Modelos Estadísticos para Escalas Ordinal y de Intervalo: Caso de una Muestra

(Statistical Models for Ordinal and Interval Scales: A Sample Case)

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

UANL, San Nicolás, N.L., México & U. Agraria de la Habana, Cuba

Resumen. Se analizan los modelos de Kolmogorov-Smirnov y el modelo de z para caso de una muestra en escalas de ordinal y de razón, respectivamente. Se presentan ejemplos reales para cada uno de estos dos modelos, detallando los procedimientos adecuados.

Palabras claves. Distribución z, escalas ordinal y de razón, modelo K-S

Abstract. Kolmogorov-Smirnov model in ordinal scale and z distribution in ratio scale for one sample case are analyzed. Real case examples are provided and step by step procedures for each model analysis are detailed and presented.

Keywords. K-S Model, ordinal and ratio scales, Z distribution

Introducción

La estadística es la ciencia que trata de verificar la validez probabilística de todos los eventos, fenómenos, procesos y objetos en espacio y tiempo. Hay dos predicciones en cualquier rama de la ciencia, el primero se trata de distinguir la diferencia entre grupos de datos y la segunda se trata de verificar el establecimiento de los patrones. Por tanto los modelos estadísticos se encargan de procedimientos para contestar estas dos predicciones. Por tanto, el reto de manejar datos estadísticos radica en saber el uso adecuado de los modelos para casos y situaciones bajo el estudio (Agresti (1990), Billingsy (1995), Brwon and Fuller (1991), Cochran and Cox (1957**,** Kuehl, 2000, Lehmann, 1981, 1998, Perez (2008**),** Schevish, 1995, 1996). Consecuentemente, es necesario conocer primero la aplicación de cada modelo y luego los procedimientos para el uso correcto de dicho modelo.

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.

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.

El objetivo de este trabajo es investigar y conocer la prueba no paramétrica de Kolmogorov-Smirnov (K-S) en escala ordinal y la prueba de z en la escala de intervalo utilizado para el caso de una muestra.

<u>Modelo de Kolmogorov-Smirnov (K-S)</u>

1. **Descripción:**
   Se utiliza para **1 muestra** en **escala ordinal.**
2. **Ecuación:**

   $D_{MAX} = | F_{AO} – F_{AE} |$

   Donde,
   D_{MAX} = Diferencia máxima
   F_{AO} = Frecuencia acumulada observada
   F_{AE} = Frecuencia acumulada esperada

Para toma de decisión, se calcula valor crítico de diferencia máxima al nivel de α = 0.05 para determinar la significancia estadística, usando la ecuación siguiente: $D(0.05) = 1.36 / \sqrt{N}$, donde, 1.36 es un valor constante y N = tamaño de la muestra.

Ejemplo: Las personas toman decisiones para votar hacia sus partidos políticos favoritos. En una encuesta de 30 personas con derecho a votar, se registran la preferencia partidista de cada uno de estas personas hacia los partidos políticos A, B, y C. (Tabla 1).

Tabla 1. Frecuencia de tendencia o preferencia hacia distintos partidos políticos.

Frecuencias	**Partidos políticos**			**Total**
	A	**B**	**C**	
Observada (F_O)	**18**	**8**	**4**	**30**
Esperadas (F_A)	**10**	**10**	**10**	**30**

3. **Plantear hipótesis:**

   Ho: Las diferencias en las F_O de preferencia partidista es debió al azar
   Ha: Las preferencias son hacia el partido A y presentan diferencias significativas en F_O

4. **Nivele de significancia:**

   $\alpha = 0.05$ para probar la hipótesis

5. Zona de rechazo:

Para todo valor de p mayor que 0.05, se rechaza la Ho. Hay que señalar que la p es la probabilidad que el efecto significativo ocurre solo por medio de aleatoriedad.

7. Procedimiento:

Los datos y el procedimiento de análisis de los mismos se encuentran en la Tabla 2.

Tabla 2. Frecuencia de tendencia o preferencia hacia distintos partidos políticos.

	Partido A	**Partido B**	**Partido C**	**Total**
F obs. (F_O)	**18**	**8**	**4**	**30**
F esp. (F_A)	**10**	**10**	**10**	**30**
F_{AO} (Acumulativa Obs.)	18/30 = 0.6	26/30 = 0.87	30/30 = 1	1
F_{AE} (Acumulativa	10/30 = 0.33	20/30 = 0.66	30/30 = 1	1

Esp.)				
$\mid F_{AO} - F_{AE} \mid$	0.27	0.21	0	

Usar la ecuación de $D_{MAX} = \mid F_{AO} - F_{AE} \mid$, y sustituir los valores en la ecuación nos arroja lo siguiente $D_{MAX} = |0.33 - 0.6| = .27$ y de aquí $D(0.05) = 1.36 / \sqrt{N} = 1.26 / \sqrt{30} = 0.248$.

8. Tomar decisión:

Valor calculado del modelo (D_{MAX}) de .27 es mayor que valor crítico D(0.05) de 0.248 y por tanto Ho se rechaza.

9. Interpretación:

La votación es mayormente hacia el partido A, la cual esta expresada por la preferencia marcada en número de votos hacia este partido derivada por la tendencia y preferencia de los votantes. Consecuentemente, la preferencia partidista reflejada por los votos no es debido a un proceso aleatorio (rechazo de Ho).

Modelo Z

1. Descripción:

Es un modelo que representa y describa la distribución normal y que se utiliza para estimar la representatividad de una muestra en escalas de intervalo y razón. La distribución z representa situaciones en donde los datos están tomados de forma aleatoria de una población grande, y donde la varianza es independiente de la media. Esta distribución es de tipo campana, simétrica y en el centro se ubican tres medidas de tendencia central: la media, la moda y la mediana los cuales tienen valores iguales.

2. Ecuación:

$Z = (X\text{i} - m) / \sigma$,
Donde,
Z es el valor estadístico de la curva normal de frecuencias
*X*i es ***cualquier*** valor de una muestra
m es el promedio de la muestra o valor representativo o típico
σ es la desviación estándar

3. Pasos:

1. Calcular la m y la σ de las observaciones de la muestra.
2. Restar la m del valor del cual (Xi) desea hacer una inferencia.
3. Dividir el resultado del paso 2 en σ, y de esta forma calcular el valor Z.
4. Localizar en la Tabla Normal el valor calculado de Z para la probabilidad de obtener la magnitud de diferencia entre la X y la m.
5. Tomar la decisión de rechazar o aceptar la Ho.

3. Ejemplo:

Los carros de dos marcas [Nacional (N) y Extranjero (F)] en la carretera frenan a una distancia según la velocidad del vehículo; a mayor velocidad, mayor será la distancia en frenar para que el vehículo se para de forma completa.
En un estudio se cuantificaron las distancias para llegar a un paro completo de 5 vehículos denominados de A hasta E (Tabla 3). En esta tabla se encuentran los datos de los vehículos en término de origen del carro (N o F), edad del carro, su velocidad al momento de iniciar a frenar y la distancia para el freno completo.. Verificar si las distancias para el frenado completo se ubican dentro del rango normal. En esta tabla en las últimas 2 columnas también se encuentran los datos (media y desviación estándar) de la velocidad y la distancia de la población.

Tabla 3. Datos del frenado de la muestra*.

Carro	**Origen**	**Edad**	**Velocidad**	**Distancia**	***m* & *σ* de la población**	
					Velocidad	**Distancia**
A	N	5	104	14.5	107.6 ± 3.64	18.7 ± 1.72
B	N	6	120	27.4	113.7 ± 3.96	20.8 ± 2.13
C	N	10	129	42.1	135.5 ± 5.49	32.5 ± 4.38
D	F	9	120	39.0	130.1 ± 5.41	29.1 ± 3.72
E	F	7	108	28.0	113.6 ± 4.32	20.8 ± 2.17

*: velocidad en km/hr, distancia en m.

Los datos vienen de 1 muestra y se presentan en la escala de razón, por tanto, se usa el modelo de Z. Si se específica la ***hipótesis alterna*** con la dirección adecuada se usa 2 colas y viceversa.

4. Plantear la hipótesis:

Ho: La diferencia entre las distancias y las velocidades de la muestra versus la población es aleatoria
Ha: Esas diferencias son significativas

5. Nivel de significancia:

$\alpha = 0.05$ para probar la hipótesis

6. Zona de rechazo:

Para todo valor de p mayor que 0.05, se rechaza la Ho

6. Procedimiento:

Carro A:
Velocidad: $Z = (X\mathrm{i} - m) / \sigma = (104 - 107.6) / 3.64 = -0.99$
Distancia: $Z = (X\mathrm{i} - m) / \sigma = (14.5 - 18.7) / 1.72 = -2.44$

Carro B:
Velocidad: $Z = (X\mathrm{i} - m) / \sigma = (120 - 113.7) / 3.96 = 1.59$
Distancia: $Z = (X\mathrm{i} - m) / \sigma = (27.4 - 20.8) / 2.13 = 3.1$

Carro C:
Velocidad: $Z = (X\mathrm{i} - m) / \sigma = (129 - 135.5) / 5.49 = -1.18$
Distancia: $Z = (X\mathrm{i} - m) / \sigma = (42.1 - 32.5) / 4.38 = 2.19$

Carro D:
Velocidad: $Z = (X\mathrm{i} - m) / \sigma = (120 - 130.1) / 5.49 = -1.87$
Distancia: $Z = (X\mathrm{i} - m) / \sigma = (39 - 29.1) / 3.72 = 2.66$

Carro E:
Velocidad: $Z = (X\mathrm{i} - m) / \sigma = (108 - 113.6) / 4.32 = -1.3$
Distancia: $Z = (X\mathrm{i} - m) / \sigma = (28 - 20.8) / 2.17 = 3.31$

7. Toma de decisión:

Un valor de Z indica el número de σ´s a un lado u otro de la media (m) de la curva normal. Ahora bien, a nivel de $\alpha = 0.05$, tenemos que 95% de los valores están dentro la media $\pm 1.96\ \sigma$´s. Por tanto todos los valores de Z de la Tabla 3 ubicados dentro de **$\pm$ 1.96σ** indican no significancia o NS (aceptación de Ho) y todos los valores de Z de la Tabla 3 ubicados afuera de **$\pm$ 1.96σ** indican significancia (*) es decir el rechazo de la hipótesis nula o Ho.

Ahora: En la Tabla Z, el valor crítico para α = 0.05 para 1 cola es 1. 645 y para 2 colas es 1.96.

Para la toma de decisiones y poder visualizar mejor las comparaciones de los valores críticos calculados y teóricos, ya que se manejan 5 carros, se puede construir la siguiente tabla (Tabla 4).

Tabla 4. Comparación de valores calculados y teóricos.

Carro	**Valores calculados de Z** **Valores de probabilidad de Z a** α = 0.05 son 1.645 y 1. 96 para 1 cola y 2 colas, respectivamente	
	Velocidad	**Distancia**
A	-0.99 NS	-2.44 *
B	1.59 NS	3.10 *
C	1.18 NS	2.19 *
D	1. 87 NS	2.66 *
E	-1.30 NS	3.31*

$: * = significativo, NS = no significativo

8. Interpretación:

Con respecto a la distancia, el carro A tiene un valor bajo, y los demás carros tienen valores superiores al rango normal, esto de forma significativa a nivel de α = 0.05. Con respecto a la velocidad, todos los carros tienen velocidades que aparentemente varían un poco de lo normal, sin embargo todas están dentro del rango normal estadísticamente a nivel α = 0.05.

9. Conclusión:

Los carros del estudio tienen distancias afuera del rango normal (de frecuencias) de una población local, pero con velocidades dentro del rango normal.

Referencias

Agresti, A. 1990. Categorical data Analysis. New York: Wiley.

Agresti, A. and B. Coull. 1998. Aproximate is better than ´Exact´ for interval estimations and differences of proportions results from adding two successes and two failures. Amer. Statist., 54: 280-288.

Berger, J.O. and T. Sellke. 1987. Testing a point null hypothesis: The irreconciability of p values and evidence. J. Amer. Statist. Assoc. 82: 112-122.

Billingsy, P. 1995. Probability and Measure. 3rd edition. New York: Wiley.
Brwon, P.J. and W.A. Fuller. 1991. Statistical analysis of measurement error models and applications. Providence, R.I.: American mathematical Society.
Cochran, W.G. and G.M. Cox. 1957. Experimental Designs. 2nd editon. New York: Wiley.
Perez, H. 2008. Estad´sitica par alas Ciencias Sociales,, del Comportamiento y de la salud. 3ª. Edición. Cengage Mexico.
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.
Schevish, M.J. 1995. Theory of Statistics. New York: Springer-Verlag.
Schevish, M.J. 1996. P values: What they are and what they are noy. Amer. Statist. 50: 203-206.