The sum of squares, SS_T, SS_E and SS_R are normally gathered in an analysis of variance (ANOVA) table.

Source	SS	d.f.	F
Regression	$SS_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$	k	$\frac{SSR}{k} / \frac{SSE}{n-k-1}$
Error	$SS_E = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	$n-k-1$	
Total	$SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$	$n-1$	

Let $\varepsilon_i \sim N(0, \sigma^2)$ and independent, $i = 1, 2, \dots, n$

The F statistics is used to test the hypothesis

$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ $\quad$ H_1: at least one β_i is different from 0

i.e. of the regression is significant

Reject H_0 if $F = \dfrac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} \geq f_{\alpha, k, n-k-1}$

eventually if. $P(F \geq F_{obs}) \leq \alpha$.

12.5 Inference in multippel linear regression

Test for the significant impact on the response given that the other variables are in the model

$H_0: \beta_j = 0$ $\quad$ $H_1: \beta_j \neq 0$

More general $H_0: \beta_j = \beta_{j0}$ $\quad$ $H_1: \beta_j \ \overset{>}{\underset{<}{\neq}} \ \beta_{j0}$

Prediction interval for a new responsevalue in x_0

Let Y_0 be the new value. We get;

$$E[Y_0 - \hat{Y}_0] = E[\underline{x}_0'\underline{\beta} + \underline{\varepsilon}] - E[\underline{x}_0'\hat{\underline{\beta}}] = \underline{x}_0'\underline{\beta} - \underline{x}_0'\underline{\beta} = 0$$

$$Var[Y_0 - \hat{Y}_0] = Var[Y_0] + Var[\hat{Y}_0] = \sigma^2 + \sigma^2 \underline{x}_0'(\underline{X}'\underline{X})^{-1}\underline{x}_0$$

such that $T = \dfrac{\hat{Y}_0 - Y_0}{S\sqrt{1 + \underline{x}_0'(\underline{X}'\underline{X})^{-1}\underline{x}_0}}$ is t-distributed with

$n-k-1$ degrees of freedom. A $100(1-\alpha)\%$ prediction interval is given by $\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-k-1}\, s\sqrt{1 + \underline{x}_0'(X'X)^{-1}\underline{x}_0}$

12.6 Choice of ~~the~~ fitted ~~value~~ model

~~Val av tilpassa modell~~

Let Y be the response and let $x_1, x_2, \ldots, x_k$ be the regressor variables. Explanatory variables with little or no impact on the response may cause the model to have bad prediction ability.

Criteria for evaluating ~~how well~~ the model fit of the model.

We have: $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i) + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$

$$SS_T = SS_E + SS_R$$

A good model should have small values for $y_i - \hat{y}_i$ and $\hat{y}_i - \bar{y} \approx y_i - \bar{y}$.

The coefficient of determination, R^2, is a measure on how much of the variation in the data that is explained by the model.

Under H_0 $\dfrac{\hat{\beta}_j - \beta_{j0}}{\hat{\sigma}\sqrt{c_{jj}}}$ is t-distributed with $n-k-1$ degrees of freedom

c_{jj}: $(j+1)$th diagonalelement in $(\underline{X}'\underline{X})^{-1}$

Let $t_{obs} = \dfrac{b_j - \beta_{j0}}{s\sqrt{c_{jj}}}$. Reject H_0 and conclude on $H_1: \beta_j \neq \beta_{j0}$

if $|t_{obs}| \geq t_{\frac{\alpha}{2}, n-k-1}$ eventually $2P(T \geq |t_{obs}|) \leq \alpha$

Confidence interval for expected response given values

$x_{10}, x_{20}, \dots, x_{k0}$ for $x_1, x_2, \dots, x_k$

Let $E[Y \mid x_{10}, \dots, x_{k0}] = \mu_{Y \mid x_{10}, \dots, x_{k0}} = \beta_0 + \beta_1 x_{10} + \dots + \beta_k x_{k0}$

Let $\underline{x}_0' = [1, x_{10}, x_{20}, \dots, x_{k0}]$, $\hat{Y}_0 = \underline{x}_0'\hat{\underline{\beta}}$

$$Var[\hat{Y}_0] = E[\underline{x}_0'(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})'\underline{x}_0] = \underline{x}_0' E[(\hat{\underline{\beta}} - \underline{\beta})(\hat{\underline{\beta}} - \underline{\beta})'] \underline{x}_0$$

$$= \sigma^2 \underline{x}_0'(\underline{X}'\underline{X})^{-1}\underline{x}_0$$

We have: $T = \dfrac{\hat{Y}_0 - \mu_{Y \mid x_{10}, \dots, x_{k0}}}{S\sqrt{\underline{x}_0'(\underline{X}'\underline{X})^{-1}\underline{x}_0}}$ is t-distributed

with $n-k-1$ degrees of freedom. $S = \sqrt{\dfrac{SSE}{n-k-1}}$

which give us that a $100(1-\alpha)\%$ confidence interval for expected response i $\underline{x}_0$ is

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-k-1}\, s\sqrt{\underline{x}_0'(\underline{X}'\underline{X})^{-1}\underline{x}_0}$$

where $\hat{y}_0 = b_0 + b_1 x_{10} + b_2 x_{20} + \dots + b_k x_{k0}$

$$R^2 \triangleq \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T}, \quad 0 \le R^2 \le 1$$

$R^2 = 0.84$ tells us that 84 % of the variation in the data is explained by the model

R^2 can always be increased by adding regression variables.

By adjusting for degrees of freedom we get.

R^2-adjusted, written as R^2_{adj}

$$R^2_{adj} = 1 - \frac{\frac{SS_E}{n-k-1}}{\frac{SS_T}{n-1}} = 1 - \frac{\hat{\sigma}^2}{\frac{SS_T}{n-1}}$$

Maximizing R^2_{adj} is equivalent to minimizing $\hat{\sigma}^2$.

It cannot always be increased by adding regression variables and can be used to decide upon how ~~meg~~ many regression-variables that should be included.