Prueba de independencia para tablas de contingencia:

$$p_{ij} = P(X = x_i, Y = y_j) = \frac{n_{i\blacksquare} n_{\blacksquare j}}{N^2}$$

$$E_{ij} = Np_{ij} = \frac{n_{i\blacksquare} n_{\blacksquare j}}{N}$$

$$X^2 = \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$v = grados\ de\ libertad = (m-1)x(k-1)$$

$$Punto\ crítico = X^2_{v,\alpha}$$

Para calcular los valores E_{ij} primero calculamos los totales por renglón y por columna:

	Republicano	Demócrata	Independiente	
Masculino	204	147	52	$n_{1\bullet} = 403$
Femenino	248	300	49	$n_{2\bullet} = 597$
	$n_{\bullet 1} = 452$	$n_{\bullet 2} = 447$	$n_{\bullet 3} = 101$	$n_{\bullet\bullet} = 1000$

De esta manera, los valores de E_{ij} son (se agregan entre paréntesis al lado de los O_{ij}):

	Republicano	Demócrata	Independiente	
Masculino	204 (182.156)	147 (180.141)	52 (40.703)	$n_{1\bullet} = 403$
Femenino	248 (269.844)	300 (266.859)	49 (60.297)	$n_{2\bullet} = 597$
	$n_{\bullet 1} = 452$	$n_{\bullet 2} = 447$	$n_{\bullet 3} = 101$	$n_{\bullet\bullet} = 1000$

H_0: género y partido de preferencia son indepe

H_A: género y partido de preferencia no son inde

Prueba de homogeneidad:

Se usa el mismo estadístico de prueba:

$$X^2 = \sum_{i=1}^{m} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

H_0: la distribución de la preferencia de votos es la misma para los géneros masculino y femenino.
H_A: la distribución de la preferencia de votos no es la misma para los géneros masculino y femenino.

En este tipo de pruebas, se puede hacer pruebas de varias cosas, para tener una mejor comprensión lea el archivo de prueba de homogeneidad.

Coeficiente de correlación lineal:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$S_{xx} = \sum {x_i}^2 - \frac{(\sum x_i)^2}{n}$$

$$S_{yy} = \sum {y_i}^2 - \frac{(\sum y_i)^2}{n}$$

$$S_{xx} = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

El coeficiente de correlación muestral tiene las siguientes interpretaciones:

Valor de r	Interpretación
$0 < r < 1$, r cercano a 1	relación lineal positiva y fuerte
$0 < r < 1$, r cercano a 0	relación lineal positiva y débil
$r \approx 0$	no existe relacion lineal
$-1 < r < 0$, r cercano a 0	relación lineal negativa y débil
$-1 < r < 0$, r cercano a -1	relación lineal negativa y fuerte

Prueba de hipótesis para la correlación:

$$H_0: \quad \rho = 0$$
$$H_A: \quad \rho \neq 0$$

$$T = \frac{r - p_0}{S_r} \rightarrow \frac{r}{S_r}$$

$$S_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

$$grados\ de\ libertad = n - 2$$

$$punto\ crítico = t_{\frac{\alpha}{2}, n-2}$$

Recta de mínimos cuadrados:

$$Y = \beta_0 + \beta_1 X$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$R^2 = r^2$$

Inferencia sobre el modelo de regresión lineal simple:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$SST = SSR - SSE$$

$$F_c = \frac{MSR}{MSE}$$

Tabla Anova:

Fuente	g.l.	Sumas de cuadrados	Cuadrados medios	F
Regresión	1	SSR	MSR	F_c
Error	$n-2$	SSE	MSE	
Total	$n-1$	SST		

$$punto\ crítico = F_{\alpha;1;n-2}$$

$$SSR = \hat{\beta}_1 S_{xy}$$

$$SST = S_{yy}$$

$$SSE = SST - SSR$$

$$\begin{aligned} H_0 &: \quad E[Y|X] = \beta_0 \\ H_A &: \quad E[Y|X] = \beta_0 + \beta_1 X \end{aligned}$$

Fuente	g.l.	Sumas de cuadrados	Cuadrados medios	F
Regresión	1	345,01	345,01	48,2964
Error	6	42,8614	7,1436	
Total	7	387,875		

Inferencia sobre los coeficientes de regresión:

$$\hat{\beta}_i \pm t_{\frac{\alpha}{2};n-2} S_{\hat{\beta}_i}$$

$$i = 0, 1$$

- $S_{\hat{\beta}_0} = S\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{nS_{xx}}}$
- $S_{\hat{\beta}_1} = \frac{S}{\sqrt{S_{xx}}}$

$$S = \sqrt{MSE}$$

$$T = \frac{\hat{\beta}_i - (\beta_i)_0}{S_{\hat{\beta}_i}}$$

$$T = \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}}$$

Predicciones basadas en la recta de regresión:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

$$\hat{y}_0 \pm t_{\frac{\alpha}{2};n-2} S_{\hat{Y}_0 - Y_0}$$

$$S_{\hat{Y}_0 - Y_0} = S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

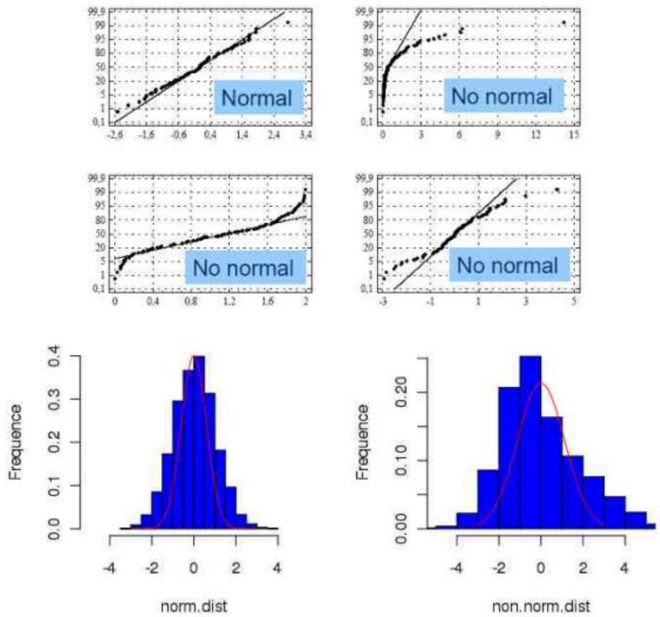
$$\hat{y}_0 \pm t_{\frac{\alpha}{2};n-2} S_{\hat{Y}_0}$$

$$S_{\hat{Y}_0} = S\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$$

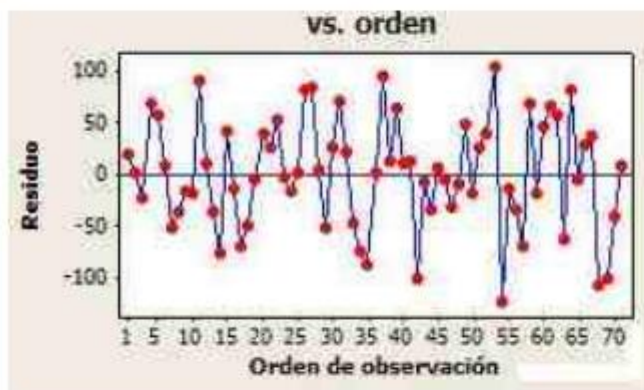
Supuesto de normalidad:

$$H_0: \text{ Los residuos se distribuyen normalmente}$$

$$H_A: \text{ Los residuos no siguen una distribución normal}$$



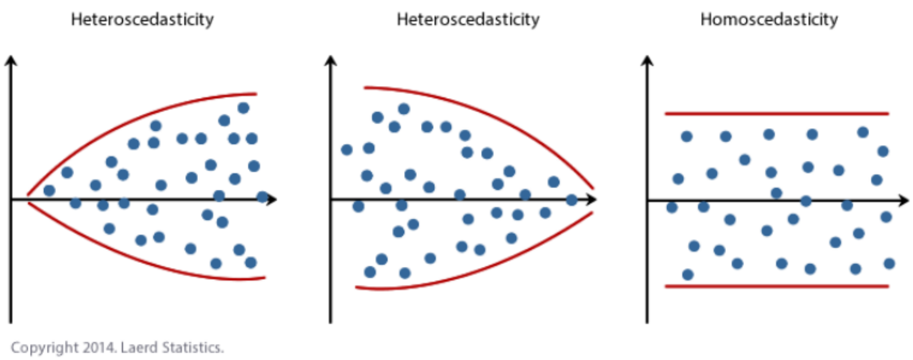
Supuesto de independencia:



$$H_0: \text{ La autocorrelación de los residuales es } 0$$

$$H_A: \text{ La autocorrelación de los residuales es diferente } 0$$

Supuesto de homocedasticidad:



Predicciones inversas:

$$\hat{X}_0 \pm t_{\alpha/2,n-2} S_{predX},$$

- La estimación puntual es $\hat{X}_0 = \frac{Y_0 - \hat{\beta}_0}{\hat{\beta}_1}$
- $S^2_{predX} = \frac{MSE}{\hat{\beta}_1^2}\left[1 + \frac{1}{n} + \frac{(\hat{X}_0 - \bar{x})^2}{S_{xx}}\right]$