Distribuciones de Probabilidad

Distribuciones Continuas

Edimer David Jaramillo - Bioestadística 1

Marzo de 2019

Distribución Exponencial (1/4)

- De gran utilidad para modelar fenómenos relacionados con tiempos de espera.
- Múltiples aplicaciones:
    - Calcular la probabilidad de que un instrumento electrónico falle en determinado período de tiempo.
    - El tiempo necesario para que ocurra un accidente de tránsito en una ruta en una vía con probabilidad P.
    - El tiempo que puede transcurrir en un servicio de urgencias para que llegue el próximo paciente.
    - Teoría de colas
    - Problemas de confiabilidad
- En un proceso Poisson donde se repite sucesivamente un experimento a intervalos de tiempo iguales, el tiempo que transcurre entre dos sucesos sigue un modelo probabilístico exponencial.

Distribución Exponencial (2/4)

Una variable aleatoria X tiene distribución Exponencial si su función de densidad es:

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

Notación:

$$X \sim Exp(\lambda)$$

- **Esperanza Matemática:**

$$E[X] = \frac{1}{\lambda}$$

- **Varianza:**

$$Var[X] = \frac{1}{\lambda^2}$$

Distribución Exponencial (3/4)

- **Ejemplo:** de registros históricos se sabe que en promedio, un rayo causa la muerte a tres personas cada año en determinado país. Obtener:
  - La probabilidad de que el tiempo hasta la próxima muerte sea menor a un año.
  - La probabilidad de que el tiempo hasta la próxima muerte sea mayor a 18 meses.

Funciones importantes:

- `rexp()`: generar números bajo la distribución Exponencial
- `dexp()`: función de densidad
- `pexp()`: probabilidad acumulada.
- `qexp()`: obtener cuantiles

Distribución Exponencial (4/4)

```
#P(X < 1)
pexp(q = 1, rate = 3, lower.tail = TRUE)
```

```
## [1] 0.9502129
```

```
#P(X > 1.5)
pexp(q = 1.5, rate = 3, lower.tail = FALSE)
```

```
## [1] 0.011109
```

```
#P(X > 1.5)
1 - pexp(q = 1.5, rate = 3, lower.tail = TRUE)
```

```
## [1] 0.011109
```

Distribución Normal (1/7)

- De gran utilidad para múltiples fenómenos de la vida real
  - Agronómicos
  - Biológicos
  - Químicos
  - Físicos
  - Antropológicos
- Centralizada en la media
- La curva tiene su máximo absoluto en μ
- La curva es simétrica a través de μ
- Se aproxima al eje horizontal sin tocarlo (curva asintótica)
- El área total bajo la curva es 1

Distribución Normal (2/7)

Una variable aleatoria X tiene distribución Normal si su función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \ con \ \sigma > 0$$

Notación:

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

- **Esperanza Matemática:**

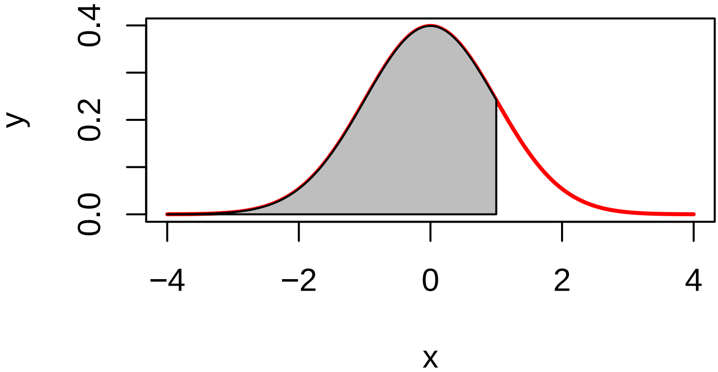
$$E[X] = \mu$$

- **Varianza:**

$$Var[X] = \sigma^2$$

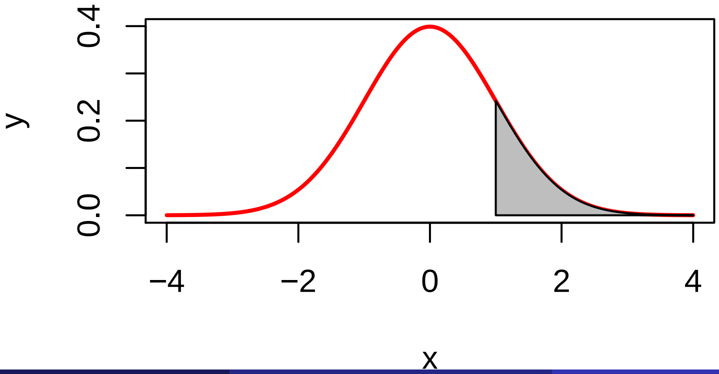
Distribución Normal (3/7)

- Obtener la probabilidad de que una variable aleatoria X tome valores menores a un valor determinado.



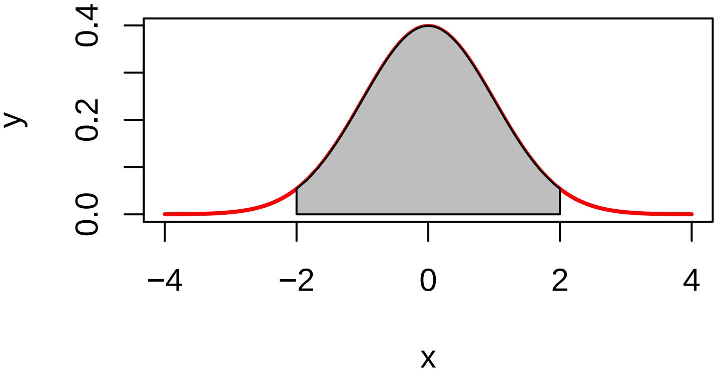
Distribución Normal (4/7)

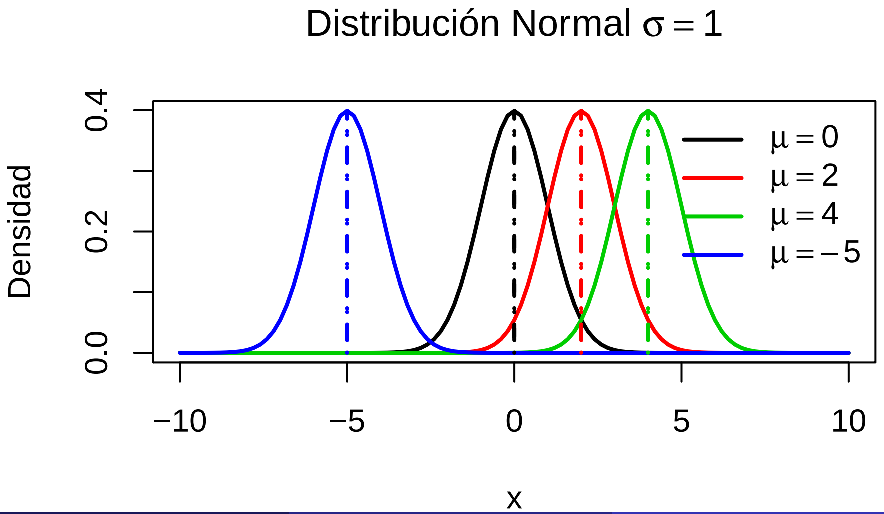
- Obtener la probabilidad de que una variable aleatoria X tome valores mayores a un valor determinado.



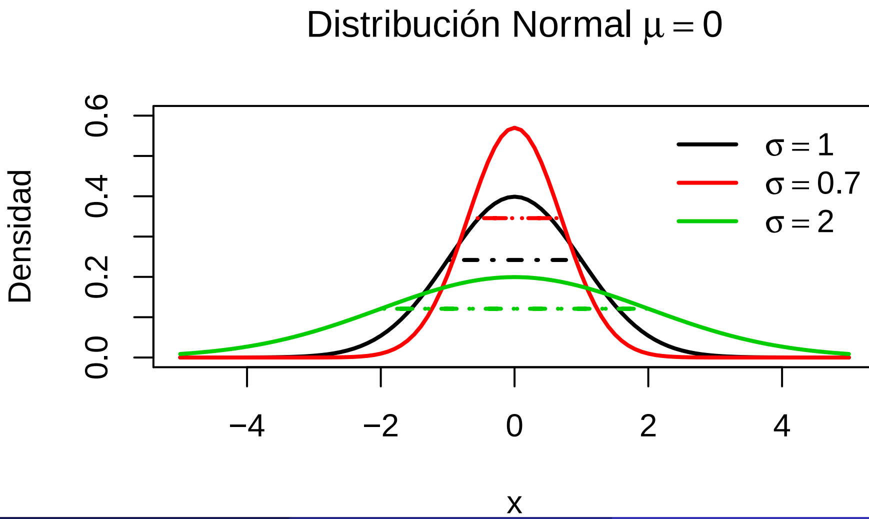
Distribución Normal (5/7)

- Obtener la probabilidad de que una variable aleatoria X tome valores entre dos valores determinados.




Distribución Normal σ = 1
Densidad
0.0
0.2
0.4
−10
−5
0
5
10
x
μ = 0
μ = 2
μ = 4
μ = −5

Distribución Normal (7/7)



Distribución Normal Estándar (1/7)

Notación:

$$Z \sim N(\mu = 0,\ \sigma^2 = 1)$$

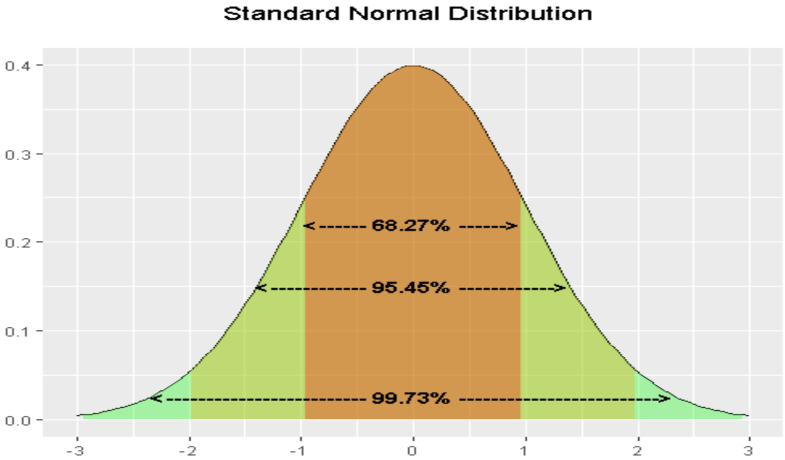
- **Estandarización:**

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

- **Tabla Z:**

Tabla Z - Distribución normal estándar.

Distribución Normal Estándar (2/7)



Distribución Normal Estándar (3/7)

- **Ejemplo:** Si X sigue una distribución normal con media igual a 10 y sigma igual a 2. ¿Cuál es la probabilidad de que la medida de la variable aleatoria X esté entre 9 y 11?

Distribución Normal Estándar (4/7)

1. **Estandarizar**

$$P(9 < X < 11) = P(\frac{9-10}{2} < \frac{x-10}{2} < \frac{11-10}{2})$$
$$P(-0.5 < z < 0.5) = P(z < 0.5) - P(z < -0.5) = 0.38292$$

Distribución Normal Estándar (5/7)

0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	$\boxed{0.6915}$	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611
0.8	0.7881	0.7910
0.9	0.8159	0.8186
1.0	0.8413	0.8438

Distribución Normal Estándar (6/7)

-1.0	0.1587	0.1562
-0.9	0.1841	0.1814
-0.8	0.2119	0.2090
-0.7	0.2420	0.2389
-0.6	0.2743	0.2709
-0.5	0.3085	0.3050
-0.4	0.3446	0.3409
-0.3	0.3821	0.3783

$$0.6915 - 0.3085 = 0.383$$

Distribución Normal Estándar (7/7)

2. **Con R**

Funciones importantes:

- `rnorm()`: generar números bajo la distribución Exponencial
- `pnorm()`: función de densidad (probabilidades)
- `qnorm()`: obtener cuantiles

```
pz_menor0.5 <- pnorm(q = 0.5)
pz_menor_menos0.5 <- pnorm(q = -0.5)
pz_menor0.5 - pz_menor_menos0.5
## [1] 0.3829249
```

Distribución Chi Cudrado (χ^2)

- Llamada también *ji cuadrada (o)* o distribución de Pearson
- Aplicación considerable en la teoría y metodología estadística
- Componente importante de las pruebas de hipótesis e inferencia estadística
- Relacionada con las distribuciones T de Student y F de Snedecor

Una variable aleatoria X tiene distribución chi cuadrado si su función de densidad es:

$$\frac{1}{2^{\frac{k}{2}}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$$

Notación:

$$X \sim \chi^2(k)$$

- **Esperanza Matemática y Varianza:**
- $E[X] = k$
- $Var[X] = 2k$

Distribución de estadísticos muestrales

Muestreo

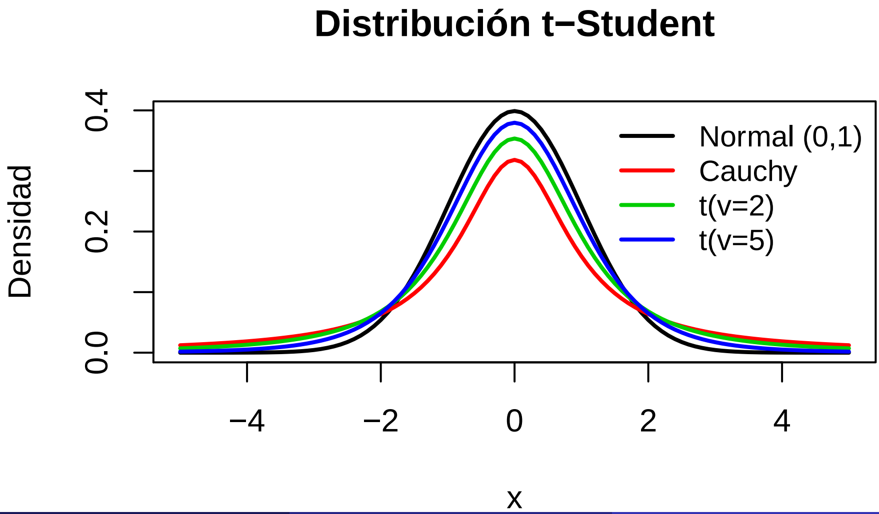
- El muestreo tiene como objetivo inferir propiedades de una población a partir de una fracción de ella, conocida como muestra.
- Desde el punto de vista estadístico el objetivo es conocer los parámetros de la distribución de la variable de interés.
- Los estadísticos muestrales sirven como aproximación (estimación) de los parámetros que caracterizan la distribución.
- Los estadísticos son desconocidos, por tanto, se consideran variables aleatorias y como tales, tienen una distribución asociada.

Distribuciones derivadas del muestreo

Distribución t de Student (1/2)

- Deriva de la distribución normal
- Surge con la dificultad de estimar la media de una población con distribución normal cuando el tamaño de muestra es pequeño.
- Notación: $X \sim t(v)$
- Es de media cero y varianza $\frac{v}{v-2}$; *con* $v > 2$
- Simétrica respecto a la media
- La varianza decrece hasta uno cuando el número de grados de libertad aumenta
- Tabla t

Distribución t de Student (2/2)



Distribución F

- Llamada la distribución F de Fisher o F de Snedecor.
- Generalmente es la distribución nula de una prueba estadística (análisis de varianza)
- Notación: $X \sim F_(k_1,\ k_2)$
- Útil en comparación de varianzas
- Tabla F

La distribución F surge como resultado de la siguiente operación entre variables aleatorias:

$$F = \frac{Y_1/k_1}{Y_2/k_2}$$

Donde:

- Y_1 y Y_2 siguen una distribución χ^2 con k_1 y k_2 grados de libertad, respectivamente.
- Y_1 y Y_2 son independientes

Ejercicios (1/5)

En un experimento de laboratorio se utilizan 10 gramos de X compuesto químico, se conoce que la duración media de un átomo de esta materia es de 140 días; obtener:

1. La probabilidad de que el compuesto desaparezca máximo a los 100 días.
2. La probabilidad de que el compuesto desaparezca en mínimo 50 días
3. Los días que transcurren hasta que haya desaparecido el 90% de este material.

```
## [1] 0.5104583
```

```
## [1] 0.6996725
```

```
## [1] 322.3619
```

Ejercicios (2/5)

- Se ha comprobado que el tiempo de vida útil de cierto tipo de marcapasos sigue una distribución exponencial con media de 16 años. Obtener:
  - La probabilidad de que a un paciente al que se le ha implantado el marcasos, deba reemplazarlo por otro antes de 20 años.
  - La probabilidad de que haya que cambiar el marcapasos máximo a los 25 años en un paciente que tiene el implante hace 5 años.
  - El tiempo transcurrido hasta que exista una probabilidad máximo del 80% de reemplazarlo.

```
Respuesta_A
## [1] 0.7134952
Respuesta_B
## [1] 0.5220042
Respuesta_C
## [1] 25.75101
```

Ejercicios (3/5)

- Asuma que la variable aleatoria Z sigue una distribución normal estándar, obtener:
  1. $P(Z \leq 1.37)$ (0.9146565)
  2. $P(Z \leq -0.86)$ (0.1948945)
  3. $P(-1.25 \leq Z \leq 0.37)$ (0.538659)
  4. $P(Z > -1.23)$ (0.8906514)

Ejercicios (4/5)

- El caudal de un canal de riego medido en m^3/seg es una variable aleatoria con distribución aproximadamente normal con media $3m^3/seg$ y desviación estándar de $0.8m^3/seg$. Con esta información, obtener la probabilidad de los siguientes eventos:
  - **Evento A:** que el caudal en un momento dado sea máximo de $2.4m^3/seg$. (0.2266274)
  - **Evento B:** que el caudal en un momento dado esté entre $2.8\ y\ 3.4m^3/seg$. (0.2901688)
  - **Evento C:** que el caudal en un momento dado sea al menos de $2m^3/seg$. (0.8943502)

Ejercicios (5/5)

- El día de floración de una hortaliza (en escala de 1 -365 días) se puede modelar con una distribución normal centrada en el 18 de agosto (día 230) y con desviación estándar de 10 días. Si desde la fecha de la floración hasta la cosecha hay un lapso de 25 días, obtener:
  1. La proporción de la cosecha que se habrá generado para el 16 de septiembre (día 259).
  2. Si se considera **primicia** a los frutos obtenidos antes del 1 de septiembre (día 244), ¿qué proporción de la cosecha se espera sea primicia?
  3. Si se considera **tardío** a los frutos obtenidos después del 20 de septiembre (día 263), ¿cuál es la proporción esperada de frutos tardíos?

```
Respuesta_A
## [1] 0.6554217
Respuesta_B
## [1] 0.1356661
Respuesta_C
## [1] 0.2118554
```