

5. DISTRIBUCIONES DE PROBABILIDADES

Dr. Edgar Acuña

<http://math.uprm.edu/~edgar>

**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 distribución 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:

$$p(x) = P[X = x], \text{ para todo } x \in R_x$$

y tiene las siguientes propiedades:

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

Cuando R_x 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 $R_x = \{0, 1, 2, 3\}$ y en particular.

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

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

X	p(x)
0	1/6
1	1/2
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 R_x , el cual consiste de enteros no negativos, entonces:

$$F(t) = p(0) + p(1) + p(2) + p(3) + \dots + p(t)$$

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 R_x , 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-μ	(x-u) ² p(x)
0	1/6	0	-1.2	.24
1	1/2	1/2	-0.2	.02
2	3/10	6/10	0.8	.192
3	1/30	1/10	1.8	.121
		μ=1.2		σ ² =0.573

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

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 $x = 0, 1, \dots, n$.

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 acumulada (*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. Haciendo uso de MINITAB

Expresar en una tabla de valores la función de probabilidad y la función de distribución acumulada de la variable aleatoria X : Número de preguntas bien contestadas por un estudiante que responde al azar un examen tipo selección múltiple que consiste de 10 preguntas, cada una con 4 alternativas de las cuales sólo una es correcta.

Solución:

Poner en una columna, llamada 'x', todos los valores posibles de la variable. La ventana session en Minotab muestra los siguientes resultados

Data Display

Row	x	P (x)	F (x)
1	0	0.056314	0.05631
2	1	0.187712	0.24403
3	2	0.281568	0.52559
4	3	0.250282	0.77588
5	4	0.145998	0.92187
6	5	0.058399	0.98027
7	6	0.016222	0.99649
8	8	0.000386	0.99997
9	9	0.000029	1.00000
10	10	0.000001	1.00000

5.3 La Distribución Normal

Es llamada también 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 μ y la desviación estándar σ 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.

Calculos de la distribución Normal en Minitab

En **MINITAB** se pueden calcular la función de densidad (*Probability density*), la función de distribución acumulada (*Cumulative probability*) y los percentiles (*Inverse cumulative probability*) de la distribución Normal para cualquier valor de la media μ y desviación estándar σ . 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 μ y desviación estándar σ 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 X_p y Z_p representen sus respectivos percentiles entonces:

$$X_p = \mu + \sigma Z_p$$

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) $P(50 < X < 80)$

Solución:

Usando **MINITAB** con **mean** = 70 y **standard deviation** = 10, se tiene que:

- a) $P(X < 60) = F(60) = .1587$
- b) $P(X > 95) = 1 - F(95) = 1 - .9938 = .0062$
- c) $P(50 < X < 80) = F(80) - F(50) = .8413 - .0228 = .8185$

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 estadísticas 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 opción *Single* como se muestra en la Figura 5.6

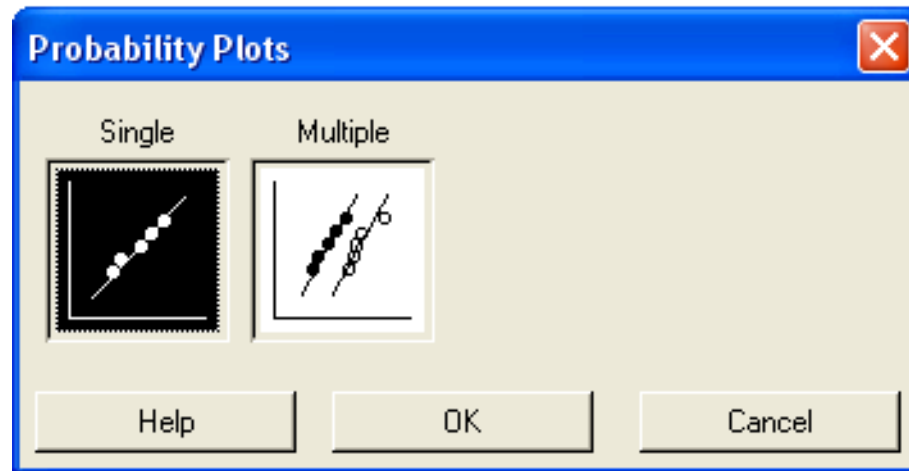


Figura 5.6. Ventana de dialogo de Probability Plots.

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					