

Simulaciones

TRABAJO PRACTICO NUMERO 2: TOMA DE DATOS

Del Rio, Lucas | 69587 | 4k3
Godoy, Federico | 62174 | 4k3
Hefty, Nicolas | 62718 | 4k3
Tolozza, Macarena | 62326 | 4k3

Año 2018 – Grupo:

INDICE

Contenidos

Enunciado	2
Realizar la toma de datos de un sistema real (mínimo 200 datos).	3
Datos Exponencial	3
Variable observada:	3
Muestra de datos:	3
Datos Poisson	4
Variable observada:	4
Muestra de Datos:	4
Resolución	5
Distribución Exponencial	5
PRUEBA DE CHI CUADRADO	5
PRUEBA DE KOLMOGOROV-SMIRNOV	7
HISTOGRAMA	7
Distribución de Poisson	8
PRUEBA DE CHI CUADRADO	8
HISTOGRAMA	9

Enunciado

Tema: Toma de datos

T.P. Nº 2

Descripción

- a) Realizar la toma de datos de un sistema real (mínimo 200 datos).
- b) Determinar a qué distribución de probabilidad se puede asimilar la misma.
- c) Realizar su histograma y gráfica.
- d) Determinar hipotéticamente el comportamiento probabilístico de la muestra.
- e) Realizar las 2 pruebas que aceptan o rechazan tal hipótesis.
- f) Repetir esta experiencia con otra muestra diferente de datos (tamaño mínimo de la muestra 200). – *Trabajado en paralelo* –

Observaciones

Pedido: Segunda semana de clase.

Entrega: Tercera semana de clase.

Realizar la toma de datos de un sistema real (mínimo 200 datos).

DATOS EXPONENCIAL

Variable observada:

Diferencia de tiempo entre entrada de las primeras 200 personas adultas en Nuevocentro shopping por la puerta de los cines, a partir de las 17 horas un día domingo.

Muestra de datos:

#	X	#	X	#	X	#	X	#	X	#	X	#	X	#	X
1	1.67	2	2.34	5	0.79	76	0.27	10	6.48	12	1.63	15	9.33	17	0.66
		2		5				10		12		15		17	12.9
2	1.3	7	2.68	2	0.52	77	8.25	2	0.53	7	6.82	2	6.54	7	6
		2		5				10		12		15		17	
3	0.32	8	3.5	3	0.62	78	0.38	3	2.36	8	0.65	3	0.52	8	1.09
		2		5				10		12		15		17	18.7
4	0.42	9	2.09	4	0.38	79	2.27	4	0.54	9	0.61	4	1.41	9	2
		3		5				10		13	10.9	15		18	21.4
5	0.67	0	3.31	5	0.77	80	0.27	5	0.6	0	1	5	0.76	0	9
		3		5				10	14.6	13	11.3	15	20.6	18	
6	0.58	1	8.44	6	1.44	81	0.49	6	1	1	6	6	6	1	9.01
	10.1	3	22.5	5				10		13		15		18	
7	1	2	2	7	0.63	82	0.25	7	2.03	2	0.77	7	6.34	2	8.54
		3		5				10		13		15		18	
8	3.97	3	0.46	8	0.26	83	4.35	8	1.39	3	2.12	8	1.11	3	1.11
		3		5				10	10.0	13		15		18	
9	0.44	4	0.66	9	0.56	84	4.2	9	2	4	2.71	9	0.39	4	1.48
1		3		6			12.8	11		13		16		18	
0	0.68	5	9.47	0	0.26	85	5	0	4.13	5	0.93	0	0.59	5	5.41
1		3		6				11	15.5	13		16		18	11.8
1	0.83	6	2.96	1	1.76	86	0.27	1	6	6	0.5	1	0.15	6	4
1		3		6			14.1	11		13		16		18	13.0
2	3.14	7	7.59	2	0.32	87	5	2	7.39	7	2.84	2	4.86	7	2
1		3		6				11		13		16		18	
3	0.53	8	13.9	3	1.31	88	6.27	3	0.62	8	2.56	3	6.06	8	9.68
1	13.3	3	23.0	6			40.5	11		13		16		18	
4	4	9	9	4	1.61	89	2	4	2.05	9	5.75	4	1.08	9	4.68

1		4		6	18.6			11		14		16		19	
5	3.56	0	0.88	5	5	90	0.93	5	1.84	0	1.11	5	7.26	0	0.37
1		4		6				11		14		16		19	
6	0.62	1	0.78	6	0.4	91	2.44	6	0.35	1	1.44	6	0.65	1	3.3
1	13.4	4		6				11		14		16		19	10.4
7	5	2	1.97	7	5.39	92	0.56	7	3.33	2	0.85	7	1.72	2	1
1		4		6				11		14		16		19	
8	0.48	3	9.69	8	3.95	93	6.53	8	2.19	3	0.91	8	1.45	3	7.88
1		4		6				11		14	30.7	16		19	
9	1.37	4	0.38	9	3.41	94	0.7	9	1.79	4	2	9	1.79	4	0.28
2		4		7				12		14		17		19	
0	0.6	5	0.71	0	2.04	95	8.54	0	3.11	5	0.94	0	0.4	5	8.3
2		4		7	22.6			12		14		17		19	
1	4.7	6	3.61	1	3	96	1.02	1	0.3	6	0.62	1	0.71	6	1.4
2		4		7				12		14		17		19	13.9
2	0.58	7	6.42	2	1.91	97	7.43	2	6.01	7	7.89	2	0.6	7	6
2	12.9	4		7				12		14		17		19	
3	5	8	3.14	3	7.01	98	0.73	3	0.31	8	1.06	3	0.55	8	1.53
2		4	14.3	7				12		14		17		19	
4	0.72	9	6	4	1.32	99	0.71	4	1.15	9	7.94	4	2.03	9	0.48
2	11.2	5		7	12.8	10		12		15		17		20	
5	4	0	0.58	5	3	0	2.51	5	0.58	0	0.57	5	0.96	0	4.61

DATOS POISSON

Variable observada:

Cantidad de personas adultas que ingresan al Nuevocentro Shopping por las puertas de entrada al cine, en un día domingo por un intervalo de 15 segundos entre las 17 y 18 horas.

Muestra de Datos:

#	X	#	X	#	X	#	X	#	X	#	X	#	X	#	X
		2						10		12		15		17	
1	6	6	4	51	1	76	7	1	3	6	5	1	2	6	8
		2						10		12		15		17	
2	4	7	2	52	4	77	8	2	3	7	0	2	2	7	5
		2	1					10		12		15		17	
3	1	8	0	53	0	78	5	3	3	8	5	3	6	8	6
		2						10		12		15		17	
4	2	9	6	54	7	79	3	4	0	9	6	4	4	9	9
		3					1	10		13		15		18	
5	1	0	2	55	5	80	1	5	2	0	2	5	5	0	5
		3					1	10		13		15		18	
6	2	1	5	56	2	81	5	6	0	1	4	6	2	1	4
		1	3					10		13		15	1	18	
7	3	2	1	57	4	82	7	7	0	2	1	7	0	2	9

		3	1					10		13	1	15		18	
8	8	3	0	58	5	83	5	8	6	3	2	8	2	3	3
		3						10		13		15		18	
9	5	4	4	59	4	84	4	9	4	4	4	9	3	4	8
1		3			1			11		13		16		18	
0	4	5	4	60	3	85	0	0	0	5	4	0	2	5	2
1		3						11		13		16		18	
1	0	6	4	61	7	86	3	1	4	6	7	1	0	6	5
1		3						11		13		16		18	
2	3	7	4	62	3	87	4	2	4	7	3	2	6	7	1
1		3						11		13		16		18	
3	1	8	2	63	4	88	7	3	4	8	4	3	4	8	3
1		3						11		13		16		18	
4	9	9	6	64	7	89	5	4	2	9	4	4	3	9	2
1	1	4						11		14		16		19	
5	2	0	4	65	4	90	0	5	5	0	5	5	2	0	5
1		4	1					11		14		16		19	
6	1	1	2	66	3	91	0	6	0	1	2	6	1	1	1
1		4						11		14		16		19	
7	4	2	5	67	0	92	7	7	0	2	5	7	2	2	7
1		4						11		14		16	1	19	
8	0	3	9	68	1	93	5	8	5	3	3	8	2	3	5
1		4						11		14		16		19	
9	0	4	5	69	0	94	7	9	7	4	2	9	3	4	8
2	1	4						12		14		17		19	
0	2	5	6	70	5	95	2	0	2	5	5	0	0	5	5
2		4						12		14		17	1	19	
1	6	6	4	71	7	96	5	1	7	6	8	1	4	6	0
2		4						12		14		17		19	
2	4	7	2	72	1	97	3	2	4	7	0	2	0	7	2
2		4	1					12		14		17	1	19	
3	6	8	0	73	5	98	3	3	0	8	6	3	5	8	8
2		4						12	1	14		17		19	
4	3	9	3	74	6	99	4	4	3	9	1	4	2	9	8
2		5				10		12		15	1	17	1	20	
5	3	0	9	75	8	0	7	5	9	0	1	5	6	0	0

Resolución

DISTRIBUCIÓN EXPONENCIAL

Dado que la probabilidad de ocurrencia de los eventos en el intervalo de tiempo es muy pequeña se estimó que correspondería una distribución exponencial.

Se plantea una hipótesis de que los datos que conforman la muestra se comportan como una distribución exponencial con:

Media: 4.47535

Lambda: 0.2234

Y un tamaño de muestra igual a 200

Para verificar si se rechaza o no la hipótesis realizaremos las pruebas de Chi cuadrado y Kolmogorov-Smirnov.

PRUEBA DE CHI CUADRADO

Sabiendo que:

N	200
Media	4,47535
Lambda	0,22344621

Valor	Máximo	40,52
	Mínimo	0,15
Cantidad de intervalos		11

Rango	4
-------	---

	Intervalo						
Rango	Desde	Hasta	Marca Clase	fo	P() c/mc	P() c/Pac	fe
4	0	4	2	134	0,5717	0,5909	118,1791
4	4	8	6	27	0,2339	0,2417	48,3476
4	8	12	10	17	0,0957	0,0989	19,7792
4	12	16	14	13	0,0391	0,0405	8,0918
4	16	20	18	2	0,0160	0,0166	3,3104
4	20	24	22	5	0,0066	0,0068	1,3543
4	24	28	26	0	0,0027	0,0028	0,5540
4	28	32	30	1	0,0011	0,0011	0,2267
4	32	36	34	0	0,0004	0,0005	0,0927
4	36	40	38	0	0,0002	0,0002	0,0379
4	40	44	42	1	0,0001	0,0001	0,0155
TOTAL				200	0,9674	0,9999	199,9893

Donde:

Rango: ValorMax/Cant de intervalos $\rightarrow 40.52/11$

Marca de Clase: Suma de límites del intervalo/2

Frecuencia observada (fo): cantidad de muestras incluidas en cada intervalo

Densidad (P() c/mc): $f(x) = \lambda \cdot e^{-\lambda \cdot x}$

Densidad Acumulada (P() c/Pac): $F(x) = 1 - e^{-\lambda \cdot x}$

Frecuencia Esperada (fe): Densidad acumulada de intervalo * tamaño de la muestra

Agrupando los valores con frecuencia esperada menor a 5 la tabla para chi cuadrado queda:

Desde	Hasta	fo	fe	c
0	4	134	118,1791	2,1180
4	8	27	48,3476	9,4259

8	12	17	19,7792	0,3905
12	16	13	8,0918	2,9772
16	44	9	5,5916	2,0777
				16,9893

Donde el Estadístico de prueba (c) es igual a la sumatoria de $((f_o - f_e)^2) / f_e$

Grado de Libertad: $5 - 2 - 1 = 2$

Para un nivel de confianza de 95% queda:

Resultado en tabla para $X^2_{0,95,2} = 5,99$

Comparando resultados $X^2_{obs} < X^2_{tabla} \rightarrow 16,9893 > 5,99$

Conclusión: La hipótesis se rechaza para esta prueba.

PRUEBA DE KOLMOGOROV-SMIRNOV

Desde	Hasta	f_o	f_e	$P_o()$	$P_e()$	$P_o() AC$	$P_e() AC$	$P_o(AC) - P_e(AC)$
0	4	134	118.1791	0.67	0.5909	0.67	0.5909	0.0791
4	8	27	48.3476	0.135	0.2417	0.8050	0.8326	0.0276
8	12	17	19.7792	0.085	0.0989	0.8900	0.9315	0.0415
12	16	13	8.0918	0.065	0.0405	0.9550	0.9720	0.0170
16	20	2	3.3104	0.01	0.0166	0.9650	0.9885	0.0235
20	24	5	1.3543	0.025	0.0068	0.9900	0.9953	0.0053
24	28	0	0.5540	0	0.0028	0.9900	0.9981	0.0081
28	32	1	0.2267	0.005	0.0011	0.9950	0.9992	0.0042
32	36	0	0.0927	0	0.0005	0.9950	0.9997	0.0047
36	40	0	0.0379	0	0.0002	0.9950	0.9999	0.0049
40	44	1	0.0155	0.005	0.0001	1.0000	0.9999	0.0001
								0.0791 MAX

Donde:

Probabilidad observada ($P_o()$): frecuencia observada del intervalo/ tamaño de la muestra

Probabilidad Esperada ($P_e()$): $P() \cdot c/Pac + P() \cdot c/Pac$ de intervalo anterior

Probabilidad observada acumulada ($P_o()AC$): $P_o() \cdot c/Pac + P_o() \cdot c/Pac$ de intervalo anterior

Probabilidad esperada acumulada ($P_e()AC$): $P_e() \cdot c/Pac + P_e() \cdot c/Pac$ de intervalo anterior

Grado de Libertad= $N-2=200$

$$K-S_{0,05,200} = \frac{1,36}{\sqrt{200}} = 0,0961$$

Comparando el valor máximo absoluto de la columna de diferencia de probabilidades con el valor obtenido en la tabla K-S obtenemos que:

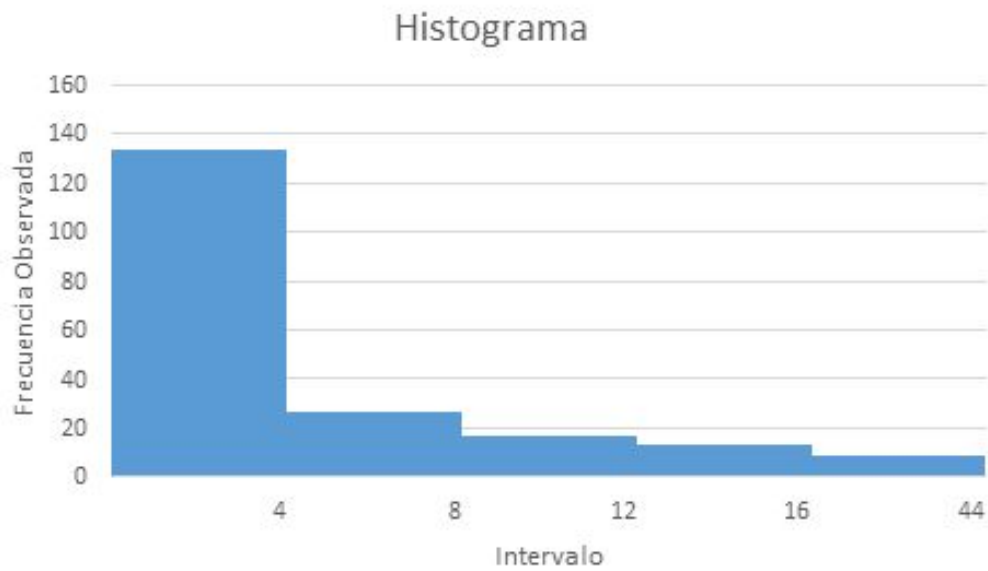
Comparando resultados $K-S_{obs} < K-S_{tabla}$

$$0,0791 < 0,0961$$

Conclusión: No se rechaza la hipótesis para esta prueba.

PODEMOS NOTAR QUE OBTENEMOS DIFERENTES CONCLUSIONES PARA CADA PRUEBA REALIZADA SOBRE LA MISMA MUESTRA DE DATOS, ENTONCES, ¿CUÁL ES VÁLIDA? AL SER UN CASO DE DISTRIBUCIONES CONTINUAS SE PREFIERE EL RESULTADO DADO POR LA PRUEBA KS (EN ESTE CASO QUE NO SE RECHAZA LA HIPÓTESIS) YA QUE ESTA PRUEBA TRABAJA DIRECTAMENTE SOBRE LAS OBSERVACIONES, MIENTRAS QUE LA PRUEBA DE CHI-CUADRADO TRABAJA SOBRE LOS DATOS AGRUPADOS.

HISTOGRAMA



DISTRIBUCIÓN DE POISSON

Se plantea una hipótesis de que los datos que conforman la muestra se comportan como una distribución de Poisson con $\lambda = 4$ y un tamaño de muestra de 200.

PRUEBA DE CHI CUADRADO

Repitiendo el mismo procedimiento que en la distribución exponencial obtenemos que:

		Valor r Max	
Lambda	4		16
Npoisson	200	Valor r Min	0

Desde	Hasta	fo	P()	fe
0	1	36	0,091578	18
2	3	45	0,341892	68
4	5	59	0,351660	70
6	7	27	0,163736	33
8	9	15	0,043001	9

10	11	6	0,007217	1
12	13	8	0,000839	0
14	15	3	0,000071	0
16	17	1	0,000005	0
			1,00	200

Donde:

La función de densidad (P()): $f(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$ Con x= límite inferior del intervalo

Realizamos la agrupación de intervalos para aquellos con frecuencia esperada menor a 5.

Desde	Hasta	fo	fe	c
0	1	36	18	17,075
2	3	45	68	7,993
4	5	59	70	1,826
6	7	27	33	1,009
8	17	33	10	50,713
				1,009

Grado de Libertad: $5-1-1=3$

Resultado en tabla para $X^2_{0,95,3} = 7,81$

Comparando resultados $X^2_{obs} < X^2_{tabla} \rightarrow 1,009 < 7,81$

Conclusión: La hipótesis se rechaza para esta prueba.

DADO QUE ESTAMOS TRABAJANDO CON UNA MUESTRA DE VARIABLES DISCRETAS NO
REALIZAMOS LA PRUEBA DE KOLMOGOROV-SMIRNOV

HISTOGRAMA

