nula.

Como sabemos que bajo estas condiciones y si H_0 (es decir $\mu = \mu_0$) es cierta

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

sigue una distribución normal estándar.

Rechazar H_0 si $\overline{X}$ es muy alta es equivalente a obtener un valor alto del estadístico de contraste

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}.$$

Entonces la regla consiste en rechazar H_0 si Z es mayor que un cierto umbral.

Sabemos que $\alpha = P(\text{Rechazar} H_0 / H_0 \text{ cierta}) = P(Z > \text{umbral}/\mu = \mu_0) = P(Z > z_{1-\alpha})$ cuando Z es una normal estándar.

Luego para que el nivel de significación del contraste sea α la regla de rechazo viene dada por la región crítica

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{1-\alpha}$$

8.5. Terminología:

- Estadístico de contraste: es el que nos permite definir una regla de rechazo de H_0.
- Región crítica o región de rechazo: es aquel rango de valores tales que si el estadístico de contraste está entre ellos se rechaza H_0.
- Región de aceptación: Es el complementario de la región crítica.

8.6. Ejemplo tabla de un contraste para la media poblacional de una población normal con varianza poblacional conocida

Condiciones:

- Población normal de media μ y varianza σ^2 conocida

Un contraste al nivel de significación α para las hipótesis:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Tiene por regla de decisión:
Rechazar H_0 si

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{1-\alpha}.$$

Definición 145 *Llamaremos valor crítico o p-valor al menor nivel de significación para el que se rechaza la hipótesis nula.*

8.7. Método de los seis pasos

Para clarificar ideas seguiremos seis pasos para resolver los contrastes de hipótesis sobre un parámetro θ que veremos en este tema:

1) Establecer la hipótesis nula H_0, por ejemplo $\theta = \theta_0$

2) Establecer la hipótesis alternativa H_1 que podrá ser $\theta > \theta_0$, $\theta < \theta_0$ o $\theta \neq \theta_0$.

3) Seleccionar un nivel de significación α

4) Seleccionar el estadístico apropiado para la prueba y establecer la región crítica o región de rechazo. Si la decisión se basa en un p-valor, como veremos, no es necesario calcular la región crítica.

5) Calcular el valor del estadístico de contraste a partir de los datos muestrales.

6) Decidir: rechazar H_0 si el valor del estadístico de contraste cae dentro de la región crítica o si el p-valor es menor o igual que el nivel de significación prefijado α; en caso contrario no rechazar H_0.

Ejemplo 146 *Una muestra aleatoria de 100 muertes registradas en un cierto país durante 1998 dio una vida promedio de 71.8 años. Suponiendo que la desviación típica poblacional es de 8.9 años, decidir si la vida promedio es, hoy en día, mayor que 70 años. Utilizar un nivel de significación del 0.05 y suponer que la duración de la vida se distribuye aproximadamente normal.*

Solución: Sigamos los seis pasos:

1) $H_0 : \mu = 70$ años.($\mu_0 = 70$) 2) $H_1 : \mu > 70$ años. 3) $\alpha = 0{,}05$ 4) Bajo estas condiciones, población normal, $\sigma^2 = 8{,}9^2$ conocida y una muestra de tamaño $n = 100$ la región crítica para estas hipótesis es:

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{1-0{,}05} = 1{,}64$$

5) Cálculo del estadístico de contraste: $\overline{x} = 71{,}8$ años, $\sigma = 8{,}9$ años. Entonces el estadístico de contraste es:

$$Z = \frac{71{,}8 - 70}{\frac{8{,}9}{\sqrt{100}}} = 2{,}02$$

6) Decisión: Como $Z = 2{,}02 > 1{,}64$ resulta que el valor del estadístico de contraste cae dentro de la región crítica, luego a partir de esta muestra no podemos aceptar (H_0) que la vida promedio es de 70 años contra que es mayor de 70 años (H_1) al nivel de significación $\alpha = 0{,}05$

Ejemplo 147 *En el ejemplo anterior calcular el p-valor e interpretarlo*

Para calcular el p valor tenemos que buscar aquel nivel de significación α más pequeño para el que se rechaza la hipótesis nula.

Para ello igualamos el valor del estadístico de contraste $Z = 2{,}02$ al umbral de la región de rechazo, es decir:

$$2{,}02 = z_{1-\alpha}$$

mirando en las tablas de la normal estándar obtenemos que $1-\alpha = 0{,}9783$ luego $\alpha = 0{,}0217$.

La interpretación de este valor es la siguiente:

Rechazaremos (H_0) que la vida promedio es de 70 años contra que es mayor de 70 años (H_1) para todos los niveles de significación $\alpha > 0{,}0217$.

Es decir la evidencia es más grande que el nivel de significación del ejemplo anterior.

Propiedad

Si en el anterior contraste utilizamos las hipótesis:

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Todavía tendríamos más evidencia para rechazar la hipótesis nula.

Entonces la región de contraste es la misma que en el caso $H_0 : \mu = \mu_0$

Damos a continuación un resumen sobre las regiones críticas para el contraste de una media de una población normal con varianza conocida para las distintas alternativas unilaterales o bilaterales:

8.8. Reglas de decisión para contraste de la media de una población normal: varianza poblacional conocida

Condiciones:

- Una muestra aleatoria simple de una población normal de media μ y varianza σ^2 conocida.

Un contraste al nivel de significación α para las hipótesis:

a)

$$\begin{cases} H_0 : \mu = \mu_0 & (\text{ o } H_0 : \mu \leq \mu_0) \\ H_1 : \mu > \mu_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{1-\alpha}.$$

b)

$$\begin{cases} H_0 : \mu = \mu_0 & (\text{ o } H_0 : \mu \geq \mu_0) \\ H_1 : \mu < \mu_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < z_{\alpha}.$$

c)

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > z_{1-\frac{\alpha}{2}} \text{ o } Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}.$$

8.9. Contraste para la media: tamaños muestrales grandes

Si no conocemos la distribución de la población o bien no es normal pero tenemos un tamaño muestral grande, podemos prescindir de la condición de normalidad de la población y aplicar las mismas reglas de rechazo para la hipótesis nula que en el caso anterior.

Además si σ^2 es desconocida se puede sustituir por la desviación típica muestral $\tilde{S}^2$. Criterio: si $n \geq 30$ podemos aplicar esta aproximación.

Ejemplo 148 *Una organización ecologista ha publicado cifras sobre el consumo anual en Kw/h de varios aparatos del hogar. Se afirma que la aspiradora consume una media de 46 Kw/h al año. Si una muestra aleatoria de 36 hogares incluidos en un estudio planeado por la asociación nacional de fabricantes de aspiradoras (ANFA) da una media muestral de 42 Kw/h al año y una desviación típica muestral de 11.9. ¿Podemos afirmar, con un nivel de significación del 5 %, a la vista de estos datos que el consumo medio es inferior a 46 kw/h año?*

Solución Nadie nos asegura que la población es normal, además σ es desconocida pero como $n = 36$ podemos utilizar las regiones de rechazo anteriores sustituyendo σ por $\tilde{s}$.

Sigamos los seis pasos:

1) $H_0 : \mu = 46$ Kw/h. ($\mu_0 = 46$)

2) $H_1 : \mu < 46$ Kw/h.

3) $\alpha = 0{,}05$

4) Bajo estas condiciones, como $n \geq 30$, σ es desconocida pero la aproximamos por $\sigma \approx \tilde{s} = 11{,}9$. Entonces la región crítica para estas hipótesis es:

$$Z = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}} < z_{0,05} = -1{,}64$$

5) Cálculo del estadístico de contraste: $\overline{x} = 42$ años, $\tilde{s} = 11{,}9$ años entonces:

$$Z = \frac{42 - 46}{\frac{11{,}9}{\sqrt{36}}} = -2{,}02$$

6) Decisión: Como $Z = -2{,}02 < -1{,}64$ resulta que el valor del estadístico de contraste cae en la región crítica, luego a partir de esta muestra rechazamos (H_0) que el consumo promedio es de 46 Kw/h contra que es menor de 46 Kw/h (H_1) al nivel de significación $\alpha = 0{,}05$.

Luego el consumo medio anual de las aspiradoras no es significativamente inferior a 46 Kw/h, por lo que podríamos rebatir los resultados de la asociación ecologista con esta muestra a este nivel de significación.

8.9.1. Reglas de decisión para el contraste de una media: Tamaños muestrales grandes

Son las misma que las de la sección 8.8 cambiando σ por $\tilde{S}$.

8.10. Contrastes para la media de una población normal: varianza poblacional desconocida

En el caso que tengamos una población normal, desconozcamos la varianza y no tengamos un tamaño muestral n grande utilizaremos (al igual que en el Tema anterior) el estadístico

$$t_{n-1} = \frac{\overline{X} - \mu_0}{\frac{\tilde{S}}{\sqrt{n}}}$$

que como sabemos sigue una distribución t de Student con $n-1$ g.l.

Las regiones críticas serán similares a las de muestras grandes pero sustituyendo los valores de la normal estándar por los correspondientes valores de la t_{n-1}.

8.10.1. Reglas de decisión para el contraste de una media de una distribución normal: varianza poblacional desconocida

Condiciones:

- Muestra aleatoria de n observaciones población normal con media μ y varianza desconocida.

Entonces una contraste al nivel de significación α para las hipótesis:

a)

$$\begin{cases} H_0 : \mu = \mu_0 \text{ o } H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$t_{n-1} = \frac{\overline{x} - \mu_0}{\frac{\tilde{S}}{\sqrt{n}}} > t_{n-1,1-\alpha}.$$

b)

$$\begin{cases} H_0 : \mu = \mu_0 \text{ o } H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$t_{n-1} = \frac{\overline{x} - \mu_0}{\frac{\tilde{S}}{\sqrt{n}}} < t_{n-1,\alpha}.$$

c)

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$t_{n-1} = \frac{\overline{x} - \mu_0}{\frac{\tilde{S}}{\sqrt{n}}} > t_{n-1,1-\frac{\alpha}{2}} \text{ o } t_{n-1} = \frac{\overline{x} - \mu_0}{\frac{\tilde{S}}{\sqrt{n}}} < t_{n-1,\frac{\alpha}{2}}.$$

Ejemplo 149 *Se espera que la resistencia en kg/m^2 de cierto material suministrado por un proveedor se distribuya normalmente con media 220. Se toma una muestra de 9 elementos, obteniéndose los siguientes resultados:*

203, 229, 215, 220, 223, 233, 208, 228, 209

Contrastar la hipótesis de que esta muestra proviene de una población con media 220 Kg/m^2 al nivel de significación del 10 %.

Solución

Calculemos los parámetros de la muestra:
$\overline{x} = \frac{203+229+215+220+223+233+208+228+209}{9} = 218{,}67$
$\tilde{s}^2 = \frac{1}{8}\left((203 - 218{,}67)^2 + (229 - 218{,}67)^2 + \ldots\right.$
$\left.\ldots + (209 - 218{,}67)^2\right) = 110{,}75$

La población es normal y como σ es desconocida y n pequeño tendremos que utilizar como estadístico de contraste la t de Student.

1) $H_0 : \mu = 220$

2) $H_1 : \mu \neq 220$

3) $\alpha = 0{,}01$; $\frac{\alpha}{2} = 0{,}05$.

4) Bajo estas condiciones, población normal, σ desconocida y una muestra de tamaño $n = 9$ pequeño, la región crítica para estas hipótesis es:

   $t_8 > t_{8,1-0{,}05} = 1{,}86$ o $t_8 =< t_{8,0{,}05} = -1{,}86$

5) Cálculo del estadístico de contraste: $\overline{x} = 218{,}67$, $\tilde{s} = \sqrt{110{,}75} = 10{,}52$ entonces: $t_{n-1} = \frac{218{,}67-220}{\frac{10{,}52}{\sqrt{9}}} = -0{,}38$

6) Decisión: Como $t_8 = -0{,}38 \not> 1{,}86$ $t_8 = -0{,}38 \not< -1{,}86$ resulta que el valor del estadístico de contraste cae fuera de la región crítica, luego a partir de esta muestra no podemos rechazar (H_0) que la resistencia en Kg/m^2 del material sea igual a 220 contra que es distinta con un nivel de significación del 10 %.

8.11. Contraste para la varianza de una distribución normal.

Como es natural basaremos los contrastes para la varianza de una población normal en el estadístico muestral $\tilde{S}^2$; más concretamente en el estadístico $\chi^2_{n-1} = \frac{(n-1)\tilde{S}^2}{\sigma^2}$, del que sabemos que sigue una distribución χ^2 con $n-1$ g.l. si la población es normal.

Claro que no conocemos el valor de σ^2 pero bajo la hipótesis nula $H_0 : \sigma = \sigma_0$ tendremos que

$\chi^2_{n-1} = \frac{(n-1)\tilde{S}^2}{\sigma_0^2}$ tendrá también una distribución χ^2 con $n-1$ g.l.

Rechazaremos H_0 si $\tilde{S}^2$ es suficientemente distinta de σ_0 es decir si $H_1 : \sigma > \sigma_0$ rechazaremos H_0 si χ^2_{n-1} es pequeño, si $H_1 : \sigma < \sigma_0$ rechazaremos H_0 si χ^2_{n-1} es grande y si $H_1 : \sigma \neq \sigma_0$ rechazaremos H_0 si χ^2_{n-1} da valores altos o bajos.

8.11.1. Reglas de decisión para el contraste de la varianza de una población normal

Condiciones:

- Muestra aleatoria de n observaciones de una población normal.

Entonces una contraste al nivel de significación α para las hipótesis:

a)
$$\begin{cases} H_0 : \sigma = \sigma_0 \text{ o } H_0 : \sigma \leq \sigma_0 \\ H_1 : \sigma > \sigma_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si
$$\chi^2_{n-1} = \frac{(n-1)\tilde{s}^2}{\sigma_0^2} > \chi^2_{n-1,1-\alpha}.$$

b)
$$\begin{cases} H_0 : \sigma = \sigma_0 \text{ o } H_0 : \sigma \geq \sigma_0 \\ H_1 : \sigma < \sigma_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si
$$\chi^2_{n-1} = \frac{(n-1)\tilde{s}^2}{\sigma_0^2} < \chi^2_{n-1,\alpha}.$$

c)
$$\begin{cases} H_0 : \sigma = \sigma_0 \\ H_1 : \sigma \neq \sigma_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si
$$\chi^2_{n-1} = \frac{(n-1)\tilde{s}^2}{\sigma_0^2} > \chi^2_{n-1,1-\frac{\alpha}{2}} \text{ o } \chi^2_{n-1} = \frac{(n-1)\tilde{s}^2}{\sigma_0^2} < \chi^2_{n-1,\frac{\alpha}{2}}.$$

Ejemplo 150 *Se han medido los siguientes valores en miles de personas para la audiencia de un programa de radio en distintos días: 521, 742, 593, 635, 788, 717, 606, 639, 666, 624. Contrastar que la varianza de la audiencia es 6400 al nivel de significación del 5 %, suponiendo que la población sea normal*

Solución:

1) $H_0 : \sigma^2 = 6400$.

2) $H_1 : \sigma^2 \neq 6400$; ya que no se especifica qué alternativa se pide.

3) Nivel de significación $\alpha = 0{,}05$.

4) Bajo estas condiciones podemos utilizar como región crítica $\chi^2_9 = \frac{(9)\tilde{S}^2}{6400} > \chi^2_{9,1-0{,}025} = 19{,}02$ o $\chi^2_9 = \frac{(9)\tilde{S}^2}{6400} < \chi^2_{9,0{,}0{,}25} = 2{,}70$

5) $\overline{x} = 653{,}10$ y $\tilde{s}^2 = 6111{,}66$ entonces $\chi^2_9 = \frac{(9)6111{,}66}{6400} = 8{,}59452$

6) Como $\chi^2_9 = 8{,}59452 \not> 19{,}02$ y $\chi^2_9 = 8{,}66514 \not< 2{,}70$ resulta que el estadístico de contraste no cae dentro de la región crítica, luego no podemos rechazar H_0 contra H_1 al nivel de significación $\alpha = 0{,}05$.

8.12. Contrastes para la proporción muestral: muestras grandes

Si denotamos por p la proporción poblacional y por $\hat{p}$ la proporción muestral hemos visto que

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

sigue, por el T.L.C, aproximadamente una distribución normal. Como es lógico no conocemos la proporción muestral, pero si suponemos que la muestra es grande podríamos aproximar, como en temas anteriores el denominador por

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

y entonces

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Como se hizo en tema anterior. Pero bajo $H_0 : p = p_0$ tenemos que $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{p_0(1-p_0)}{n}}$:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

sigue teniendo aproximadamente una distribución normal estándar si n es grande. De forma similar al contraste de la media podemos definir las siguientes regiones críticas al nivel de significación α para las distintas hipótesis alternativas.

8.12.1. Reglas de decisión para el contraste de una proporción muestral: tamaño muestral grande

Condiciones:

- Muestra aleatoria simple de tamaño grande n procedente de una población con proporción poblacional de la característica de interés p, y proporción muestral de la misma $\hat{p}$

Entonces una contraste al nivel de significación α para las hipótesis:

a)

$$\begin{cases} H_0 : p = p_0 & (\text{ o } H_0 : p \leq p_0) \\ H_1 : p > p_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{1-\alpha}$$

b)

$$\begin{cases} H_0 : p = p_0 & (\text{ o } H_0 : p \geq p_0) \\ H_1 : p < p_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z_\alpha.$$

c)

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

Tiene por regla de decisión:

Rechazar H_0 si

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{1-\frac{\alpha}{2}} \text{ o } \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z_{\frac{\alpha}{2}}.$$

Ejemplo 151 *Un constructor afirma que se pone preinstalación de aire acondicionado en el 70 % de todas las viviendas que se están construyendo en una determinada ciudad. ¿Estaríamos de acuerdo con la afirmación del constructor si en una investigación se obtiene en muestra aleatoria de tamaño 100 resulta que 53 de las casas tienen preinstalación de aire acondicionado? Utilizar un nivel de significación* $\alpha = 0{,}01$*. Calcular el p-valor del contraste.*

Solución Seguiremos los seis pasos:

1) $H_0 : p = 0{,}7$

2) $H_1 : p \neq 0{,}7$

3) $\alpha = 0{,}01$ luego $\frac{\alpha}{2} = 0{,}005$

4) Región crítica
$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{1-0{,}005} = 2{,}57$ o $\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < z_{0{,}005} = -2{,}57$

5) $\hat{p} = \frac{53}{100} = 0{,}53$
$Z = \frac{0{,}53-0{,}7}{\sqrt{0{,}7(0{,}3)100}} = -3{,}7$

6) $Z = -3{,}7 \not> 2{,}57$ pero $Z = -3{,}7 < -2{,}57$ luego el estadístico de contraste está en la región crítica por lo tanto no podemos aceptar la afirmación del constructor al nivel de significación $\alpha = 0{,}05$

Calculemos el p valor. Lo haremos por el lado izquierdo que es por donde antes alcanzaremos el valor $Z = -3{,}7$ como tenemos que $z_{0{,}0001} = -3{,}7$ entonces el p-valor es

$\frac{\alpha}{2} = 0{,}0001$ por lo tanto el p-valor es $\alpha = (2)(0{,}0001) = 0{,}0002$. Es decir la afirmación del constructor dista mucho de ser cierta.

8.13. Pruebas de bondad de ajuste

A lo largo de este tema se han tratado contrastes estadísticos de hipótesis sencillos para distintos parámetros y en muchos casos se ha supuesto que la población era normal. Consideraremos ahora una prueba para determinar si una población tiene una determinada distribución teórica.

La prueba se basará en la diferencia entre las frecuencias observadas en la muestra y las frecuencias que se obtendrían con la distribución hipotética.

Veamos el ejemplo más sencillo. Queremos saber si un dado está bien balanceado, es decir si la distribución teórica del dado es $P_X(x) = \frac{1}{6}$ para $x = 1, 2, 3, 4, 5, 6$.

Supongamos que lanzamos el dado 120 veces y anotamos cada uno de los resultados. En teoría si el dado no está cargado esperaríamos obtener 20 veces cada resultado. Los resultados de la muestra se dan en la siguiente tabla:

	Valor obtenido en el lanz.					
Frecuencia	1	2	3	4	5	6
Observada (o_i)	20	22	17	18	19	24
Esperada (e_i) si H_0 es cierta	20	20	20	20	20	20

Tenemos que “medir” de alguna manera la “distancia” entre los resultados observados y los teóricos.

Como vemos en la tabla tenemos que comparar $k = 6$ valores.

8.13.1. Un contraste de bondad de ajuste: distribución totalmente conocida

Supongamos que tenemos n ($n \leq 25$ o 30) observaciones de las que se calculan sus frecuencias observadas en k *clases* (que debe ser $k \geq 5$ y queremos contrastar que sigue una distribución totalmente conocida, es decir conocemos la forma de la distribución de contraste y todos sus parámetros:

$$\begin{cases} H_0 : \text{La población tiene esta distribución} \\ H_1 : \text{La población tiene otra distribución} \end{cases}$$

Entonces el estadístico de contraste es:

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

y tiene aproximadamente una distribución χ^2 con $k-1$ g.l. si todas las frecuencias esperadas superan 5 (en caso contrario se pueden reagrupar los datos reduciéndose los grados de libertad)

La regla de rechazo al nivel de confianza α es:

Rechazar H_0 si:

$$\chi^2_{k-1} > \chi^2_{k-1,1-\alpha}$$

Ejemplo 152 *¿Podemos afirmar al nivel de significación $\alpha = 0{,}05$ que el dado del ejemplo anterior está bien balanceado (y por lo tanto conocemos su distribución y sus parámetros) a la vista de la muestra?*

Solución:

Bajo estas condiciones conocemos completamente la distribución teórica, $k = 6$ y las frecuencia absolutas teóricas son superiores a 5 entonces:

$$\chi^2_{k-1} = \chi^2_5 = \frac{(20-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(24-20)^2}{20} = 1{,}7 \not> \chi^2_{5,\alpha}$$

No podemos rechazar H_0 al nivel de significación $\alpha = 0{,}05$.

Ejemplo 153 *La información que la marca de pilas AAA aporta en el exterior de su envoltorio dice: "`La duración media de las pilas AAA sigue una distribución normal de media 3.5 horas y desviación típica 0.7 horas siempre que se conserven en un sitio seco y fresco`". La inspección de AENOR toma una muestra aleatoria de pilas obteniéndose los siguientes resultados:*

Límites de la clase	o_i
1.45 - - 1.95	*2*
1.95 - - 2.45	*1*
2.45 - - 2.95	*4*
2.95 - - 3.45	*15*
3.45 - - 3.95	*10*
3.95 - - 4.45	*5*
4.45 - - 4.95	*3*

¿A la vista de estos datos podemos afirmar que la información del fabricante es cierta al nivel de significación del 5 %?

Solución:

El contraste es:

$$\begin{cases} H_0: & \text{La distribución de duración es normal} \\ & \text{con } \mu = 3{,}5 \text{ y } \sigma = 0{,}7 \\ H_1: & \text{La distribución es cualquier otra} \end{cases}$$

Vamos ha realizar el contraste de bondad de ajuste, para ello tenemos que calcular las frecuencias esperadas. La muestra es de tamaño $n = 40$. Sea X=duración de una pila escogida al azar. Entonces:

$$P(1{,}95 \leq X \leq 2{,}45/H_0) = P(1{,}95 \leq X \leq 2{,}45/X \text{sigue una distribución normal con} \mu = 3{,}5 \text{ y } \sigma =$$

$$P(\frac{1{,}95 - 3{,}5}{0{,}7} \leq Z < \frac{2{,}45 - 3{,}5}{0{,}7}) = F_Z(-1{,}5) - F_Z(-2{,}21) \approx (1 - 0{,}9332) - (1 - 0{,}9864) = 0{,}0532.$$

Entonces la frecuencia esperada entre 40 pilas para el intervalo 1.95 - - 2.45 es $e_1 = (40)(0{,}0532) = 2{,}128 \approx 2{,}1$ (nota: en este último cálculo se suele aproximar al primer decimal)

De forma análoga se calculan los demás e_i y se obtienen los siguientes resultados:

Límites de la clase	o_i		e_i	
menor que 1.95	2		0.5	
1.95 - - 2.45	1		2.1	
2.45 - - 2.95	4	7	5.9	8.5
2.95 - - 3.45	15	15	10.3	10.3
3.45 - - 3.95	10	10	10.7	10.7
3.95 - - 4.45	5		7.0	
mayor que 4.45	3	8	3.5	10.5

Donde se observa que las frecuencia esperadas de los dos primeros intervalos y el del último no superan 5. Así que agrupamos los tres primeros intervalos en uno y los dos últimos también, de forma que las frecuencias esperadas y observadas quedan como en las segundas columnas.

Entonces $k = 4$ y el estadístico de contraste es

$$\chi^2_{k-1} = \chi^2_3 = \frac{(7 - 8{,}5)^2}{8{,}5} + \frac{(15 - 10{,}3)^2}{10{,}3} + \frac{(10 - 10{,}7)^2}{10{,}7} + \frac{(8 - 10{,}5)^2}{10{,}5} = 3{,}05$$

Ahora como tenemos que $\chi^2_{k-1} = 3{,}05 \not> \chi^2_{k-1,1-\alpha} = \chi_{3,1-0{,}05} = 7{,}815$ no hay razón para rechazar la hipótesis nula al nivel de significación $\alpha = 0{,}05$.

Nota: Como se observa en la tabla el primer y último intervalo se consideran con toda la cola de la probabilidad.

8.13.2. Un contraste de bondad de ajuste: algún parámetro poblacional desconocido

Si queremos contrastar si una población tiene una distribución por ejemplo normal, Poisson. y no conocemos los parámetros de estas distribuciones por ejemplo en la normal no conocemos μ o σ o ambas, el contraste que tenemos que realizar es similar al anterior pero el número de grados del libertad del estadístico de contraste será $k-m-1$ donde k es el número de categorías y m es el número de parámetros que se estiman.

Ejemplo 154 *Durante la segunda guerra mundial se dividió el mapa de Londres en cuadrículas de 0.25 Km^2 y se contó el número de bombas caídas en cada cuadrícula durante un bombardeo alemán. Los resultados fueron:*

num. impactos en la cuad.(x_i)	*0*	*1*	*2*	*3*	*4*	*5*
frecuencia (o_i)	*229*	*211*	*93*	*35*	*7*	*1*

Si realmente los bombardeos no seguían un plan prefijado la distribución del número de bombas en cada cuadrícula tendría que ser una $Po(\lambda)$. Contrastar esta hipótesis al nivel de significación $\alpha = 0{,}05$.

Solución: Sabemos el tipo de distribución pero no conocemos el parámetro λ lo tendremos que estimar por

$\lambda = \frac{\sum_{i=0}^{5} x_i o_i}{\sum_{i=0} o_i} = \frac{535}{576} = 0{,}929$

Calculemos las frecuencias esperadas e_i cuando la distribución de X=número de bombas por cuadrícula es una $Po(0{,}929)$, como sabemos que

$$P_X(x_i) = P(X = x_i) = \frac{0{,}929^{x_i}}{x_i!} e^{-0{,}929} = p_i$$

por lo tanto

x_i	0	1	2	3	4	5≥
p_i	.395	.367	.17	.053	0.012	0.003
$e_i = p_i \cdot 576$	227.5	211.4	97.9	30.5	6.9	1.7

Tendremos que agrupar las dos últimas columnas. En resumen:

x_i	0	1	2	3	$4 \geq$
o_i	229	211	93	35	8
e_i	227.5	211.4	97.9	30.5	8.6

Entonces tenemos que el número de clases es $k = 5$ y el número de parámetros estimados es $m = 1$

Entonces $\chi^2_{k-m-1} = \chi^2_3 = 0{,}961692$ y como $\chi^2_{3,1-0{,}05} = 7{,}815$ por lo tanto $0{,}961692 \not> 7{,}815$ no podemos rechazar la hipótesis nula con este nivel de significación. Por lo tanto podemos afirmar que el bombardeo era aleatorio y que no estaba dirigido a objetivos militares.

8.13.3. Prueba de bondad de ajuste de Kolgomorov-Smirnov (K-S)

El contrate de Kolgomorov - Smirnov que es conocido como K-S. Dada una ley de distribución continua F el test K-S contrata las siguientes hipótesis:

$$\begin{cases} H_0 : \text{La distribución de la muestra sigue la ley de distribución} F(x) \\ H_1 : \text{no sigue esa ley de distribución} \end{cases}$$

En principio la ley de distribución F puede ser cualquier distribución continua: normal, exponencial, uniforme, etc. Aunque en casos particulares, por ejemplo normalidad existen mejora de este test (por ejemplo para normalidad para algunos tamaños muestrales se debe aplicar el test de Kolgomorov-Smirnov-Lilliefors[1])

Este contraste parte de una muestra aleatoria de un cierta variable X: $x_1, x_2, \ldots, x_n$. A la muestra ordenada la denotaremos por : $x_{(1)} \leq x_{(2)}, \ldots, \leq x_{(n)}$

entonces podemos definir la función de distribución muestral (empírica) de la variable X para su muestra de tamaño n a la que denotaremos por F_n donde

[1]Ver: Daniel Peña,Sánchez Rivera. "*Estadísitica Modelos y métodos. 1 Fundamentos* ". Segunda Edición. Ed. Alianza Universidad Textos. 1991. Pág.369

$$F_n(X) = \begin{cases} 0 & \text{si } x \leq x_{(1)} \\ \frac{k}{n} & \text{si } x_{(k)} \leq x \leq x_{(k+1)} \\ 1 & \text{si } x \geq x_n \end{cases}$$

Definimos la máxima discrepancia para cada observación como:

$$D_n(x_h) = \text{máx}\{|F_n(x_{h-1}) - F(x_h)|\,, |F_n(x_h) - F(x_h)|\}$$

Ya haora definimos el estadístico D_n como la mayor de las discrepancias:

$$D_n = \max_{h=1,\ldots,n} D_n(x_h)$$

La regal de decisión es rechazar H_0 alp niverl α si

$$D_n \geq D_{n,\alpha}$$

Donde $D_{n,\alpha}$ es el cuantil de la distribución del test de Kolgomorv- Smirnov que podréis encontrar en las tablas.

Ejemplo 155 *Consideremos una sere de tiempo de vida de un cierto componente electrónico:*

$$16, 8, 9, 12, 6, 11, 20, 7, 2, 24$$

Vamos a contrastar si proviene de una distribución exponencial. Estimaremos primero el parámetro λ de la exponencial por la media muestral $\overline{x} = 11{,}5$

$$\begin{cases} H_0 : \textit{los datos provienen de una exp}(\frac{1}{11,5}) \\ H_1 : \textit{siguen otra distribución} \end{cases}$$

La muestra ordenada y los cálculos necesarios se muestran en la sguiente tabla:

h	x_h	$F_n(x)$	$F(x) = 1 - e^{-x/11,5}$	$F_n(x_{h-1}) - F(x)\|$	$F_n(x_{h-1}) - F(x)\|$	máx
1	*2*	*0.1*	*0.16*	*0.16*	*0.06*	*0.16*
2	*6*	*0.2*	*0.41*	*0.31*	*0.21*	*0.31*
3	*7*	*0.3*	*0.46*	*0.26*	*0.16*	*0.26*
4	*8*	*0.4*	*0.50*	*0.2*	*0.1*	*0.2*
5	*9*	*0.5*	*0.54*	*0.14*	*0.04*	*0.14*
6	*11*	*0.6*	*0.62*	*0.12*	*0.02*	*0.12*
7	*12*	*0.7*	*0.65*	*0.05*	*0.15*	*0.15*
8	*16*	*0.8*	*0.75*	*0.05*	*0.05*	*0.5*
9	*20*	*0.9*	*0.82*	*0.02*	*0.08*	*0.08*
10	*24*	*1*	*0.88*	*0.02*	*0.12*	*0.12*
						$D_n = 0{,}31$

Como,mirando en la tablas del test K-S se tiene que $D_{10,0,01} = 490$ y $D_{10} = 0{,}31 \not> 4{,}90$ rechazamos que estos datos se ajusten a esa exponencial al nivel $\alpha = 0{,}01$.

8.14. Contraste de las medias de dos poblaciones normales o tamaños muestrales grandes

Comenzaremos los contrates de dos parámetros con el caso en que tengamos dos muestras aleatoria de una misma misma variable e independientes entre si. Que las muestras son independientes quiere decir que la selección de los individuos a observar es independiente.

Supondremos pues dos muestras aleatoria independientes de tamaños n_1 y n_2 y medias μ_1 y μ_2 respectivamente.

Así tendremos una muestra será $x_{11}, x_{12}, \ldots x_{1n_1}$ y la otra $x_{21}, x_{22}, \ldots, x_{2n_2}$. Denotaremos por $\overline{X}_1$ y $\overline{X}_2$ son las medias aritméticas de cada muestra.

La hipótesis nula que se contrasta es $H_0 : \mu_1 - \mu_2 = 0$ aunque se suele escribir como $H_0 : \mu_1 = \mu_2$.

Podemos aplicar este test para poblaciones que tengan distribución normal o bien si n_1 y n_2 son suficientemente grandes.

El test aplicar depende de si las varianzas (σ_1^2 y σ_2^2 respectivamente) son conocidas o no. En el caso que sean desconocidas hay dos variantes del test: que aceptemos que las varianzas son iguales o que son distintas. Esta diferencia radica sólo en utilizar una fórmula distinta para estimar la varianza muestral pues en el caso en que sean iguales podemos aprovechar las dos muestras para estimar la varianza de la población.

Los estadísticos de contraste, las regiones críticas y los distintos intervalos de confianza se encuentra en las tablas de resúmenes de los contrates.

Ejemplo 156 *Supongamos que queremos estudiar los tiempos de ejecución de un algoritmo en dos tipos de sistemas operativos. Para ello disponemos de dos muestras de ordenadores independientes de tamaños* $n_1 = n_2 = 20$. *Supongamos que los tiempos siguen aproximadamente una distribución normal y que las desviaciones típicas son conocidas* $\sigma_1 = 1$ *y* $\sigma_2 = 2$.

Los resultados de la muestra del primer sistema son:

$$10{,}54, 10{,}73, 9{,}11, 8{,}07, 10{,}56, 9{,}87, 9{,}52, 8{,}34, 9{,}83, 8{,}11, 11{,}14, 10{,}11, 7{,}6, 11{,}13, 10{,}95, 9{,}48, 9{,}31, 11{,}8$$

mientras que los resultados de la muestra del segundo sistema son:

$$11{,}93, 14{,}24, 11{,}06, 11{,}2, 12{,}31, 12{,}92, 13{,}61, 15{,}03, 13{,}06, 11{,}47, 12{,}8, 14{,}55, 10{,}46, 15{,}21, 12{,}58, 11{,}76,$$

Se tiene que las medias en cada una de las poblaciones son $\overline{X}_1 = 9{,}789$ *y* $\overline{X}_2 = 12{,}385$. *Ahora necesitamos estimar la varianza muestral en este caso se utiliza la fórmula* $\tilde{S} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$; *con nuestros datos se obtiene que* $\tilde{S} = \sqrt{\frac{1}{20} + \frac{4}{20}} = 0{,}5$

El estadístico de contraste es

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\tilde{S}}$$

en nuestro caso vale $Z = \frac{9{,}789 - 12{,}385}{0{,}5} = -5{,}192$

si contrastamos contra $H_1 : \mu_1 \neq \mu_2$ *entonces la región crítica del contraste para* $\alpha = 0{,}05$ *es*

$$Z < z_{\frac{\alpha}{2}} \text{ o } Z > z_{1-\frac{\alpha}{2}}$$

Para este nivel de significación $z_{1-\frac{\alpha}{2}} = z_{1-0{,}025} = z_{0{,}975} = 1{,}96$ *y por lo tanto* $z_{\frac{\alpha}{2}} = z_{0{,}025} = -1{,}96$

Concluimos que no podemos aceptar que los tiempos de ejecución tiene medias iguales contra que las tiene distintas al nivel de significación $\alpha = 0{,}05$

Ahora podemos calcular un intervalo de confianza del 95 % para la diferencia de medias $\mu_1 - \mu_2$, mirando en las tablas de los resúmenes de los test, tenemos que el intevalo pedido es

$$\left(\overline{X}_1 - \overline{X}_2 + z_{\frac{\alpha}{2}}\tilde{S}, \overline{X}_1 - \overline{X}_2 + z_{1-\frac{\alpha}{2}}\tilde{S}\right)$$

Que en nuestro caso, para $\alpha = 0{,}05$ es

$$(9{,}789 - 12{,}385 + z_{0{,}025} \cdot 0{,}5, 9{,}789 - 12{,}385 + z_{0{,}975} \cdot 0{,}5) = (-2{,}596 - 1{,}96 \cdot 0{,}5, -2{,}596 + 1{,}96 \cdot 0{,}5$$

Por lo tanto un intervalo de confianza al nivel del 95 % para $\mu_1 - \mu_2$ es

$$(-3{,}576, -1{,}616)$$

Notemos que en este caso el cero no se encuentra en el intervalo de confianza.

Por último calculemos el p-valor para el contraste bilateral será el valor de α tal que

$$z_{\frac{\alpha}{2}} = -5{,}192$$

de donde (utilizando R pnorm(0.025) o en su caso las tablas) se tiene que

$\frac{\alpha}{2} = 1{,}040235e - 07$ *y por lo tanto* $\alpha = 2{,}08047e - 07$. *Utilizando las tablas hubiéramos concluido que el p-valor es prácticamente cero.*

Se deja como ejercicio el cálculo de los dos contrastes bilaterales, sus p-valores y los intervalos de confianza unilaterales.

Ejemplo 157 *Con los mismos datos que en el ejemplo anterior pero suponiendo ahora que las varianza no son conocidas los test cambian. En primer lugar tendremos que estimar la varianza de otra forma. Lo podemos hacer de dos maneras: suponiendo que las varianzas poblacionales son iguales o que son distintas. En el primero de los casos se obtiene un estadístico que sigue la distribución t de Student y en el segundo otra vez un estadístico Z con distribución normal.*

Necesitamos $\tilde{S}_1$ y $\tilde{S}_2$ las cuasi desviaciones típicas muestrales que valen $1{,}201323$ *y* $1{,}579462$ *respectivamente*

Supongamos las varianzas iguales entonces el estadístico de contraste para la hipótesis nula bilateral es

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\tilde{S}_{1,2}}$$

Donde la desviación típica muestral se estima por

$$\tilde{S}_{1,2} = \sqrt{\frac{(n_1-1)\tilde{S}_1^{\,2} + (n_2-1)\tilde{S}_2^2}{n_1+n_2-2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

El estadístico t sigue una distribución
$t_{n_1+n_2-2}$*, es decir t de Student con n_1+n_2-2 grados de libertad.*
Luego en nuestro caso $\tilde{S}_{1,2} = \sqrt{\frac{(20-1)(1{,}201323)^2+(20-1)(1{,}579462)^2}{20+20-2} \cdot \left(\frac{1}{20}+\frac{1}{20}\right)} = 0{,}4437272$
Entonces el valor del estadístico de contraste es $t = \frac{9{,}789-12{,}385}{0{,}4437272} = -5{,}850441$
La región crítica es, rechazar H_0 si

$$t \leq t_{n_1+n_2-2,\frac{\alpha}{2}} \text{ o } t \geq t_{n_1+n_2-2,1-\frac{\alpha}{2}}$$

Para el nivel de significación $\alpha = 0{,}05$ tenemos que $t_{n_1+n_2-2,1-\frac{\alpha}{2}} = t_{38,0{,}975}$ mirando las tablas de las t de Student aproximamos por $t_{40,0{,}975} = 2{,}021075$ que es el valor más cercano (con R y la instrucción qt(0.975,38) se obtiene $2{,}024394$*). Ahora tenemos que $t_{38,0{,}025} = -t_{38,0{,}975}$ y lo aproximamos por* $-t_{40,0{,}975} = -2{,}021075$

Así rechazamos H_0 ya que $t = -5{,}850441 < -2{,}021075 \approx t_{38,0{,}025}$.

Para el cálculo del p-valor tendríamos que igualar $t_{30,\frac{\alpha}{2}} = t = -5{,}850441$ con R hacemos $pt(-5{,}850441, 38)$ y se obtiene que $\frac{\alpha}{2} = 4{,}565589e-07$ luego el p-valor es $2 \cdot 4{,}565589e-07 = 9{,}131178e-07$ que es muy próximo a cero.

Por último el intervalo de confianza para la diferencia de las medias $\mu_1 - \mu_2$ es (donde $m = n_2 + n_2 - 2$)

$$(\overline{X}_1 - \overline{X}_2 + t_{m,\frac{\alpha}{2}}\tilde{S}_{1,2}, \overline{X}_1 - \overline{X}_2 + t_{m,1-\frac{\alpha}{2}}\tilde{S}_{1,2})$$

Que en nuestro caso es

$$(9{,}789-12{,}385-2{,}021075\cdot 0{,}4437272, 9{,}789-12{,}385+2{,}021075\cdot 0{,}4437272) = (-3{,}492806, -1{,}699194)$$

Se deja como ejercicio el cálculo de los dos contrastes bilaterales, sus p-valores y los intervalos de confianza unilaterales.

También se deja como ejercicio el caso en el que las varianzas son distintas.

Ejemplo 158 *Como ejercicio resolver utilizando las tablas de contrastes el ejercicio anterior en el caso de varianzas desconocidas y distintas.*

8.15. Contraste de dos proporciones

El test de contraste de dos proporciones se enfrenta a la comparación del parámetro p de probabilidad de éxito en dos poblaciones Bernoulli de parámetros p_1 y p_2, independientes de tamaños n_1 y n_2.

Este test es parecido al de dos medias y sólo se puede aplicar con tamaños muestrales grandes.

Tendremos dos muestras y sus correspondientes proporciones muestrales $\hat{p}_1$ y $\hat{p}_2$. El estadístico de contraste es

$$Z = \frac{\overline{p}_1 - \overline{p}_2}{\sqrt{\overline{p} \quad \overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Donde $\overline{p} = \frac{n_1\overline{p}_1 + n_2\overline{p}_2}{n_1+n_2}$ y $\overline{q} = 1 - \overline{p}$

El estadístico Z sigue una ley normal estándar.

La región de rechazo frente a la alternativa bilateral es, rechazar H_0 al nivel α si :

$$Z < Z_{\frac{\alpha}{2}} \text{ o } Z > Z_{1-\frac{\alpha}{2}}$$

El intervalo de confianza para la diferencia de proporciones poblacionales $p_1 - p_2$ al nivel $(1-\alpha)\cdot 100\,\%$ es

$$\left(\overline{p}_1 - \overline{p_2} + z_{\frac{\alpha}{2}}\sqrt{\overline{pq}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \overline{p}_1 - \overline{p_2} - z_{\frac{\alpha}{2}}\sqrt{\overline{pq}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}\right)$$

Ejemplo 159 *Se toman dos muestras de usuarios de internet de Mallorca y otra del resto de las islas. Se quiere saber si la proporción de usuarios de GNU Linux es igual o distinta entre las dos muestras. La muestra de Mallorca tiene tamaño* 100 *y resultaron 20 usuarios mientras que la de Menorca tiene tamaño 50 y resultaron 12 usuarios de GNU Linux. Como ejercicio contrastar la hipótesis de igualdad de proporciones al nivel de significación* 0,05*, calcular el p-valor y los intervalos de confianza con el mismo valor de* α*.*

8.16. Muestras dependientes

Hasta ahora hemos considerado que las muestras de las dos poblaciones de las que teníamos que contrastar su media o su varianza eran elegidas de forma que fueran independientes. Otra caso distinto es cuando las dos muestras corresponden a los mismos individuos o a individuos emparejados por algún factor determinante. Ejemplos de este factor son pares de gemelos univitelinos, máquinas absolutamente clónicas u otros emparejamientos que se puedan considerar aceptables para el diseño del experimento, como coeficiente intelectual, por peso, por edad, por ideología etc....

En estos casos se habla de un diseño de datos dependientes o emparejados[2]. En este caso el contraste más común corresponde a restar los valores de las muestras para cada individuo y realizar un contraste para averiguar si la media de las diferencias o proporciones es cero.

Lo más importante de este caso es aprender que hay diferentes maneras de realizar un diseño experimental para contrastar una hipótesis. Este diseño debe haber sido fijado, justificadamente, antes de realizar la experiencia, es decir antes de la recogida de datos. En caso contrario deberíamos realizar sólo un estudio descriptivo[3], pues los datos no corresponderían a un diseño experimental del que conozcamos la prueba de contaste de hipótesis adecuada a contrastar. Desde este punto de vista es mucho más importante, como no podía ser de otra manera, que los datos correspondan a un modelo del cual sepamos contrastar hipótesis y que estas hipótesis respondan a la decisión que se desea tomar.

Para saber los contrates que se proponen consultad la tabla de contrastes de hipótesis. A continuación se presentan dos ejemplos. El primero es un contraste de medias.

Ejemplo 160 *Un ayuntamiento dispone de aulas de informáticas en varios centros culturales. Estas aulas disponen de ordenadores similares, de las mismas características. Estos ordenadores antiguos necesitan una actualización consistente en instalar una nueva versión del sistema operativo que utilizaban y que es mucho más* pesada *pues requiere más memoria. Tenemos dos posibilidades ampliar la memoria 512 Kb o un Mb (hay más opciones que no contemplaremos como cambiar de sistema operativo por uno más ligero o cambiar los ordenadores por otros más potentes). Lo que resulta obvio es que una mayor aumento de memoria es más caro. Lo que se nos pide es decidir si el aumento de memoria de medio mega es significativamente mejor que el de un mega. El decisor político necesita saber esto antes de plantearse la*

[2]En inglés:*paired* y por lo tanto se habla de *paired test*

[3]Podemos hacer uno inferencial, pero dejando claro que los datos no son exactamente datos experimentales que respondan a muestras aleatorias

opción de otro sistema operativo u otras opciones. Supongamos que el coste de las licencias es cero; por ejemplo el ayuntamiento tiene una contrato por un número de licencias ilimitado.

Así que vamos a varios de los centros culturales y retiramos las memoria de los ordenadores. Así podemos probar a instalar, en un centro de prueba, las dos posibles ampliaciones de memoria. Instalamos una y realizamos una prueba de rendimiento que da como resultado unos tiempos de ejecución. Esta prueba se realiza en los mismos ordenadores y para cada una de las configuraciones de memoria. Los resultados obtenidos son:

ordenador i	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
antes	*8.1*	*11.9*	*11.4*	*12.9*	*9.0*	*7.2*	*12.4*	*6.9*	*8.9*	*8.3*
después	*6.9*	*6.7*	*8.3*	*8.6*	*18.9*	*7.9*	*7.4*	*8.7*	*7.9*	*12.4*
$d_i =$ *antes-después*	*1.2*	*5.2*	*3.1*	*4.3*	*-9.9*	*-0.7*	*5.0*	*-1.8*	*1.0*	*-4.1*

Se tiene que $\overline{d} = 0{,}33$ *y* $\tilde{S}_d = 4{,}72$*. Contrastar la igualdad de medias con el test que corresponda y que podéis encontrar en las tablas de contraste de hipótesis. Calcular el p-valor y un intervalo de confianza del* 95 % *para la diferencia de las dos medias.*

El estadístico de contraste es $t = \frac{\overline{d}}{\tilde{S}_d/\sqrt{n}} = \frac{0{,}33}{4{,}72/\sqrt{10}} = 0{,}22$

Así para la región critica bilateral el p-valor es el α *tal que :*

$$t_{9,1-\frac{\alpha}{2}} = 0{,}22$$

mirando en las tablas de la t de Student se obtiene que $t_{9,0,5871} \approx t_{9,0,9} = 1{,}383029$ *(es una aproximación muy mala).*

por lo tanto $1 - \frac{\alpha}{2} = 0{,}9$ *y* $\alpha = 0{,}2$*. Con R sería pt(9,0.22) que nos da* $0{,}7720373$ *y el p-valor exacto sería la solución de la ecuación* $1 - \frac{\alpha}{2} = 0{,}7720373$*. Como ejercicio calcular el intervalo de confianza para la diferencia de medias*

El siguiente es un ejemplo de contraste de proporciones en la misma población antes y después de un un "evento".

Ejemplo 161 *Un periódico de distribución sólo por internet quiere saber si sus lectores se interesan en la sección de tecnología. Para ello selecciona una muestra de* 1000 *lectores y les pregunta si siguen la sección de tecnología obteniéndose 35 lectores interesados. Al mismo tiempo se encarga una reorganización de la sección de tecnología introduciendo más contenidos. Una vez* *realizado este cambio se realiza la misma pregunta a los mismos lectores. Los resultados fueron.*

		Después	
		Sí	*No*
Antes	*Sí*	*300*	*10*
	No	*100*	*590*

Utilizar el contraste apropiado (mirar tablas de contrastes hipótesis). Calcular el p-valor y un intervalo de confianza del 95 % *para la diferencia de las dos proporciones.*

8.17. Comparación de dos varianzas

Hemos visto la necesidad de comparar dos varianzas como paso previo a una comparación de medias de muestras independientes, aun que también puede tener sentido en si misma.

El contraste corresponde con una hipótesis nula $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ contra las alternativas unilaterales y bilateral habituales. El intervalo de confianza que se encuentra en las tablas de contrastes es para el cociente, no para la diferencia de varianzas. Así que si vamos a aceptar la igualdad de varianzas este intervalo debe contener a 1.

El estadístico de contraste sigue una ley de distribución F de Fisher y tiene por parámetros dos grados de libertad.

Ejemplo 162 *En el ejemplo de los ordenadores suponer que tenemos suficiente equipo y memoria para realizarlo en dos centros culturales de forma independiente. Se trata de contrastar la diferencia de las medias de los rendimientos en el caso que se determine según sean las varianzas iguales o distintas. Para ello se debe primero contrastar la igualdad de varianzas. Hacerlo como ejercicio.*