Formulario Métodos de Regresión. Grado en Estadística y Empresa.

Regresión lineal simple	Regresión lineal múltiple
Modelo: $y = \beta_0 + \beta_1 x + u$, donde $u \sim N(0,\sigma^2)$	Modelo: $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$, donde $u \sim N(0,\sigma^2)$ En notación matricial: $Y = X \beta + U$, donde $U \sim NM_n(0, \sigma^2 I_n)$, $X = [1\ \ X_1\ \ \dots\ \ X_k]$ es una matriz $nx(k+1)$ de regresores, $\beta = (\beta_0, \beta_1, \dots, \beta_k)'$ es el vector de parámetros, Y es la variable respuesta.
Estimación: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, donde $\hat{\beta}_1 = \frac{s_{xy}}{s_x^2}$, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ $e_i = y_i - \hat{y}_i$, $s_R^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$	Estimación: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$, $s_R^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k-1}$
Propiedades de los estimadores: $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{n\, s_x^2}\right)$, $\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2}{n}\left(1 + \frac{\bar{x}^2}{s_x^2}\right)\right)$, $\frac{(n-2)\, s_R^2}{\sigma^2} \sim \chi^2_{n-2}$	Propiedades de los estimadores: $\hat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}\right)$, $\frac{\mathbf{e}'\mathbf{e}}{\sigma^2} \sim \chi^2_{n-k-1}$
Contraste de regresión: $H_0: \beta_1 = 0$ $F = \frac{VE}{VNE/(n-2)} \sim F(1, n-2)$ $VE = n\, \hat{\beta}_1^2\, s_x^2$, $VNE = (n-2)\, s_R^2$, $VT = n\, s_y^2$	Contraste de regresión: $H_0: \beta_1 = \dots = \beta_k = 0$ $F = \frac{VE/k}{VNE/(n-k-1)} \sim F(k, n-k-1)$ $VE = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{y}^2$, $VNE = \mathbf{e}'\mathbf{e} = \mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$, $VT = \mathbf{Y}'\mathbf{Y} - n\bar{y}^2$
Correlación: $R^2 = \frac{VE}{VT}$	Correlación: $R^2 = \frac{VE}{VT}$, $\tilde{R}^2 = 1 - \frac{VNE/(n-k-1)}{VT/(n-1)}$
Predicción: $\hat{y}_h \sim N\left(m_h, \frac{\sigma^2}{\hat{n}_h}\right)$, donde $\hat{n}_h = \frac{1}{v_h} = \frac{n}{1 + (x_h - \bar{x})^2 / s_x^2}$ $e_h = y_h - \hat{y}_h \sim N\left(0, \sigma^2\left(1 + \frac{1}{\hat{n}_h}\right)\right)$, donde $\hat{n}_h = \frac{1}{v_h} = \frac{n}{1 + (x_h - \bar{x})^2 / s_x^2}$	Predicción: $\hat{y}_h \sim N\left(m_h, \frac{\sigma^2}{\hat{n}_h}\right)$, donde $\hat{n}_h = \frac{1}{v_{hh}} = \frac{1}{\mathbf{x_h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_h}}$ $e_h = y_h - \hat{y}_h \sim N\left(0, \sigma^2\left(1 + \frac{1}{\hat{n}_h}\right)\right)$, donde $\hat{n}_h = \frac{1}{v_{hh}} = \frac{1}{\mathbf{x_h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_h}}$
Diagnosis del modelo: $e_i = y_i - \hat{y}_i$ Puntos palanca: $v_i > 4/n$, Puntos influyentes: $D(i) \geq F(2, n-2)$, donde $D(i) = \frac{r_i^2}{2}\left(\frac{v_i}{1 - v_i}\right)$ es la distancia de Cook, con $r_i = \frac{e_i}{s_R\sqrt{1 - v_i}}$	Diagnosis del modelo: $\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}}$ Puntos palanca: $v_{ii} > 2(k+1)/n$, Puntos influyentes: $D(i) \geq F(k+1, n-k-1)$, donde $D(i) = \frac{r_i^2}{k+1}\left(\frac{v_{ii}}{1 - v_{ii}}\right)$ es la distancia de Cook, con $r_i = \frac{e_i}{s_R\sqrt{1 - v_{ii}}}$