

**Universidad de los Andes
Facultad de Ciencias
Departamento de Biología**

**FUNDAMENTOS DE
BIOESTADÍSTICA**



Samuel Segnini

Mérida, Venezuela, 2008

PREFACIO

En términos generales, la fobia a los números pareciera ser una condición natural en nosotros, los biólogos. La mayoría somos presa de la angustia si se trata de manejar estadísticamente datos más allá del simple cálculo de la media aritmética y la desviación estándar. La situación no es mejor si solicitamos la ayuda de algún experto, puesto que se nos dificulta, por un lado expresarle lo que queremos hacer y por otro lado entenderlo. Nos quedamos perplejos cuando sin más preámbulo, el especialista nos suelta preguntas como éstas: ¿Son independientes las muestras? ¿Están normalizados los datos? ¿Son homogéneas las varianzas?. Peor es la situación, si decidimos resolver nuestros problemas sin ayuda y sustituimos al experto por una computadora y usamos uno de los tantos programas de métodos estadísticos existentes. Si logramos seleccionar el método adecuado entre la jerga de nombres de las pruebas ofrecidas, como Anova, Prueba de T para una muestra, Prueba de T para muestras apareadas, Prueba de Wilcoxon, Modelos Lineales, Test de Kruskal-Wallis, etc, seguidamente nos enfrentamos con un problema no menos complicado como es interpretar el resumen de resultados que aparece en la pantalla después de correr el programa. Por supuesto que el mismo está lleno de palabras, símbolos, números y letras griegas que no sabemos interpretar. Ante tal situación recurrimos a los textos de estadística buscando, infructuosamente la mayoría de las veces, algo que se parezca al jeroglífico de resultados que sin la menor pizca de piedad nos arrojó el computador. Después de repetir este proceso unas cuantas veces y cada vez enterarnos que nos equivocamos al escoger el método o que interpretamos mal la información producida, aceptamos que debemos aprender de la estadística lo suficiente para poder examinar en forma elemental los resultados propios o de extraños, y para poder comunicarnos en forma clara con los expertos cuando requerimos de su ayuda. Algunas de las situaciones antes planteadas las sufrí personalmente, especialmente por el hecho de haberme dedicado a la Ecología, una de las disciplinas biológicas que más hace uso de la estadística. Esta circunstancia me obligó a estudiarla y a trajinar con sus métodos. La experiencia adquirida me llevó a dedicarme a la enseñanza de los fundamentos de los métodos estadísticos a los estudiantes de la Licenciatura en Biología de la Universidad de Los Andes. En todos esos años de docencia he venido elaborado notas escritas de los aspectos teóricos de la estadística, así como ejemplos y ejercicios aplicados a las distintas áreas de la biología, con la esperanza de contribuir a mejorar el conocimiento y comprensión de esta materia por los estudiantes. Son estos manuscritos los que han servido de base para escribir este texto en el cual se ha enfatizado la aplicación práctica de los métodos estadísticos, resaltando lo esencial de su fundamento, su utilidad y limitaciones. Intentamos aproximarnos intuitivamente a la teoría estadística, con un mínimo de rigurosidad matemática, ilustrando sus fundamentos con ejemplos prácticos sin recurrir a su demostración formal. Esperamos que el mismo permita al lector obtener una formación teórica muy básica y familiarizarse con los métodos bioestadísticos más corrientes y útiles. Debo finalizar expresando mi más profundo agradecimiento a todos los estudiantes de los cursos de estadística de la Licenciatura en Biología y de los Postgrados en Química Analítica, en Química Aplicada y en Ecología Tropical de la Facultad de Ciencias de la Universidad de Los Andes, en Mérida, Venezuela, con quienes he compartido el placer de aprender y enseñar por muchos años, además de haberme impulsado a publicar estos apuntes sobre los fundamentos de la bioestadística.

Samuel Segnini

CONTENIDO

CAPÍTULO 1. NOCIONES BÁSICAS DE PROBABILIDAD 1

1.1. INTRODUCCIÓN	1
1.2. FENÓMENOS DETERMINÍSTICOS Y ALEATORIOS.....	1
1.2.1. Modelos determinísticos	2
1.2.2. Modelos aleatorios	2
1.2.3. Experimentos aleatorios	3
1.2.4. Espacio muestral	3
1.3. Espacio muestral discreto y continuo.....	4
1.3.1. Evento o Suceso	5
1.3.2. Evento elemental	5
1.3.3. Evento complemento.....	5
1.3.4. Evento vacío.....	5
1.4. ALGEBRA DE LOS EVENTOS	6
1.4.1. Unión de eventos.....	6
1.4.2. Intersección de eventos	6
1.4.3. Diagramas de Venn	7
1.5. CONCEPCIONES DE LA PROBABILIDAD	7
1.5.1. Asignación objetiva de la probabilidad.....	7
1.5.2. Probabilidad a priori.....	8
1.5.3. Probabilidad a posteriori	9
1.5.4. Asignación subjetiva de la probabilidad	10
1.6. DESARROLLO AXIOMÁTICO DE LA PROBABILIDAD	11
1.6.1. Axioma de positividad	11
1.6.2. Axioma de certidumbre.....	11
1.6.3. Axioma de la adición.....	11
1.7. REGLAS PARA EL CÁLCULO DE PROBABILIDADES	12
1.7.1. Regla del conjunto vacío (Teorema 1)	12
1.7.2. Regla de adición para eventos mutuamente excluyentes (Teorema 2)	12
1.7.3. Regla de adición para eventos solapados (Teorema 3)	13
1.7.4. Regla de la complementación (Teorema 4).....	14
1.8. CALCULO DE PROBABILIDADES	15
1.9. PROBABILIDAD CONDICIONAL	20
1.10. EVENTOS INDEPENDIENTES	24
1.11. EJERCICIOS.....	27

CAPÍTULO 2. VARIABLES ALEATORIAS..... 32

2.1. INTRODUCCIÓN	32
2.2. DEFINICIÓN DE VARIABLE ALEATORIA.....	32
2.3. VARIABLES ALEATORIAS DISCRETAS Y CONTINUAS.....	35

2.3.1. Variables aleatorias discretas	35
2.3.2. Variables aleatorias continuas	35
2.4. FUNCIÓN DE PROBABILIDAD DE UNA VARIABLE ALEATORIA	36
2.4.1. Función de probabilidad de una variable aleatoria discreta	36
2.4.1.1. Parámetros de la distribución de una variable aleatoria discreta	38
2.4.1.1.1. Valor promedio	39
2.4.1.1.2. Varianza	40
2.4.2. Función de probabilidad de una variable aleatoria continua.....	42
2.4.2.1. Parámetros de la distribución de una variable continua.....	44
2.5. FUNCIÓN DE DISTRIBUCIÓN ACUMULADA	44
2.5.1. Función acumulada para variables discretas	45
2.5.2. Función acumulada para variables continuas.....	48
2.6. EJERCICIOS.....	51

CAPITULO 3. ALGUNAS DISTRIBUCIONES DE PROBABILIDAD 55

3.1. INTRODUCCIÓN	55
3.2. MODELOS PROBABILÍSTICOS DISCRETOS	55
3.2.1. Modelo de Bernoulli	55
3.2.2. Modelo Binomial.....	57
3.2.2.1. Distribución de probabilidades del modelo binomial	60
3.2.2.2. Función de probabilidad acumulada para el modelo binomial	62
3.2.2.3. Valor Esperado y Varianza de la Distribución Binomial	64
3.2.3. Modelo de Poisson	64
3.2.3.1. Distribución de probabilidades del modelo de Poisson	64
3.2.3.2. Función de probabilidad acumulada para el modelo de Poisson	68
3.2.3.3. Relación entre los modelos Binomial y Poisson.....	69
3.2.3.4. Valor Esperado y Varianza de la Distribución de Poisson	70
3.3. MODELOS PROBABILÍSTICOS CONTINUOS	71
3.3.1. Modelo Normal	71
3.3.1.1. Transformación de una variable X en la variable Z.....	75
3.4. EJERCICIOS.....	77

CAPITULO 4. DISTRIBUCIONES MUESTRALES 85

4.1. INTRODUCCIÓN	85
4.2. ALGUNOS CONCEPTOS IMPORTANTES	86
4.2.1. Universo, población y muestra.....	86
4.2.2. Parámetros y estadísticos	88
4.3. MUESTRAS Y MUESTREO	89
4.3.1. Muestra representativa	89
4.3.2. Muestreo aleatorio.....	90
4.3.3. Muestreo estratificado	91
4.4. DISTRIBUCIONES MUESTRALES.....	92

4.4.1. Distribución de la media muestral.....	92
4.4.1.1. Propiedad Reproductiva de la distribución normal.....	93
4.4.1.2. Teorema del límite central	95
4.4.1.3. El teorema de límite central y los fenómenos biológicos	96
4.4.2. Distribución de la diferencia de medias muestrales	96
4.4.2.1. La diferencia de medias muestrales y el Teorema del Límite Central.....	98
4.5. EJERCICIOS.....	99

CAPITULO 5. ESTIMACIÓN 104

5.1. INTRODUCCIÓN	104
5.2. ESTIMACIÓN PUNTUAL	104
5.2.1. Estimador insesgad.....	105
5.2.2. Estimador consistente.....	106
5.2.3. Estimador eficiente.....	106
5.3. ESTIMACIÓN POR INTERVALO	107
5.3.1. Intervalo de confianza para una media poblacional	107
5.3.1.1. Caso 1: Muestreo en una población distribuida normalmente y con varianza conocida	107
5.3.1.2. Caso 2: Muestreo a partir de una población distribuida normalmente, con varianza desconocida y tamaño de muestra grande ($n \geq 30$).....	112
5.3.1.3. Caso 3: Muestreo a partir de una población distribuida normalmente, con varianza desconocida y tamaño de muestra pequeño ($n < 30$).....	112
5.3.1.3.1. Distribución de T.....	113
5.3.1.4. Caso 4: Muestreo a partir de una población con distribución desconocida, varianza conocida y tamaño de muestra grande ($n \geq 30$).....	114
5.3.1.5. Caso 5: Muestreo a partir de una población con distribución y varianza desconocidas y tamaño de muestra grande ($n \geq 30$).....	115
5.3.1.6. Caso 6: Muestreo a partir de una población con distribución desconocida y tamaño de muestra pequeño ($n \geq 30$).....	116
5.3.2. Intervalo de confianza para la diferencia de dos medias poblacionales.....	116
5.3.2.1. Caso 1: Muestreo a partir de poblaciones distribuidas normalmente y con varianzas conocidas.....	116
5.3.2.2. Otros casos	119
5.3.2.3. Intervalos de confianza para $\mu_{x_2} - \mu_{x_1}$ y el estadístico t.....	119
5.3.2.3.1. Varianzas iguales.....	120
5.3.2.3.2. Varianzas diferentes	121
5.4. EJERCICIOS	123

CAPITULO 6. PRUEBA DE HIPÓTESIS 127

6.1. INTRODUCCIÓN	127
6.2. LA PRUEBA DE HIPÓTESIS: UN PROCEDIMIENTO DE DECISIÓN	127
6.3. PROCEDIMIENTO GENERAL PARA LA PRUEBA DE HIPÓTESIS	130
6.3.1. Hipótesis.....	130

6.3.2. Nivel de significación.....	134
6.3.2.1. Errores de tipos I y II	135
6.3.3. Estadístico de prueba.....	137
6.3.4. Zona de aceptación.....	138
6.3.5. Cómputos	140
6.3.6. Decisión.....	140
6.3.7. Conclusión.....	141
6.4. PRUEBA DE HIPÓTESIS PARA UNA MEDIA POBLACIONAL.....	142
6.4.1. Prueba de hipótesis para una media poblacional cuando la muestra proviene de una población distribuida normalmente y con varianza conocida	142
6.4.2. Prueba de hipótesis para una media poblacional cuando la muestra proviene de una población distribuida normalmente, con varianza desconocida y tamaño de muestra grande ($n \geq 30$)	144
6.4.3. Prueba de hipótesis para una media poblacional cuando la muestra proviene de una población distribuida normalmente, con varianza desconocida y tamaño de muestra pequeño ($n < 30$).....	146
6.4.4. Prueba de hipótesis para una media poblacional cuando la muestra proviene de una población con distribución no normal y tamaño de muestra grande ($n \geq 30$)	147
6.5. PRUEBA DE HIPÓTESIS PARA DOS MEDIAS POBLACIONALES.....	151
6.5.1. Prueba de hipótesis para dos medias poblacionales cuando las muestras provienen de poblaciones distribuidas normalmente y con varianzas conocidas.....	151
6.5.2. Prueba de hipótesis para dos medias poblacionales cuando las muestras provienen de poblaciones distribuidas normalmente, con varianzas desconocidas y tamaño de muestras grandes ($n_1, n_2 \geq 30$).....	153
6.5.3. Prueba de hipótesis para dos medias poblacionales cuando las muestras provienen de poblaciones distribuidas normalmente, con varianzas desconocidas y tamaño de muestras pequeñas ($n_1, n_2 < 30$)	155
6.5.4. Prueba de hipótesis para dos medias poblacionales cuando las muestras provienen de poblaciones con distribución no normal y tamaño de muestras grandes ($n_1, n_2 \geq 30$)	160
6.5.5. Prueba de hipótesis para dos medias poblacionales usando observaciones apareadas	163
6.6. PRUEBA DE HIPÓTESIS PARA DOS VARIANZAS POBLACIONALES	167
6.7. EJERCICIOS.....	174

CAPITULO 7: INTRODUCCIÓN AL ANÁLISIS DE VARIANZA 184

7.1. INTRODUCCIÓN	184
7.2. FUNDAMENTOS DEL ANÁLISIS DE VARIANZA (ANDEVA).....	185
7.2.1. ¿Una o varias pruebas de hipótesis?.....	186
7.2.2. Varianza dentro de los grupos.....	187
7.2.3. Varianza entre grupos.....	187
7.2.4. Prueba de hipótesis para dos varianzas	188
7.2.5. Cálculo de la varianza dentro de los grupos.....	190
7.2.6. Cálculo de la varianza entre grupos	191
7.2.7. Prueba de hipótesis para las varianzas	191

7.3. PARTICIÓN DE LA SUMA TOTAL DE CUADRADOS.....	193
7.4. NOTACIÓN BÁSICA Y CÁLCULOS NECESARIOS.....	195
7.4.1. Sumas de cuadrados (SC).....	196
7.4.2. Cuadrados medios (CM).....	196
7.4.3. Grados de libertad.....	197
7.5. ANÁLISIS DE VARIANZA DE UN FACTOR.....	197
7.6. COMPARACIÓN MÚLTIPLE DE MEDIAS.....	203
7.6.1. Prueba de la Mínima Diferencia Significativa (MDS).....	204
7.6.2. Prueba de Tukey o de la Diferencia Verdaderamente Significativa (DVS).....	208
7.7. MODELO Y SUPUESTOS BÁSICOS DEL ANÁLISIS DE VARIANZA.....	209
7.7.1. Modelo lineal.....	209
7.7.2. Supuestos básicos.....	211
7.7.2.1. Aditividad.....	211
7.7.2.2. Aleatoriedad.....	212
7.7.2.3. Independencia.....	212
7.7.2.4. Homogeneidad de las Varianzas.....	213
7.7.2.5. Normalidad.....	214
7.8. ANÁLISIS DE VARIANZA DE DOS FACTORES.....	214
7.9. ¿UNO O DOS FACTORES?.....	220
7.10. ANÁLISIS DE VARIANZA DE DOS FACTORES CON REPETICIONES.....	222
7.11. FUNDAMENTO DEL ANDEVA DE DOS FACTORES CON OBSERVACIONES REPETIDAS.....	224
7.11.1. Varianza residual.....	225
7.11.2. Variación añadida.....	226
7.11.2.1. Variación añadida debido al factor A.....	227
7.11.2.2. Variación añadida debido al factor B.....	227
7.11.2.3. Variación añadida debida a la interacción.....	228
7.11.3. Partición de la suma de cuadrados total.....	229
7.11.4. Tabla de resultados.....	229
7.11.5. Significación de las varianzas o cuadrados medios.....	229
7.11.5.1. Significación de la varianza debida al Factor A.....	229
7.11.5.2. Significación de la varianza debida al Factor B.....	230
7.11.5.3. Significación de la varianza debida a la Interacción.....	230
7.12. PROCEDIMIENTO GENERAL PARA EL ANDEVA DE DOS FACTORES CON OBSERVACIONES REPETIDAS.....	230
7.13. EJERCICIOS.....	235

CAPITULO 8: ASOCIACIÓN ENTRE DOS VARIABLES..... 246

8.1. INTRODUCCIÓN.....	246
8.2. NATURALEZA DE LA ASOCIACIÓN ENTRE VARIABLES.....	246
8.3. FORMAS DE EVALUAR UNA ASOCIACIÓN.....	247
8.3.1. Evaluación gráfica.....	247
8.3.2. Evaluación Estadística.....	248
8.4. ANÁLISIS DE CORRELACIÓN.....	248
8.4.1. Coeficiente de Correlación.....	252

8.4.2. Estimación de ρ_{xy}	252
8.4.3. Prueba de significación para r	256
8.4.3.1. Prueba de hipótesis para $\rho = 0$	256
8.4.3.2. Prueba de hipótesis para $\rho = \rho_0$	258
8.4.3.3. Prueba de hipótesis para dos coeficientes de correlación ($\rho_1 = \rho_2$)	260
8.5. ANÁLISIS DE REGRESIÓN	261
8.5.1. Buscando la mejor ecuación de una recta	262
8.5.1.1. Método de los mínimos cuadrados	263
8.5.1.1.1. Cálculo de a y b por el método de los mínimos cuadrados	264
8.5.2. Modelos de regresión lineal	267
8.5.2.1. Modelo I	268
8.5.2.2. Modelo II	269
8.5.2.3. Modelo I vs. Modelo II	270
8.5.3. Estimación de α y β	270
8.5.4. Evaluación de la regresión	272
8.5.4.1. Prueba de hipótesis para β	272
8.5.4.2. Análisis de varianza	274
8.5.5. Bondad del ajuste por regresión lineal	278
8.5.6. Comprobación del modelo	279
8.5.7. Formas de utilizar la ecuación de regresión	280
8.5.7.1. Predicción de y para un dado valor x_i	281
8.5.7.2. Estimación de $\mu_{y/x}$ para un dado valor x_i	286
8.5.8. Comparando dos rectas de regresión	287
8.5.9. Advertencias para el uso de la ecuación de regresión	290
8.6. EJERCICIOS	290
 BIBLIOGRAFÍA CONSULTADA	 295
 APÉNDICE A: Prueba de normalidad de Shapiro-Wilks.....	 A1
 APÉNDICE B: Transformación de datos	 B1
 APÉNDICE C: Tablas Estadísticas.....	 C1