

MMS075 Exercise class 1 formula sheet

András Bálint, andras.balint@chalmers.se

January 22, 2020

Formula for simple linear regression:

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

Observed data in form of (predictor, response) pairs:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

i -th predicted response and residual:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad e_i = y_i - \hat{y}_i$$

Residual squared at observation i :

$$e_i^2 = (y_i - \hat{y}_i)^2$$

Residual sum of squares:

$$\text{RSS} = e_1^2 + e_2^2 + \dots + e_n^2$$

Least squares coefficient estimates:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

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

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n}$$

Estimating standard error of the random error term by residual standard error:

$$\text{RSE} = \sqrt{\frac{1}{n-2} \text{RSS}} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Estimating the standard error of coefficients:

$$\text{SE}(\hat{\beta}_0) = \text{RSE} \times \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

$$\text{SE}(\hat{\beta}_1) = \text{RSE} \times \sqrt{\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Confidence intervals for the coefficients:

$$\hat{\beta}_0 \pm 2 \cdot \text{SE}(\hat{\beta}_0)$$

$$\hat{\beta}_1 \pm 2 \cdot \text{SE}(\hat{\beta}_1)$$

Note: for the precise confidence intervals, 2 should be replaced by the 97.5% quantile of the t_{n-2} distribution.

Proportion of variability in the response that is explained by the predictor:

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

Total sum of squares:

$$\text{TSS} = \sum_{i=1}^n (y_i - \bar{y})^2$$

Correlation of X and Y :

$$r = \text{Cor}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Correlation and R^2 :

$$R^2 = r^2$$