

ANÁLISIS ESTADÍSTICO DE DATOS

USANDO MINITAB

Tercera Edición

EDGAR ACUÑA FERNANDEZ

**UNIVERSIDAD DE PUERTO RICO
RECINTO UNIVERSITARIO DE MAYAGUEZ
DEPARTAMENTO DE MATEMATICAS
e-mail:edgar@math.uprm.edu
homepage:math.uprm.edu/~edgar**

©2010 Derechos Reservados por Edgar Acuña. Prohibida su reproducción sin permiso del autor

MINITAB es marca registrada de Minitab Inc.

WINDOWS es marca registrada de Microsoft Corporation.

*A mi familia de Puerto Rico y Perú
por todo su apoyo*

PREFACIO

Muchas personas piensan que lo que se estudia en un curso introductorio de Estadística es simplemente hacer gráficas, construir tablas numéricas y reportar algunos resultados de encuestas. A todo lo anterior los estudiantes a nivel universitario le añaden el cálculo de un gran número de fórmulas y el uso de varias tablas que contienen una gran cantidad de números; y en consecuencia no ven los beneficios de tomar el curso. El objetivo general de este texto es mostrar al estudiante que tomar un curso de Estadística es de gran utilidad en su formación profesional, mas hoy en día cuando la tecnología nos permite guardar una gran variedad de información y el acceso a ella es cada vez mas fácil.

La definición más simple de la Estadística es que, es la ciencia donde se aprende a partir de los datos tomados. Aquí aprender se está usando en el sentido de sacar conclusiones acerca de la población a partir de la muestra tomada. Por lo tanto; es de vital importancia que la muestra que se está tomando tenga propiedades óptimas; es decir, que la distribución de los datos sea simétrica con un pico en el centro, tenga poca variabilidad y no posea valores atípicos. Lamentablemente, la mayoría de los textos introductorios de Estadística dedican pocas páginas a analizar la bondad de los datos recolectados para sacar conclusiones confiables. Por consiguiente, casi nada se menciona de las consecuencias de usar una muestra que tiene malas características.

Lo que persigue este texto es que el estudiante dedique más tiempo a la interpretación de las gráficas y de los resultados obtenidos por computadoras para poder sacar sus conclusiones y no pierda tiempo memorizando y evaluando fórmulas usando una calculadora, que aunque se han mejorado muchísimo no tienen la capacidad de una computadora. En muchas partes del texto se mostrará al estudiante como detectar la presencia de datos anormales y evaluar el efecto producido por ellos, pero los métodos para remediar estos problemas no serán discutidos aquí, pues son material de cursos de Estadística de nivel más avanzado.

El estudiante debe tomar en cuenta que tomar un curso de estadística es distinto a tomar una clase de matemáticas. La respuesta a una pregunta en matemáticas es única, mientras que en estadística, el estudiante debe aplicar muchas veces su propio criterio en el momento de llegar a una conclusión. En casi todos los ejemplos del texto se ha tratado de dar una interpretación de los resultados para que el estudiante aprenda, emulándolos.

El programa estadístico usado en este texto es **MINITAB**, uno de los de mayor antigüedad y uno de los pocos que ha sobrevivido al paso del tiempo. Comenzó a usarse en 1972 en la Universidad del Estado de Pensylvania. **MINITAB** es muy fácil de aprender y es el paquete estadístico más usado a nivel mundial en la enseñanza de los cursos introductorios de Estadística. Está disponible para distintos tipos de computadoras y diversos sistemas operativos.

Este texto sólo se refiere al uso de **MINITAB** en computadoras personales bajo el sistema operativo WINDOWS XP, pero su uso es similar en Windows 2000, 95/98 y Windows NT. Aquí se incluye el material que se usa en un curso introductorio de Estadística de dos semestres, el cual sólo requiere conocimiento de Precálculo. Los primeros 6 capítulos del texto y las dos primeras secciones del capítulo 7 son cubiertos en el primer semestre, el material restante es cubierto en el segundo semestre. El autor ha usado versiones de este texto en la Universidad de Puerto Rico, Recinto de Mayaguez durante los últimos 15 años y ha usado **MINITAB** desde 1983.

Este texto es más que un manual de **MINITAB**, porque si bien se explica el uso del mismo para la mayoría de los procedimientos estadísticos, se dedica muchas páginas a la explicación de los métodos a ser usados y a la interpretación de los resultados. En algunos casos se trata de explicar el origen de los métodos y en otros no, debido más que todo a que no se requiere un curso de cálculo como pre-requisito. Por otro lado; este texto no cubre todos los aspectos de **MINITAB**. Las opciones de Estadística multivariada, Series de tiempo, Control de calidad y algunas otras más no son tratadas en este texto.

Cambios con respecto a la segunda edición

En esta edición se ha usado la versión 14 de **MINITAB**. Sin embargo; la mayoría de las instrucciones también se aplican a la versiones 12 y 13 usadas en las dos ediciones anteriores. Los principales cambios son en la parte gráfica del capítulo 3. También se han modificado algunas ventanas de diálogo a través de la mayoría de los capítulos.

Por otro lado se ha incrementado las explicaciones de los métodos estadísticos y se ha añadido más ejercicios al texto. Se ha incluido también una tabla de la normal estándar y un índice de términos usados en el texto

Conjunto de Datos

Los conjuntos de datos usados en el texto están disponible en la página de internet del libro, cuya dirección es www.math.uprm.edu/~edgar/libro.html . En caso de tener problemas de acceder la página de internet, mandar un mensaje electrónico a cualquiera de las siguientes direcciones: edgar.acuna@upr.edu, edgar@math.uprm.edu ó edgar@cs.uprm.edu.

Los archivos de datos están en formato ASCII para que puedan ser usados también con otro tipo de programa estadísticos o con EXCEL.

Soluciones de ejercicios.

Las soluciones de los problemas impares del texto están disponibles en la página de internet del curso.

Agradecimientos

El autor desea agradecer profundamente a la Dra. Frida Coaquira Nina, por sus valiosas contribuciones al contenido y presentación de este texto, a la profesora Caroline Rodríguez por sus contribuciones a la parte gráfica del capítulo 3, y al profesor Jose Neville Diaz quién detectó varios errores en el texto.

Finalmente, el autor también desea agradecer el apoyo de **MINITAB** Inc. al proveerle el software en forma gratuita. Para contactar a **MINITAB** escribir a la dirección 3081 Enterprise Drive, State College, PA 16801 USA, enviar un e-mail a sales@minitab.com, o acceder su página de internet: www.minitab.com.

CONTENIDO

	Página
1. INTRODUCCIÓN.....	1
1.1 Conceptos Estadísticos Básicos.....	1
1.2 Definición de la Estadística.....	2
1.3 División de la Estadística.....	2
1.4 Tipos de Datos.....	3
1.5 Técnicas de Muestreo.....	4
1.6 Maneras de recolectar datos.....	5
2. UNA INTRODUCCIÓN A MINITAB	6
2.1 Entrando a MINITAB	6
2.2 Las Ventanas de MINITAB.....	6
2.3 La Barra de Menús.....	8
2.3.1 El menú File.....	9
2.3.2 El menú Edit.....	10
2.3.3 El menú Data.....	11
2.3.4 El menú Calc.....	14
2.3.5 El menú Stat.....	16
2.3.6 El menú Graph.....	18
2.3.7 El menú Editor.....	19
2.3.8 El menú Tools.....	21
2.3.9 El menú Window.....	22
2.3.10 El menú Help.....	23
2.4 Entrada y manipulación de datos en MINITAB	25
2.4.1 Entrando los datos usando la hoja de trabajo.....	25
2.4.2 Entrando los datos usando comandos en la ventana session.....	27
2.5 Mostrando los datos.....	27
2.6 Guardando los datos.....	28
2.7 Abriendo un archivo de datos.....	28
2.8 Imprimiendo los datos.....	29
2.9 Usando Proyectos.....	29
3. ESTADÍSTICA DESCRIPTIVA.....	31
3.1 Organización de datos Cuantitativos Discretos.....	32
3.1.1 Tabla de Frecuencias.....	32
3.1.2 El plot de puntos (Dotplot).....	34
3.1.3 Gráfica de Línea.....	36
3.2. Organización de datos Cuantitativos Continuos.....	37
3.2.1 Tablas de frecuencias-Histogramas en modo texto.....	37
3.2.2 Histograma en modo gráfico.....	38
3.3 Presentación de datos cualitativos.....	43
3.3.1 Gráficas de Barras.....	44
3.3.2 Gráficas Circulares.....	46
3.4 Gráfica de “Stem-and-Leaf” (Tallo y Hojas).....	49
3.5 Cálculo de Medidas Estadísticas.....	53
3.5.1 Medidas de Centralidad.....	53
3.5.2 Medidas de Variabilidad.....	56

3.5.3	Medidas de Posición.....	58
3.5.4	Cálculo de Medidas estadísticas Usando MINITAB	59
3.6	El Diagrama de Caja (“Boxplot”).....	62
3.7	Organización y presentación de Datos Bivariados.....	64
3.7.1	Datos Bivariados categóricos.....	64
3.7.2	Una variable es cualitativa y la otra cuantitativa.....	72
3.7.3	Datos Bivariados continuos.....	73
3.8	El Coeficiente de Correlación.....	75
3.9	Una introducción a Regresión Lineal.....	78
Ejercicios		82
4.	CONCEPTOS BÁSICOS DE PROBABILIDADES	90
4.1	Espacio Muestral y Eventos.....	90
4.1.1	Experimentos aleatorios y espacios muestrales.....	90
4.1.2	Eventos	92
4.1.3	Relaciones entre eventos.....	93
4.2	Métodos para asignar Probabilidades.....	95
4.2.1	Método Axiomático.....	95
4.2.2	Método Clásico.....	99
4.2.3	Método Frecuencial.....	101
4.2.4	Estimando la probabilidad de ocurrencia de un evento.....	102
4.2.5	Método Subjetivo.....	105
4.3	Probabilidad Condicional.....	105
4.3.1	La regla del producto.....	107
4.3.2	Probabilidad Total y la regla de Bayes.....	109
4.4	Eventos Independientes.....	115
4.5	Aplicación de Técnicas de Conteo al cálculo de Probabilidades.....	118
4.5.1	Regla Multiplicativa del conteo.....	118
4.5.2	Permutaciones.....	119
4.5.3	Combinaciones.....	121
Ejercicios		127
5.	DISRIBUCIONES DE PROBABILIDADES	130
5.1	Variables Aleatorias.....	130
5.1.1	Función de Probabilidad de una Variable Aleatoria Discreta	131
5.1.2	Función de Distribución Acumulativa.....	132
5.1.3	Valor Esperado y Varianza de una variable aleatoria discreta.....	134
5.2	Distribución Binomial.....	136
5.3	Distribución Normal.....	140
5.4	Cotejando si hay Normalidad.....	146
5.5	Simulando datos de una distribución dada.....	148
Ejercicios		150
6.	DISTRIBUCIONES MUESTRALES	153
6.1	Distribución de la Media Muestral cuando la población es Normal	153
6.2	El teorema del Limite Central	154
6.3	Distribución de la Proporción Muestral	158

Ejercicios	160
7. INFERENCIA ESTADÍSTICA	161
7.1 Inferencia acerca de la Media Poblacional (varianza conocida).....	162
7.2 Inferencia acerca de la Media Poblacional (varianza desconocida).....	172
7.3 Inferencia acerca de Proporciones.....	176
7.4 Inferencia acerca de la Varianza Poblacional.....	180
7.4.1 La distribución Ji-Cuadrado.....	180
7.4.2 Intervalo de Confianza para la Varianza Poblacional.....	182
7.4.3 Prueba de hipótesis para la Varianza Poblacional.....	185
7.5 Comparación de las varianzas de dos poblaciones.....	188
7.6 Comparación entre dos medias poblacionales usando muestras independientes.....	192
7.7 Comparando medias de dos poblaciones usando muestras pareadas.....	198
7.8 Comparando dos proporciones.....	201
Ejercicios	207
8. ANÁLISIS DE DATOS CATEGÓRICOS	210
8.1 Prueba de Independencia y Homogeneidad.....	210
8.2 Medidas de Asociación.....	217
8.3 La prueba de Bondad de Ajuste.....	221
Ejercicios	226
9. REGRESIÓN LINEAL	229
9.1 Regresión Lineal Simple.....	229
9.1.1 Interpretación de Parámetros.....	232
9.2 Inferencia en Regresión Lineal.....	232
9.2.1 Inferencia acerca de los Coeficientes de Regresión.....	233
9.2.2 El análisis de varianza para Regresión Lineal Simple.....	234
9.2.3 El Coeficiente de Determinación.....	235
9.2.4 Intervalo de Confianza para el valor medio de Y e Intervalo de Predicción.....	235
9.3 Análisis de Residuales.....	238
9.4 Modelos No lineales y Transformaciones.....	242
9.4.1 Regresión Cuadrática.....	243
9.4.2 Modelos No lineales que pueden ser transformados en lineales.....	244
9.5 Regresión Lineal Múltiple.....	246
9.6 Inferencia en Regresión Lineal Múltiple.....	249
9.6.1 Prueba de hipótesis de que cada coeficiente de regresión es cero.....	249
9.6.2 Prueba de hipótesis de que todos los coeficientes de regresión son ceros.....	249
9.6.3 Prueba de hipótesis para un subconjunto de coeficientes de regresión.....	250
9.7 Selección de variables.....	253
9.7.1 Los métodos "stepwise".....	253
9.7.2 Método de los mejores subconjuntos.....	268
Ejercicios	271
10. DISEÑOS EXPERIMENTALES	279
10.1 Diseños de Clasificación Simple.....	279
10.2 Comparaciones Múltiples.....	286
10.3 Diseños de Clasificación Doble.....	291

10.4 Modelos con Interacción.....	298
Ejercicios.....	301
11. PRUEBAS NOPARAMÉTRICAS.....	305
11.1 Pruebas Noparamétricas para una sola muestra.....	305
11.1.1 Prueba de los Signos.....	305
11.1.2 La Prueba de Rangos con signos de Wilcoxon.....	307
11.2 Pruebas Noparamétricas para muestras pareadas.....	309
11.3 La Prueba de Mann-Withney para dos muestras independientes.....	311
11.4 La Prueba de Kruskal-Wallis para comparar mas de dos grupos.....	313
11.5 El coeficiente de Correlación de Spearman	316
Ejercicios.....	317
Apéndice A. Tabla de probabilidades acumuladas de la distribución normal estándar.	320
Apéndice B. Tabla de percentiles de la distribución t.....	322
Índice.....	323