# 5. DISTRIBUCIONES DE PROBABILIDADES


**Dr. Edgar Acuña**

**http://academic.uprm.edu/eacunaf**

**UNIVERSIDAD DE PUERTO RICO**

**RECINTO UNIVERSITARIO DE MAYAGUEZ**

# DISTRIBUCIONES DE PROBABILIDADES

Se introducirá el concepto de **variable aleatoria**, cuya importancia radica en introducir modelos matemáticos en el cálculo de probabilidades. Luego, se considerarán las distribuciones de probabilidades de variables aleatorias discretas con su media y varianza respectiva. Se discutirá en detalle la distribución binomial Se considerará el estudio de la distribucion Normal, la cual es de crucial importancia para el proceso de Inferencia Estadística.

# 5.1 Variables Aleatorias

Una variable aleatoria es aquella que asume sus valores de acuerdo a los resultados de un experimento aleatorio. Usualmente se representa por las últimas letras del alfabeto: X, Y o Z. Una variable aleatoria X es una función cuyo dominio es la colección de eventos del espacio muestral S y cuyo rango $R_x$, es un subconjunto de los números reales.

Ejemplos de variables aleatorias:

- X: La suma que aparece al lanzar un par de dados.
- Y: El número de caras que aparecen al lanzar una moneda tres veces.
- Z: El número de errores que se encuentran en la página de un libro.
- T: El tiempo de vida de la componente de un sistema
- W: El tiempo de espera para ser atendido en un banco

# Ejemplo 5.1

De una caja que contiene 5 bolas numeradas del 1 al 5 se extraen 3 bolas una por una y sin reposición. Entonces **X**: El mayor de los tres números sacados, es una variable aleatoria.

El espacio muestral es:

S = {(1,2,3), (1,2,4), (1,2,5), (1,3,4), (1,3,5), (1,4,5), (2,3,4), (2,3,5), (2,4,5), (3,4,5)}

y la variable aleatoria X asume los valores: 3, 4 y 5. Por ejemplo,

$$X(2,3,4) = 4$$

Si el rango de valores $R_x$ de la variable aleatoria X es finito o infinito enumerable entonces se dice que es una **variable aleatoria discreta**. Si su rango de valores $R_x$ es infinito no enumerable entonces se dice que es una **variable aleatoria continua**.

# 5.1.1. Función de probabilidad de una variable aleatoria discreta

Si X es una variable aleatoria discreta con rango de valores $R_x$ entonces, su función de probabilidad se define por:

$p(x) = P[X = x]$, para todo x  $R_x$

y tiene las siguientes propiedades:

- $p(x) > 0$ y
- $\Sigma\, p(x) = 1$, $x \in R_x$

Cuando Rx no contiene muchos valores es más conveniente expresar p(x) en una tabla de valores, la cual es llamada tabla de función de probabilidad.

# Ejemplo 5.4.

De un lote que contiene 10 artículos, de los cuales 4 son dañados se extraen al azar y sin reposición 3. Se define la variable aleatoria X: Número de artículos dañados que hay en la muestra. Hallar la función de probabilidad de X.

**Solución:** En este caso el rango de valores de X es Rx = {0, 1, 2, 3} y en particular.

p(2) = Prob(sacar 2 dañados) = $\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}$ , y en general p(x) = $\frac{\binom{4}{x}\binom{6}{3-x}}{\binom{10}{3}}$, para x = 0,1,2,3.

Las combinaciones se muestran en la siguiente tabla de función de probabilidad:

| X | p(x) |
|---|---|
| 0 | 1/6 |
| 1 | ½ |
| 2 | 3/10 |
| 3 | 1/30 |

# 5.1.2. Función de distribución acumulativa

Sea X una variable aleatoria discreta con función de probabilidad p(x) y rango de valores $R_x$, entonces su función de distribución acumulativa se define por:

$$F(t) = P(X \leq t) = \sum_{x \leq t} p(x)$$

t es cualquier número real. En particular, si t es un valor que está en Rx , el cual consiste de enteros no negativos, entonces:

F(t) = p(0) + p(1) + p(2) + p(3) +…+ p(t)

**Ejemplo 5.5.** Hallar la función de distribución acumulativa para el Ejemplo anterior.

**Solución:**

| x | F(x) |
|---|---|
| 0 | 1/6 |
| 1 | 4/6 |
| 2 | 29/30 |
| 3 | 1 |

La gráfica de una función de distribución acumulativa es creciente y del tipo escalonado, con saltos en los puntos que están en el rango de valores y cuya magnitud es igual al valor de la función de probabilidad en dicho punto. Más formalmente tiene la siguiente propiedad:

**Propiedad.** *La relación entre la función de distribución de probabilidad y la función de distribución acumulativa está dada por:*

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

para todo valor de x en el rango de valores de la variable aleatoria.

# 5.1.3 Valor Esperado y Varianza de una Variable Aleatoria Discreta

Sea X una variable aleatoria discreta con función de probabilidad $p(x)$ y rango de valores Rx, entonces su Valor Esperado o Media se define como el número:

$$\mu = E(X) = \sum_{x} xp(x)$$

La suma es sobre todos los valores $x$ que están en $R_x$.

La **Varianza** de una variable aleatoria discreta x con función de probabilidad p(x) y media μ se define por:

$$\sigma^2 = \sum (x-\mu)^2 p(x)$$

Donde la suma es sobre todos los valores del rango de X.

**Ejemplo 5.10.** Hallar la media y varianza del número de artículos dañados del Ejemplo 5.4.

**Solución:**

| x | p(x) | Xp(x) | X-$\mu$ | $(x-u)^2p(x)$ |
|---|---|---|---|---|
| 0 | 1/6 | 0 | -1.2 | .24 |
| 1 | ½ | ½ | -0.2 | .02 |
| 2 | 3/10 | 6/10 | 0.8 | .192 |
| 3 | 1/30 | 1/10 | 1.8 | .121 |
| | | $\mu$=1.2 | | $\sigma^2$=0.573 |

Otra formas del calcular la varianza es $\sigma^2 = \sum x^2p(x)-\mu^2$.

# 5.2 La Distribución Binomial.

Un experimento es llamado de Bernoulli, si satisface las siguientes características:

- En cada repetición puede ocurrir sólo una de dos maneras, una de ellas es llamada *Exito* y la otra *Fracaso*.
- La probabilidad de *Exito*, representada por $p$, debe permanecer constante cuando el experimento es repetido muchas veces.
- Las repeticiones de los experimentos deben ser independientes entre sí.

La función de probabilidad de una binomial es:

para $x = 0, 1, \ldots, n$.

El valor de $p(x)$ para diversos valores de $n$ y $p$ aparece en tablas de todo texto básico de Estadística.

$$p(x) == \binom{n}{x} p^{x} (1-p)^{n-x}$$

# Calculos en Minitab para la funcion de distribucion Binomial

En **MINITAB** se pueden calcular la función de probabilidad (***Probability***), la función de distribución acumalada (***Cumulative probability***) y los percentiles ***(Inverse cumulative probability***) de la distribución Binomial para cualquier valor de $n$ y $p$. Para esto hay que seguir la secuencia **Calc** ▸ **Probability Distributions** ▸ **Binomial**.

# Ejemplo 5.13

- Un estudiante que responde al azar un examen tipo selección múltiple que consiste de 10 preguntas, cada una con 5 alternativas de las cuales sólo una es correcta. Calcular la probabilidad de que el estudiante:

a)Tenga exactamente 3 preguntas buenas.

b)Tenga 6 ó menos preguntas buenas.

c)Tenga por lo menos 4 buenas.

# Ejemplo 5.14

- La prueba ELISA es usada para detectar la presencia de anticuerpos al virus del SIDA.  ELISA,  detecta que hay anticuerpos presentes en el 95 por ciento de  los casos de que la muestra de sangre está contaminada con el virus del SIDA.  Suponga que entre las muchas muestras que pasan por un Banco de Sangre hay 12 que están contaminadas con SIDA. Cual es la probabilidad de que ELISA
- a)  detecte 9 de estos casos?
- b)  detecte por lo menos 2 de estos casos?
- c)  5 de los casos NO sean detectados ?
- d)  Por lo menos 4 casos no sean detectados por ELISA?
-

# Ejemplo 5.15

El Departamento de Salud ha determinado que el 10% de los puertorriqueños son zurdos.  Se  elige al azar 9 estudiantes de una escuela en Puerto Rico. ¿Cuál es la probabilidad de que:

- Exactamente 2 de ellos sean zurdos?
- No mas de 5 sean zurdos?
- Exactamente 6 de ellos sean diestros?
- Por lo menos 7 de ellos sean diestros?

# Media y Varianza de una Binomial

La media o valor esperado de una Binomial es np

La varianza de una Binomial es npq , donde q=1-p es la probabilidad de fracaso.

En el ejemplo 5.15 el numero esperado de zurdos es (9) ((.1)=.9 un zurdo entre 9

# 5.3 La Distribución Normal

Es llamada tambien Distribución Gaussiana en honor a K. Gauss, es una distribución del tipo continuo. Su comportamiento es reflejado por la Curva Normal que es la gráfica de la siguiente ecuación.

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

Donde la media $\mu$ y la desviación estándar $\sigma$ son los parámetros de la distribución.

Si una variable aleatoria X tiene una distribución Normal, para calcular la probabilidad de que X caiga entre dos valores a y b entonces, hallar el área debajo de la curva entre a y b, y se puede hacer por el proceso de Cálculo llamado Integración.

# Distribución Normal Estandar

A la distribucion Normal que tiene media 0 y desviacion estandar 1 se le llama normala estandar y se representa por Z

Las areas debado de la curva normal estandar estan en tablas.

# Ejemplo

Hallar las siguientes probabilidades:

a) $P\ (Z < 2)$
b) $P\ (Z > 1.54)$
c) $P\ (1.25 < Z < 2.10)$
d) $P\ (Z < -1.33)$
e) $P\ (Z>-.1.95)$
f) $P\ (-1.33 < Z < 2.15)$

# Calculos de la distribucion Normal en Minitab

En **MINITAB** se pueden calcular la función de densidad (***Probability density***), la función de distribución acumalada (***Cumulative probability***) y los percentiles ***(Inverse cumulative probability***) de la distribución Normal para cualquier valor de la media $\mu$ y desviación estándar $\sigma$. No se requiere transformación a una normal estándar. Para esto hay que seguir la secuencia **Calc** ▸ **Probability Distributions** ▸ **Normal**.

# Estandarización de una Normal

Dada una variable aleatoria X distribuida Normalmente con media $\mu$ y desviación estándar $\sigma$ entonces puede ser convertida a una normal estándar mediante el proceso de estandarización, definido por $Z = (X - \mu)/\sigma$, donde $X \sim N(\mu, \sigma 2)$. Además si Xp y Zp representen sus respectivos percentiles entonces:

$$Xp = \mu + \sigma Zp$$

# Fórmulas para calcular área debajo de la curva normal

F representa la distribución acumulada de la distribución Normal, es decir el área acumulada a la izquierda del valor dado

a) $P(X < a) = F(a)$

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

c) $P(X > b) = 1 - F(b)$

# Ejemplo 5.17

Si X es una población Normal con media $\mu = 70$ y $\sigma = 10$. Hallar las siguientes probabilidades:

a) $P(X < 60)$

b) $P(X > 95)$

c) $P(50 < X < 80)$

**d) $P(75<x<90)$**

# Ejemplo

Se estima que un conductor conduce un promedio de 12,400 millas al año, con una desviación estándar de 3,800 millas. Asumiendo normalidad del millaje recorrido. Calcular la probabilidad de que en el próximo año el conductor conduzca:

- Más 12,100 millas pero menos que 13,200 millas
- Más de 15,000 millas.

# Ejemplo

En el estudio Framingham acerca de factores que afectan las enfermedades cardíacas se hizo un seguimiento por un período de 16 años a una gran cantidad de hombres sanos. Se encontró que inicialmente la distribución de los niveles de colesterol de los hombres era Normal con media $\mu = 224$ y con desviación estándar $\sigma = 48$

a) Una persona con un colesterol menor de 200 es considerada como una con bajo riesgo de tener complicaciones cardíacas. ¿Qué porcentaje de hombres tendrán bajo riesgo?

b) Si el colesterol de la persona es mayor de 250 entonces tendrá problemas cardiacos en el futuro. ¿Qué porcentaje de hombres tendrán problemas cardiacos?

c) Los hombres que tienen el 5% más alto de colesterol serán sometidos a una dieta, para bajarle su colesterol y evitar que tenga problemas cardiacos en el futuro. ¿Cuál será el nivel de colesterol máximo permitido para NO someterse a la dieta?

# 5.4 Cotejando si hay Normalidad

Cuando se trata de sacar conclusiones acerca de la población usando los datos de la muestra, se asume generalmente que los datos de la población se distribuyen de forma normal. Como no se conocen todos los elementos de la población, se deben usar los datos de la muestra para verificar si efectivamente la población es Normal. Existen varias pruebas estadisticas para verificar Normalidad.

En **MINITAB**, primero se elige la opción ***Basic Statistics*** de **Stat** y luego ***Normality Test*** del submenú que aparece.

# Continuación…

Si los puntos caen cerca de una línea, entonces se dice que hay **Normalidad**. En **MINITAB** este plot es obtenido siguiendo la secuencia **Graph** ▸ **Probability Plot**. En la ventana que aparece elegir la opcion ***Single*** como se muestra en la Figura 5.6

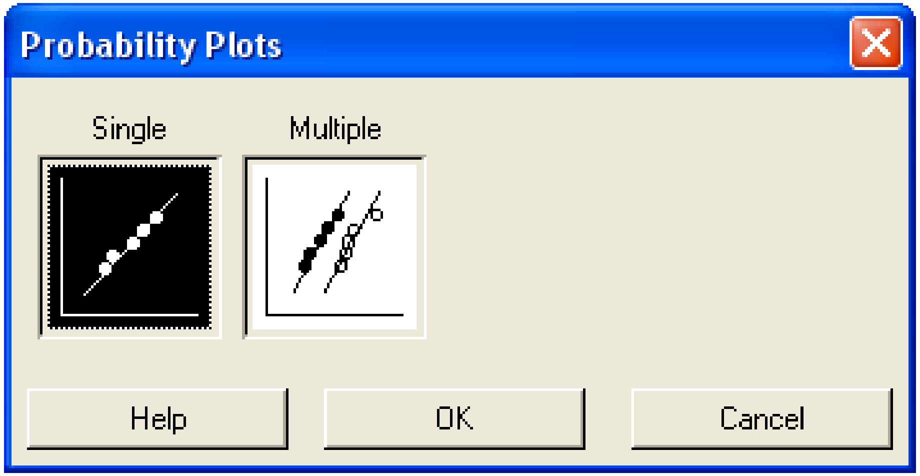


Figura 5.6. Ventana de dialogo de Probability Plots.

# 5.5 Simulando datos de una distribución conocida

Muchas veces se hace difícil conseguir datos reales para corroborar un método estadístico, una manera de resolver dicho problema es hacer que la computadora produzca mediante simulación dichos datos.

**MINITAB** tiene una lista grande de distribuciones conocidas, que pueden ser simuladas, esta lista se puede ver seleccionando **Random Data** en el menú **Calc**.

# Ejemplo 5.22.

Supongamos que deseamos simular 30 notas de una población normal que tiene media 70 y desviación estándar 10.

**Solucion:**

Los datos aparecen con 4 decimales, pero si se elige la opción ***Format column*** del menú **Editor**, se puede definir que el número de decimales sean cero para que los datos salgan enteros. Los datos generados aparecen en la ventana **session** como sigue:

```
Data Display

C1
        80        80        77        75        54        69        53
        79        81        64        73        64        69        84
        60        95        71        63        58        65        79
        69        64        87        75        95        58        68
        63        81
```