Repetition week 9

Diagnostic checks of regression models.

Multicollinearity

$$r_{ij} = \frac{\sum_{k=1}^{n}\left(x_{ik} - \bar{x}_i\right)\left(x_{jk} - \bar{x}_j\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{ik} - \bar{x}_i\right)^2}\sqrt{\sum_{k=1}^{n}\left(x_{jk} - \bar{x}_j\right)^2}}$$

Variance Inflation factor

$$VIF_j = \frac{1}{1 - R_j^2},\ VIF_j < 10\ \forall j$$

Influential Observations

$$h_{ii} > \frac{2(k+1)}{n}$$

Cook's distance

$$D_i = \frac{\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}\right)^T \boldsymbol{X}^T \boldsymbol{X}\left(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)}\right)}{(k+1)\hat{\sigma}^2},\ i = 1,2,\mathrm{K}\ ,n$$

Study of residuals

$$e_i = y_i - \hat{y}_i$$

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}}$$ (standardized or studentized)

$$\partial_i = Y_i - \hat{Y}_{i,-i} = \frac{Y_i - \hat{Y}_i}{1 - h_{ii}}$$ (press residuals)

$$\frac{Y_i - \hat{Y}_i}{S_{(-i)}\sqrt{1 - h_{ii}}} \sim t_{n-k-2}$$ (r-studentized or external studentized)

Important

Outliers, heterogeneity, misspecification, normal distributions, correlation over time.

Transformations

$Z = g(Y) \approx g(\mu) + g'(\mu)(Y - \mu)$ $Z = g(Y) \approx g(\mu) + g'(\mu)(Y - \mu)$

$$SD(Y) \propto \sqrt{\mu}, \quad g(Y) = \sqrt{Y} = Y^{0.5}$$

$$SD(Y) \propto \mu, \quad g(Y) = \ln(Y) = Y^{0}$$

$$SD(Y) \propto \mu^2, \quad g(Y) = \frac{1}{Y} = Y^{-1}$$

Box-Cox transformation: Y^{λ}

Confidence intervals/ellipsoids

Single coefficients

$$100(1-\alpha)\%: b_j \pm t_{\alpha/2,n-k-1}\hat{\sigma}\sqrt{c_{jj}}$$

Vector of coefficients

$$100(1-\alpha)\%: (\boldsymbol{b}-\boldsymbol{\beta})^T \boldsymbol{X}^T \boldsymbol{X} (\boldsymbol{b}-\boldsymbol{\beta}) \leq (k+1)\hat{\sigma}^2 f_{\alpha,k+1,n-k-1}$$

Expected response

$$\hat{y}_0 \pm t_{\alpha/2,n-k-1}\hat{\sigma}\sqrt{\boldsymbol{x}_0 (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_0}$$

Prediction interval

$$\hat{y}_0 \pm t_{\alpha/2,n-k-1}\hat{\sigma}\sqrt{1 + \boldsymbol{x}_0 (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{x}_0}$$