Sociedad Ecuatoriana de Estadística

“Análisis de Encuestas por Muestreo con R”

Repaso de probabilidad y variables aleatorias

Andrés Peña M.

a.pena@rusersgroup.com

Diciembre 2021

Tabla de contenidos

1. Probabilidad y variables aleatorias

Probabilidad y variable aleatoria

Dado un espacio muestral Ω, una familia de eventos A y una medida de probabilidad $P: A \rightarrow [0,1]$ es posible definir un **Espacio de Probabilidad** denotado por (Ω, A , P).

ESPACIO MUESTRAL

- •FINITO
- •INFINITO
  - . NUMERABLE
  - .NO NUMERABLE

Dado un espacio de probabilidad (Ω, A, P), una **variable aleatoria** es toda función que asocia un número real a cada resultado del espacio muestral.

$$X: \Omega \rightarrow \mathbb{R}$$

Variables aleatorias discretas

Una variable aleatoria **discreta** es una v.a. cuyos valores posibles constituyen un conjunto finito, o bien se pueden listar en una secuencia infinita en la que hay un primer elemento, un segundo elemento, etcétera.

La función de probabilidad, llamada en este caso **función de cuantía,** indica una probabilidad para cada valor posible de la variable aleatoria.

$$\mathrm{p}(x) = \mathrm{P}(\mathrm{X} = x)$$

- Si x_i es un valor particular de X entonces $p(x_i) > 0$.
- Si x_i no es un valor particular de X entonces $p(x_i) = 0$

Si la variable asume *k* valores distintos $x_1, x_2, \ldots, x_k$ la suma de las probabilidades debe ser igual a uno.

$$\sum_{i=1}^{k} p(x_i) = 1$$

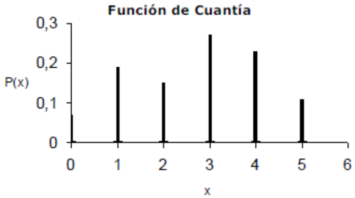
Variables aleatorias discretas

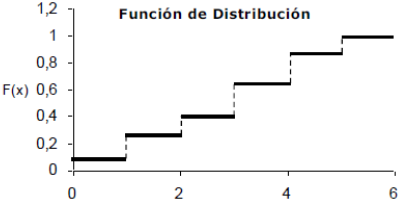
A partir de la función de cuantía es posible encontrar la **función de distribución:**

$$F(a) = P(X \leq a) = \sum_{x_i \leq a} p(x_i)$$

También podemos calcular la probabilidad de un evento a≤ x ≤ b

$$P(b \leq X \leq a) = \sum_{a \leq x_i \leq b} p(x_i)$$





Variables aleatorias continuas

Las variables aleatorias **continuas** se asocian con espacios muestrales infinitos no numerables, permitiendo asignar a cada valor un punto único en un intervalo de números reales.

Se dice que X es una variable aleatoria continua si existe una función, llamada **función de densidad** de X, que satisface las siguientes condiciones:

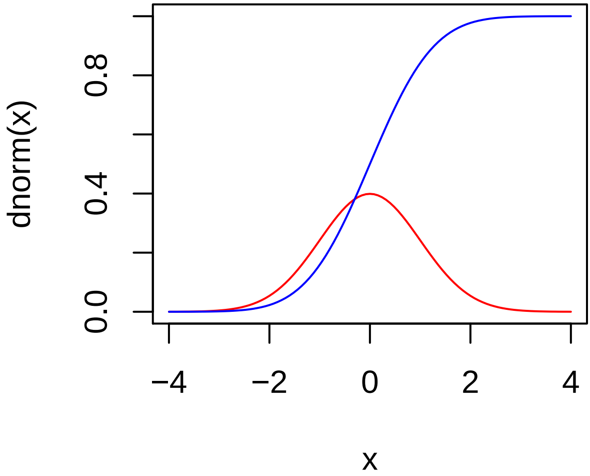
$$1) \quad f(X) \geq 0$$

$$2) \quad \int_{-\infty}^{\infty} f(X)\, dx = 1$$

La **función de distribución** será:

$$F(a) = P(x \leq a) = \int_{-\infty}^{a} f(x) d(x)$$

Variables aleatorias continuas



Esperanza y Varianza de variables aleatorias

Esperanza: es el valor promedio que se presentará si el experimento se repite muchas veces

V.A.D.

$$E(x) = \sum_{i=1}^{k} x_i \, p(x_i) \quad x = 1, 2, \ldots, k$$

V.A.C.

$$E(x) = \int_{1}^{k} x \, f(x) \, d(x) \quad 1 < x < k$$

Varianza: mide la dispersión de los valores de la variable respecto de la esperanza, si el experimento se repite muchas veces.

V.A.D.

$$V(x) = \sum_{1}^{k} [x_i - E(x)]^2 \, p(x)$$

V.A.C.

$$V(x) = \int_{1}^{k} [x - E(x)]^2 \, f(x) \, d(x)$$

Función de densidad, distribución, cuantiles en R

- Función de densidad $<-f(x)$:

```
dnorm(x, mean = 0, sd = 1)
```

- Función de distribución $<-F(x)=P(X\leq x)$ o también $P(X>x)=1-F(x)$:

```
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)
```

- Función cuantil:

```
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)
```

- Generador de muestra aleatoria con los parámetros dados:

```
rnorm(n, mean = 0, sd = 1)
```

Modelos de probabilidad simples

Modelos de Probabilidad

- **Para variables aleatorias discretas:** Bipuntual, Binomial, Hipergeométrica, Binomial Negativa y Poisson.

- **Para variables aleatorias continuas:** Normal, Exponencial y Uniforme

Variables aleatorias discretas

Modelo Bipuntual: donde cada prueba puede concluir en sólo dos resultados excluyentes, uno de los cuales se llama éxito y otro fracaso.

$$\Omega = \{E, F\} \qquad P(E) = p$$

$$P(F) = 1 - p$$

Si en una población dicotómica se selecciona una observación al azar, se genera una observación muestral, que es una variable aleatoria con distribución bipuntual.

$$p(x) = p^{x}(1-p)^{1-x} \qquad \mathrm{x} = 0,1$$

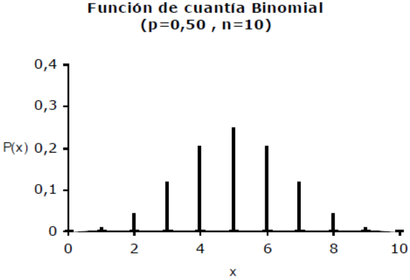
Variables aleatorias discretas

Modelo Binomial: se realizan n pruebas bipuntuales. La variable aleatoria se define como número de éxitos en n pruebas independientes (con reemplazo).

$$p(x) = C_x^n p^x (1-p)^{n-x} \qquad x = 0, 1, 2, 3, \ldots, n$$

$$E(x) = np$$

$$V(x) = np(1-p)$$



Variables aleatorias discretas

Modelo Binomial – Ejemplo:

Un club nacional de automovilistas comienza una campaña telefónica con el propósito de aumentar el número de miembros. Con base en experiencia previa, se sabe que una de cada 20 personas que reciben la llamada se une al club. Si en un día 25 personas reciben la llamada telefónica ¿cuál es la probabilidad de que por lo menos dos de ellas se inscriban al club?

$$p(x \geq 2) = 1 - p(x \leq 1)$$

$$p(x \geq 2) = 1 - \left[C_{25}^{0} (0,05)^{0} (0,95)^{25} + C_{25}^{1} (0,05)^{1} (0,95)^{24} \right]$$

$$p(x \geq 2) = 0,3576$$

x=0,1,2,...,25

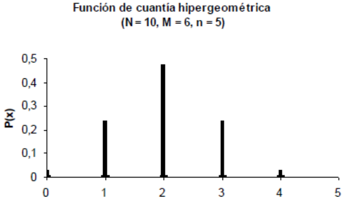
Variables aleatorias discretas

Modelo Hipergeométrico: se realizan n pruebas sin reemplazo. La variable aleatoria se define como número de éxitos observados en las n pruebas no independientes.

$$P(x) = \frac{C_M^x . C_{N-M}^{n-x}}{C_N^n} \qquad x = 0,1,2,3...,\min(n, M)$$

$$E(x) = n\frac{M}{N}$$

$$V(x) = n\ \frac{M}{N}\ \frac{N-M}{N}\ \frac{N-n}{N-1}$$

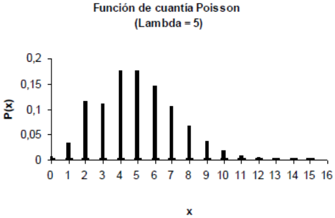


Variables aleatorias discretas

Modelo Poisson: se realizan n pruebas independientes a una velocidad constante en el tiempo o en el espacio. La variable aleatoria se define como número de éxitos en un intervalo de longitud fija.

$$p(x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \qquad x = 0,1,2,3,\ldots$$

$$E(x) = \lambda$$
$$V(x) = \lambda$$



Una variable con distribución Binomial se aproxima a Poisson cuando la muestra es grande y la probabilidad de existo es pequeña.

$$p(x) = C_n^x p^x (1-p)^{1-x}$$

$$Lim\ p(x) = \frac{e^{-\lambda} \lambda^{x}}{x!} \qquad x = 0,1,2,\ldots$$

$$Si\ \mathrm{p} \rightarrow 0 \ , \quad \mathrm{n} \rightarrow \infty$$

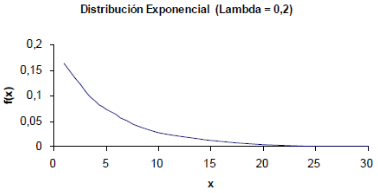
Variables aleatorias continuas

Distribución Exponencial: Cuando se quiere modelar el lapso en un intervalo (de tiempo o espacio) entre dos eventos Poisson, los que se presentan en forma independiente y a una frecuencia constante.
O cuando se tiene una variable con comportamiento exponencial

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

$$E(x) = \frac{1}{\lambda}$$

$$V(x) = \frac{1}{\lambda^2}$$



Variables aleatorias continuas

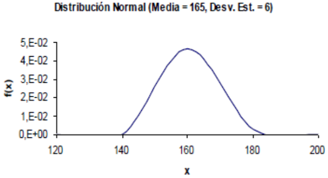
Distribución Normal:

• Numerosas distribuciones empíricas tienen aprox. distribución normal
• Sirve como una buena aproximación para las variables discretas.
•La distribución del estadístico media muestral converge a la normal

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

$$E(x)=\mu$$

$$V(x)=\sigma^2$$



Variables aleatorias continuas

Distribución Normal estandarizada

Se trabaja con la variable z, que es la variable aleatoria normal estandarizada

$$z = \frac{x - \mu}{\sigma}$$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

$$E(z) = 0$$

$$V(z) = 1$$

Variables aleatorias continuas

Distribución Normal - Ejemplo:
Se conoce que la inteligencia medida (CI) por medio de pruebas es normal con media de 100 y desviación estándar de 10. ¿cuál es la probabilidad de que un alumno seleccionado al azar tenga un CI mayor a 100?

$$P(x > 100) = 1 - F(100)$$

$$P(x > 100) = 1 - \int_{0}^{100} \frac{1}{10\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-100)^2}{10^2}} \qquad 0<x<100$$

$$P(z > (\frac{100\text{-}100}{10})) = P(z > 0) = 0{,}50$$

Gracias!!!