Daena: International Journal of Good Conscience. 8(3)92-102. Octubre 2013. ISSN 1870-557X

Modelos Estadísticos Paramétricos y No Paramétricos para dos Muestras

(Statistical Parametric and Nonparametric Models for two samples)

Guillen, A., M.H. Badii, J.L. Abreu & H. Rodríguez*

Resumen. Se describen y discuten con ejemplos reales y prácticos los modelos de tablas de contingencia en la escala nominal, el de Kolmogorov-Smirnov (escala nominal), Mann-Whitney (escala ordinal) y t de Student (escala razón) presentando ejemplos típicos para cada uno de estos tres modelos. Los pasos de procedimiento para realizar el análisis y la interpretación de los resultados de cada modelo se describen y discuten de manera detallada.

Palabras claves. Kolmogorov-Smirnov, Mann-Whitney, tablas de contingencia, t de Student

Abstract. Contingency table and Kolmogorov-Smirnov models both in nominal scale, mann-Whitney model in ordinal scale, and t student model in ratio scale are analyzed and real case examples are given for each model. The procedural steps for the analysis and the interpretation of each and all three models are provided in detailed form.

Keywords. Contingency table, Kolmogorov-Smirnov, Mann-Whitney, t de Student

Introducción

Cuando se maneja dos grupos y se compara entre ellos para establecer diferencia o similitud, el investigador debe en primer lugar estar consciente de tipo de datos que se utiliza, es decir, debe determinar la escala correcta para los datos del estudio. Por esto, hay que tener una noción clara de la definición de la variable y los diferentes tipos de escalas para la utilización de las variables. Una variable es una característica de un elemento de la población. Debido a que existen diferentes tipos de rasgos, también hay distintas escalas para expresar las variables que se definen de forma breve a continuación (Hettmansperger and McKean, 1998; Huber, 1981; Johnson and Kotz, 1972; Johnson et al., 1995).

Escala Nominal: en donde se mide los atributos y/o las cualidades del elemento, los cuales pueden ser numéricos o no numéricos. Esta escala proporciona información muy general sobre los elementos y se trata de igualdad y desigualdad de los elementos.

Escala Ordinal: Aquí se ordenen o se ranquean los diferentes niveles (seleccionadas de forma subjetiva en una escala) de los atributos de los elementos los cuales de nuevo pueden ser numérico o no numérico. Esta escala contiene los rasgos de la escala anterior y por ende nos proporciona mayor cantidad de información que la escala nominal. La ordenación tiene sentido en esta escala. Estas dos escalas se utilizan en las investigaciones de tipo cualitativo. Escala Intervalo: Se trata de medir la magnitud del elemento, es de tipo solamente numérico y tiene dos rasgos siguientes: a) hay distancias constantes entre los valores adyacentes de la variable, y b) tiene un cero no real, es decir convencional. Escala Razón: Tiene todos los rasgos de la escala intervalo con la excepción de que en la escala razón, sí se utiliza un cero verdadero y por tanto en esta escala la noción de relacionar tiene sentido. Estos dos últimas escalas se usan para investigaciones de tipo cuantitativo.

Cuando se hace comparación entre dos grupos, aparte de determinar el tipo correcto de escala a utilizar, el investigador, también tiene que establecer si los datos de los dos grupos bajo del estudio están relacionadas (dependen entre sí) o no relacionadas (independientes el uno del otro) (kendall and Stuart, 1979, Kirk, 1982; Kueh, 2000; LehMANN, 1981, 1998; McPherson, 1990; Neter et al., 1993)

El objetivo de este trabajo es investigar la predicción sobre la diferencia entre dos grupos de datos por medio de los modelos de Kolmogorov-Smirnov en escala nominal, Man-Whitney en escala ordinal y t de Student en escala de razón.

Tablas de contingencia

1. Descripción:

Se utiliza para 2 o más muestras independientes con varias variables en escala Nominal.

2. Ecuación:

$$\mathbf{X^2 = \sum [(O - E)^2 / E]}$$

Donde,
X^2 = Valor calculado del modelo
$\sum$ = Signo de sumatorio
O = Frecuencia observada
E = Frecuencia esperada o teórica
gl = K -1 , donde, K = número de muestras o grupos
Ex = ($\sum$F)($\sum$C) / GT
Ex = Frecuencia esperada de la celda “X”
$\sum$H = Suma de hilera
$\sum$C = Suma de columna

GT = Gran total

Ejemplo.

Para medir el efecto (grado de mejoría) de 2 tratamientos de enseñanza, se dividen de manera al azar los 58 alumnos en **2** grupos de 33 que reciben un método nuevo de enseñanza y 25 alumnos que reciben el método tradicional de enseñanza. En esta investigación, los resultados, es decir, el número de alumnos con mejoría, lo cual indica un aumento significativo en la calificación versus no mejoría (sin aumento significativo en la calificación del alumno) se encuentran en la Tabla 1.

Tabla 1. Nivel de aumento en la calificación del alumno con método nuevo y método tradicional (C = Columna, H = Hilera)

	Aumento en calificación	**No aumento en calificación**	**∑H**
Método Nuevo	25	8	33
Método tradicional	16	9	25
∑C	41	17	**GT 58**

3. Plantear hipótesis:

Ho: Las diferencias en la mejora de calificación es un efecto aleatorio
Ha: El método nuevo tiene mejor efecto en la mejora de calificación que el método tradicional

4. Nivele de significancia:

$\alpha = 0.05$ para probar la hipótesis

5. Zona de rechazo:

Para todo valor menor que $p = .05$ ($p \leq .05$) se rechaza la Ho y viceversa.

6. Procedimiento:

Primero, hay que calcular los valores esperados:

$Ex_1 = 33 * 41 / 58 = 23.33$
$Ex_2 = 33 * 17 / 58 = 9.67$
$Ex_3 = 25 * 41 / 58 = 17.67$
$Ex_4 = 25 * 17 / 58 = 7.33$

$X^2 = \sum [(O - E)^2 / E] = [(25 - 23.33)^2 / 23.33 + \ldots + (9 - 7.33)^2 / 7.33] =$ 0.946

gl = (C -1) (F -1) = (2 -1) (2- 1) = 1

El gl es igual a 1, y el valor tabulado para 3 gl es igual a 3.841

7. Tomar decisión:

Valor calculado del modelo de 0.946 lo cual es menor que el valor crítico de 3.841 y
por tanto, Ho se Acepta.

8. Interpretación:

Los 2 métodos de enseñanza tienen el mismo efecto en la mejora de calificación, y por tanto, el hecho de que el método nuevo aparentemente está causando mayor mejora en la calificación, en realidad esto se debe a un efecto aleatorio, es decir, no hay diferencia entre las 2 métodos de enseñanza en término de la mejora de las calificaciones.

Kolmogorov-Smirnov

1. Descripción:

Se utiliza para 2 muestra en escala ordinal.

2. Ecuación:

$D_{MAX} = | F_{1AO} - F_{2AO} |$

Donde,
D_{MAX} = Diferencia máxima
F_{1AO} = Frecuencia acumulada observada para muestra 1
F_{2AO} = Frecuencia acumulada observada para muestra 2

Para toma de decisión, se calcula valor crítico de diferencia máxima al nivel de α = 0.05 y para determinar la significancia estadística, se usa la ecuación: D(.05) = 1.36

/ $\sqrt{[(N_1+N_2)/N_1N_2]}$, donde, 1.36 es un valor constante y N_1 = tamaño de la muestra 1, N_2 = tamaño de la muestra 2.

Ejemplo.

Se supone que debido a la exigencia regulatoria, los carros fabricados y con procedencia de América del norte y la Unión Europea en contraste con los carros procedentes de otros regiones pasan las pruebas de de emisión de CO2. En un estudio de 150 carros con procedencias de 5 regiones del mundo, se notó que la mitad probaron la prueba de emisión de CO2 y la otra mitad reprobaron esta ensayo. Verifique si la probación está condicionada por la procedencia (Tabla 2).

Tabla 2. Distribución (de número) en grupos aprobados y reprobados en base procedencia*.

Procedencia	**Aprobados**	**Reprobados**
	F1o	**F2o**
Africa	3	34
América Latina	7	25
Asia	12	8
América del norte	25	5
Europa	28	3
Total	**75**	**75**

En la Tabla 2a, se observan las diferencias entre las observaciones acumuladas (**F1oa y F2oa**) las dos columnas aprobados y reprobados.

Tabla 2a. Distribución (de número) en grupos aprobados y reprobados en base procedencia*.

Procedencia	**Aprobados** (F1oa)	**Reprobados** (F2oa)	**|F1oa–F2oa|**
	F1o	**F2o**	
Africa	0.040	0.453	0.413
América Latina	0.100	0.786	***<u>0.653</u>***
Asia	0.293	0.893	0.6
América del norte	0.627	0.960	0.33
Europa	1.000	1.000	0
Total	**75**	**75**	

*: Fo= F. observada, Foa= F. observada acumulada, 1= aprobado, 2 = reprobado.

Frecuencias esperadas: Carros aprobados

3/75 = 0.040

10/75 = 0.10

22/75 = 0.293
47/75 = 0.627
75 / 75 = 1.000

Frecuencias esperadas: Carros reprobados
34/75 = 0.453
59/75 = 0.786
67/75 = 0.0.893
72/75 = 0.0.960
75 / 75 = 1.000

3. Plantear hipótesis:

Ho: Las diferencias en las F1oa y F2oa es aleatoria e independiente de la procedencia
Ha: Las diferencias en las F1oa y F2oa dependen de la procedencia

4. Nivele de significancia:

$\alpha = 0.05$ para probar la hipótesis

5. Procedimiento:

Usar la ecuación de $D_{MAX} = | F_{AO} – F_{AE} |$, y sustituir los valores en la ecuación nos arroja lo siguiente $D_{MAX} = |.783 - .13| = 0.653$ y de aquí $D(0.05) = 1.36 / \sqrt{[(N_1+N_2)/N_1N_2]} = 1.36 / \sqrt{[(75+75) /75*75]} = 0.222$.

6. Tomar decisión:

Valor calculado del modelo (D_{MAX}) de .653 es mayor que valor crítico D(0.05) de 0.222 y por tanto Ho se rechaza.

7. Interpretación:

Las diferencias en la aprobación de en la prueba de emisión del gas de CO2 están condicionadas significativamente (nivel de 95% de confianza) por la procedencia de los carros.

Modelo de Mann-Whitney

1. Descripción:

Se usa este modelo en caso de muestras independientes en una escala ordinal y cuando NO se cumplen los requisitos de t de Student.

2. Ecuación:

$$\boldsymbol{U_1 = n_1 n_2 + [n_1 (n_1 +1) / 2] - \sum R_1}$$
$$\boldsymbol{U_2 = n_1 n_2 + [n_2 (n_2 +1) / 2] - \sum R_2}$$

Donde,

U1 & U2: Valores de la Prueba de Mann-Whitney
n_1: Tamaño de la muestra del grupo 1
n_2: Tamaño de la muestra del grupo 2
$\sum R_1$: Sumatoria de los rangos del grupo 1
$\sum R_2$: Sumatoria de los rangos del grupo 2

Ejemplo:

Demostrar que un método de enseñanza es mejor que otro, Aplicar los 2 métodos a un grupo de 10 niños, (5 en cada grupo) y registrar las calificaciones de cada uno de los 5 niños de cada grupo (Tabla 3).

Tabla 3. Calificaciones de 2 grupos de niños bajo cada método de enseñanza.

Método	**Calificaciones**					**$\sum$R**
Tradicional	80	85	25	70	90	
Rangos del grupo 1	4	5	1	3	6	19
Experimental	95	100	93	110	45	
Rango del grupo 2	8	9	7	10	2	36

3. Hipótesis:

Ho: las diferencias en las calificaciones entre 2 métodos se deben al azar
Ha: Las calificaciones del método experimental son más altas que otro método.

4. Nivel de significancia:

$\alpha = 0.05$

5. Zona de rechazo:

Para todo valor de p Menor que 0.05 ($p<0.05$) se rechaza la Ho.

6. Procedimiento:

$U_1 = n_1\, n_2 + [n_1\, (n_1 +1) / 2] - \sum R_1 = 5 * 5 + 5(5+1) /2 - 19 = 21$
$U_2 = n_1\, n_2 + [n_2\, (n_2 +1) / 2] - \sum R_2 = 5 * 5 + 5(5+1) /2 - 36 = 4$

De estos valores seleccionar el más pequeño (4) y se compara este valor con la Tabla en donde el valor crítico de la tabla de Mann-Whitney es igual a $p = 0.043$

7. Toma de decisión:

El valor calculado de 0.043 es menor que 0.05 ($P<.05$), y por tanto se rechaza la Ho.

8. Interpretación:

Hay diferencia estadísticamente significativa entre las calificaciones debido a los 2 métodos y de hecho el método experimental es más efectivo.

Modelo de t de Student

1. Descripción:

Condiciones de este modelo:

1. Escala intervalo: intervalos entre las mediciones sean equivalentes, y que pueden realizar todas las operaciones permisibles.
2. Observaciones sean independientes y que proceden de una muestra aleatoria de una población con distribución normal.
3. Homogeneidad de varianzas (la génesis del error sea el mismo), pero en los casos de biología, medicina, salud y en término general en la ciencia natural es difícil que esto ocurre, por tanto se puede checar la homogeneidad de varianza por medio de la prueba de F_{MAX}.

2. Ecuación:

$$t = (m_1 - m_2) \;/\; (\sigma_1^2/n_1 + \sigma_2^2/n_2)^{1/2}$$

Donde,

t = valor del modelo de t
m_1 = La media de la muestra 1
m_2 = La media de la muestra 2
σ_1^2 = Varianza de la muestra 1
σ_2^2 = Varianza de la muestra 1
n_1 = Tamaño de la muestra 2
n_2 = Tamaño de la muestra 2

Ejemplo:

Se compara la talla de 20 niños de 5 años de edad (divididos en 2 grupos de 10) de 2 niveles socioeconómicos y se especula que hay diferencia en tallas debido al factor socioeconómico (Tabla 4).

Tabla 4. Tallas de 20 niños de 2 niveles socioeconómicos.

Nivel socio económico bajo		**Nivel socio económico bajo**	
I	Talla (cm)	I	Talla (cm)
1	101	1	103
2	102	2	105
3	100	3	104
4	104	4	106
5	102	5	108
6	102	6	100
7	99	7	108
8	103	8	104
9	97	9	105
10	99	10	107
$n_1 = 10$	$\sum X = 1{,}009$	$n_2 = 10$	$\sum X = 1{,}050$
$m_1 = 100.9$	$\sum X^2 = 101{,}849$	$m_2 = 105.0$	$\sum X^2 = 110{,}304$
	$\sigma_1^2 = 2.13^2$		$\sigma_2^2 = 2.45^2$

3. Hipótesis:

Ho: Las diferencias en las tallas entre 2 grupos es debido al azar
Ha: Las diferencias en las tallas entre 2 grupos es debido factor socioeconómico y se supone que el nivel socioeconómico favorece una mayor altura.

4. Nivel de significancia:

$\alpha = 0.05$

5. Zona de rechazo:

Para todo valor de p Menor que 0.05 ($p<0.05$) se rechaza la Ho.

6. Procedimiento:

$$\mathbf{t = (m_1 - m_2) \; / \; (\sigma_1^2/n_1 + \sigma_2^2/n_2)^{1/2}}$$

$t = (100.9 - 105) / [(2.13^2 /10) + (2.45^2/10)]^{1/2} = - 3.99$
$gl = (n_1 + n_2) - 2 = (10 + 10) - 2 = 18$

7. Toma de decisión:

Valor de t de Tabla con 18 gl a nivel de p .05 es igual a 2.101, el valor de t calculado es mayor que el valor de t tabulado y por tanto se rechaza la hipótesis nula o Ho.

8. Interpretación:

Existe diferencia estadísticamente significativa entre las tallas de niños de dos niveles socioeconómicos distintos.

Referencias

Hettmansperger, T.P. and J.W. McKean. 1998. Robust Nonparametric Statistical Methods. New York: Wiley.
Huber, P.J. 1981. Robust Statistics. New York: Wiley.
Hohnson, N.L. and S. Kotz. 1972. Distributions in Statistics. New York: Wiley.
Johnson, N.L., S. Kotz and N. Balakrishnan. 1995. Continuous Univariate Distributions. Vol. 2. 2nd. Edition. New York: Wiley.
Johnson, N.L., S. Kotz and N. Balakrishnan. 1995. Univariate Discrete Distributions. Vol. 2. 2nd. Edition. New York: Wiley.

Kendall, M. and A. Stuart. 1979. The Advanced Theory of Statistics. 4th edition. New York. Macmillan.
Kirk, R.E. 1982. Experimental Design. 2ed edition. Pacific Grove, CA: Brooks.
Kuehl, R.O. 2000. Design of Experiments. 2ed edition. Pacific Grove, CA: Duxbury.
Lehmann, E.L. 1981. Testing Statistical hypothesis. Ed edition. New York: Wiley.
Lehmann, E.L., 1998. Theory of Point estimation. 2ed edition. New York: Springer-Verlag.
McPherson, G. 1990. Statisitcs in Scientific Investigations. New York: Springer-Verlag.
Neter, J., W. Wasserman and G.A. Whitmore. 1993. Applied Statistics. Boston: Allyn & Bacon.

***Autores**

Guillen, A., M.H. Badii, J.L. Abreu & H. Rodríguez. UANL, San Nicolás, N.L., México & U. Agraria de la Habana, Cuba