Pruebas de Hipótesis

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases.

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos.

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$.

Pruebas de Hipótesis

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$. Cada N_i será una v.a. binomial con parámetros n y p_i con $i = 1, 2 \cdots, k$.

Pruebas de Hipótesis

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$. Cada N_i será una v.a. binomial con parámetros n y p_i con $i = 1, 2 \cdots, k$. El número esperado de ensayos en la categoría i será $E[N_i] = n\, p_i\,;\ i = 1, 2, \cdots, k\,.$

Pruebas de Hipótesis

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$. Cada N_i será una v.a. binomial con parámetros n y p_i con $i = 1, 2 \cdots, k$. El número esperado de ensayos en la categoría i será $E\,[N_i] = n\,p_i\,;\; i = 1, 2, \cdots, k\,.$

Se desea establecer si los datos observados provienen de cierta distribución especial con f.d.a. $F_0(x)$.

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$. Cada N_i será una v.a. binomial con parámetros n y p_i con $i = 1, 2 \cdots, k$. El número esperado de ensayos en la categoría i será $E[N_i] = n\,p_i\,;\; i = 1, 2, \cdots, k\,.$

Se desea establecer si los datos observados provienen de cierta distribución especial con f.d.a. $F_0(x)$. Las hipótesis a probar son:

Pruebas de Hipótesis

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$. Cada N_i será una v.a. binomial con parámetros n y p_i con $i = 1, 2 \cdots, k$. El número esperado de ensayos en la categoría i será $E[N_i] = n\, p_i\,;\; i = 1, 2, \cdots, k$.

Se desea establecer si los datos observados provienen de cierta distribución especial con f.d.a. $F_0(x)$. Las hipótesis a probar son:

$H_0 : F(x) = F_0(x)$ vs $H_a : F_0(x)$ no es la c.d.f. asociada a la muestra .

Pruebas de Hipótesis

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$. Cada N_i será una v.a. binomial con parámetros n y p_i con $i = 1, 2 \cdots, k$. El número esperado de ensayos en la categoría i será $E[N_i] = n\, p_i\,;\; i = 1, 2, \cdots, k$.

Se desea establecer si los datos observados provienen de cierta distribución especial con f.d.a. $F_0(x)$. Las hipótesis a probar son:

$$H_0 : F(x) = F_0(x) \quad \text{vs} \quad H_a : F_0(x) \text{ no es la c.d.f. asociada a la muestra .}$$

Si F_0 está claramente especificada, es posible conocer valores particulares para los p_i y asi obtener $E[N_i]$.

Pruebas de Bondad de Ajuste

Suponga que se tiene un experimento multinomial, es decir una serie de ensayos (n ensayos) idénticos e independientes y k posibles categorías ó clases. Sea p_i la probabilidad de clasificar un sujeto u objeto en la categoría i y sea N_i el número de sujetos u objetos que caen en la categoría i de los n ensayos. Se tiene que $N_1 + \cdots + N_k = n$. Cada N_i será una v.a. binomial con parámetros n y p_i con $i = 1, 2 \cdots, k$. El número esperado de ensayos en la categoría i será $E\,[N_i] = n\,p_i\,;\; i = 1, 2, \cdots, k$.

Se desea establecer si los datos observados provienen de cierta distribución especial con f.d.a. $F_0(x)$. Las hipótesis a probar son:

$H_0 : F(x) = F_0(x)$ vs $H_a : F_0(x)$ no es la c.d.f. asociada a la muestra .

Si F_0 está claramente especificada, es posible conocer valores particulares para los p_i y asi obtener $E\,[N_i]$. En otro caso, los p_i deberán ser estimados y en vez de tener $E\,[N_i]$, se tiene una estimación dada por $n\,\hat{p}_i$.

Pruebas de Bondad de Ajuste

En resumen tenemos:

Pruebas de Bondad de Ajuste

En resumen tenemos:

clase	1	2		k	total
Frec observada	n_1	n_2	$\ldots$	n_k	n
Probabilidad	p_1	p_2	$\ldots$	p_k	1
Frec esperada	$n\,p_1$	$n\,p_2$		$n\,p_k$	n

Pruebas de Bondad de Ajuste

En resumen tenemos:

clase	1	2		k	total
Frec observada	n_1	n_2	$\dots$	n_k	n
Probabilidad	p_1	p_2	$\dots$	p_k	1
Frec esperada	$n\,p_1$	$n\,p_2$		$n\,p_k$	n

Si $n\,p_i \geq 5$; $i = 1, 2 \dots, k$, la variable aleatoria:

Pruebas de Bondad de Ajuste

En resumen tenemos:

clase	1	2		k	total
Frec observada	n_1	n_2	$\ldots$	n_k	n
Probabilidad	p_1	p_2	$\ldots$	p_k	1
Frec esperada	$n\,p_1$	$n\,p_2$		$n\,p_k$	n

Si $n\,p_i \geq 5 \;\; ; \;\; i = 1, 2 \ldots, k$, la variable aleatoria:

$$X = \sum_{i=1}^{k} \frac{(N_i - n\,p_i)^2}{n\,p_i} \quad \sim \chi^2(k-1)\,.$$

Pruebas de Bondad de Ajuste

En resumen tenemos:

clase	1	2		k	total
Frec observada	n_1	n_2	$\ldots$	n_k	n
Probabilidad	p_1	p_2	$\ldots$	p_k	1
Frec esperada	$n\,p_1$	$n\,p_2$		$n\,p_k$	n

Si $n\,p_i \geq 5 \;\; ; \;\; i = 1, 2 \ldots, k$, la variable aleatoria:

$$X = \sum_{i=1}^{k} \frac{(N_i - n\,p_i)^2}{n\,p_i} \quad \sim \chi^2(k-1) \, .$$

Observe que:

Pruebas de Bondad de Ajuste

En resumen tenemos:

clase	1	2		k	total
Frec observada	n_1	n_2	$\dots$	n_k	n
Probabilidad	p_1	p_2	$\dots$	p_k	1
Frec esperada	$n\,p_1$	$n\,p_2$		$n\,p_k$	n

Si $n\,p_i \geq 5$; $i = 1, 2 \dots, k$, la variable aleatoria:

$$X = \sum_{i=1}^{k} \frac{(N_i - n\,p_i)^2}{n\,p_i} \quad \sim \chi^2(k-1) \,.$$

Observe que:

$$X = \sum_{i=1}^{k} \frac{(N_i - n\,p_i)^2}{n\,p_i} = \sum_{\text{todas las celdas}} \frac{(f.obs - f.esp)^2}{f.esp} \sim \chi^2(k-1) \,.$$

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

$$H_0 : p_i = p_{i\,0}\ ;\ i = 1,\, 2,\, \cdots\, ,\, k \quad \text{vs} \quad H_a : \exists_j \text{ tal que } p_j \neq p_{j\,0}\ .$$

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

$$H_0 : p_i = p_{i\,0}\ ;\ i = 1,\, 2,\, \cdots\, ,\, k \quad \text{vs} \quad H_a : \exists_j \text{ tal que } p_j \neq p_{j\,0}\ .$$

En este caso es estadístico de prueba es de la forma:

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

$$H_0 : p_i = p_{i\,0}\ ;\ i = 1,\,2,\,\cdots\,,\,k \quad \text{vs} \quad H_a : \exists_j \text{ tal que } p_j \neq p_{j\,0}\ .$$

En este caso es estadístico de prueba es de la forma:

$$X_C = \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(k-1)\ .$$

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

$$H_0 : p_i = p_{i\,0}\ ;\ i = 1,\, 2,\, \cdots\,,\, k \quad \text{vs} \quad H_a : \exists_j \text{ tal que } p_j \neq p_{j\,0}\ .$$

En este caso es estadístico de prueba es de la forma:

$$X_C = \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(k-1)\ .$$

La región de rechazo es de la forma:

Pruebas de Hipótesis

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

$$H_0 : p_i = p_{i\,0}\ ;\ i = 1,\, 2,\, \cdots ,\, k \quad \text{vs} \quad H_a : \exists_j \text{ tal que } p_j \neq p_{j\,0}\ .$$

En este caso es estadístico de prueba es de la forma:

$$X_C = \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(k-1)\ .$$

La región de rechazo es de la forma:

$$R.C = \{\, X_C \,|\, X_C > \chi^2_\alpha(k-1)\,\} \quad ; \quad \text{para } \alpha \text{ dado}\ .$$

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

$$H_0 : p_i = p_{i\,0} \;;\; i = 1,\, 2,\, \cdots ,\, k \quad \text{vs} \quad H_a : \exists_j \text{ tal que } p_j \neq p_{j\,0} \;.$$

En este caso es estadístico de prueba es de la forma:

$$X_C = \sum_{i=1}^{k} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(k-1) \;.$$

La región de rechazo es de la forma:

$$R.C = \{\, X_C \,|\, X_C > \chi^2_\alpha(k-1) \,\} \quad ; \quad \text{para } \alpha \text{ dado} \;.$$

Finalmente, el Valor P se obtiene como: $V_p = P\left(\, \chi^2_\alpha(k-1) > X_C \,\right)$.

Pruebas de Bondad de Ajuste

Suponga que se desea establecer valores particulares para los p_i; es decir, se quiere probar un juego de hipótesis de la forma:

$$H_0 : p_i = p_{i\,0} \; ; \; i = 1,\, 2,\, \cdots\, ,\, k \quad \text{vs} \quad H_a : \exists_j \text{ tal que } p_j \neq p_{j\,0} \; .$$

En este caso es estadístico de prueba es de la forma:

$$X_C = \sum_{i=1}^{k} \frac{\left(n_i - n\,p_i\right)^2}{n\,p_i} \sim \chi^2(k-1) \; .$$

La región de rechazo es de la forma:

$$R.C = \{\, X_C \,|\, X_C > \chi^2_\alpha(k-1)\,\} \quad ; \quad \text{para } \alpha \text{ dado} \; .$$

Finalmente, el Valor P se obtiene como: $V_p = P\left(\, \chi^2_\alpha(k-1) > X_C \,\right)$.
El cálculo con de probabilidades con la χ^2 es similar al de la Distribución t.

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara.

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Cara	1	2	3	4	5	6
Fre.obs (n_i)	85	94	108	112	98	103
Frec.esp $(n\,p_i)$	100	100	100	100	100	100

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Cara	1	2	3	4	5	6
Fre.obs (n_i)	85	94	108	112	98	103
Frec.esp $(n\,p_i)$	100	100	100	100	100	100

Si el dado No está cargado, significa que cada cara tendrá la misma probabilidad de aparecer al lanzar el dado, es decir, cada cara tendrá una probabilidad igual a $1/6$.

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Cara	1	2	3	4	5	6
Fre.obs (n_i)	85	94	108	112	98	103
Frec.esp ($n\, p_i$)	100	100	100	100	100	100

Si el dado No está cargado, significa que cada cara tendrá la misma probabilidad de aparecer al lanzar el dado, es decir, cada cara tendrá una probabilidad igual a $1/6$. Basta con que una de estas probabilidades no sea $1/6$ para considerar que el dado está cargado.

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Cara	1	2	3	4	5	6
Fre.obs (n_i)	85	94	108	112	98	103
Frec.esp $(n\,p_i)$	100	100	100	100	100	100

Si el dado No está cargado, significa que cada cara tendrá la misma probabilidad de aparecer al lanzar el dado, es decir, cada cara tendrá una probabilidad igual a $1/6$. Basta con que una de estas probabilidades no sea $1/6$ para considerar que el dado está cargado. Así las cosas, las hipótesis a plantear son:

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Cara	1	2	3	4	5	6
Fre.obs (n_i)	85	94	108	112	98	103
Frec.esp $(n\,p_i)$	100	100	100	100	100	100

Si el dado No está cargado, significa que cada cara tendrá la misma probabilidad de aparecer al lanzar el dado, es decir, cada cara tendrá una probabilidad igual a $1/6$. Basta con que una de estas probabilidades no sea $1/6$ para considerar que el dado está cargado. Así las cosas, las hipótesis a plantear son:

$$H_0:\ p_i = \frac{1}{6}\ ,\ i = 1,\, 2,\, \cdots\ ,\, 6 \quad vs \quad H_a:\ \exists_j \ \text{tal que}\ \ p_j \neq \frac{1}{6}\,.$$

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Cara	1	2	3	4	5	6
Fre.obs (n_i)	85	94	108	112	98	103
Frec.esp ($n\,p_i$)	100	100	100	100	100	100

Si el dado No está cargado, significa que cada cara tendrá la misma probabilidad de aparecer al lanzar el dado, es decir, cada cara tendrá una probabilidad igual a $1/6$. Basta con que una de estas probabilidades no sea $1/6$ para considerar que el dado está cargado. Así las cosas, las hipótesis a plantear son:

$$H_0 : \; p_i = \frac{1}{6}\;,\; i = 1,\,2,\,\cdots\,,\,6 \quad vs \quad H_a : \exists_j \;\text{ tal que }\; p_j \neq \frac{1}{6}\,.$$

Si X denota la variable aleatoria: resultado en el lanzamiento del dado, la hipótesis H_0 especifica que la distribución de X es una uniforme discreta; es decir, $p(x) = 1/6 \;\;,\;\; x \;=\; 1,\cdots,6\,.$

Ejemplo 101

Para determinar si un dado cúbico está cargado o no, éste es lanzado 600 veces y se anota cuantas veces aparece cada cara. Los resultados observados fueron:

Cara	1	2	3	4	5	6
Fre.obs (n_i)	85	94	108	112	98	103
Frec.esp ($n\,p_i$)	100	100	100	100	100	100

Si el dado No está cargado, significa que cada cara tendrá la misma probabilidad de aparecer al lanzar el dado, es decir, cada cara tendrá una probabilidad igual a $1/6$. Basta con que una de estas probabilidades no sea $1/6$ para considerar que el dado está cargado. Así las cosas, las hipótesis a plantear son:

$$H_0:\ p_i = \frac{1}{6}\,,\ i = 1,\,2,\,\cdots,\,6 \quad vs \quad H_a:\ \exists_j \text{ tal que } p_j \neq \frac{1}{6}\,.$$

Si X denota la variable aleatoria: resultado en el lanzamiento del dado, la hipótesis H_0 especifica que la distribución de X es una uniforme discreta; es decir, $p(x) = 1/6\ ,\ x = 1, \cdots, 6$. Si H_0 se asume cierta, las frecuencias esperadas serán $n\,p_i = 600\,\left(\frac{1}{6}\right) = 100 \geq 5$.

Pruebas de Hipótesis

El estadístico de Prueba en este caso está dado por:

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \,.$$

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \ .$$

En este caso se tiene que $X_c = 4.82$.

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \; .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por:

Pruebas de Hipótesis

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \, .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por: $R.C = \{X_C \,|\, X_C > \chi^2_{0.05}(5)\}$.

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5)\ .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por: $R.C = \{X_C \,|\, X_C > \chi^2_{0.05}(5)\,\}$. Como $\chi^2_{0.05}(5) = 11.07$, entonces

Pruebas de Hipótesis

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \ .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por: $R.C = \{ X_C \,|\, X_C > \chi^2_{0.05}(5) \}$. Como $\chi^2_{0.05}(5) = 11.07$, entonces $R.C = \{ X_C \,|\, X_C > 11.07 \}$.

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \ .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por: $R.C = \{ X_C \mid X_C > \chi^2_{0.05}(5) \}$. Como $\chi^2_{0.05}(5) = 11.07$, entonces $R.C = \{ X_C \mid X_C > 11.07 \}$.

Como $X_C \notin R.C$, entonces no se rechaza H_0, lo que indica que la información suministrada no es suficiente para concluir que el dado está cargado, en este caso, la evidencia está más a favor de indicar que el dado No está cargado.

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \; .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por: $R.C = \{ X_C \,|\, X_C > \chi^2_{0.05}(5) \}$. Como $\chi^2_{0.05}(5) = 11.07$, entonces $R.C = \{ X_C \,|\, X_C > 11.07 \}$.

Como $X_C \notin R.C$, entonces no se rechaza H_0, lo que indica que la información suministrada no es suficiente para concluir que el dado está cargado, en este caso, la evidencia está más a favor de indicar que el dado No está cargado. El Valor P de esta prueba está dado por:

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5) \ .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por: $R.C = \{ X_C \,|\, X_C > \chi^2_{0.05}(5) \}$. Como $\chi^2_{0.05}(5) = 11.07$, entonces $R.C = \{ X_C \,|\, X_C > 11.07 \}$.

Como $X_C \notin R.C$, entonces no se rechaza H_0, lo que indica que la información suministrada no es suficiente para concluir que el dado está cargado, en este caso, la evidencia está más a favor de indicar que el dado No está cargado. El Valor P de esta prueba está dado por:

$$V_p = P\left(\chi^2(5) > 4.82\right) = 0.43824 \ .$$

El estadístico de Prueba en este caso está dado por:

$$X_C = \sum_{i=1}^{6} \frac{(n_i - 100)^2}{100} \sim \chi^2(5)\ .$$

En este caso se tiene que $X_c = 4.82$. Si se fija $\alpha = 0.05$, la región de rechazo está dada por: $R.C = \{X_C \,|\, X_C > \chi^2_{0.05}(5)\}$. Como $\chi^2_{0.05}(5) = 11.07$, entonces $R.C = \{X_C \,|\, X_C > 11.07\}$.

Como $X_C \notin R.C$, entonces no se rechaza H_0, lo que indica que la información suministrada no es suficiente para concluir que el dado está cargado, en este caso, la evidencia está más a favor de indicar que el dado No está cargado. El Valor P de esta prueba está dado por:

$$V_p = P\left(\chi^2(5) > 4.82\right) = 0.43824\ .$$

Este valor confirma la decisión de no rechazar H_0 .

Ejemplo 102

Un usuario de cierto tipo de calculadora tiene la idea de que el generador de dígitos de la misma, tiene sesgo, es decir, algunos dígitos se generan con mayor frecuencia que otros.

Ejemplo 102

Un usuario de cierto tipo de calculadora tiene la idea de que el generador de dígitos de la misma, tiene sesgo, es decir, algunos dígitos se generan con mayor frecuencia que otros. para verificarlo genera con su calculadora 1000 dígitos aleatorios.

Pruebas de Hipótesis

Ejemplo 102

Un usuario de cierto tipo de calculadora tiene la idea de que el generador de dígitos de la misma, tiene sesgo, es decir, algunos dígitos se generan con mayor frecuencia que otros. para verificarlo genera con su calculadora 1000 dígitos aleatorios. Los resultados obtenidos fueron:

Ejemplo 102

Un usuario de cierto tipo de calculadora tiene la idea de que el generador de dígitos de la misma, tiene sesgo, es decir, algunos dígitos se generan con mayor frecuencia que otros. para verificarlo genera con su calculadora 1000 dígitos aleatorios. Los resultados obtenidos fueron:

Dígito	0	1	2	3	4	5	6	7	8	9
Frec Obs	108	108	95	91	89	106	92	107	110	94
F Espe	100	100	100	100	100	100	100	100	100	100

Ejemplo 102

Un usuario de cierto tipo de calculadora tiene la idea de que el generador de dígitos de la misma, tiene sesgo, es decir, algunos dígitos se generan con mayor frecuencia que otros. para verificarlo genera con su calculadora 1000 dígitos aleatorios. Los resultados obtenidos fueron:

Dígito	0	1	2	3	4	5	6	7	8	9
Frec Obs	108	108	95	91	89	106	92	107	110	94
F Espe	100	100	100	100	100	100	100	100	100	100

¿Con base en esta información, ¿Qué se puede concluir sobre el generador de dígitos aleatorios de la calculadora?

Solución

Si la calculadora genera los dígitos de manera aleatoria, sin sesgo, esto implicaría que la probabilidad de generar cualquiera de los 10 dígitos sería la misma, es decir, $1/10$.

Ejemplo 102

Un usuario de cierto tipo de calculadora tiene la idea de que el generador de dígitos de la misma, tiene sesgo, es decir, algunos dígitos se generan con mayor frecuencia que otros. para verificarlo genera con su calculadora 1000 dígitos aleatorios. Los resultados obtenidos fueron:

Dígito	0	1	2	3	4	5	6	7	8	9
Frec Obs	108	108	95	91	89	106	92	107	110	94
F Espe	100	100	100	100	100	100	100	100	100	100

¿Con base en esta información, ¿Qué se puede concluir sobre el generador de dígitos aleatorios de la calculadora?

Solución

Si la calculadora genera los dígitos de manera aleatoria, sin sesgo, esto implicaría que la probabilidad de generar cualquiera de los 10 dígitos sería la misma, es decir, $1/10$. Así las cosas, basta con que uno de los dígitos, tenga una probabilidad diferente de $1/10$, para considerar que el generador de la calculadora está sesgado.

Las hipótesis a probar en esta caso son:

Pruebas de Hipótesis

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10} \,,\; i = 0, \cdots , 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10} \,.$$

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10}\,,\ i = 0, \cdots, 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10}\,.$$

Bajo H_0, la frecuencia Esperada es $n\,p_i = 1000\left(\frac{1}{10}\right) = 100 \geq 5\,.$

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10} \, , \; i = 0, \cdots , 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10} \, .$$

Bajo H_0, la frecuencia Esperada es $n\,p_i = 1000\left(\frac{1}{10}\right) = 100 \geq 5$. El estadístico de prueba estará dado por:

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10}\ ,\ i = 0, \cdots, 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10}\ .$$

Bajo H_0, la frecuencia Esperada es $n\, p_i = 1000 \left(\frac{1}{10}\right) = 100 \geq 5$. El estadístico de prueba estará dado por:

$$X_c = \sum_{i=0}^{9} \frac{(n_i - 100)^2}{100} \sim \chi^2(9)\ .$$

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10}\ ,\ i = 0, \cdots, 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10}\ .$$

Bajo H_0, la frecuencia Esperada es $n\,p_i = 1000\left(\frac{1}{10}\right) = 100 \geq 5$. El estadístico de prueba estará dado por:

$$X_c = \sum_{i=0}^{9} \frac{(n_i - 100)^2}{100} \sim \chi^2(9)\ .$$

Usando los datos contenidos en la tabla anterior se tiene que $X_C = 6.4$.

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10} \ , \ i = 0, \cdots , 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10} \ .$$

Bajo H_0, la frecuencia Esperada es $n\,p_i = 1000\left(\frac{1}{10}\right) = 100 \geq 5$. El estadístico de prueba estará dado por:

$$X_c = \sum_{i=0}^{9} \frac{(n_i - 100)^2}{100} \sim \chi^2(9) \ .$$

Usando los datos contenidos en la tabla anterior se tiene que $X_C = 6.4$.

El Valor P de esta prueba se obtiene como:

Pruebas de Hipótesis

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10}\ ,\ i = 0, \cdots, 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10}\ .$$

Bajo H_0, la frecuencia Esperada es $n\,p_i = 1000\left(\frac{1}{10}\right) = 100 \geq 5$. El estadístico de prueba estará dado por:

$$X_c = \sum_{i=0}^{9} \frac{(n_i - 100)^2}{100} \sim \chi^2(9)\ .$$

Usando los datos contenidos en la tabla anterior se tiene que $X_C = 6.4$.

El Valor P de esta prueba se obtiene como: $V_p = P(\chi^2(9) > 6.4 = 0.6993$.

Las hipótesis a probar en esta caso son:

$$H_0 : p_i = \frac{1}{10} \ , \ i = 0, \cdots , 9 \quad vs \quad H_a : \exists_j \text{ tal que } p_j \neq \frac{1}{10} \ .$$

Bajo H_0, la frecuencia Esperada es $n\, p_i = 1000 \left(\frac{1}{10}\right) = 100 \geq 5$. El estadístico de prueba estará dado por:

$$X_c = \sum_{i=0}^{9} \frac{(n_i - 100)^2}{100} \sim \chi^2(9) \ .$$

Usando los datos contenidos en la tabla anterior se tiene que $X_C = 6.4$.

El Valor P de esta prueba se obtiene como: $V_p = P(\chi^2(9) > 6.4 = 0.6993$.

Dado que esta probabilidad es muy grande, no se puede rechazar H_0, con lo que se concluye que la evidencia muestral no es suficiente para afirmar que el generador de dígitos de la calculadora esté sesgado, más bien, apoya la hipótesis de que el generador de la calculadora es aleatorio.

Pruebas de Hipótesis

Consideraciones

Es muy común que en este tipo de procedimientos de Bondad de Ajuste, aunque esté claramente especificada cuál es la Hipótesis Nula, es decir, está claramente identificada la Distribución $F_0(x)$, está contenga parámetros desconocidos.

Consideraciones

Es muy común que en este tipo de procedimientos de Bondad de Ajuste, aunque esté claramente especificada cuál es la Hipótesis Nula, es decir, está claramente identificada la Distribución $F_0(x)$, está contenga parámetros desconocidos. Por ejemplo, se desea probar que la distribución $F_0(x)$ corresponde a un modelo $bin(n.\, p)$, pero se desconoce el parámetro p, o la distribución es $p(\lambda)$, pero se deconoce λ, o la distribución es $n(\mu,\, \sigma^2)$, pero se desconocen μ y σ^2.

Pruebas de Hipótesis

Consideraciones

Es muy común que en este tipo de procedimientos de Bondad de Ajuste, aunque esté claramente especificada cuál es la Hipótesis Nula, es decir, está claramente identificada la Distribución $F_0(x)$, está contenga parámetros desconocidos. Por ejemplo, se desea probar que la distribución $F_0(x)$ corresponde a un modelo $bin(n. p)$, pero se desconoce el parámetro p, o la distribución es $p(\lambda)$, pero se deconoce λ, o la distribución es $n(\mu, \sigma^2)$, pero se desconocen μ y σ^2.

Observe que el estadístico de Prueba usado hasta ahora, requiere la especificación de todas la probabilidades p_i, para poderse calcular.

Pruebas de Hipótesis

Consideraciones

Es muy común que en este tipo de procedimientos de Bondad de Ajuste, aunque esté claramente especificada cuál es la Hipótesis Nula, es decir, está claramente identificada la Distribución $F_0(x)$, está contenga parámetros desconocidos. Por ejemplo, se desea probar que la distribución $F_0(x)$ corresponde a un modelo $bin(n.\, p)$, pero se desconoce el parámetro p, o la distribución es $p(\lambda)$, pero se deconoce λ, o la distribución es $n(\mu,\, \sigma^2)$, pero se desconocen μ y σ^2.

Observe que el estadístico de Prueba usado hasta ahora, requiere la especificación de todas la probabilidades p_i, para poderse calcular. Si la distribución $F_0(x)$, contiene parámetros desconocidos, estos deben ser previamente estimados y posteriormente los p_i, se obtendrían a partir de dichas estimaciones.

Consideraciones

Es muy común que en este tipo de procedimientos de Bondad de Ajuste, aunque esté claramente especificada cuál es la Hipótesis Nula, es decir, está claramente identificada la Distribución $F_0(x)$, está contenga parámetros desconocidos. Por ejemplo, se desea probar que la distribución $F_0(x)$ corresponde a un modelo $bin(n. p)$, pero se desconoce el parámetro p, o la distribución es $p(\lambda)$, pero se deconoce λ, o la distribución es $n(\mu, \sigma^2)$, pero se desconocen μ y σ^2.

Observe que el estadístico de Prueba usado hasta ahora, requiere la especificación de todas la probabilidades p_i, para poderse calcular. Si la distribución $F_0(x)$, contiene parámetros desconocidos, estos deben ser previamente estimados y posteriormente los p_i, se obtendrían a partir de dichas estimaciones.

En ese caso el estadístico de Prueba, Bajo H_0, tendrá los mismos grados de libertad pero restando un número igual al de los parámetros estimados.

Consideraciones

Es muy común que en este tipo de procedimientos de Bondad de Ajuste, aunque esté claramente especificada cuál es la Hipótesis Nula, es decir, está claramente identificada la Distribución $F_0(x)$, está contenga parámetros desconocidos. Por ejemplo, se desea probar que la distribución $F_0(x)$ corresponde a un modelo $bin(n.\,p)$, pero se desconoce el parámetro p, o la distribución es $p(\lambda)$, pero se deconoce λ, o la distribución es $n(\mu,\, \sigma^2)$, pero se desconocen μ y σ^2.

Observe que el estadístico de Prueba usado hasta ahora, requiere la especificación de todas la probabilidades p_i, para poderse calcular. Si la distribución $F_0(x)$, contiene parámetros desconocidos, estos deben ser previamente estimados y posteriormente los p_i, se obtendrían a partir de dichas estimaciones.

En ese caso el estadístico de Prueba, Bajo H_0, tendrá los mismos grados de libertad pero restando un número igual al de los parámetros estimados. Veamos un ejemplo para ilustrar esta situación.

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no.

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no. A pesar de desconocer la proporción diaria de defectuosos en la empresa, se tiene la idea de que el número de artículos defectuosos de los 4 seleccionados es Binomial.

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no. A pesar de desconocer la proporción diaria de defectuosos en la empresa, se tiene la idea de que el número de artículos defectuosos de los 4 seleccionados es Binomial. Para verificarlo, se repite este experimento durante 100 días y cada vez se registra el número de defectuosos entre los 4 seleccionados.

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no. A pesar de desconocer la proporción diaria de defectuosos en la empresa, se tiene la idea de que el número de artículos defectuosos de los 4 seleccionados es Binomial. Para verificarlo, se repite este experimento durante 100 días y cada vez se registra el número de defectuosos entre los 4 seleccionados. Los resultados se muestran en la siguiente tabla.

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no. A pesar de desconocer la proporción diaria de defectuosos en la empresa, se tiene la idea de que el número de artículos defectuosos de los 4 seleccionados es Binomial. Para verificarlo, se repite este experimento durante 100 días y cada vez se registra el número de defectuosos entre los 4 seleccionados. Los resultados se muestran en la siguiente tabla.

No Defectuosos	0	1	2	3	4
Frec Obs	11	15	43	22	9

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no. A pesar de desconocer la proporción diaria de defectuosos en la empresa, se tiene la idea de que el número de artículos defectuosos de los 4 seleccionados es Binomial. Para verificarlo, se repite este experimento durante 100 días y cada vez se registra el número de defectuosos entre los 4 seleccionados. Los resultados se muestran en la siguiente tabla.

No Defectuosos	0	1	2	3	4
Frec Obs	11	15	43	22	9

¿Usando los resultados anteriores, determine si en efecto el número de artículos defectuosos sigue una distribución Binomial, con $n = 4$.

Solución

Sea $X_1, \cdots, X_{100}$ una muestra aleatoria que representa el número de defectuosos observados diariamente durante 100 días.

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no. A pesar de desconocer la proporción diaria de defectuosos en la empresa, se tiene la idea de que el número de artículos defectuosos de los 4 seleccionados es Binomial. Para verificarlo, se repite este experimento durante 100 días y cada vez se registra el número de defectuosos entre los 4 seleccionados. Los resultados se muestran en la siguiente tabla.

No Defectuosos	0	1	2	3	4
Frec Obs	11	15	43	22	9

¿Usando los resultados anteriores, determine si en efecto el número de artículos defectuosos sigue una distribución Binomial, con $n = 4$.

Solución

Sea $X_1, \cdots, X_{100}$ una muestra aleatoria que representa el número de defectuosos observados diariamente durante 100 días. Según el enunciado se desea probar las hipótesis:

Ejemplo 103

De la producción diaria de una empresa se seleccionan de manera aleatoria 4 artículos y se examinan uno a uno para verificar si son defectuosos o no. A pesar de desconocer la proporción diaria de defectuosos en la empresa, se tiene la idea de que el número de artículos defectuosos de los 4 seleccionados es Binomial. Para verificarlo, se repite este experimento durante 100 días y cada vez se registra el número de defectuosos entre los 4 seleccionados. Los resultados se muestran en la siguiente tabla.

No Defectuosos	0	1	2	3	4
Frec Obs	11	15	43	22	9

¿Usando los resultados anteriores, determine si en efecto el número de artículos defectuosos sigue una distribución Binomial, con $n = 4$.

Solución

Sea $X_1, \cdots, X_{100}$ una muestra aleatoria que representa el número de defectuosos observados diariamente durante 100 días. Según el enunciado se desea probar las hipótesis:

$$H_0 : X_i \sim b(4,\, p) \quad \text{vs} \quad H_a : X_i \text{ no se distribuye } b(4,\, p)\ .$$

Pruebas de Hipótesis

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud.

Pruebas de Hipótesis

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4, p)$.

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4, p)$. Así, la función de verosimiltud para p está dada por:

Pruebas de Hipótesis

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4, p)$. Así, la función de verosimiltud para p está dada por:

$$L(p) = \prod_{i=1}^{100} \binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} = \prod_{i=1}^{100} \binom{4}{X_i} p^{\sum_{i=1}^{100} X_i} (1-p)^{4*100 - \sum_{i=1}^{100} X_i} .$$

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4, p)$. Así, la función de verosimiltud para p está dada por:

$$L(p) = \prod_{i=1}^{100} \binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} = \prod_{i=1}^{100} \binom{4}{X_i} p^{\sum\limits_{i=1}^{100} X_i} (1-p)^{4*100-\sum\limits_{i=1}^{100} X_i} .$$

$$\ell(p) = \sum_{i=1}^{100} ln\binom{4}{X_i} + \left(\sum_{i=1}^{100} X_i\right) ln(p) + \left(4*100 - \sum_{i=1}^{100} X_i\right) ln(1-p) .$$

Pruebas de Hipótesis

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4, p)$. Así, la función de verosimiltud para p está dada por:

$$L(p) = \prod_{i=1}^{100} \binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} = \prod_{i=1}^{100} \binom{4}{X_i} p^{\sum\limits_{i=1}^{100} X_i} (1-p)^{4*100-\sum\limits_{i=1}^{100} X_i} .$$

$$\ell(p) = \sum_{i=1}^{100} ln\binom{4}{X_i} + \left(\sum_{i=1}^{100} X_i\right) ln(p) + \left(4 * 100 - \sum_{i=1}^{100} X_i\right) ln(1-p) .$$

Al derivar esta última ecuación e igualar a cero se obtiene que $\hat{p} = \frac{1}{4*100} \sum\limits_{i=1}^{100} X_i$.

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4, p)$. Así, la función de verosimiltud para p está dada por:

$$L(p) = \prod_{i=1}^{100} \binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} = \prod_{i=1}^{100} \binom{4}{X_i} p^{\sum\limits_{i=1}^{100} X_i} (1-p)^{4*100-\sum\limits_{i=1}^{100} X_i} .$$

$$\ell(p) = \sum_{i=1}^{100} ln\binom{4}{X_i} + \left(\sum_{i=1}^{100} X_i\right) ln(p) + \left(4*100 - \sum_{i=1}^{100} X_i\right) ln(1-p) .$$

Al derivar esta última ecuación e igualar a cero se obtiene que $\hat{p} = \frac{1}{4*100} \sum\limits_{i=1}^{100} X_i$.

Es fácil verificar que en efecto $\hat{p}$ es un máximo.

Pruebas de Hipótesis

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4,\, p)$. Así, la función de verosimiltud para p está dada por:

$$L(p) = \prod_{i=1}^{100} \binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} = \prod_{i=1}^{100} \binom{4}{X_i} p^{\sum\limits_{i=1}^{100} X_i} (1-p)^{4*100 - \sum\limits_{i=1}^{100} X_i} .$$

$$\ell(p) = \sum_{i=1}^{100} ln\binom{4}{X_i} + \left(\sum_{i=1}^{100} X_i\right) ln(p) + \left(4*100 - \sum_{i=1}^{100} X_i\right) ln(1-p) .$$

Al derivar esta última ecuación e igualar a cero se obtiene que $\hat{p} = \frac{1}{4*100} \sum\limits_{i=1}^{100} X_i$.

Es fácil verificar que en efecto $\hat{p}$ es un máximo. Para los datos recopilados se tiene que $\hat{p} = \frac{203}{400} = 0.5075$.

Pruebas de Hipótesis

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4, p)$. Así, la función de verosimiltud para p está dada por:

$$L(p) = \prod_{i=1}^{100} \binom{4}{X_i} p^{X_i} (1-p)^{4-X_i} = \prod_{i=1}^{100} \binom{4}{X_i} p^{\sum\limits_{i=1}^{100} X_i} (1-p)^{4*100 - \sum\limits_{i=1}^{100} X_i} .$$

$$\ell(p) = \sum_{i=1}^{100} ln\binom{4}{X_i} + \left(\sum_{i=1}^{100} X_i\right) ln(p) + \left(4*100 - \sum_{i=1}^{100} X_i\right) ln(1-p) .$$

Al derivar esta última ecuación e igualar a cero se obtiene que $\hat{p} = \frac{1}{4*100} \sum\limits_{i=1}^{100} X_i$.

Es fácil verificar que en efecto $\hat{p}$ es un máximo. Para los datos recopilados se tiene que $\hat{p} = \frac{203}{400} = 0.5075$. Luego, bajo H_0 cierta, se tiene que $X_i \sim bin(4, 0.5075)$.

Como p es desconocido, bajo H_0 cierta es posible estimar a p usando el método de máxima verosimilitud. Bajo H_0 cierta, se tiene que $X_i \sim b(4,\, p)$. Así, la función de verosimiltud para p está dada por:

$$L(p) = \prod_{i=1}^{100} \binom{4}{X_i} p^{X_i} \, (1-p)^{4-X_i} = \prod_{i=1}^{100} \binom{4}{X_i} \, p^{\sum\limits_{i=1}^{100} X_i} \, (1-p)^{4*100-\sum\limits_{i=1}^{100} X_i} \, .$$

$$\ell(p) \, = \sum_{i=1}^{100} ln\binom{4}{X_i} + \left(\sum_{i=1}^{100} X_i\right) ln(p) + \left(4*100 - \sum_{i=1}^{100} X_i\right) ln(1-p) \, .$$

Al derivar esta última ecuación e igualar a cero se obtiene que $\hat{p} = \frac{1}{4*100} \sum\limits_{i=1}^{100} X_i$. Es fácil verificar que en efecto $\hat{p}$ es un máximo. Para los datos recopilados se tiene que $\hat{p} = \frac{203}{400} = 0.5075$. Luego, bajo H_0 cierta, se tiene que $X_i \sim bin(4,\, 0.5075)$. De esta manera se pueden calcular las respectivas probabilidades, $p_i = P(X = i)$, $i = 0,\, 1,\, 2,\, 3,\, 4$.

$$p_0 = P(X=0) = \binom{4}{0}(0.5075)^0(0.4925)^{4-0} = 0.0589$$

$$p_1 = P(X=1) = \binom{4}{1}(0.5075)^1(0.4925)^{4-1} = 0.2425$$

$$p_2 = P(X=2) = \binom{4}{2}(0.5075)^2(0.4925)^{4-2} = 0.3748$$

$$p_3 = P(X=3) = \binom{4}{3}(0.5075)^3(0.4925)^{4-3} = 0.2575$$

$$p_4 = 1 - p(X=0) - p(X=1) - p(X=2) - p(X=3) = 0.0663\ .$$

Pruebas de Hipótesis

$$p_0 = P(X=0) = \binom{4}{0}(0.5075)^0\,(0.4925)^{4-0} = 0.0589$$

$$p_1 = P(X=1) = \binom{4}{1}(0.5075)^1\,(0.4925)^{4-1} = 0.2425$$

$$p_2 = P(X=2) = \binom{4}{2}(0.5075)^2\,(0.4925)^{4-2} = 0.3748$$

$$p_3 = P(X=3) = \binom{4}{3}(0.5075)^3\,(0.4925)^{4-3} = 0.2575$$

$$p_4 = 1 - p(X=0) - p(X=1) - p(X=2) - p(X=3) = 0.0663\ .$$

La tabla de datos quedaría finalmente así:

$$p_0 = P(X = 0) = \binom{4}{0} (0.5075)^0 \, (0.4925)^{4-0} = 0.0589$$

$$p_1 = P(X = 1) = \binom{4}{1} (0.5075)^1 \, (0.4925)^{4-1} = 0.2425$$

$$p_2 = P(X = 2) = \binom{4}{2} (0.5075)^2 \, (0.4925)^{4-2} = 0.3748$$

$$p_3 = P(X = 3) = \binom{4}{3} (0.5075)^3 \, (0.4925)^{4-3} = 0.2575$$

$$p_4 = 1 - p(X = 0) - p(X = 1) - p(X = 2) - p(X = 3) = 0.0663 \; .$$

La tabla de datos quedaría finalmente así:

No Defectuosos	0	1	2	3	4
Frec Obs (n_i)	11	15	43	22	9
p_i	0.0589	0.2425	0.3748	0.2575	0.0663
Esperada ($n\,p_i$)	5.89	24.25	37.48	25.75	6.63

$$p_0 = P(X = 0) = \binom{4}{0} (0.5075)^0 (0.4925)^{4-0} = 0.0589$$

$$p_1 = P(X = 1) = \binom{4}{1} (0.5075)^1 (0.4925)^{4-1} = 0.2425$$

$$p_2 = P(X = 2) = \binom{4}{2} (0.5075)^2 (0.4925)^{4-2} = 0.3748$$

$$p_3 = P(X = 3) = \binom{4}{3} (0.5075)^3 (0.4925)^{4-3} = 0.2575$$

$$p_4 = 1 - p(X = 0) - p(X = 1) - p(X = 2) - p(X = 3) = 0.0663 \; .$$

La tabla de datos quedaría finalmente así:

No Defectuosos	0	1	2	3	4
Frec Obs (n_i)	11	15	43	22	9
p_i	0.0589	0.2425	0.3748	0.2575	0.0663
Esperada ($n\,p_i$)	5.89	24.25	37.48	25.75	6.63

En este caso la estimación del parámetro p hace que la distribución del Estadístico de prueba pierda un grado de libertad.

$$p_0 = P(X = 0) = \binom{4}{0} (0.5075)^0 \, (0.4925)^{4-0} = 0.0589$$

$$p_1 = P(X = 1) = \binom{4}{1} (0.5075)^1 \, (0.4925)^{4-1} = 0.2425$$

$$p_2 = P(X = 2) = \binom{4}{2} (0.5075)^2 \, (0.4925)^{4-2} = 0.3748$$

$$p_3 = P(X = 3) = \binom{4}{3} (0.5075)^3 \, (0.4925)^{4-3} = 0.2575$$

$$p_4 = 1 - p(X = 0) - p(X = 1) - p(X = 2) - p(X = 3) = 0.0663 \, .$$

La tabla de datos quedaría finalmente así:

No Defectuosos	0	1	2	3	4
Frec Obs (n_i)	11	15	43	22	9
p_i	0.0589	0.2425	0.3748	0.2575	0.0663
Esperada ($n\,p_i$)	5.89	24.25	37.48	25.75	6.63

En este caso la estimación del parámetro p hace que la distribución del Estadístico de prueba pierda un grado de libertad. Observe que $n\,p_i \geq 5$. Así:

$$p_0 = P(X = 0) = \binom{4}{0} (0.5075)^0 \, (0.4925)^{4-0} = 0.0589$$

$$p_1 = P(X = 1) = \binom{4}{1} (0.5075)^1 \, (0.4925)^{4-1} = 0.2425$$

$$p_2 = P(X = 2) = \binom{4}{2} (0.5075)^2 \, (0.4925)^{4-2} = 0.3748$$

$$p_3 = P(X = 3) = \binom{4}{3} (0.5075)^3 \, (0.4925)^{4-3} = 0.2575$$

$$p_4 = 1 - p(X = 0) - p(X = 1) - p(X = 2) - p(X = 3) = 0.0663 \, .$$

La tabla de datos quedaría finalmente así:

No Defectuosos	0	1	2	3	4
Frec Obs (n_i)	11	15	43	22	9
p_i	0.0589	0.2425	0.3748	0.2575	0.0663
Esperada ($n\,p_i$)	5.89	24.25	37.48	25.75	6.63

En este caso la estimación del parámetro p hace que la distribución del Estadístico de prueba pierda un grado de libertad. Observe que $n\,p_i \geq 5$. Así:

$$X_C = \sum_{i=0}^{4} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5 - 1 - 1) \, .$$

Pruebas de Hipótesis

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$.

Pruebas de Hipótesis

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$.

Pruebas de Hipótesis

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$. Como esta probabilidad es pequeña, se rechaza H_0 con mucha seguridad y se concluye que, según la información suministrada, el número de artículos defectuosos en la muestra de 4 tiene una distribución que NO es binomial.

Pruebas de Hipótesis

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$. Como esta probabilidad es pequeña, se rechaza H_0 con mucha seguridad y se concluye que, según la información suministrada, el número de artículos defectuosos en la muestra de 4 tiene una distribución que NO es binomial.

Es importante resaltar que esta prueba se debe usar SOLO en el caso de variables categóricas o discretas.

Pruebas de Hipótesis

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$. Como esta probabilidad es pequeña, se rechaza H_0 con mucha seguridad y se concluye que, según la información suministrada, el número de artículos defectuosos en la muestra de 4 tiene una distribución que NO es binomial.

Es importante resaltar que esta prueba se debe usar SOLO en el caso de variables categóricas o discretas. Existe otro conjunto de pruebas de hipótesis que aplican para variables continuas. Entre las más conocidas están 'Kolmogorov-Smirnov', 'Shapiro-Wilks', 'Cramer Von Misses', 'Jarque Bera', 'Prueba de Asimetría y Kurtosis de Mardia', entre otras.

Pruebas de Hipótesis

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$. Como esta probabilidad es pequeña, se rechaza H_0 con mucha seguridad y se concluye que, según la información suministrada, el número de artículos defectuosos en la muestra de 4 tiene una distribución que NO es binomial.

Es importante resaltar que esta prueba se debe usar SOLO en el caso de variables categóricas o discretas. Existe otro conjunto de pruebas de hipótesis que aplican para variables continuas. Entre las más conocidas están 'Kolmogorov-Smirnov', 'Shapiro-Wilks', 'Cramer Von Misses', 'Jarque Bera', 'Prueba de Asimetría y Kurtosis de Mardia', entre otras. El siguiente ejemplo muestra porqué no es adecuado usar esta prueba para datos continuos.

Ejemplo 104

Un investigador afirma que los resultados obtenidos en una prueba de Matemáticas, a estudiantes de primer semestre, se puede modelar usando una distribución Normal.

Pruebas de Hipótesis

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$. Como esta probabilidad es pequeña, se rechaza H_0 con mucha seguridad y se concluye que, según la información suministrada, el número de artículos defectuosos en la muestra de 4 tiene una distribución que NO es binomial.

Es importante resaltar que esta prueba se debe usar SOLO en el caso de variables categóricas o discretas. Existe otro conjunto de pruebas de hipótesis que aplican para variables continuas. Entre las más conocidas están 'Kolmogorov-Smirnov', 'Shapiro-Wilks', 'Cramer Von Misses', 'Jarque Bera', 'Prueba de Asimetría y Kurtosis de Mardia', entre otras. El siguiente ejemplo muestra porqué no es adecuado usar esta prueba para datos continuos.

Ejemplo 104

Un investigador afirma que los resultados obtenidos en una prueba de Matemáticas, a estudiantes de primer semestre, se puede modelar usando una distribución Normal. Para verificarlo se toma una muestra de 50 estudiantes que presentaron la prueba y se registran los respectivos puntajes obtenidos.

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$. Como esta probabilidad es pequeña, se rechaza H_0 con mucha seguridad y se concluye que, según la información suministrada, el número de artículos defectuosos en la muestra de 4 tiene una distribución que NO es binomial.

Es importante resaltar que esta prueba se debe usar SOLO en el caso de variables categóricas o discretas. Existe otro conjunto de pruebas de hipótesis que aplican para variables continuas. Entre las más conocidas están 'Kolmogorov-Smirnov', 'Shapiro-Wilks', 'Cramer Von Misses', 'Jarque Bera', 'Prueba de Asimetría y Kurtosis de Mardia', entre otras. El siguiente ejemplo muestra porqué no es adecuado usar esta prueba para datos continuos.

Ejemplo 104

Un investigador afirma que los resultados obtenidos en una prueba de Matemáticas, a estudiantes de primer semestre, se puede modelar usando una distribución Normal. Para verificarlo se toma una muestra de 50 estudiantes que presentaron la prueba y se registran los respectivos puntajes obtenidos. Estos fueron:

Usando los datos de la tabla anterior se tiene que $X_C = 10.2$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(3) > 10.2\right) = 0.0169$. Como esta probabilidad es pequeña, se rechaza H_0 con mucha seguridad y se concluye que, según la información suministrada, el número de artículos defectuosos en la muestra de 4 tiene una distribución que NO es binomial.

Es importante resaltar que esta prueba se debe usar SOLO en el caso de variables categóricas o discretas. Existe otro conjunto de pruebas de hipótesis que aplican para variables continuas. Entre las más conocidas están 'Kolmogorov-Smirnov', 'Shapiro-Wilks', 'Cramer Von Misses', 'Jarque Bera', 'Prueba de Asimetría y Kurtosis de Mardia', entre otras. El siguiente ejemplo muestra porqué no es adecuado usar esta prueba para datos continuos.

Ejemplo 104

Un investigador afirma que los resultados obtenidos en una prueba de Matemáticas, a estudiantes de primer semestre, se puede modelar usando una distribución Normal. Para verificarlo se toma una muestra de 50 estudiantes que presentaron la prueba y se registran los respectivos puntajes obtenidos. Estos fueron:

Pruebas de Hipótesis

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

$$H_0 : X \sim n(\mu,\, \sigma^2) \quad vs \quad H_1 : X \text{ No se distribuye } n(\mu,\, \sigma^2)\ .$$

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

$$H_0 : X \sim n(\mu, \sigma^2) \quad vs \quad H_1 : X \text{ No se distribuye } n(\mu, \sigma^2) \ .$$

En este caso μ y σ^2 son desconocidas.

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

$$H_0 : X \sim n(\mu,\, \sigma^2) \quad vs \quad H_1 : X \text{ No se distribuye } n(\mu,\, \sigma^2)\ .$$

En este caso μ y σ^2 son desconocidas. Los respectivos MLE para μ y σ^2 son $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S_n = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$.

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

$$H_0 : X \sim n(\mu,\, \sigma^2) \quad vs \quad H_1 : X \text{ No se distribuye } n(\mu,\, \sigma^2)\ .$$

En este caso μ y σ^2 son desconocidas. Los respectivos MLE para μ y σ^2 son $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S_n = \frac{1}{n} \sum\limits_{i=1}^{n} \left(X_i - \bar{X}\right)^2$. Los valores obtenidos son $\hat{\mu} = 73.82$ y $\hat{\sigma}^2 = 203.3876$. Así, las hipótesis a probar se pueden resumir en:

Pruebas de Hipótesis

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

$$H_0 : X \sim n(\mu, \sigma^2) \quad vs \quad H_1 : X \text{ No se distribuye } n(\mu, \sigma^2)\,.$$

En este caso μ y σ^2 son desconocidas. Los respectivos MLE para μ y σ^2 son $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S_n = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$. Los valores obtenidos son $\hat{\mu} = 73.82$ y $\hat{\sigma}^2 = 203.3876$. Así, las hipótesis a probar se pueden resumir en:

$$H_0 : X \sim n(73.82,\, 203.3876) \quad vs \quad H_1 : X \text{ No se distribuye } n(73.82,\, 203.3876)\,.$$

Pruebas de Hipótesis

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

$$H_0 : X \sim n(\mu, \sigma^2) \quad vs \quad H_1 : X \text{ No se distribuye } n(\mu, \sigma^2) .$$

En este caso μ y σ^2 son desconocidas. Los respectivos MLE para μ y σ^2 son $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S_n = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2$. Los valores obtenidos son $\hat{\mu} = 73.82$ y $\hat{\sigma}^2 = 203.3876$. Así, las hipótesis a probar se pueden resumir en:

$$H_0 : X \sim n(73.82, 203.3876) \quad vs \quad H_1 : X \text{ No se distribuye } n(73.82, 203.3876) .$$

Para usar este estadístico de prueba, se requiere crear categorías de puntajes. Una manera de hacerlo es construyendo una tabla de frecuencias.

Pruebas de Hipótesis

30	35	42	52	58	60	60	62	66	67	67	68	68	68	68	68	68
71	72	72	73	74	74	75	76	76	78	78	78	79	80	80	82	83
86	88	90	90	92	93	94	94	96	97							

¿Qué se puede concluir acerca de la hipótesis del investigador?

Solución

Si X denota la variable aleatoria Puntaje obtenido en la prueba, se desean probar las hipótesis:

$$H_0 : X \sim n(\mu, \sigma^2) \quad vs \quad H_1 : X \text{ No se distribuye } n(\mu, \sigma^2) \, .$$

En este caso μ y σ^2 son desconocidas. Los respectivos MLE para μ y σ^2 son $\hat{\mu} = \bar{X}$ y $\hat{\sigma}^2 = S_n = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$. Los valores obtenidos son $\hat{\mu} = 73.82$ y $\hat{\sigma}^2 = 203.3876$. Así, las hipótesis a probar se pueden resumir en:

$$H_0 : X \sim n(73.82, 203.3876) \quad vs \quad H_1 : X \text{ No se distribuye } n(73.82, 203.3876) \, .$$

Para usar este estadístico de prueba, se requiere crear categorías de puntajes. Una manera de hacerlo es construyendo una tabla de frecuencias. Usando la regla de Sturguess se obtiene la siguiente tabla de frecuencias.

Clase	$X \leq 39$	$39 < X \leq 49$	$49 < X \leq 59$	$60 < X \leq 69$
Frecuencia	2	1	2	12
p_i	p_1	p_2	p_3	p_4

Clase	$70 < X \leq 79$	$80 < X \leq 89$	$X > 89$
Frecuencia	16	9	8
p_i	p_5	p_6	p_7

Pruebas de Hipótesis

Clase	$X \leq 39$	$39 < X \leq 49$	$49 < X \leq 59$	$60 < X \leq 69$
Frecuencia	2	1	2	12
p_i	p_1	p_2	p_3	p_4

Clase	$70 < X \leq 79$	$80 < X \leq 89$	$X > 89$
Frecuencia	16	9	8
p_i	p_5	p_6	p_7

Para Calcular las frecuencias esperadas es necesario obtener las probabilidades en cada categoría.

Pruebas de Hipótesis

Clase	$X \leq 39$	$39 < X \leq 49$	$49 < X \leq 59$	$60 < X \leq 69$
Frecuencia	2	1	2	12
p_i	p_1	p_2	p_3	p_4

Clase	$70 < X \leq 79$	$80 < X \leq 89$	$X > 89$
Frecuencia	16	9	8
p_i	p_5	p_6	p_7

Para Calcular las frecuencias esperadas es necesario obtener las probabilidades en cada categoría. Bajo H_0 cierta se tiene que:

Clase	$X \leq 39$	$39 < X \leq 49$	$49 < X \leq 59$	$60 < X \leq 69$
Frecuencia	2	1	2	12
p_i	p_1	p_2	p_3	p_4

Clase	$70 < X \leq 79$	$80 < X \leq 89$	$X > 89$
Frecuencia	16	9	8
p_i	p_5	p_6	p_7

Para Calcular las frecuencias esperadas es necesario obtener las probabilidades en cada categoría. Bajo H_0 cierta se tiene que:

$$p_1 = P(X \leq 39) = P\left(\frac{X-\mu}{\sigma} \leq \frac{39-73.82}{\sqrt{203.3876}}\right) = P(Z < -2.44) = 0.0073$$

$$p_2 = P(39 < X \leq 49) = P\left(\frac{39-73.82}{\sqrt{203.3876}} < \frac{X-\mu}{\sigma} \leq \frac{49-73.82}{\sqrt{203.3876}}\right)$$

$$= P(-2.44 < Z \leq -1.74) = 0.0336$$

$$p_3 = P(49 < X \leq 59) = P(-1.74 < Z \leq -1.04) = 0.1082$$

$$p_4 = P(59 < X \leq 69) = P(-1.04 < Z \leq -0.34) = 0.2178$$

$$p_5 = P(69 < X \leq 79) = P(-0.34 < Z \leq 0.36) = 0.2736$$

$$p_6 = P(79 < X \leq 89) = P(0.36 < Z \leq 1.06) = 0.2149$$

$$p_7 = 1 - \sum_{i=1}^{6} p_i = 0.1446\ .$$

Pruebas de Hipótesis

Con esta información se obtiene la tabla requerida para evaluar el Estadístico de Prueba.

Con esta información se obtiene la tabla requerida para evaluar el Estadístico de Prueba. Observe que algunas de las categorías no cumplen el supuesto de que $n\,p_i \geq 5$.

Con esta información se obtiene la tabla requerida para evaluar el Estadístico de Prueba. Observe que algunas de las categorías no cumplen el supuesto de que $n\,p_i \geq 5$. Por esa razón se agrupa de manera conveniente y se obtiene una nueva tabla.

Con esta información se obtiene la tabla requerida para evaluar el Estadístico de Prueba. Observe que algunas de las categorías no cumplen el supuesto de que $n\,p_i \geq 5$. Por esa razón se agrupa de manera conveniente y se obtiene una nueva tabla.

Clase	Fre-obs	Probabilidad	Frec-Esperada
$X \leq 39$	2	0.0073	0.365
$39 < X \leq 49$	1	0.0336	1.68
$49 < X \leq 59$	2	0.1082	5.41
$59 < X \leq 69$	12	0.2178	10.89
$69 < X \leq 79$	16	0.2736	13.68
$79 < X \leq 89$	9	0.2149	10.745
$X > 89$	8	0.1446	7.23

Clase	Fre-obs	Probabilidad	Frec-Esperada
$X \leq 59$	5	0.1491	7.455
$59 < X \leq 69$	12	0.2178	10.89
$69 < X \leq 79$	16	0.2736	13.68
$79 < X \leq 89$	9	0.2149	10.745
$X > 89$	8	0.1446	7.23

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

Pruebas de Hipótesis

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Observe que se han restado 2 a los grados de libertad, en compensación a que debimos estimar dos parámetros, para poder obtener las p_i.

Pruebas de Hipótesis

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Observe que se han restado 2 a los grados de libertad, en compensación a que debimos estimar dos parámetros, para poder obtener las p_i. Realizando los respectivos cálculos se obtiene que $X_C = 1.56$.

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Observe que se han restado 2 a los grados de libertad, en compensación a que debimos estimar dos parámetros, para poder obtener las p_i. Realizando los respectivos cálculos se obtiene que $X_C = 1.56$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(2) > 1.56\right) = 0.4584\,.$

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Observe que se han restado 2 a los grados de libertad, en compensación a que debimos estimar dos parámetros, para poder obtener las p_i. Realizando los respectivos cálculos se obtiene que $X_C = 1.56$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(2) > 1.56\right) = 0.4584$. Dado que esta probabilidad es grande, no se debe rechazar H_0, con lo que se concluye que la información suministrada apoya la hipótesis de que los puntajes obtenidos en dicha prueba se distribuyen normales.

Pruebas de Hipótesis

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Observe que se han restado 2 a los grados de libertad, en compensación a que debimos estimar dos parámetros, para poder obtener las p_i. Realizando los respectivos cálculos se obtiene que $X_C = 1.56$. El Valor P de esta prueba se obtiene como $Vp = P\left(\chi^2(2) > 1.56\right) = 0.4584$. Dado que esta probabilidad es grande, no se debe rechazar H_0, con lo que se concluye que la información suministrada apoya la hipótesis de que los puntajes obtenidos en dicha prueba se distribuyen normales.

A pesar de que la prueba aparentemente indica normalidad en los puntajes, esta prueba no es concluyente, sobre todo cuando acepta H_0.

Pruebas de Hipótesis

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Observe que se han restado 2 a los grados de libertad, en compensación a que debimos estimar dos parámetros, para poder obtener las p_i. Realizando los respectivos cálculos se obtiene que $X_C = 1.56$. El Valor P de esta prueba se obtiene como $V\,p = P\left(\chi^2(2) > 1.56\right) = 0.4584$. Dado que esta probabilidad es grande, no se debe rechazar H_0, con lo que se concluye que la información suministrada apoya la hipótesis de que los puntajes obtenidos en dicha prueba se distribuyen normales.

A pesar de que la prueba aparentemente indica normalidad en los puntajes, esta prueba no es concluyente, sobre todo cuando acepta H_0. Esto debido a que en el proceso se cambia la naturaleza de la Variable y se transforma en categorías.

Pruebas de Hipótesis

Usando la información de la última Tabla y dado que $n\,p_i \geq 5$, entonces el estadístico de prueba estará dado por:

$$X_C = \sum_{i=1}^{5} \frac{(n_i - n\,p_i)^2}{n\,p_i} \sim \chi^2(5-1-2)\,.$$

Observe que se han restado 2 a los grados de libertad, en compensación a que debimos estimar dos parámetros, para poder obtener las p_i. Realizando los respectivos cálculos se obtiene que $X_C = 1.56$. El Valor P de esta prueba se obtiene como $V p = P\left(\chi^2(2) > 1.56\right) = 0.4584$. Dado que esta probabilidad es grande, no se debe rechazar H_0, con lo que se concluye que la información suministrada apoya la hipótesis de que los puntajes obtenidos en dicha prueba se distribuyen normales.

A pesar de que la prueba aparentemente indica normalidad en los puntajes, esta prueba no es concluyente, sobre todo cuando acepta H_0. Esto debido a que en el proceso se cambia la naturaleza de la Variable y se transforma en categorías. Esta prueba no es potente para Aceptar H_0 y no se recomienda usar en datos continuos.

Pruebas de Hipótesis

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**.

Pruebas de Hipótesis

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**. Este Test es muy potente sobre todo cuando los tamaños de muestra no son muy grandes.

Pruebas de Hipótesis

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**. Este Test es muy potente sobre todo cuando los tamaños de muestra no son muy grandes. Cuando se tienen muchos datos Test como el de Kolmogorov-Smirnov, Cramer Von Misses, funciona muy bien.

Pruebas de Hipótesis

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**. Este Test es muy potente sobre todo cuando los tamaños de muestra no son muy grandes. Cuando se tienen muchos datos Test como el de Kolmogorov-Smirnov, Cramer Von Misses, funciona muy bien. Usando el Test de Shapiro Wilks, se comprueba que en efecto los Puntajes NO se distribuyen de manera Normal.

Pruebas de Hipótesis

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**. Este Test es muy potente sobre todo cuando los tamaños de muestra no son muy grandes. Cuando se tienen muchos datos Test como el de Kolmogorov-Smirnov, Cramer Von Misses, funciona muy bien. Usando el Test de Shapiro Wilks, se comprueba que en efecto los Puntajes NO se distribuyen de manera Normal. En R se puede realizar esta prueba para los 50 puntajes observados.

Pruebas de Hipótesis

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**. Este Test es muy potente sobre todo cuando los tamaños de muestra no son muy grandes. Cuando se tienen muchos datos Test como el de Kolmogorov-Smirnov, Cramer Von Misses, funciona muy bien. Usando el Test de Shapiro Wilks, se comprueba que en efecto los Puntajes NO se distribuyen de manera Normal. En R se puede realizar esta prueba para los 50 puntajes observados. Se usa la función *shapiro.test*:

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**. Este Test es muy potente sobre todo cuando los tamaños de muestra no son muy grandes. Cuando se tienen muchos datos Test como el de Kolmogorov-Smirnov, Cramer Von Misses, funciona muy bien. Usando el Test de Shapiro Wilks, se comprueba que en efecto los Puntajes NO se distribuyen de manera Normal. En R se puede realizar esta prueba para los 50 puntajes observados. Se usa la función *shapiro.test*:

```
> eje104 <-c(30,35,42,52,58,60,...,90,92,93,94,94,96,97)
> eje104
 [1] 30 35 42 52 58 60 60 62 66 67 67 68 68 68 68 68 68
[18] 70 70 70 71 72 72 73 74 74 75 76 76 78 78 78 79 80
[35] 80 82 83 84 84 85 86 88 90 90 92 93 94 94 96 97
> shapiro.test(eje104)

        Shapiro-Wilk normality test

data:  eje104
W = 0.9366, p-value = 0.009907
```

Como $Vp = 0.0099$, indica que se puede rechazar H_0 con mucha seguridad y concluir, que según los puntajes observados, estos no se distribuyen normalmente.

Una prueba muy utilizada para probar si un conjunto de datos proviene de una distribución normal, es el **Test de Shapiro Wilks**. Este Test es muy potente sobre todo cuando los tamaños de muestra no son muy grandes. Cuando se tienen muchos datos Test como el de Kolmogorov-Smirnov, Cramer Von Misses, funciona muy bien. Usando el Test de Shapiro Wilks, se comprueba que en efecto los Puntajes NO se distribuyen de manera Normal. En R se puede realizar esta prueba para los 50 puntajes observados. Se usa la función *shapiro.test*:

```
> eje104 <-c(30,35,42,52,58,60,...,90,92,93,94,94,96,97)
> eje104
 [1] 30 35 42 52 58 60 60 62 66 67 67 68 68 68 68 68 68
[18] 70 70 70 71 72 72 73 74 74 75 76 76 78 78 78 79 80
[35] 80 82 83 84 84 85 86 88 90 90 92 93 94 94 96 97
> shapiro.test(eje104)

        Shapiro-Wilk normality test

data:  eje104
W = 0.9366, p-value = 0.009907
```

Como $Vp = 0.0099$, indica que se puede rechazar H_0 con mucha seguridad y concluir, que según los puntajes observados, estos no se distribuyen normalmente.

En la siguiente figura se muestra la densidad ajustada a los puntajes.

Pruebas de Hipótesis

En la siguiente figura se muestra la densidad ajustada a los puntajes. Observe que su forma dista mucho de ser simétrica y por lo tanto un Modlo Normal, no es adecuado para explicar el comportamiento de los puntajes obtenidos en el exámen..

Pruebas de Hipótesis

En la siguiente figura se muestra la densidad ajustada a los puntajes. Observe que su forma dista mucho de ser simétrica y por lo tanto un Modlo Normal, no es adecuado para explicar el comportamiento de los puntajes obtenidos en el exámen.. El código en R usado es:

Pruebas de Hipótesis

En la siguiente figura se muestra la densidad ajustada a los puntajes. Observe que su forma dista mucho de ser simétrica y por lo tanto un Modlo Normal, no es adecuado para explicar el comportamiento de los puntajes obtenidos en el exámen.. El código en R usado es:

plot(density(puntajes), xlab="Puntaje", main="Densidad para Puntajes", ylab=).

Pruebas de Hipótesis

En la siguiente figura se muestra la densidad ajustada a los puntajes. Observe que su forma dista mucho de ser simétrica y por lo tanto un Modlo Normal, no es adecuado para explicar el comportamiento de los puntajes obtenidos en el exámen.. El código en R usado es:

plot(density(puntajes), xlab="Puntaje", main="Densidad para Puntajes", ylab=).