

---

# Práctica 3. Variables aleatorias

---

## 1. Muestras aleatorias de las distribuciones usuales

Como ya se ha visto anteriormente, en *Minitab* podemos generar datos de distribuciones usuales utilizando la opción **Calc**⇒**Random Data**. Esta opción permite generar una muestra de datos de cualquier columna de la hoja de datos actualmente abierta o de una de las distribuciones de probabilidad que aparecen listadas.

En primer lugar, vamos a crear una nueva hoja de datos que llevará por nombre **Probabilidad.mtw**. A continuación, vamos a crear una columna, en dicha hoja de datos, que lleve por nombre '**100 datos de N(5,2)**' y que contenga 100 datos aleatorios procedentes de una distribución  $\mathcal{N}(5, 2)$  (Normal de media 5 y desviación típica 2). Para ello, seleccionamos **Calc**⇒**Random Data**⇒**Normal**; en **Number of rows of data to generate** tecleamos **100**; en **Store in column** tecleamos el nombre '**100 datos de N(5,2)**'; en **Mean** tecleamos **5** y en **Standard deviation** ponemos un **2**.

A continuación vamos a hacer el histograma, con la curva Normal superpuesta, de la muestra aleatoria obtenida en la columna '**100 datos de N(5,2)**'. Para ello, recordemos que hay que seleccionar la opción **Graph**⇒**Histogram**. En el cuadro de diálogo resultante elegimos **With Fit**. En el siguiente cuadro de diálogo, en **Graph variables** seleccionamos, de la lista de variables que tenemos a la izquierda, la columna '**100 datos de N(5,2)**' y pulsamos en **OK**. En la representación gráfica podemos apreciar que el histograma está cerca de la curva Normal superpuesta, lo cual es lógico puesto que hemos creado una muestra de una distribución Normal. También podemos ver, en la leyenda que aparece en la parte superior derecha del gráfico, que la media de la muestra obtenida se aproxima a 5 y la desviación típica se aproxima a 2.

## 2. Función de densidad y función de probabilidad

*Minitab* puede calcular el resultado de la función de densidad (o de la función de probabilidad) para un valor concreto o para una lista de valores. Para ello hay que elegir la opción **Calc**⇒**Probability Distributions** y a continuación el nombre de la variable aleatoria: Chi-square (chi-cuadrado de Pearson), Normal, F (de Snedecor), t (de Student), etc.

Dentro del cuadro de diálogo que aparecerá hay que seleccionar **Probability Density** (para las distribuciones continuas) o **Probability** (para las distribuciones discretas).

Para entender mejor el interés de esta opción, vamos a determinar los resultados de la función de densidad de una distribución  $\mathcal{N}(0, 1)$  (Normal Estándar) para una lista de valores que vamos a crear (todos los números comprendidos entre -4 y 4, con un incremento de 0,01). Luego haremos la representación gráfica de esta función de densidad. Para ello se procede de la siguiente manera:

- a) Mediante la opción **Calc**⇒**Make Patterned Data**⇒**Simple Set of Numbers** crearemos una nueva columna que denominaremos 'x de -4 a 4' y que contendrá todos los números comprendidos entre el -4 y el 4 con un incremento de 0,01. Podemos comprobar que en la columna 'x de -4 a 4' hay 801 números.
- b) En otra columna se calculan los resultados de la función de densidad de la variable aleatoria Normal Estándar para cada valor de la columna 'x de -4 a 4'. Para hacerlo, se selecciona **Calc**⇒**Probability Distributions**⇒**Normal**; se activa **Probability density**; en **Mean** y en **Standard deviation** se deja lo que aparece por defecto (cero y uno, respectivamente); en **Input column** se selecciona, de la lista de variables de la izquierda, la columna 'x de -4 a 4' y en **Optional storage** se teclea el nombre de la columna que contendrá los resultados de la función de densidad; por ejemplo, 'f(x) N(0,1)'.
- c) Finalmente, para representar gráficamente la función de densidad de la variable aleatoria Normal Estándar se elige la opción **Graph**⇒**Scatterplot**, después se elige **With connect line**. En el siguiente cuadro de diálogo, en **Y variables** se selecciona, de la lista de variables de la izquierda, la columna 'f(x) N(0,1)' y en **X variables** se selecciona la columna 'x de -4 a 4'. Sería conveniente quitar los puntos del gráfico, dejando sólo la línea de conexión, para lo cual se hace doble *clic* sobre la curva, en **Attributes**⇒**Symbols** se marca la opción **Custom** y en **Type** se selecciona **None** (buscando hacia arriba). Luego se hace un *clic* dentro del gráfico, pero no sobre la curva.

Ahora vamos a calcular los resultados de la función de probabilidad de la distribución discreta  $\mathcal{B}(200, 0,4)$  (Binomial de parámetros  $n = 200$  y  $p = 0,4$ ), vamos a hacer su representación grafica y vamos a averiguar el valor de la media de dicha variable aleatoria discreta. Para ello procedemos de la siguiente manera:

- 1) Mediante la opción **Calc**⇒**Make Patterned Data**⇒**Simple Set of Numbers** crearemos una nueva columna que denominaremos 'x de 0 a 200' y que contendrá todos los resultados posibles de la distribución  $\mathcal{B}(200, 0,4)$ , que, como sabemos, son: 0, 1, 2, ..., 200.
- 2) Calculamos los resultados de la función de probabilidad de  $\mathcal{B}(200, 0,4)$  para todos y cada uno de los valores de la columna 'x de 0 a 200'. Para ello, seleccionamos la opción **Calc**⇒**Probability Distributions**⇒**Binomial**; activamos **Probability**; en **Numbers of trials** tecleamos 200; en **Event probability** tecleamos 0,4; en **Input column** elegimos, de la lista de variables de la izquierda, la columna 'x de 0 a 200' y en **Optional storage** tecleamos el nombre de la columna que contendrá los resultados de la función de probabilidad; por ejemplo, 'p(x) B(200,0,4)'.
- 3) Ahora vamos a hacer la representación gráfica bidimensional que tiene en el eje horizontal los resultados de la columna 'x de 0 a 200' y en el eje vertical los resultados de la columna 'p(x) B(200,0,4)'. Para ello, se selecciona la opción **Graph**⇒**Scatterplot**, después se elige **With connect line**. En el siguiente cuadro de diálogo, en **Y variables** se selecciona, de la lista de variables de la izquierda, la columna 'p(x) B(200,0,4)' y en **X variables** se selecciona la columna 'x de 0 a 200'. Como ya hemos dicho anteriormente, sería conveniente quitar los puntos del gráfico, dejando sólo la línea de conexión.

Se puede comprobar que esta representación gráfica se aproxima mucho a la curva de densidad de una distribución Normal, lo cual se debe a lo siguiente: cuando  $n$  es grande y  $p$  no se acerca a 0 ni a 1, entonces  $\mathcal{B}(n, p)$  se aproxima a  $\mathcal{N}(np, \sqrt{npq})$ , siendo  $q = 1 - p$ .

- 4) También vamos a calcular la media teórica de la distribución  $\mathcal{B}(200, 0,4)$ . Recordemos que la media de una distribución discreta es  $E(X) = \sum x_i \cdot p(x_i)$ . Por tanto, usamos la opción **Calc**⇒**Calculator**. En **Store result in variable** tecleamos el nombre de la columna que contendrá los

resultados de los productos  $x_i \cdot p(x_i)$ ; por ejemplo, 'x p(x)'; en **Expression** ponemos (empleando la lista de variables y la calculadora de dicho cuadro de diálogo) 'x de 0 a 200' \* 'p(x) B(200,0,4)'. Ahora tenemos que calcular la suma de todos los resultados de la columna 'x p(x)', para lo cual elegimos la opción **Calc**  $\Rightarrow$  **Column Statistic**; activamos **Sum**; en **Input variable** seleccionamos, de la lista de variables de la izquierda, la columna 'x p(x)' y dejamos desactivada la opción **Store result in**. En la ventana de sesión podemos ver el resultado de la media, que es igual a  $E(X) = n \cdot p = 200 \cdot 0'4 = 80$ .

### 3. Función de distribución (probabilidad acumulada)

Para calcular el resultado de la función de distribución de una variable aleatoria,  $F(t) = P(X \leq t)$ , hay que elegir la opción **Calc**  $\Rightarrow$  **Probability Distributions** y a continuación el nombre de la variable aleatoria. Dentro del cuadro de diálogo que aparece hay que seleccionar **Cumulative Probability**.

Por ejemplo, vamos a calcular la probabilidad  $P(X \leq -1'36)$ , siendo  $X$  una variable aleatoria Normal Estándar. Como  $P(X \leq -1'36) = F(-1'36)$ , para calcular su resultado seleccionamos la opción **Calc**  $\Rightarrow$  **Probability Distributions**  $\Rightarrow$  **Normal**; activamos **Cumulative Probability**; en **Mean** y en **Standard deviation** dejamos lo que aparece por defecto (cero y uno, respectivamente). No activamos la opción **Input column** sino la opción **Input constant**, en donde colocamos el valor **-1,36**. Podemos almacenar el resultado en una constante tecleando en el recuadro **Optional storage** una  $K$  seguida de un número o poniendo un nombre a dicho resultado. Si no rellenamos el recuadro **Optional storage**, el resultado aparece en la ventana de sesión. Se puede comprobar que la probabilidad pedida es  $P(X \leq -1'36) = F(-1'36) = 0'086915$ .

Si queremos calcular probabilidades de los tipos  $P(X > a)$ ,  $P(a < X < b)$ , etc., tenemos que utilizar lápiz y papel, y aplicar las propiedades de la probabilidad para llegar a expresiones en las que sólo aparezcan probabilidades del tipo  $P(X \leq x)$  (función de distribución), pues éstas son las que calcula **Minitab**. No tenemos que olvidar, por ejemplo, que si  $X$  es una variable aleatoria continua, entonces  $P(X = a) = 0$  para todo  $a$ , por lo que se cumplen las siguientes igualdades:  $P(X \leq x) = P(X < x)$ ,  $P(X \geq x) = P(X > x)$ ,  $\dots$ . Pero si  $X$  es una variable aleatoria discreta, las probabilidades  $P(X \leq x)$  y  $P(X < x)$  no son (en general) iguales.

Como ya hemos dicho, cuando  $n$  es grande y  $p$  no se acerca a 0 ni a 1, entonces  $\mathcal{B}(n, p)$  se aproxima a  $\mathcal{N}(np, \sqrt{npq})$ , siendo  $q = 1 - p$ . Vamos a poder observarlo con el siguiente ejemplo:

Sea  $X$  una variable aleatoria  $\mathcal{B}(200, 0'4)$  y sea  $Y$  una variable aleatoria Normal de media 80 y desviación típica 6'928203. Vamos a comprobar (mediante una representación gráfica conjunta) que las funciones de distribución de ambas variables son muy parecidas. La solución es la siguiente:

- a) Calculamos los resultados de la función de distribución de  $\mathcal{B}(200, 0'4)$  para todos y cada uno de los valores de dicha columna 'x de 0 a 200'. Para ello, seleccionamos la opción **Calc**  $\Rightarrow$  **Probability Distributions**  $\Rightarrow$  **Binomial**; activamos **Cumulative probability**; en **Numbers of trials** tecleamos **200**; en **Event probability** tecleamos **0,4**; en **Input column** elegimos, de la lista de variables de la izquierda, la columna 'x de 0 a 200' y en **Optional storage** tecleamos el nombre de la columna que contendrá los resultados de la función de distribución de la Binomial; por ejemplo, 'F(x) B(200,0,4)'
- b) Calculamos los resultados de la función de distribución de  $\mathcal{N}(80, 6'928203)$  para los mismos valores de  $x$ , es decir, para los valores de la columna 'x de 0 a 200'. Para ello, se elige **Calc**  $\Rightarrow$  **Probability Distributions**  $\Rightarrow$  **Normal**; se activa **Cumulative probability**; en **Mean** se teclea **80**; en

**Standard deviation** se pone **6,928203**; en **Input column** elegimos, de la lista de variables de la izquierda, la columna '**x de 0 a 200**' y en **Optional storage** tecleamos el nombre de la columna que contendrá los resultados de la función de distribución de la Normal; por ejemplo, '**F(x) N(80,6,9)**'.

- c) Ahora vamos a superponer, en un mismo gráfico, las dos funciones de distribución. Para ello, se selecciona la opción **Graph**⇒**Scatterplot**⇒**With connect line**. En el cuadro de diálogo que aparece, junto al **1** en **Y variables** seleccionamos la columna '**F(x) B(200,0,4)**' y en **X variables** seleccionamos la columna '**x de 0 a 200**', y junto al **2** en **Y variables** seleccionamos la columna '**F(x) N(80,6,9)**' y en **X variables** seleccionamos otra vez la columna '**x de 0 a 200**'. Luego pulsamos **Multiple graphs** y en el cuadro de diálogo resultante activamos **Overlay on the same graph**. Como ya hemos dicho anteriormente, sería conveniente quitar los puntos del gráfico, dejando sólo la línea de conexión.

## 4. Inversa de la función de distribución (percentiles)

En ocasiones, en lugar de querer calcular probabilidades de sucesos, se desea justamente lo contrario, conocer el valor  $x$  que hace que la probabilidad del suceso ( $X \leq x$ ) sea igual a un valor determinado  $p$ ; es decir, hallar  $x$  para que se cumpla  $P(X \leq x) = p$ ; esto no es más que calcular percentiles de variables aleatorias. Para calcular el resultado de los percentiles de una variable aleatoria hay que elegir la opción **Calc**⇒**Probability Distributions** y a continuación el nombre de la variable aleatoria. Dentro del cuadro de diálogo que aparece hay que seleccionar **Inverse cumulative probability**.

Por ejemplo, vamos a calcular el valor  $x$  que verifica  $P(X \leq x) = 0'98$ , cuando  $X \equiv \chi_{20}^2$  (chi-cuadrado de Pearson con 20 grados de libertad). Para ello seleccionamos la opción **Calc**⇒**Probability Distributions**⇒**Chi-Square**. En el cuadro de diálogo activamos **Inverse cumulative probability**. Dejamos lo que aparece por defecto (cero) en **Noncentrality parameter**. En **Degrees of freedom** tecleamos **20**. No activamos la opción **Input column** sino la opción **Input constant**, en donde colocamos el valor **0,98**. Podemos almacenar el resultado en una constante tecleando en el recuadro **Optional storage** una  $K$  seguida de un número o poniendo un nombre a dicho resultado. Si no rellenamos el recuadro **Optional storage**, el resultado aparece en la ventana de sesión. Se puede comprobar que el valor  $x$  que verifica  $P(X \leq x) = 0'98$  es  $35'0196$ ; es decir,  $P(X \leq 35'0196) = 0'98$ , siendo  $X \equiv \chi_{20}^2$ .

Si queremos calcular el valor  $a$  tal que las probabilidades de los tipos  $P(X > a)$ ,  $P(|X| < a)$ ,  $P(|X| > a)$ , etc., sean iguales a un cierto resultado, tenemos que utilizar lápiz y papel, y aplicar las propiedades de la probabilidad para llegar a expresiones en las que sólo aparezcan ecuaciones del tipo  $P(X \leq x) = p$  (percentiles), pues éstas son las que calcula *Minitab*.

## 5. Ejercicios propuestos

- 3.1. Genera 10000 datos aleatorios procedentes de una variable aleatoria Binomial de parámetros  $n = 50$  y  $p = 0'25$ . Calcula la media de esta columna de datos aleatorios.
- 3.2. Haz la representación gráfica de la función de probabilidad de una variable aleatoria Binomial de parámetros  $n = 50$  y  $p = 0'25$ .
- 3.3. Haz la representación gráfica de la función de distribución de una variable aleatoria Binomial de parámetros  $n = 50$  y  $p = 0'25$ .

**3.4.** Sea  $X$  una variable aleatoria Binomial de parámetros  $n = 50$  y  $p = 0'25$ . Calcula:

- a)  $P(X = 10)$ .
- b)  $P(X \leq 12)$ .
- c)  $P(X \geq 3)$ .
- d)  $P(X < 5)$ .
- e)  $P(X > 7)$ .
- f)  $P(10 < X < 20)$ .
- g)  $P(10 \leq X < 20)$ .
- h)  $P(10 < X \leq 20)$ .
- i)  $P(10 \leq X \leq 20)$ .

**3.5.** Si  $Z$  es una variable Normal Estándar, determina:

- a)  $P(Z \leq 2'21)$ .
- b)  $P(Z < 3'47)$ .
- c)  $P(Z \leq -1'75)$ .
- d)  $P(Z > 2'46)$ .
- e)  $P(Z \geq 3'24)$ .
- f)  $P(Z > -3'08)$ .
- g)  $P(1'12 \leq Z \leq 2'68)$ .
- h)  $P(-0'85 < Z < 1'27)$ .
- i)  $P(-2'97 < Z \leq -1'33)$ .

**3.6.** Si  $X$  es una variable Normal con media  $8'46$  y desviación típica  $1'14$ , halla:

- a)  $P(X \leq 9'11)$ .
- b)  $P(X < 12'33)$ .
- c)  $P(X \leq 6'41)$ .
- d)  $P(X > 10'52)$ .
- e)  $P(X \geq 12'61)$ .
- f)  $P(X > 4'01)$ .
- g)  $P(6'11 \leq X \leq 11'91)$ .
- h)  $P(7'53 < X < 10'33)$ .
- i)  $P(5'05 \leq X < 6'83)$ .

**3.7.** Halla el valor de los siguientes cuantiles:

- a)  $Z_{0'58}$ .
- b)  $Z_{0'42}$ .
- c)  $Z_{0'999}$ .

d)  $Z_{0'001}$ .

- 3.8.** Genera 10000 datos aleatorios procedentes de una distribución chi-cuadrado de Pearson con 100 grados de libertad. Calcula la media de esta columna de datos aleatorios. Haz un histograma de los datos aleatorios generados, con la curva Normal superpuesta. ¿Puedes extraer alguna conclusión?
- 3.9.** Haz la representación gráfica de la función de densidad de una variable aleatoria chi-cuadrado de Pearson con 100 grados de libertad. Los valores del eje horizontal pueden ser todos los comprendidos entre 0 y 200 con un incremento de 0'1.
- 3.10.** Haz la representación gráfica de la función de distribución de una variable aleatoria chi-cuadrado de Pearson con 100 grados de libertad. Los valores del eje horizontal pueden ser todos los comprendidos entre 0 y 200 con un incremento de 0'1.
- 3.11.** Calcula el valor de los siguientes cuantiles:
- a)  $\chi_{6,0'01}^2$ .
  - b)  $\chi_{6,0'99}^2$ .
  - c)  $\chi_{72,0'975}^2$ .
- 3.12.** Sea  $X$  una variable aleatoria que sigue una distribución chi-cuadrado de Pearson con 15 grados de libertad. Determina el valor de  $a$  que verifica la siguiente igualdad:
- a)  $P(X \leq a) = 0'05$ .
  - b)  $P(X > a) = 0'99$ .
- 3.13.** Calcula el valor de los siguientes cuantiles:
- a)  $t_{26,0'9}$ .
  - b)  $t_{26,0'1}$ .
  - c)  $t_{75,0'8}$ .
- 3.14.** Sea  $X$  una variable aleatoria que sigue una distribución  $t$  de Student con 20 grados de libertad. Determina el valor de  $a$  que verifica la siguiente igualdad:
- a)  $P(X \leq a) = 0'99$ .
  - b)  $P(X \geq a) = 0'25$ .
- 3.15.** Calcula el valor de los siguientes cuantiles:
- a)  $F_{8,6,0'975}$ .
  - b)  $F_{25,50,0'01}$ .
  - c)  $F_{45,35,0'01}$ .
- 3.16.** Sea  $X$  una variable aleatoria que sigue una distribución  $F$  de Snedecor con 10 grados de libertad en el numerador y 8 grados de libertad en el denominador. Determina el valor de  $a$  que verifica la siguiente igualdad:
- a)  $P(X < a) = 0'9$ .
  - b)  $P(X > a) = 0'05$ .