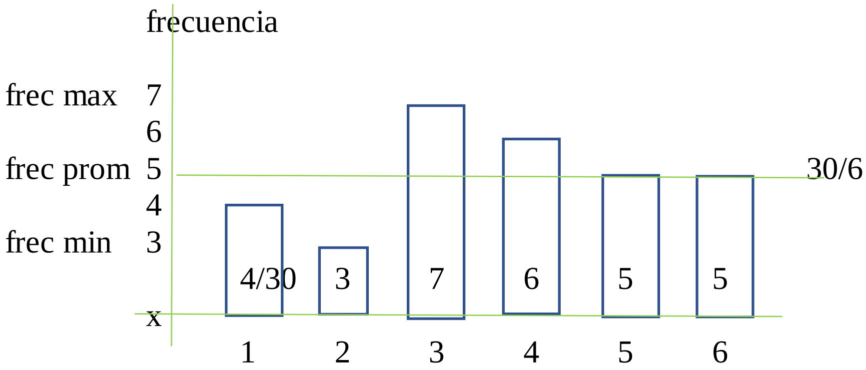
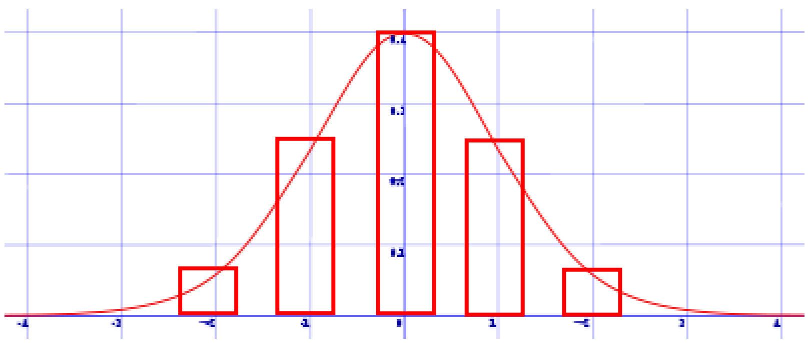
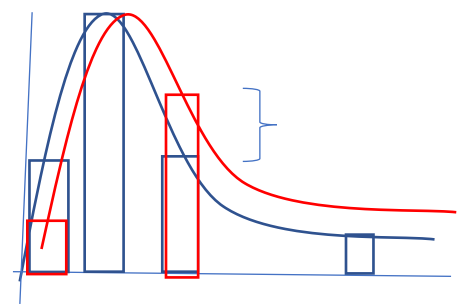
De una muestra con 30 valores obtenidos, se elabora la distribución de frecuencias relativas, el histograma y se calcula la media aritmética.



distribución Normal u otro comportamiento.





Prueba de la **Ji cuadrada** (χ^2)

Para probar que ciertos valores (reales) siguen alguna distribución de probabilidad f(x) (teórica) se utilizan las pruebas de bondad de ajuste: **Ji cuadrada,** si la muestra en tiene más de 100 datos, en caso contrario se usa la prueba de **Kolmogorov-Smirnov.**

Para probar estadísticamente que un conjunto de datos empiricos o de una muestra, no **difieren significativamente** de aquellos valores que se esperan de una distribución teórica específica, podemos considerar las pruebas de bondad de ajuste.

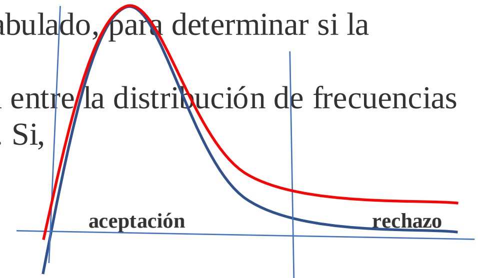
Una medida de la discrepancia que existe entre la frecuencia esperada y la observada, la proporciona el estimador χ^2 (ji-cuadrada).

El estimador lo proporciona: $\chi^2 = \sum_{i=1}^{i=q} (fo_i - fe_i)^2 / fe_i$; donde:

$\sum$ = Suma de todas las clases desde i=1 hasta q
fo_i = frecuencia observada en la clase i
fe_i = frecuencia esperada en la clase i

Si **χ^2 = 0,** la frecuencia teórica y observada concuerdan exactamente.
χ^2 > **0, mientras mayor sea, mayor es la discrepancia.**

Si $\sum\chi^2 > \mathbf{0}$, debemos de comparar el valor observado de χ^2_o con el valor tabulado, para determinar si la variación es aleatoria.



En la práctica, la hipótesis nula (Ho) es: no existe diferencia significativa entre la distribución de frecuencias observadas y la distribución teórica específica con los mismos parámetros. Si,

$\chi^2_{\text{observada}} < \chi^2_{\mathbf{teo}(1-\alpha),\,\nu}$

La prueba χ^2 considera las siguientes condiciones:

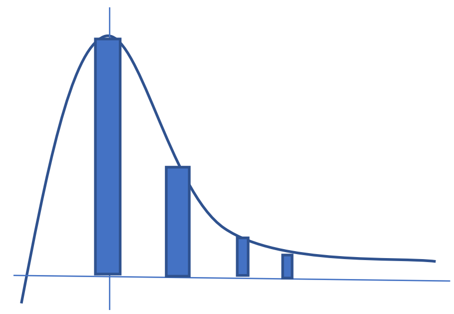
1. Se deben usar conteos o números reales y que $n > 100$
2. Las frecuencias observadas para cada intervalo de clase deben de equivaler a cinco o más (en caso contrario se agrupan con las clases siguientes o anteriores).
3. Los grados de libertad (ν) se dan por la relación: $\nu = q-1$, donde q es el número de clases.

Ejemplo. Para el caso de **consultas telefónicas** recibidas en un consultorio cada hora (datos discretos), se pueden calcular sus parámetros de la siguiente manera:

Media aritmética ($\bar{X}$)= $\sum_{i=1}^{i=q} x_i f_i / n$; $S^2 = ((\sum_{i=1}^{i=q} x_i^2 f_i) - n\bar{X}^2) / (n-1)$

La $\sum$ es desde el primer evento (o clase) x_i, hasta el evento n.

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
0	315	0	0
1	142	142	142
2	40	80	160
3	~~9~~ 12	27 40	81
4	2	8	32
5	1	5	25
	509	262	440



$\bar{X} = 262 / 509 = 0.5147$ $S^2 = (440 - (509)(0.5147)^2) / (509-1) = 0.6007$

Deseamos probar que la distribución de las consultas telefónicas recibidas en un consultorio, provienen de una distribución **Poisson**, con un nivel de confianza de 0.95 o sea que $\mathbf{(1-\alpha)}$.=0.95, y $\mathbf{(\alpha)= 0.05}$

La distribución Poisson es:

$$f(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!},$$

Considerando que: $f(k;\lambda) = f(x;\lambda)$ es la probabilidad de que ocurra x. $\boldsymbol{e}$ es la constante con valor 2.71828… y λ es la media para esta distribución poisson, que es igual a la media aritmética más la varianza dividido entre 2.

Se había calculado: media aritmética, Como ambas están cercanas, calculamos y obtenemos: ($\bar{X}$) = 0.5147 y Varianza (S^2) = 0.6007

$\lambda = (0.5147 + 0.6007) / 2 = 0.5577$

Establecemos la Ho: No existe diferencia significativa entre los datos observados y los que proporciona una distribución poisson con $\lambda = 0.5577$.
A continuación se elabora la siguiente tabla para calcular $\chi^2_{\text{observada.}}$

La columna de la probabilidad del evento n se calcula usando la distribución poisson:
$P(0) = [(0.5577)^0 (2.71828)^{-(0.5577)}] / 0! = 0.5725$

$P(1) = [(0.5577)^1 (2.71828)^{-(0.5577)}] / 1! = 0.3192$

$P(2) = [(0.5577)^2 (2.71828)^{-(0.5577)}] / 2! = 0.089$, etc.

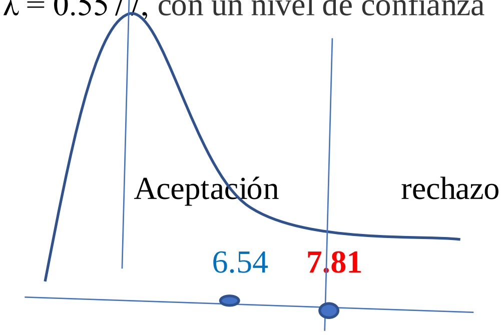
Como la frecuencia esperada de las clases 5 y 4 no llegan a cinco, se unen a la clase 3 y la tabla queda:

xi	$f_{i\ (0bservada)}$	Prob poisson	$f_{i(esperada)}$	$\chi^2 = (fo_i - fe_i)^2 / fe_i$
0	315	0.5725	.5725*509=291	1.5154
1	142	0.3192	.3192*509=162	2.4691
2	40	0.089	45	0.5555
3 o mas	12	0.0165	8	2
	509		509	6.54

$\chi^2_{\text{observada.}} = 6.54$

En las tabla de la χ^2, determinamos el valor de tabulado de $\chi^2_{\mathbf{teo(\alpha), v}}$, entonces $\chi^2_{\mathbf{teo(.05), 3}} = \mathbf{7.8147}$

Como $\chi^2_{\text{observada}} < \chi^2_{\mathbf{teo(\alpha), v}}$ **, entonces no se rechaza Ho,** es decir no existe diferencia significativa entre los datos observados y los que proporcionaría una distribución Poisson con $\lambda = 0.5577$, con un nivel de confianza de 0.95.



Ejemplo: para la función **exponencial**: $f(x)= \lambda e^{-\lambda x}$ $\lambda = \mu = .5147$

Establecemos la Ho: No existe diferencia significativa entre los datos observados de las llamadas telefónicas y los que proporciona una distribución exponencial con $\lambda = 0.5147$.

A continuación se elabora la siguiente tabla para calcular $\chi^2_{\text{observada}}$.

La columna de la probabilidad teorica del evento n se calcula usando la distribución exponencial:

$P(0) = [(0.5147)\ (2.71828)^{-(0.5147*0)}\ ] = 0.5147$

$P(1) = [(0.5147)\ (2.71828)^{-(0.5147*1)}\ ] = 0.3076$

$P(2) = [(0.5147)\ (2.71828)^{-(0.5147*2)}\ ] = 0.1838$, etc.

$P(4) = [(0.5147)\ (2.71828)^{-(0.5147*4)}\ ] = 0.$

$P(4) = [(0.5147)\ (2.71828)^{-(2.0588)}\ ] = 0.$

$P(4) = [(0.5147)\ (0.1276)\ ] = 0.0656$

$P(5) = [(0.5147)\ (2.71828)^{-(0.5147*5)}\ ] = 0.$

xi	$f_{i\ (0\text{bservada})}$	Prob exponencial	$f_{i(\text{esperada})}$ REDONDEO	$\chi^2 = (fo_i - fe_i)^2 / fe_i$
0	315	0.5147	.5147*509=262	9.5419
1	142	0.3076	.3076*509=157	1.4331
2	40	0.1838	94	31.0212
3	~~9~~ 12	0.1098	56 109	86.3211
4	2	0.0656	33 53	
5	1	0.0392	20	
	509		509	

χ^2 esp(.05, 3)= 7.8147 < **χ^2 obs=128.3173**

la distribución de probabilidad de las consultas telefónicas recibidas en un consultorio, **no provienen** de una distribución **exponencial**, con un nivel de confianza de 0.95

ACEPT RECHAZO

7.81 128.3173

Distribución Normal.

Se llama **distribución normal**, **distribución de Gauss** o **distribución gaussiana**, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

La gráfica de su Función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro, la media. Esta curva se conoce como campana de Gauss.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la ingente cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- caracteres morfológicos de individuos como la estatura;
- caracteres fisiológicos como el efecto de un fármaco;
- caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos;
- caracteres psicológicos como el cociente intelectual;
- nivel de ruido medido en un ambiente.
- Errores cometidos al medir ciertas magnitudes, etc.
- Se dice que una variable aleatoria continua X sigue una distribución normal de parámetros μ y σ y se denota $X \sim N(\mu, \sigma)$ si su función de densidad está dada por:

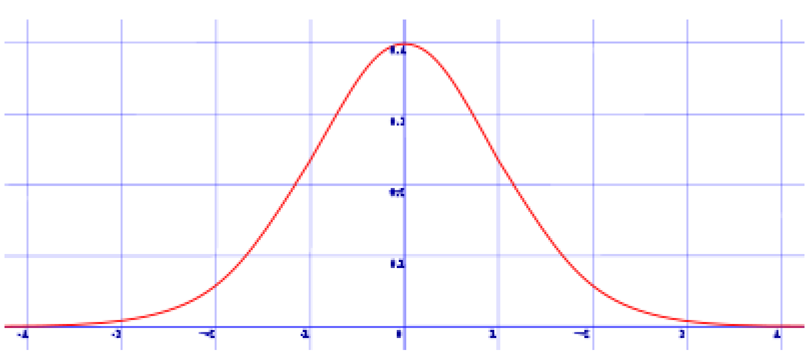
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

donde μ (mu) es la media y σ (sigma) es la desviación típica (σ^2 es la varianza).

Se llama **distribución normal "estándar"** a aquella en la que sus parámetros toman los valores $\mu = 0$ y $\sigma = 1$. En este caso la función de densidad tiene la siguiente expresión:

$$f(x) = f_{0,1}(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}, \quad x \in \mathbb{R},$$

Su gráfica se muestra abajo y con frecuencia se usan tablas para el cálculo de los valores de su distribución.



$E(x) = \mu$ y $Var(x) = \sigma^2$

Con la variable aleatoria x distribuida normalmente, se obtiene Z que también se distribuye Normalmente entre [0,1]:

$$Z = (x - \mu) / \sigma ,$$

Ejemplo de *variables DISCRETAS*. La siguiente tabla es la distribución de frecuencias de la **producción** semanal del producto x, recabadas durante 120 semanas. Determinar la frecuencia relativa y elaborar la distribución de frecuencias relativas para los datos observados.

x (producción semanal)	f(x) = frecuencia (Cantidad de semanas con esta producción)	frecuencia relativa x / total de observaciones
Menos de 45	1	1 / 120 = 0.008
46 - 55	1	1 / 120 = 0. 008
56 - 65	3	3 / 120 = 0.025
66 - 75	7	7 / 120 = 0.058
76 - 85	11	11 / 120 = 0.092
86 - 95	21	21 / 120 = 0.175
96 - 105	28	28 / 120 = 0.234
106 - 115	16	16 / 120 = 0.134
116 - 125	22	22 / 120 = 0. 183
126 - 135	7	7 / 120 = 0.058
136 - 145	1	1 / 120 = 0. 008
146 o más	2	2 / 120 = 0. 017
	120	1.000

Después de obtener el gráfico de la distribución de frecuencias relativas observadas, se compara visualmente con las gráficas de las distribuciones de frecuencias teóricas y se seleccionan las posibles distribuciones de probabilidad a partir de la cual se derivó.

Después de identificar una o más distribuciones teóricas, se procede a determinar los parámetros para luego realizar las pruebas estadísticas necesarias.
Usualmente los parámetros se pueden determinar a partir de la media aritmética y de la varianza (s^2).

Para el caso de la **producción** semanal del producto x (datos discretos), los parámetros son:

Media aritmética ($\bar{X}$)= $\sum_{i=1}^{i=q} x_i f_i / n$; $s^2 = ((\sum_{i=1}^{i=q} x_i^2 f_i) - n \bar{X}^2) / (n-1)$

La $\sum$ es desde el primer evento (o clase) x_i, hasta el evento n.
Donde M_i es la marca de clase del intervalo i.

$M_i = x_i$	f_i	$M_i f_i$	$M_i^2 f_i$
40.5	1	40.5	1 640.25
50.5	1	50.5	2 550.25
60.5	3	181.5	10 980.75

70.5	7	493.5	34 791.75
80.5	11	885.5	71 282.75
90.5	21	1 900.5	171 995.25
100.5	28	2 814.5	282 807.00
110.5	16	1 678.0	195 364.00
120.5	22	2 651.0	319 445.50
130.5	7	913.5	119 211.75
140.5	1	140.5	19 740.25
150.5	2	301.0	45 300.5
	120	12 140.0	1 275 110.00

$\bar{X}$ = 12 140 / 120 = 101.1666 ; S^2 = (1275110 – (120)(101.1666)^2) / (120 – 1) = 394.5234; S=19.8626

A continuación se elabora la siguiente tabla para calcular $\chi^2_{observada.}$

La columna de la probabilidad teórica del evento o clase i, se calcula usando la distribución normal:
Z_{56} = (55-101.1666) / (19.8626) = -2.2739
Z_{66} = (65-101.1666) / (19.8626) = -1.7704
Z_{76} = (75-101.1666) / (19.8626) = -1.3173
Z_{86} = (85-101.1666) / (19.8626) = -.8139
Z_{96} = (95-101.1666) / (19.8626) = -.3104, etc.

						REDONDEO	
$M_i = x_i$	$f_{i\,(0bs)}$	Límites de clase	normal: z	P(Z)	P(Xi)	$f_{i(esperada)}$	χ^2 = (fo_i - fe_i)^2 / fe_i
60.5	5	56	-2.2739	0.0116	0.0268	4	0.25
70.5	7	66	-1.7704	0.0384	0.054	7	0.125
80.5	11	76	-1.267	0.1038	0.1198	16	1.0666
90.5	21	86	-0.7635	0.2236	0.1738	23	0.1739
100.5	18	96	-0.2601	0.3974	0.1974	27	3.0
110.5	16	106	0.2433	0.4052	0.1756	24	2.6666
120.5	22	116	0.7468	0.2296	0.124	17	1.4705
130.5	10	126	1.2502	0.1056	0.061	8	0.0
		135	1.7033	0.0446			

χ^2 esp(.05, 7)= 14.07 > **χ^2 obs=9.8359**

la distribución de probabilidad de la producción de la empresa, provienen de una distribución **normal**, con un nivel de confianza de 0.95

De la siguiente muestra, determine cual es su comportamiento con una confiabilidad del 95%.

Clase i	**marca**	**Frec obs**	**Z lim inf**	**Prob Z**	**Prob teo normal (i)**	**Frec esp (i)**	**X^2**

10-30	**20**	**10**	**-2.4145**	**0.0080**	**0.0655**	**7.6635**	**.7124**
30-50	**40**	**22**	**-1.4587**	**0.0735**	**0.235**	**27.495**	**1.0982**
50-70	**60**	**50**	**-0.5025**	**0.3085**	**0.3651**	**42.7167**	**1.2418**
70-90	**80**	**25**	**0.4535**	**0.3264**	**0.2456**	**28.7352**	**.4855**
90-110	**100**	**10**	**1.4095**	**0.0808**	**0.0717**	**8.3889**	**.3094**

Z_{110}= 2.3655 0.0091

n=117 x=7080/117=60.5128 s^2= [275 200 -248 453.5486]/(n-1)=437.66 s= 20.9205

χ^2 esp(.05, 4)= 9.49 > **χ^2 obs=3.8473**

conclusión; la muestra se comporta normal.