

Distribuciones en R-Commander

1. Pasos preliminares

Empezaremos por:

1. Ejecutar R.
2. Cambiar el directorio de trabajo al directorio que hayamos escogido (por ejemplo Mis Documentos)
3. Limpiar el entorno de trabajo de cualquier objeto heredado de una sesión anterior:
`rm(list=ls())`
4. Comprobamos los objetos que tenemos definidos con la instrucción `ls()`.
5. Lanzamos R-Commander con la instrucción :

`library(Rcmdr)`

NOTA: *en el caso en que se nos cierre R-Commander sin cerrarse R, para volver a ejecutarlo, necesitamos poner en la línea de comandos de R, la instrucción `Commander()`.*

2. Menú Distribuciones en R-Commander

En el menú Distribuciones de Rcommander, tenemos acceso a los comandos asociados a la **representación**, el **cálculo** o la **simulación** de distintos modelos clásicos de distribuciones, clasificados según si corresponden a variables aleatorias discretas o continuas.

2.1. Representación gráfica de funciones de densidad, de probabilidad o de distribución acumulada

En el menú submenú asociado a cada modelo concreto, escogemos la opción de gráfica, para obtener una representación de la función de densidad (para va. continua), puntual de probabilidad (para v.a. discreta) o la función de distribución acumulada.

Ejercicio. Para tres de los modelos de v.a. continua que propone R-Commander:

1. Obtener una gráfica de la densidad.
2. Comentar las características de cada densidad (¿qué valores puede tomar la variable? ¿sirve para modelizar fenómenos simétricos o asimétricos? ¿presenta mucha dispersión? ¿dónde está el centro aprox. de los datos?)
3. ¿Qué parámetros tiene el modelo?
4. ¿Cuál es el efecto de variar el o los parámetros del modelo?

Ejercicio. Para tres de los modelos de v.a. discreta que propone R-Commander:

1. Obtener una gráfica de la función puntual de probabilidad. ¿Qué se ha representado en el eje Oy?
2. Comentar las características de cada función puntual de probabilidad (¿qué valores puede tomar la variable? ¿sirve para modelizar fenómenos simétricos o asimétricos? ¿presenta mucha dispersión? ¿dónde está el centro aprox. de los datos?)
3. ¿Cuál es valor que es más probable que tome la variable?
4. ¿Qué parámetros tiene el modelo?
5. ¿Cuál es el efecto de variar el o los parámetros del modelo?

2.2. Cálculo de probabilidades o cuantiles con R-Commander

R-Commander dispone de los algoritmos de cálculo de probabilidades y de cuantiles asociados a varios modelos discretos o continuos.

2.2.1. Ejemplo: cálculo de probabilidades asociadas a una variable Normal

En el menú **Distribuciones >Continuas >Normal**, escogemos la opción **Probabilidades Normales**.

Fijamos la media y desviación típica de la variable Normal que nos interesa, y si queremos introducir más de un valor en el recuadro **Valor(es) de la variable**, podemos hacerlo usando un vector, por ejemplo:

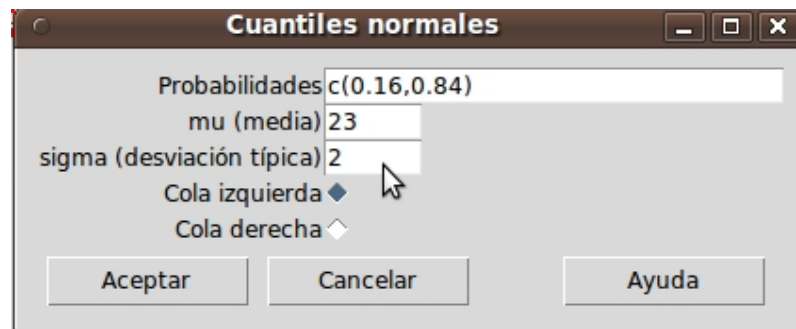
`c(280,300)`

2.2.2. Ejemplo: cálculo de cuantiles asociados a una variable Normal

Al igual que en el párrafo anterior, podemos calcular cuantiles asociados a una variable normal, escogiendo la opción **Cuantiles normales**.

En el recuadro **probabilidades**, tenemos la opción de especificar un valor entre 0 y 1, o un vector de valores, de manera similar que en el párrafo anterior. Si quiero calcular

los cuantiles 16% y 84% de una distribución Normal con media 23 y desviación típica 2, uso:



2.2.3. Ejercicios

1. Si X sigue una distribución *normal* de media 10 y desviación típica 2, se pide:
(a) $P(X \leq 12)$; (b) $P(9 \leq X \leq 13)$ (c) $P(|X - 10| \leq 1)$;
2. Estudiamos la resistencia de piezas de cemento a medida que va aumentando el tiempo de fraguado. En un instante dado, decidimos modelar la resistencia de una pieza escogida al azar por una distribución Normal con media 25 y desviación típica 3.
 - a) ¿Qué proporción de piezas tienen una resistencia superior a 28?
 - b) Queremos encontrar el valor de la resistencia por debajo del cual se encuentra el 97.5% de las piezas. ¿Podeis proporcionar este valor?
 - c) Si escojo una caja al azar, ¿cuál es la probabilidad de que contenga menos de 3 piezas con una resistencia superior a 28?
3. La dimensión de ciertas piezas de hormigón sigue una distribución normal de media 150 y desviación típica 0.4. Sabiendo que se consideran aceptables todas aquellas piezas cuya longitud se encuentre dentro del intervalo (149'2 , 150'4). Determinar:
 - a) El porcentaje de piezas defectuosas.
 - b) Proporcionar un intervalo centrado en 150 que contenga el 95% de los valores de la dimensión.
 - c) Estamos interesados en reducir la variabilidad de los valores de la dimensión, es decir que queremos reducir σ para que el 99.7% de estos valores estén comprendidos entre 147 y 153. ¿Qué valor de σ debemos conseguir?
4. Una determinada empresa dedicada a la fabricación de cemento-cola ha adquirido una máquina de envasado. Según los datos que le suministra el fabricante de

la envasadora, el peso de una bolsa una vez envasada la cantidad deseada de cemento cola sigue una distribución Normal cuyo promedio es de 2000 gramos y su desviación típica es de 50 gr.

Un saco de cemento-cola se considera defectuoso si su peso final es inferior a 1950 gr.

- a) Determinar la proporción de sacos defectuosos que producirá esta envasadora.
 - b) ¿Cuál es la proporción de sacos con peso inferior a 1900g?
 - c) Un palet tiene 500 sacos. ¿Cuál es la distribución del peso del palet?
5. En un gran almacén, el número de clientes que llegan a una caja **cada 60 minutos** puede modelarse como un proceso de Poisson de media 8.
- a) Calcular la probabilidad de que, en una hora, lleguen al menos 8 clientes a una caja determinada.
 - b) Calcular la probabilidad de que, en una hora, lleguen entre 6 y 9 clientes (ambos incluidos) a la cola...

2.3. Simulación de variables aleatorias con R-Commander

Para cada una de las distribuciones disponibles en los submenús **Distribuciones continuas** y **Distribuciones discretas**, es posible obtener valores simulados, escogiendo la instrucción **Muestras de la distribución** Para ello, tendremos que especificar el número de muestras que queremos extraer, el número de observaciones en cada muestra, así como los parámetros de la distribución.

2.3.1. Primer ejemplo: simulación de proceso de inspección de la calidad

Contexto: Queremos simular la extracción de muestras de 5 piezas en una producción, y la comprobación de su adecuación. Introducimos la variable X que toma el valor 1 si la pieza es correcta y 0 si la pieza es defectuosa.

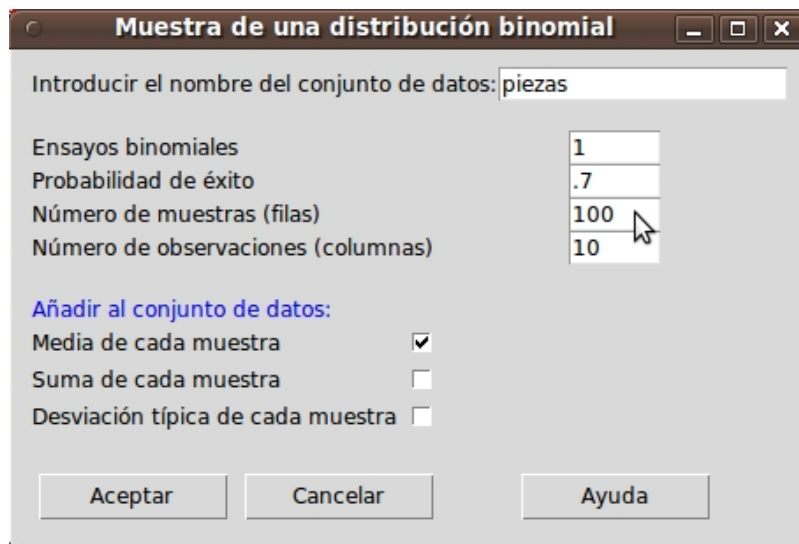
Fijamos que la proporción de piezas correctas en la producción es de 0.7, y vamos a simular 100 muestras de 10 piezas cada una.

En R-Commander, usamos **Distribuciones >Distribuciones discretas >Distribución binomial >Muestras de una distribución binomial** Debemos rellenar los distintos campos:

1. Decidimos del nombre del conjunto que se creará: por ejemplo **piezas**.
2. Especificamos el número de ensayos binomiales. En nuestro caso, la variable X corresponde a una realización del experimento dicotómico, por lo tanto, fijamos este número a 1.

3. Especificamos la probabilidad de éxito: en nuestro caso 0.7
4. Especificamos el número de muestras: en nuestro caso 100
5. Especificamos el número de observaciones por muestra: en nuestro caso 10.
6. Pedimos además que se añada una columna al conjunto que se creará, y que contendrá la media muestral de cada muestra (en nuestro caso será la proporción de piezas correctas...)

En resumen, tenemos las siguientes elecciones:



Al dar a Aceptar, se crea el conjunto de datos piezas. Podemos cambiarle el nombre a la columna `mean` y llamarla por ejemplo `pgorro`.

Realizamos un histograma de `pgorro`. ¿Le parece adecuada la aproximación Normal a la distribución de \hat{p} (ver transparencias “Introducción a la inferencia”)? ¿Qué podríamos hacer para mejorar esta aproximación?

2.3.2. Segundo ejemplo: simulación de valores de una distribución normal

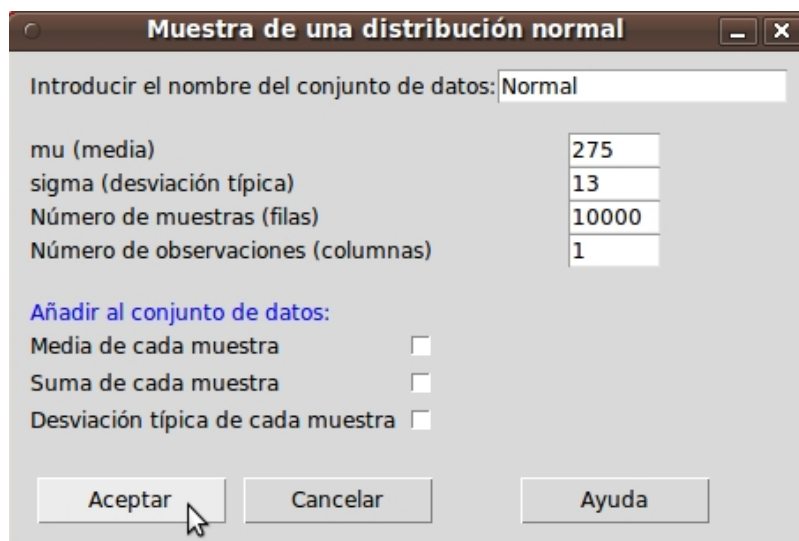
Objetivo: Queremos simular 10000 valores de una distribución normal con media 275 y desviación típica igual a 13. Es decir que queremos 10 000 muestras que contengan cada una una única observación...

En R-Commander, usamos `Distribuciones >Distribuciones continuas >Distribución normal >Muestras de una distribución normal` Debemos rellenar los distintos campos:

1. Decidimos del nombre del conjunto que se creará: por ejemplo `Normal`.
2. Especificamos la media En nuestro caso, será igual a 275.

3. Especificamos la desviación típica: en nuestro caso 13
4. Especificamos el número de muestras: en nuestro caso 10000
5. Especificamos el número de observaciones por muestra: en nuestro caso 1.
6. Desactivamos la opción que permite calcular la media muestral (cada muestra sólo comporta un valor...).

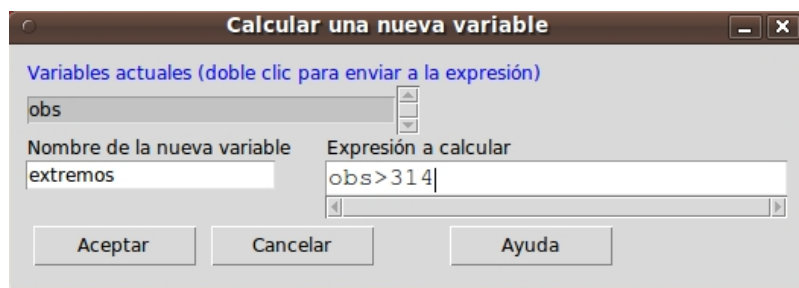
En resumen, tenemos las siguientes elecciones:



A continuación realizamos un histograma de los 10000 valores obtenidos. ¿Se va pareciendo a una campana de Gauss?

Finalmente, vamos a comprobar cuántas observaciones hemos obtenido que sean mayores de $\mu + 3\sigma$, es decir mayores de 314. Para ello, creamos una nueva variable en el conjunto de dato que sea de tipo Lógico. Para cada fila del conjunto, valdrá TRUE si la observación es mayor de 314, y FALSE en caso contrario.

Usamos la instrucción **Calcular una nueva variable** en el menú **Datos > Modificar variables del conjunto activo** . Creamos una nueva variable llamada **extremos** que sea igual a `obs > 314`:



Sólo nos queda pedirle a R que nos devuelva un resumen del conjunto de datos activo, para fijarnos en el número de veces que la variable `extremos` toma el valor TRUE. Yo he obtenido 10 valores TRUE, entre el total de 10 000... ¿Es este valor coherente con la probabilidad de que una Normal con media 275 y desviación típica sea mayor que 314?...