Regression analysis in matrix form - summary

- The model:

  $$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

  Where:

  $$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix},\ \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix},\ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix},\ \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

  In multiple regression the matrix $\mathbf{X}$ and the vector $\boldsymbol{\beta}$ are:

  $$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & x_{32} & \cdots & x_{k2} \\ 1 & x_{13} & x_{23} & x_{33} & \cdots & x_{k3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} & \cdots & x_{kn} \end{pmatrix},\ \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}.$$

  The dimensions of the above vectors and matrices are:

  $$\mathbf{Y} : n \times 1,\ \mathbf{X} : n \times (k+1),\ \boldsymbol{\beta} : (k+1) \times 1,\ \boldsymbol{\epsilon} : n \times 1.$$

  Where k is the number of independent variables in the model. In simple regression $k = 1$ and the dimensions of $\mathbf{X}$ is $n \times 2$, and β is 2×1.

- Least squares estimates:

  $$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

  The matrix $(\mathbf{X}'\mathbf{X})^{-1}$ has dimensions $(k+1) \times (k+1)$.

- Variance-covariance matrix of $\hat{\beta}$:

  $$\text{cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

- Fitted values:

  $$\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}.$$

  Where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the so called hat matrix with dimensions $n \times n$.
  Note: $\mathbf{H} = \mathbf{H}'$, and $\mathbf{H}\mathbf{H} = \mathbf{H}$ (symmetric and idempodent).

- Variance-covariance matrix of the fitted values:

  $$\text{cov}(\hat{\mathbf{Y}}) = \sigma^2\mathbf{H}.$$

  Therefore the variance of the i_{th} fitted value is: $\text{var}(\hat{y}_i) = \sigma^2 h_{ii}$, where h_{ii} is the i_{th} diagonal element of the hat matrix.

- Residuals:

  $$\mathbf{e} = (\mathbf{I} - \mathbf{H})\mathbf{Y}.$$

- Variance-covariance matrix of the residuals:

  $$\text{cov}(\mathbf{e}) = \sigma^2(\mathbf{I} - \mathbf{H}).$$

  Therefore the variance of the i_{th} residual is: $\text{var}(e_i) = \sigma^2(1 - h_{ii})$, where $1 - h_{ii}$ is the i_{th} diagonal element of the matrix $\mathbf{I} - \mathbf{H}$.

- The i_{th} diagonal element of the hat matrix:

  $$h_{ii} = \mathbf{x}_\mathbf{i}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_\mathbf{i},$$

  Where, $\mathbf{x}_\mathbf{i}'$ is the i_{th} row of the $\mathbf{X}$ matrix.

- Estimation of σ^2:
  An unbiased estimator of σ^2 is s_e^2 given b:

  $$s_e^2 = \frac{SSE}{n-k-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-k-1} = \frac{\mathbf{Y'Y} - \hat{\boldsymbol{\beta}}\mathbf{'X'Y}}{n-k-1}.$$

- Inferences on $\boldsymbol{\beta}$:
  Suppose we want to test the hypothesis:
  $H_0 : \beta_1 = 0$
  $H_a : \beta_1 \neq 0$

  Test statistic:

  $$t = \frac{\hat{\beta}_1 - \beta_1}{s_e\sqrt{\mathbf{a'}(\mathbf{X'X})^{-1}\mathbf{a}}}.$$

  Where $\mathbf{a}$ is a $(k+1) \times 1$ vector (length is equal to the number of parameters to be estimated $\beta_0, \beta_1, \cdots, \beta_k$).

  $\mathbf{a} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. This vector will extract from the matrix $(\mathbf{X'X})^{-1}$ the element needed for var$(\hat{\beta}_1)$.

  Reject H_0 if $t > t_{\frac{\alpha}{2};n-k-1}$ or $t < -t_{\frac{\alpha}{2};n-k-1}$.

- Similarly:
  $H_0 : \beta_1 - \beta_2 = 0$
  $H_a : \beta_1 - \beta_2 \neq 0$

  Now the vector $\boldsymbol{a}$ is:

  $\mathbf{a} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. This vector will extract from the matrix $(\mathbf{X'X})^{-1}$ the elements needed for var$(\hat{\beta}_1 - \hat{\beta}_2)$.

  Test statistic:

  $$t = \frac{\hat{\beta}_1 - \hat{\beta}_2 - (\beta_1 - \beta_2)}{s_e\sqrt{\mathbf{a'}(\mathbf{X'X})^{-1}\mathbf{a}}}.$$

  Reject H_0 if $t > t_{\frac{\alpha}{2};n-k-1}$ or $t < -t_{\frac{\alpha}{2};n-k-1}$.

- Testing the overall significance of the model:
  $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$
  H_a : At least one $\beta_i \neq 0$
  Test statistic:

  $$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE}.$$

  This follows $F_{k,n-k-1}$. Therefore reject H_0 if $F > F_{\alpha;k,n-k-1}$.

- Confidence interval for the expectation of a predicted value:
  Let $\mathbf{x'_g} = (1, x_{1g}, x_{2g}, \cdots, x_{kg})$. Then the predicted $\hat{y}_g$ at $\mathbf{x'_g}$ is:

  $$\hat{y}_g = \mathbf{x'_g}\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{1g} + \hat{\beta}_2 x_{2g} + \cdots + \hat{\beta}_k x_{kg}.$$

  with

  mean $E(\hat{y}_g) = \mathbf{x'_g}\boldsymbol{\beta}$ and

  variance var$(\hat{y}_g) = \sigma^2\mathbf{x'_g}(\mathbf{X'X})^{-1}\mathbf{x_g}$.

  We replace σ^2 with s_e^2 to have the following confidence interval for $E(y_g)$:

  $$\hat{y}_g \pm t_{\frac{\alpha}{2};n-k-1}s_e\sqrt{\mathbf{x'_g}(\mathbf{X'X})^{-1}\mathbf{x_g}}.$$

- Coefficient of determination R^2:

  $$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}.$$