

PRUEBA DE BONDAD DE AJUSTE
NOTAS DE CLASES
PROFESOR: Carlos Alberto Márquez Fernández
GUÍA 2

Objetivo:

- Determinar la distribución de probabilidad que puede modelar un conjunto de datos de muestra

Objetivos Específicos

- Aplicar la pruebas de bondad de ajuste de Chi cuadrado para determinar cuando un conjunto de datos sigue una distribución teórica
- Aplicar la pruebas de bondad de ajuste de Kolmogoró Smirnov para determinar cuando un conjunto de datos sigue una distribución teórica

Introducción

En los sistemas reales regularmente nos encontramos con variables cuyo comportamiento es aleatorio, y son susceptibles de ser modeladas por variables de entrada de un modelo estocástico, estas variables requieren de un tratamiento estadístico para su generación de manera artificial, el cual se realiza usualmente por medio de un modelo teórico de distribución de probabilidad, es así como las pruebas de bondad de ajuste es una buena herramienta para determinar el comportamiento de un conjunto de datos

En muchas ocasiones cuando se está simulando un sistema, las variables controlables del modelo son estocásticas, las variables de entradas las cuales tienen un comportamiento aleatorio son muestreadas con el objetivo de tener un conjunto de datos sobre dicha variable aleatoria y encontrar el modelo de distribución de probabilidad que pueda representar la serie de datos producidos de la muestra.

En otras palabras lo que se desea es probar la hipótesis que un modelo de probabilidad teórico (normal, exponencial, poisson etc) en particular será un modelo satisfactorio de la población en estudio.

Este ajuste de los datos a un modelo de distribución de probabilidad se puede realizar por medio de las pruebas estadísticas más conocidas como pruebas de bondad de ajuste tales como la **chi cuadrado y la de Kolmogor-smirnov**.

Debe tenerse en cuenta que cuando a una serie de datos se le aplica cualquiera de las pruebas de bondad y se encuentra que ningún modelo teórico se puede ajustar al serie de datos, se trabaja entonces con el modelo empírico (que no es modelo estándar conocido)

PRUEBA DE BONDAD DE AJUSTE CHI CUADRADO χ^2

El procedimiento de la prueba requiere una muestra aleatoria de tamaño n proveniente de la población cuya distribución de probabilidad es desconocida. Estas n observaciones se pueden distribuir en k intervalos de clases y pueden ser representadas en histogramas.

La prueba se puede utilizar tanto para distribuciones discretas como para distribuciones continuas

La prueba se puede sintetizar en los siguientes pasos.

1. Se colocan los n datos históricos (muestrales) en una tabla de frecuencia de la siguiente manera:
 - a. Se busca en cuantos intervalos de clases se puede distribuir los datos en estudio lo cual se puede hacer $m = \sqrt{n}$ o alternativamente es muy común utilizar las encontrar el número de intervalos se aplica la **regla de sturges**:
 $m = 1 + 3,3 \log n$ donde n es el número de datos
 - b. Luego encontramos el **rango** el cual es la diferencia entre el mayor valor y el menor valor.

$$R = X_{\max} - X_{\min}$$

- c. Amplitud de cada intervalo está dado por:

$$A = \frac{\text{Rango}}{\text{Número de intervalos}}$$

- d. se obtienen las frecuencias observadas en cada intervalos se calcula la media, la varianza y las desviación estándar.
2. Se propone una distribución de probabilidad una distribución de probabilidad de acuerdo con la tabla de frecuencia o con la curva que muestre un histograma o polígono de frecuencia.
3. Con la distribución propuesta, se calcula la frecuencia esperada para cada uno de los intervalos (FE_i) de la siguiente manera:
 - Si la variable es continua se halla mediante la integración de la distribución propuesta y luego se multiplica por el número total de datos.
 - Si la variable es continua se utiliza de modelo matemático de la distribución propuesta y se evalúan todas la categorías y luego se multiplica por el número total de datos.

4. Se calcula el estadístico de prueba

$$C = \sum_{i=1}^m \frac{(FE_i - FO_i)^2}{FE_i}$$

Nota: El estadístico de prueba tiene distribución Chi cuadrado con, m-k-1 grados de libertad, siempre que las frecuencias esperadas sean 5 o más para todas las categorías

5. Si el estimador C es menor o igual al valor correspondiente X^2 con m-k-1 grados de libertad (K= números de parámetros estimados de la distribución propuesta estimada por los estadísticos muestrales) y a un nivel de confiabilidad de $1-\alpha$, entonces no se puede rechazar la hipótesis de que los datos siguen la distribución que se propuso.

DETERMINACIÓN DE LOS GRADOS DE LIBERTAD EN UNA PRUEBA DE BONDAD DE AJUSTE.

Si queremos calcular el número apropiado de grados de libertad en una prueba de bondad de ajuste chi cuadrado, contamos el número de clases (representados por m) para los cuales hemos comparados las frecuencias observadas y las esperadas, entonces aplicamos la regla (m-1) y luego se resta un grado adicional de libertad para cada parámetro de la población que tenga que ser estimado de los datos de la muestra. Debe notarse que esta regla es igual a la que tenemos en el punto 5. m-k-1

RECOMENDACIONES IMPORTANTES

Un aspecto que debe notarse en la aplicación de este procedimiento de prueba es el relacionando con la magnitud de las frecuencias esperadas. Si estas frecuencias son muy pequeñas, entonces el estadístico χ^2 no reflejará el alejamiento entre lo observado y lo esperado, si no solo la pequeña magnitud de las frecuencias esperadas. No hay ningún acuerdo general con respecto al valor mínimo de las frecuencias esperadas, pero los valores 3, 4 y 5 son los que mas se utilizan como mínimos. Algunos autores sugieren que la frecuencia esperada puede ser tan pequeña como 1 o 2, con tal que muchas de ellas sean mayores que 5. Si se espera que una frecuencia sea demasiado pequeña, entonces puede combinarse con la frecuencia esperada en un intervalo de clase adyacente. Las frecuencias observadas correspondientes también se combinan, por lo que m debe disminuirse en uno.

No es necesario que los intervalos de clases tengan el mismo ancho.

La prueba de bondad de ajuste chi cuadrado tal vez nos sea el mejor procedimiento cuando al variable es continua.

PRUEBA CHI CUADRADO PARA UNA DISTRIBUCIÓN DE PROBABILIDAD BINOMIAL

En un estudio diseñado para determinar la tolerancia de los pacientes a un nuevo analgésico, 100 médicos seleccionaron cada uno una muestra de 25 pacientes para participar en el estudio. Cada paciente después de haber tomado el nuevo analgésico durante un periodo especificado, fue interrogado para saber si prefería éste o el que había tomado regularmente con anterioridad.

El interés consiste en determinar si los datos son compatible con la hipótesis de que extrajeron de una población que sigue una distribución Binomial.

Aplicación 2

Analicemos una distribución discreta

Para analizar el número de artículos defectuosos en una fabrica en la ciudad de Medellín, tomamos una muestra aleatoria de $n=60$ artículos y se observó el número de defectuosos y se obtuvieron los siguientes resultados:

Números de artículos defectuosos	Frecuencia Observada
0	32
1	15
2	9
3	4
Total n	60

Se propuso una distribución de poissón ya que esta se define para $x=0,1,2,3,\dots$ mediante el siguiente modelo.

$$P(X) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Al proponer esta distribución es claro que debemos conocer la media, la cual la estimamos a partir de los datos de la muestra. La estimación del número promedio de artículos defectuosos, es el promedio maestro $\lambda = (32 \times 0 + 15 \times 1 + 9 \times 2 + 4 \times 3) / 60 = 0.75$

$$P(X = 0) = \frac{0.75^0 e^{-0.75}}{0!} = 0.472$$

$$P(X = 1) = \frac{0.75^1 e^{-0.75}}{1!} = 0.354$$

$$P(X = 2) = \frac{0.75^2 e^{-0.75}}{2!} = 0.133$$

$$P(X \geq 3) = 1 - (p_0 + p_1 + p_2) = 0.041$$

Las frecuencias esperadas se calculan multiplicando el tamaño de la muestra $n=60$ Por las probabilidades p_i , esto es, $FE_i = n * p_i$

Nº de defectos	prob. Poisson	FE
0	0,472	28,32
1	0,354	21,24
2	0,133	7,98
3(o más)	0,041	2,46

Puesto que la frecuencia esperada el la última celda es menor que 5 se combinan las dos últimas celdas:

Nº de defectos	Fo	FE
0	32	28,32
1	15	21,24
2(o más)	13	10,44

La estadística de prueba Chi cuadrado tiene $m-k-1=3-1-1=1$ grados de libertad, debido a que la media de poisson fue estimada a partir de los datos de muestra.

Al aplicarse le procedimiento para la prueba $\alpha = 0.05$, de la siguiente manera.

1. La variable de interés es la forma de la distribución de los defectos en las tarjetas de circuito impreso.
2. H_0 : El número de defectos en la tarjetas de circuito impreso tiene una distribución de Poisson.
3. H_a : El número de defectos en la tarjetas de circuito impreso no tiene una distribución de Poisson.
4. $\alpha = 0.05$
5. El estadístico de prueba es:

$$X^2 = \sum_{i=1}^k \frac{(Fo - FE)^2}{FE}$$

6. Se rechaza H_0 si $X^2 > X^2_{0.05,1}=3.84$

7. Cálculos

$$X_0^2 = \frac{(32 - 28.32)^2}{28.32} + \frac{(15 - 21.24)^2}{21.24} + \frac{(13 - 10.44)^2}{10.44} = 2.94$$

8. Conclusiones:

$$X_0^2 < X^2_{0.05,1}$$

$2.94 < 3.84$ No es posible rechazar no es posible rechazar la hipótesis nula de que la distribución de los defectos en las tarjetas de circuito impreso sigue una distribución de Poisson

Problema 2

Aplicación de la distribución de Poisson

El gerente de un banco de la ciudad de Medellín busca asignar la cantidad de cajas a la cantidad de tal manera que se brinde un buen nivel de servicio y al mismo tiempo mantener un nivel razonable en el costo total de mano de obra.

Para llevar a cabo el estudio una variable importante es la llegadas de los clientes, sobre la cual realizamos un muestreo, luego se analiza si la distribución de probabilidad de Poisson puede modelar este conjunto de datos.

Para llevar a cabo el proceso de muestreo definiremos las entradas de los clientes en términos de la cantidad de clientes que entran a banco durante intervalos de cinco minutos. En consecuencia, las hipótesis nula y alternativa siguientes son las adecuada para el estudio del banco.

H_0 : La cantidad de clientes que llegan al banco durante intervalos de 5 minutos tiene una distribución de Poisson de probabilidades

H_a : La cantidad de clientes que llegan al banco durante intervalos de 5 minutos no tiene una distribución de Poisson de probabilidades.

Para probar la hipótesis de una distribución de Poisson para la cantidad de entradas un empleado del banco selecciona al azar una muestra de 128 intervalos de 5 minutos, en días hábiles y durante un periodo de tres semana, puede ser en las horas de la mañana. Para cada intervalo se anota la cantidad de clientes que llegan. Y se determina la cantidad de intervalos de 5 minutos donde no hubo entradas, la cantidad de intervalos de 5 minutos donde con una entrada, la cantidad de intervalos de 5 minutos donde con dos entradas, y así sucesivamente. Los datos los podemos representar en un atabla de frecuencias.

Cantidad de cilientes	Fo
0	2
1	8
2	10
3	12
4	18
5	22
6	22
7	16
8	12
9	6
Total	128

Cantidad de cilientes	f(x)	FE=f(X)*128
0	0,0067	0,8625
1	0,0337	4,3123
2	0,0842	10,7807
3	0,1404	17,9679
4	0,1755	22,4598
5	0,1755	22,4598
6	0,1462	18,7165
7	0,1044	13,3689
8	0,0653	8,3556
9	0,0363	4,6420
10 o más	0,0318	4,0740

Las categorías de frecuencias menores de 5 no causan dificultad, porque se puede sumar con categorías adyacentes para satisfacer el requisito de “cuando menos 5” para la frecuencia esperada. En este caso combinaremos 0 y 1 en una sola categoría, y además combinaremos 9 con “10 o más” para formar otra categoría. Con ello satisfacemos la regla de de una frecuencia esperada de mínima de 5.

Cantidad de clientes	Fo	FE	$\frac{(Fo - FE)^2}{FE}$
0 o 1	10	5,1747	4,4994
2	10	10,7807	0,0565
3	12	17,9679	1,9822
4	18	22,4598	0,8856
5	22	22,4598	0,0094
6	22	18,7165	0,5760
7	16	13,3689	0,5178
8	12	8,3556	1,5896
9 o más	6	8,7160	0,8463
Total	128	128,0000	10,9628

Finalmente
$$\chi^2 = \sum_{i=1}^k \frac{(Fo - FE)^2}{FE} = 10.9628$$

Para determinar los grados de libertad adecuados, asociado con esta prueba de bondad de ajuste tiene m-k-1 grados de libertad, siendo m la cantidad de categorías, k la cantidad de parámetros poblacionales estimados con los datos de muestra. Para nuestro caso tenemos entonces m-k-1=9-1-1=7 grados de libertad para la distribución chi cuadrado en el estudio del banco.

Al aplicarse le procedimiento para la prueba $\alpha = 0.05$, de la siguiente manera.

Número de unidades con defecto	Número de muestras
0	138
1	53
2 ó más	9

1. La variable de interés es el número de clientes que entran

al banco

2. H_0 : El número de clientes que entran al banco tiene una distribución de Poisson.
3. H_a : El número de clientes que entran al banco no tiene una distribución de Poisson.
4. $\alpha = 0.05$
5. El estadístico de prueba es:

$$\chi^2 = \sum_{i=1}^k \frac{(F_o - F_E)^2}{F_E}$$

6. Se rechaza H_0 si $\chi^2 > \chi^2_{0.05,7} = 3.84$
7. Los cálculos se realizan similar al ejercicio anterior

8. Conclusiones:

$$\chi_0^2 < \chi^2_{0.05,1}$$

10.9766 < 14.07 No es posible rechazar no es posible rechazar la hipótesis nula de que las entradas de clientes puede modelarse por medio de Poisson

Ejemplo:

Si un ingeniero de control de calidad toma una muestra de 10 neumáticos que salen de una línea de ensamblaje y él desea verificar sobre la base de los datos que siguen, los números de llantas con defectos observadas en 200 días, si es cierto que el 5% de todos los neumáticos tienen defecto; es decir, si el muestrea una población binomial con $n = 10$ y $p = 0.05$

Establecer la hipótesis

Ho: La población es binomial

Ha: La población no es binomial

Establecer la estadística de prueba

$$\chi^2 = \sum_{i=1}^k \frac{[f_{o_i} - f_{e_i}]^2}{f_{e_i}}$$

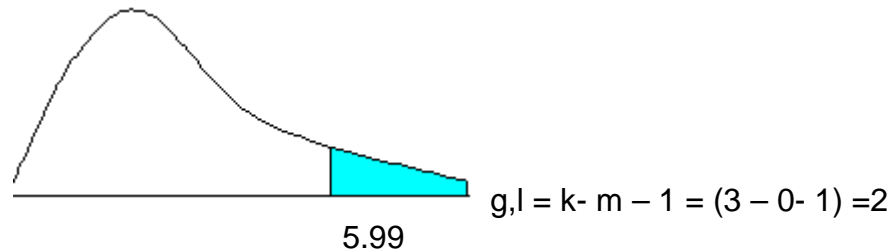
O_i = Valor observado en la i-ésimo dato.

E_i = Valor esperado en la i-ésimo dato.

K = Categorías o celdas.

m = Parámetros

1. 3. Definir el nivel de significancia y la zona de rechazo



Nivel de significancia = 0.05

Zona de rechazo = $\{ \chi^2 / \chi^2 \geq 5.99 \}$

m = 0 porque no se necesita estimar ningún parámetro

$$\chi^2 = \sum_{i=1}^k \frac{[f_{o_i} - f_{e_i}]^2}{f_{e_i}}$$

Calculamos el estadístico de prueba

Para poder calcular las frecuencias esperadas tenemos que calcular las probabilidades utilizaremos la fórmula de la binomial

$$f(x) = P(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

donde $n = 10$ $p = 0.05$

$$f(0) = \binom{10}{0} 0.05^0 (1-0.05)^{10-0} = 0.599$$

$$f(1) = \binom{10}{1} 0.05^1 (1-0.05)^{10-1} = 0.315$$

y la probabilidad de 2 ó más = $1.0 - 0.599 - 0.315 = 0.086$

Ahora ya podemos encontrar las frecuencias esperadas:

$$200 (0.599) = 119.8 \qquad 200(0.315) = 63 \qquad 200 (0.086) = 17.2$$

Número de unidades con defecto	Número de muestras Observadas	Valor Esperado
0	138	119,8
1	53	63
2 ó más	9	17,2
Total	200	200

Al aplicar la formula se tiene:

$$\chi^2 = \frac{(138-119.8)^2}{119.8} + \frac{(53-63.0)^2}{63} + \frac{(9-17.2)^2}{17.2} = 8.26$$

Como 8.26 es mayor que 5.99, se rechaza la hipótesis nula con un nivel de significancia de 0.05.

Conclusión

Se concluye que el porcentaje verdadero de neumáticos con defecto no es el 5%.

Ejemplo. Si el número de errores que comete una secretaria al transcribir un documento es una variable aleatoria que tiene una distribución de Poisson. Se reviso 440 transcripciones hechas por ella y arrojo los siguientes resultados:

Variable = números de errores

Número de errores	Frecuencia
0	18
1	53
2	103
3	107
4	82
5	46
6	18
7	10
8	2
9	1

Probar si los datos de los errores se ajustan a una distribución de Poisson.
Use $\alpha = 0.05$.

H_0 : La población se comporta como una distribución de Poisson

H_a : La población no se comporta como una distribución de Poisson.

El estadístico de prueba que usaremos es:

$$\chi^2 = \sum \frac{(f_{o_i} - f_{E_i})^2}{f_{E_i}}$$

Nivel de significación $\alpha = 0,05$

$$gl. = 9-1-1=7$$

$$\chi^2_{0,05;7} = 14,067$$

$$\text{Región de rechazo} = \{\chi^2 \mid \chi^2 \geq 14,067\}$$

Para poder calcular las frecuencias esperadas tenemos que calcular las probabilidades utilizaremos la formula de la distribución de Poisson:

$$p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Como no se conoce la media de la distribución de Poisson la estimamos con la media de los datos, que es 3,04 luego, $\lambda = 3,04$

errores	Frecuencia observada	Prob.	Frecuencia esperada
0	18	0,0478	21,032
1	53	0,1454	63,976
2	103	0,2210	97,240
3	107	0,2239	98,516
4	82	0,1702	74,888
5	46	0,1034	45,496
6	18	0,0524	23,056
7 ó mas	13	0,0359	15,796
Total	440	1	440

Aplicamos los datos al estadístico de prueba

$$\chi^2 = \frac{(18-21,032)^2}{21,032} + \frac{(53-63,976)^2}{63,976} + \dots + \frac{(13-15,796)^2}{15,796} = 5,6766$$

Como 6,7566 es menor 14,076 no se rechaza la hipótesis nula.

Conclusión: La población se comporta como una distribución de Poisson con media 3,04

Ejemplo 3.

El Departamento de Psicología, basándose en informaciones anteriores, al final del semestre antepasado, el 80% de los alumnos aprobaron todas las materias inscritas, un 10% aprobó la mitad, un 6% reprobó todas las materias y un 4% se retiró. Al final del semestre pasado el departamento selecciono a 400 alumnos, resultado 287 aprobaron todas las asignaturas, 49 aprobaron la mitad, 30 reprobaron todas las asignaturas y 34 se retiraron. ¿Podemos concluir, a raíz de los resultados, que la información del semestre antepasado se ha vuelto a repetir el semestre pasado?

Hipótesis nula: de que los porcentajes del semestre pasado son los mismos que en el semestre antepasado.

Atributos	Datos observados	Probabilidad	Datos esperados
Aprobó todo	287	0,80	320
Aprobó la mitad	49	0,10	40
Reprobó todo	30	0,06	24
Se retiró	34	0,04	16
Total	400	1	400

$$\chi^2 = 27,178$$

Como tenemos 4 categorías y ningún parámetro estimado los grados de libertad serán: 4-0-1= 3

$$\chi^2_{0,05;3} = 12,84$$

Como 27,178 es mayor que 12,84 se rechaza la hipótesis nula.

Conclusión: Los porcentajes no se repitieron el semestre pasado

PRUEBA DE BONDAD DE AJUSTE DE KOLMOGOROV SMIRNOV

- Es aplicable solamente a variables aleatorias continuas
- Comparar la gráfica de la distribución empírica acumulada con la correspondiente gráfica de la función de densidad acumulada de la distribución teórica propuesta.
- Si hay un acercamiento entre las gráficas existe una probabilidad de que la distribución teórica se ajusta a los datos.

El hecho de que utiliza la distribución de probabilidad acumulada la hace un poco más eficiente que la prueba anterior

La metodología de la prueba es la siguiente:

1. Se colocan los n datos históricos en una tabla de frecuencias con $m = \sqrt{n}$ intervalos o utilizando la fórmula de de Struges:
 $K = 1 + 3.3 \log n$; donde n es el número de datos de la muestra.

1.1 Encuentre la amplitud del intervalo de clase por medio de la siguiente relación

$$Amplitud = \frac{Rango}{k}$$

1.2 Para cada intervalo se tendrá la frecuencia observada i (FO_i). Se calcula la media y la varianza de los datos

2. Se encuentra la probabilidad observada (PO_i), dividiendo la frecuencia observada de cada intervalo por el número total de datos.
3. Se calcula la probabilidad acumulada observada de cada intervalo (PAO_i) del paso 2
4. Se propone una distribución de probabilidad de acuerdo con la forma de tabla de frecuencia obtenida en 1. O con la gráfica de los datos.
5. Con la función acumulada de la distribución propuesta, se calcula la probabilidad esperada acumulada para cada intervalo (PEA_i) mediante la integración de la distribución propuesta.

6. Se calcula la probabilidad acumulada (PAE_i) para cada intervalo de clase.
7. Se calcula el valor absoluto entre la diferencia de PAO y PAE para cada intervalo y se selecciona la máxima diferencia, llamándola MD
8. El estimador MD se comporta con un valor limite correspondiente a la (tabla que contiene los valores críticos de kolmogorov-Smirnov). Con n datos y a un nivel de confianza de $1 - \alpha$. Si el estimador MD es menor o igual al valor limite de la tabla, entonces se acepta la hipótesis de que la información histórica sigue la distribución propuesta.

EJERCICIOS DE APLICACIÓN CUANDO SE PROPONE UNA DISTRIBUCIÓN DE PROBABILIDAD UNIFORME.

Un muestreo realizado sobre la demanda de televisores en un almacén del centro de Medellín durante 40 días tiene el siguiente comportamiento.

2	1	5	8	9	10	11	2	3	4
2	5	3	4	6	6	2	13	11	8
8	9	8	10	4	5	7	8	9	11
7	12	5	11	6	10	12	13	13	12

1. Hallamos el rango $R = X_{\max} - X_{\min}$

$$R = 13 - 1 = 12$$

$$k = 1 + 3.3 \log 41$$

2. Encontramos el valor de $k = 6.3$ y se aproxima a 6

$$\text{Luego } k = 6$$

3. Hallar la amplitud $A = \frac{R}{K} = \frac{12}{6} = 2$

4. El limite inferior es 1 y al superior se le suma la amplitud y así sucesivamente, como se nota en la tabla de frecuencias
5. Luego se completa la tabla de frecuencias
6. Se quiere proponer la hipótesis que los datos en estudio siguen una distribución uniforme de lo cual procedemos de la siguiente manera:
 - a. Integramos la función de densidad de la distribución uniforme

La función de densidad de una variable aleatoria uniforme es : $f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{para otros valores} \end{cases}$

La función de distribución acumulada se encuentra integrando $f(x)$ de la siguiente manera :

$$F(x) = \int_1^{LS} \frac{1}{b-a} dx = \frac{1}{b-a} \int_1^{LS} dx = \frac{1}{b-a} X \Big|_1^{LS} = \frac{LS-1}{12-1}$$

$$\text{luego } F(x) = \frac{LS-1}{12}$$

Con esta formula se encuentra la POA (Probabilidad observada acumulada) para cada intervalo de la s

$$\text{Para el primer intervalo } POA = \frac{3-1}{12} = 0.167$$

$$\text{Para el segundo intervalo } POA = \frac{5-1}{12} = 0.333 \text{ y así sucesivamente}$$

LS = Limite superior de cada intervalo

Li	Ls	Fo	FOA	POA	PEA	$ POA - PEA $
1	3	7	7	0,175	0,167	0,0167
3	5	7	14	0,35	0,333	0,025
5	7	5	19	0,475	0,5	0,0083
7	9	8	27	0,675	0,667	0,0167
9	11	7	34	0,85	0,833	0
11	13	6	40	1	1	0

Podemos observar que la máxima diferencia de la columna $|POA - PEA|$ es 0.025 La cual al ser comparada con la tabla de los valores críticos de kolmogorov –Smirnov el cual es $d_{5\%,40} = 0.2150$, con la cual se cumple la hipótesis, luego no se rechaza la hipótesis que este conjunto de datos se pueden modelar por medio de una distribución uniforme entre 1 y 13 televisores demandados por días A un nivel de confianza de del 95%.

Luego que se ha encontrado que los datos se pueden modelar por una distribución uniforme se busca entonces generar variables aleatorias uniformes como entradas estocásticas para un modelo de simulación a través la formula

$$X = a + (b - a) * R$$

$a = \text{Limite inferior}$

$b = \text{Limite superior}$

$R = \text{Un número aleatorio uniforme}(0,1)$

Para estimar los parámetros a y b utilizamos los métodos de los momentos en base a los datos históricos.

Como se tienen dos momentos se deben usar los dos primeros de la media y la varianza. Es decir, igualamos la media y la varianza de la población (μ y σ^2) a la media de la muestra \bar{X} y S^2

Se sabe que la media para una distribución uniforme $\mu = \frac{a+b}{2} = \bar{X}$ y que la varianza es $\sigma^2 = \frac{(b-a)^2}{12} = S^2$

Despejando a y b de las dos expresiones anteriores:

De la primera ecuación despejamos b , $b = 2\bar{X} - a$

$$\frac{(b-a)^2}{12} = S^2$$

De la segunda despejamos b , $(b-a)^2 = 12S^2$

$$b-a = 2\sqrt{3S^2}$$

$$b = a + 2\sqrt{3S^2}$$

$$a + 2\sqrt{3S^2} = 2\bar{X} - a$$

Ahora igualamos las dos ecuaciones $2a = 2\bar{X} - 2\sqrt{3S^2}$

$$a = \bar{X} - \sqrt{3S^2}$$

En resumen la ecuaciones para estimar a y b requieren que se estime la media y la varianza de la muestra

$$\begin{aligned} a &= \bar{X} - \sqrt{3S^2} \\ b &= a + 2\sqrt{3S^2} \end{aligned}$$

EJERCICIOS DE APLICACIÓN CUANDO SE PROPONE UNA DISTRIBUCIÓN DE PROBABILIDAD EXPONENCIAL.

Un muestreo realizado sobre la demanda de café en el eje cafetero a través del tiempo, medida por cargas es

0	1	10	7	4	5	6	6	8	4
2	9	7	12	5	6	3	8	1	16
8	3	2	11	4	19	13	0	7	6
21	19	0	0	5	2	2	2	20	2
2	5	2	2	3	0	3	15	4	12

7. Hallamos el rango $R = X_{\max} - X_{\min}$

$$R = 20 - 0 = 20$$

$$k = 1 + 3.3 \log 50$$

8. Encontramos el valor de $k = 6.6$ y se aproxima a 7

$$\text{Luego } k = 7$$

9. Hallar la amplitud $A = \frac{R}{K} = \frac{20}{7} = 2.8$

$$A = 3$$

Li	Ls	Mc	Fo	FOA	POA	PEA	$ POA - PEA $
0	3	1,5	20	20	0,4	0,39	0,00653979
3	6	4,5	12	32	0,64	0,63	0,00789052
6	9	7,5	7	39	0,78	0,78	0,00314024
9	12	10,5	4	43	0,86	0,86	0,00465657
12	15	13,5	2	45	0,9	0,92	0,01790882
15	18	16,5	1	46	0,92	0,95	0,03020844
>	18	19,5	4	50	1	1	0

La $D_m = 0,030208$ se compara con $d_{5\%,50} = 0.2150$ lo cual indica que los datos de café en el eje cafetero siguen una distribución exponencial con media de $\lambda = 6$ toneladas/días. Es decir,

Demanda $\sim \text{Exp} (\lambda = 6)$ toneladas/días

Recordemos que el parámetro que define la distribución de probabilidad exponencial es la media λ la cual se encuentra cuando los datos están

agrupados de la siguiente manera:

$$\bar{X} = \frac{\sum Mc_i * Fo}{n}$$

Mc_i = Marca de clases

Fo = Frecuencia observada

n = Tamaño de la muestra

$$s^2 = \frac{\sum (Mc_i - \bar{X})^2 * Fo}{n}$$

Mc_i = Marca de clases

\bar{X} = media muestral

Fo = Frecuencia observada

n = Tamaño de la muestra

Con esta formula se puede comprobar que $\bar{X} = 6$ y se puede demostrar por el estimador de máxima verosimilitud de λ es la media muestral \bar{X} , es decir $\lambda = \bar{X}$

Luego $\lambda = 6$

Recuérdese que al integrar la función de densidad exponencial

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{para } x \geq 0 \\ 0 & \text{para otros valores} \end{cases}$$

Se obtiene

$$F(x) = \int_0^{LS} \frac{1}{\lambda} e^{-\frac{LS}{\lambda}} dx = -\int_0^{LS} \frac{1}{\lambda} e^{-\frac{LE}{\lambda}} dx = -e^{-\frac{LE}{\lambda}} \Big|_0^{LS} = -(\ell^{-\frac{LE}{\lambda}} - \ell^0) = 1 - \ell^{-\frac{LE}{\lambda}}$$

$$\text{Luego } F(x) = 1 - \ell^{-\frac{LS}{\lambda}}$$

Con la cual encontramos PEA

Para nuestro ejemplo

$$F(x) = 1 - \ell^{-\frac{LS}{6}} \quad \text{Para cada intervalo}$$

EJERCICIOS DE APLICACIÓN CUANDO SE PROPONE UNA DISTRIBUCIÓN DE PROBABILIDAD NORMAL.

Un almacén encuentra que el tiempo de entrega de los pedidos por parte de su proveedor es aleatorio, el proveedor alega que siempre entrega su mercancía en 7 días o menos. La evitar el debate el gerente del almacén hizo un muestreo de las últimas entregas y obtuvo la siguiente información:

Demuestre a partir de la prueba de bondad so los datos se pueden modelar a partir de una distribución de probabilidad normal.

Hallamos la media y la varianza muestrales aplicando las formulas para datos agrupados utilizados en el ejercicio anterior.

$$\bar{X} = 8,4 \quad Y \quad S=3$$

Li	Ls	Mc	Fo	FOA	POA	Z	PEA	$ POA - PEA $
0	1	0,5	0	0	0	-2,5	0,00637	0,00637181
2	3	2,5	2	2	0,04	-1,8	0,03436	0,00135306
4	5	4,5	8	10	0,18	-1,1	0,12515	0,05342231
6	7	6,5	12	22	0,39	-0,5	0,31597	0,07689158
8	9	8,5	20	42	0,75	0,19	0,57597	0,17402612
10	11	10,5	10	52	0,93	0,86	0,80572	0,12285495
12	13	19,5	3	55	0,98	1,53	0,93734	0,04480127
14	15	19,5	1	56	1	2,2	0,98622	0,01378158
16	17	19,5	0	56	1	2,87	0,99797	0,00202614

56

La $DM=0,17403$ se compara con $d_{5\%,56} = 0,183382$, y ya que DM es menor, entonces no se puede rechazar la hipótesis de que los días de entrega de los pedidos siguen una distribución normal con media de $\bar{X} = 8,4$ Y $S=3$ con un nivel de confianza de 95%

Tiempo de entrega $\sim N(8.4,3)$ días