5. Linear Regression

Outline. .......................................................... 2

Simple linear regression ........................................ 3
  Linear model .................................................... 4
  Linear model .................................................... 5
  Linear model .................................................... 6
  Small residuals .................................................. 7
  Minimize $\sum \hat{e}_i^2$ ....................................... 8
  Properties of residuals ........................................ 9
  Regression in R ............................................... 10

How good is the fit? ............................................ 11
  Residual standard error ...................................... 12
  $R^2$ ............................................................ 13
  $\tilde{R}^2$ ..................................................... 14
  Analysis of variance ......................................... 15
  $r$ ............................................................... 16

Multiple linear regression ..................................... 17
  $\geq 2$ independent variables. .............................. 18
  Statistical error ............................................. 19
  Estimates and residuals ..................................... 20
  Computing estimates ........................................ 21
  Properties of residuals ..................................... 22
  $R^2$ and $\tilde{R}^2$ ........................................... 23

Ozone example .................................................. 24
  Ozone example ................................................ 25
  Ozone data ................................................... 26
  R output ....................................................... 27

Standardized coefficients .................................... 28
  Standardized coefficients .................................. 29
  Using hinge spread ......................................... 30
  Interpretation ............................................... 31
  Using st.dev. ............................................... 32
  Interpretation ............................................... 33
Outline

- We have seen that linear regression has its limitations. However, it is worth studying linear regression because:
  - Sometimes data (nearly) satisfy the assumptions.
  - Sometimes the assumptions can be (nearly) satisfied by transforming the data.
  - There are many useful extensions of linear regression: weighted regression, robust regression, nonparametric regression, and generalized linear models.
- How does linear regression work? We start with one independent variable.

Simple linear regression

Linear model

- Linear statistical model: \( Y = \alpha + \beta X + \epsilon \).
- \( \alpha \) is the intercept of the line, and \( \beta \) is the slope of the line. One unit increase in \( X \) gives \( \beta \) units increase in \( Y \). (see figure on blackboard)
- \( \epsilon \) is called a statistical error. It accounts for the fact that the statistical model does not give an exact fit to the data.
- Statistical errors can have a fixed and a random component.
  - Fixed component: arises when the true relation is not linear (also called lack of fit error, bias) - we assume this component is negligible.
  - Random component: due to measurement errors in \( Y \), variables that are not included in the model, random variation.
Linear model

- Data \((X_1, Y_1), \ldots, (X_n, Y_n)\).
- Then the model gives: \(Y_i = \alpha + \beta X_i + \epsilon_i\), where \(\epsilon_i\) is the statistical error for the \(i\)th case.
- Thus, the observed value \(Y_i\) almost equals \(\alpha + \beta X_i\), except that \(\epsilon_i\), an unknown random quantity is added on.
- The statistical errors \(\epsilon_i\) cannot be observed. Why?
- We assume:
  - \(E(\epsilon_i) = 0\) for all \(i = 1, \ldots, n\)
  - \(\text{Var}(\epsilon_i) = \sigma^2\) for all \(i = 1, \ldots, n\)
  - \(\text{Cov}(\epsilon_i, \epsilon_j) = 0\) for all \(i \neq j\)

The population parameters \(\alpha\), \(\beta\), and \(\sigma\) are unknown. We use lower case Greek letters for population parameters.

- We compute estimates of the population parameters: \(\hat{\alpha}\), \(\hat{\beta}\), and \(\hat{\sigma}\).
- \(\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i\) is called the fitted value. (see figure on blackboard)
- \(\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i)\) is called the residual.
- The residuals are observable, and can be used to check assumptions on the statistical errors \(\epsilon_i\).
- Points above the line have positive residuals, and points below the line have negative residuals.
- A line that fits the data well has small residuals.
Small residuals

- We want the residuals to be small in magnitude, because large negative residuals are as bad as large positive residuals.
- So we cannot simply require \(\sum \hat{e}_i = 0\).
- In fact, any line through the means of the variables - the point \((\bar{X}, \bar{Y})\) - satisfies \(\sum \hat{e}_i = 0\) (derivation on board).
- Two immediate solutions:
  - Require \(\sum |\hat{e}_i|\) to be small.
  - Require \(\sum \hat{e}_i^2\) to be small.
- We consider the second option because working with squares is mathematically easier than working with absolute values (for example, it is easier to take derivatives). However, the first option is more resistant to outliers.
- Eyeball regression line (see overhead).

Minimize \(\sum \hat{e}_i^2\)

- SSE stands for Sum of Squared Error.
- We want to find the pair \((\hat{\alpha}, \hat{\beta})\) that minimizes \(SSE(\alpha, \beta) := \sum(Y_i - \alpha - \beta X_i)^2\).
- Thus, we set the partial derivatives of \(RSS(\alpha, \beta)\) with respect to \(\alpha\) and \(\beta\) equal to zero:
  - \[\frac{\partial SSE(\alpha, \beta)}{\partial \alpha} = \sum(-1)(2)(Y_i - \alpha - \beta X_i) = 0\]
  - \[\Rightarrow \sum(Y_i - \alpha - \beta X_i) = 0\]
  - \[\frac{\partial SSE(\alpha, \beta)}{\partial \beta} = \sum(-X_i)(2)(Y_i - \alpha - \beta X_i) = 0\]
  - \[\Rightarrow \sum X_i(Y_i - \alpha - \beta X_i) = 0\]
- We now have two normal equations in two unknowns \(\alpha\) and \(\beta\). The solution is (derivation on board, Section 1.3.1 of script):
  - \[\hat{\beta} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}\]
  - \[\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}\]
Properties of residuals

- \[ \sum \hat{e}_i = 0 \], since the regression line goes through the point \((\bar{X}, \bar{Y})\).
- \[ \sum X_i \hat{e}_i = 0 \] and \[ \sum Y_i \hat{e}_i = 0 \]. \( \Rightarrow \) The residuals are uncorrelated with the independent variables \(X_i\) and with the fitted values \(\hat{Y}_i\).
- Least squares estimates are uniquely defined as long as the values of the independent variable are not all identical. In that case the numerator \(\sum (X_i - \bar{X})^2 = 0\) (see figure on board).

Regression in R

- `model <- lm(y ~ x)`
- `summary(model)`
- Coefficients: `model$coef` or `coef(model)`
  (Alias: coefficients)
- Fitted mean values: `model$fitted` or `fitted(model)`
  (Alias: fitted.values)
- Residuals: `model$resid` or `resid(model)`
  (Alias: residuals)
- See R-code Davis data
How good is the fit?

Residual standard error

- Residual standard error: $\hat{\sigma} = \sqrt{\frac{SSE}{n - 2}} = \sqrt{\frac{\sum \hat{\epsilon}_i^2}{n-2}}$.
- $n - 2$ is the degrees of freedom (we lose two degrees of freedom because we estimate the two parameters $\alpha$ and $\beta$).
- For the Davis data, $\hat{\sigma} \approx 2$. Interpretation:
  - on average, using the least squares regression line to predict weight from reported weight, results in an error of about 2 kg.
  - If the residuals are approximately normal, then about $2/3$ is in the range $\pm 2$ and about $95\%$ is in the range $\pm 4$.

$R^2$

- We compare our fit to a null model $Y = \alpha' + \epsilon'$, in which we don’t use the independent variable $X$.
- We define the fitted value $\hat{Y}_i' = \hat{\alpha}'$, and the residual $\hat{\epsilon}_i' = Y_i - \hat{Y}_i'$.
- We find $\hat{\alpha}'$ by minimizing $\sum (\hat{\epsilon}_i')^2 = \sum (Y_i - \hat{\alpha}')^2$. This gives $\hat{\alpha}' = \bar{Y}$.
- Note that $\sum (Y_i - \hat{Y}_i)^2 = \sum \hat{\epsilon}_i^2 \leq \sum (\hat{\epsilon}_i')^2 = \sum (Y_i - \bar{Y})^2$ (why?).
$R^2$

- $TSS = \sum (\hat{\epsilon}_i)^2 = \sum (Y_i - \bar{Y})^2$ is the total sum of squares: the sum of squared errors in the model that does not use the independent variable.
- $SSE = \sum \epsilon_i^2 = \sum (Y_i - \hat{Y}_i)^2$ is the sum of squared errors in the linear model.
- Regression sum of squares: $RegSS = TSS - SSE$ gives reduction in squared error due to the linear regression.
- $R^2 = RegSS/TSS = 1 - SSE/TSS$ is the proportional reduction in squared error due to the linear regression.
- Thus, $R^2$ is the proportion of the variation in $Y$ that is explained by the linear regression.
- $R^2$ has no units ⇒ doesn’t change when scale is changed.
- ‘Good’ values of $R^2$ vary widely in different fields of application.

Analysis of variance

- $\sum (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y}) = 0$ (will be shown later geometrically)
- $RegSS = \sum (\hat{Y}_i - \bar{Y})^2$ (derivation on board)
- Hence,

$$
\begin{align*}
TSS &= SSE + RegSS \\
\sum (Y_i - \bar{Y})^2 &= \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \bar{Y})^2
\end{align*}
$$

This decomposition is called analysis of variance.
Correlation coefficient \( r = \pm \sqrt{R^2} \) (take positive root if \( \hat{\beta} > 0 \) and take negative root if \( \hat{\beta} < 0 \)).

- \( r \) gives the strength and direction of the relationship.

- Alternative formula: \( r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}} \).

- Using this formula, we can write \( \hat{\beta} = r \frac{SD_Y}{SD_X} \) (derivation on board).

- In the 'eyeball regression', the steep line had slope \( \frac{SD_Y}{SD_X} \), and the other line had the correct slope \( r \frac{SD_Y}{SD_X} \).

- \( r \) is symmetric in \( X \) and \( Y \).

- \( r \) has no units \( \Rightarrow \) doesn’t change when scale is changed.

Multiple linear regression

\( \geq 2 \) independent variables

- \( Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \epsilon \) (see Section 1.1.2 of script)

- This describes a plane in 3-dimensional space \( \{X_1, X_2, Y\} \) (see figure):
  - \( \alpha \) is the intercept
  - \( \beta_1 \) is the increase in \( Y \) associated with a one-unit increase in \( X_1 \) when \( X_2 \) is held constant
  - \( \beta_2 \) is the increase in \( Y \) for a one-unit increase in \( X_2 \) when \( X_1 \) is held constant.
Statistical error

- Data: \((X_{11}, X_{12}, Y_1), \ldots, (X_{n1}, X_{n2}, Y_n)\).
- \(Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i\), where \(\epsilon_i\) is the statistical error for the \(i\)th case.
- Thus, the observed value \(Y_i\) equals \(\alpha + \beta_1 X_{i1} + \beta_2 X_{i2}\), except that \(\epsilon_i\), an unknown random quantity, is added on.
- We make the same assumptions about \(\epsilon\) as before:
  - \(E(\epsilon_i) = 0\) for all \(i = 1, \ldots, n\)
  - \(\text{Var}(\epsilon_i) = \sigma^2\) for all \(i = 1, \ldots, n\)
  - \(\text{Cov}(\epsilon_i, \epsilon_j) = 0\) for all \(i \neq j\)
- Compare to assumptions in section 1.2 of script.

Estimates and residuals

- The population parameters \(\alpha, \beta_1, \beta_2,\) and \(\sigma\) are unknown.
- We compute estimates of the population parameters: \(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2\) and \(\hat{\sigma}\).
- \(\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}\) is called the fitted value.
- \(\hat{\epsilon}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\alpha} + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2})\) is called the residual.
- The residuals are observable, and can be used to check assumptions on the statistical errors \(\epsilon_i\).
- Points above the plane have positive residuals, and points below the plane have negative residuals.
- A plane that fits the data well has small residuals.
Computing estimates

- The triple \((\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)\) minimizes \(SSE(\alpha, \beta_1, \beta_2) = \sum \hat{\epsilon}_i^2 = \sum (Y_i - \alpha - \beta_1 X_{i1} - \beta_2 X_{i2})^2\).
- We can again take partial derivatives and set these equal to zero.
- This gives three equations in the three unknowns \(\alpha, \beta_1\) and \(\beta_2\). Solving these normal equations gives the regression coefficients \(\hat{\alpha}, \hat{\beta}_1\) and \(\hat{\beta}_2\).
- Least squares estimates are unique unless one of the independent variables is invariant, or independent variables are perfectly collinear.
- The same procedure works for \(p\) independent variables \(X_1, \ldots, X_p\). However, it is then easier to use matrix notation (see board and section 1.3 of script).
- In R: `model <- lm(y ~ x1 + x2)`

Properties of residuals

- \(\sum \hat{\epsilon}_i = 0\)
- The residuals \(\hat{\epsilon}_i\) are uncorrelated with the fitted values \(\hat{Y}_i\) and with each of the independent variables \(X_1, \ldots, X_p\).
- The standard error of the residuals \(\hat{\sigma} = \sqrt{\sum \hat{\epsilon}_i^2/(n - p - 1)}\) gives the “average” size of the residuals.
- \(n - p - 1\) is the degrees of freedom (we lose \(p + 1\) degrees of freedom because we estimate the \(p + 1\) parameters \(\alpha, \beta_1, \ldots, \beta_p\)).
$R^2$ and $\tilde{R}^2$

- $TSS = \sum (Y_i - \bar{Y})^2$.
- $SSE = \sum (Y_i - \hat{Y}_i)^2 = \sum \hat{\epsilon}_i^2$.
- $RegSS = TSS - SSE = \sum (\hat{Y}_i - \bar{Y})^2$.
- $R^2 = RegSS/TSS = 1 - SSE/TSS$ is the proportion of variation in $Y$ that is captured by its linear regression on the $X$’s.
- $R^2$ can never decrease when we add an extra variable to the model. Why?
- Adjusted $R^2$: $\tilde{R}^2 = 1 - \frac{SSE/(n-p-1)}{TSS/(n-1)}$ penalizes $R^2$ when there are extra variables in the model.
- $R^2$ and $\tilde{R}^2$ differ very little if sample size is large.

Ozone example

- Data from Sandberg, Basso, Okin (1978):
  - $SF = $ Summer quarter maximum hourly average ozone reading in parts per million in San Francisco
  - $SJ = $ Same, but then in San Jose
  - $YEAR = $ Year of ozone measurement
  - $RAIN = $ Average winter precipitation in centimeters in the San Francisco Bay area for the preceding two winters
- Research question: How does SF depend on YEAR and RAIN?
- Think about assumptions: Which one may be violated?
Ozone data

<table>
<thead>
<tr>
<th>YEAR</th>
<th>RAIN</th>
<th>SF</th>
<th>SJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>1965</td>
<td>18.9</td>
<td>4.3</td>
<td>4.2</td>
</tr>
<tr>
<td>1966</td>
<td>23.7</td>
<td>4.2</td>
<td>4.8</td>
</tr>
<tr>
<td>1967</td>
<td>26.2</td>
<td>4.6</td>
<td>5.3</td>
</tr>
<tr>
<td>1968</td>
<td>26.6</td>
<td>4.7</td>
<td>4.8</td>
</tr>
<tr>
<td>1969</td>
<td>39.6</td>
<td>4.1</td>
<td>5.5</td>
</tr>
<tr>
<td>1970</td>
<td>45.5</td>
<td>4.6</td>
<td>5.6</td>
</tr>
<tr>
<td>1971</td>
<td>26.7</td>
<td>3.7</td>
<td>5.4</td>
</tr>
<tr>
<td>1972</td>
<td>19.0</td>
<td>3.1</td>
<td>4.6</td>
</tr>
<tr>
<td>1973</td>
<td>30.6</td>
<td>3.4</td>
<td>5.1</td>
</tr>
<tr>
<td>1974</td>
<td>34.1</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>1975</td>
<td>23.7</td>
<td>2.1</td>
<td>2.7</td>
</tr>
<tr>
<td>1976</td>
<td>14.6</td>
<td>2.2</td>
<td>2.1</td>
</tr>
<tr>
<td>1977</td>
<td>7.6</td>
<td>2.0</td>
<td>2.5</td>
</tr>
</tbody>
</table>

R output

```r
> model <- lm(sf ~ year + rain)
> summary(model)

Call: lm(formula = sf ~ year + rain)

Residuals:
     Min      1Q  Median      3Q     Max
-0.61072 -0.20317  0.06129  0.16329  0.51992

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  388.412083  49.573690  7.835  1.41e-05 ***
  year       -0.195703   0.025112 -7.793  1.48e-05 ***
    rain      0.034288   0.009655  3.551  0.00526 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3224 on 10 degrees of freedom
Multiple R-Squared: 0.9089,     Adjusted R-squared: 0.8906
F-statistic: 49.87 on 2 and 10 DF,  p-value: 6.286e-06
```
Standardized coefficients

- We often want to compare coefficients of different independent variables.
- When the independent variables are measured in the same units, this is straightforward.
- If the independent variables are not commensurable, we can perform a limited comparison by rescaling the regression coefficients in relation to a measure of variation:
  - using hinge spread
  - using standard deviations

Using hinge spread

- Hinge spread = interquartile range (IQR)
- Let $IQR_1, \ldots, IQR_p$ be the IQRs of $X_1, \ldots, X_p$.
- We start with $Y_i = \hat{\alpha} + \hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_p X_{ip} + \hat{\epsilon}_i$.
- This can be rewritten as: $Y_i = \hat{\alpha} + \left( \hat{\beta}_1 IQR_1 \right) \frac{X_{i1}}{IQR_1} + \cdots + \left( \hat{\beta}_p IQR_p \right) \frac{X_{ip}}{IQR_p} + \hat{\epsilon}_i$.
- Let $Z_{ij} = \frac{X_{ij}}{IQR_j}$, for $j = 1, \ldots, p$ and $i = 1, \ldots, n$.
- Let $\hat{\beta}^*_j = \hat{\beta}_j IQR_j$, $j = 1, \ldots, p$.
- Then we get $Y_i = \hat{\alpha} + \hat{\beta}^*_1 Z_{i1} + \cdots + \hat{\beta}^*_p Z_{ip} + \hat{\epsilon}_i$.
- $\hat{\beta}^*_j = \hat{\beta}_j IQR_j$ is called the standardized regression coefficient.
Interpretation

- Interpretation: Increasing $Z_j$ by 1 and holding constant the other $Z_\ell$'s ($\ell \neq j$), is associated, on average, with an increase of $\hat{\beta}_j^*$ in $Y$.
- Increasing $Z_j$ by 1, means that $X_j$ is increased by one IQR of $X_j$.
- So increasing $X_j$ by one IQR of $X_j$ and holding constant the other $X_\ell$'s ($\ell \neq j$), is associated, on average, with an increase of $\hat{\beta}_j^*$ in $Y$.

- Ozone example:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff.</th>
<th>Hinge spread</th>
<th>Stand. coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>-0.196</td>
<td>6</td>
<td>-1.176</td>
</tr>
<tr>
<td>Rain</td>
<td>0.034</td>
<td>11.6</td>
<td>0.394</td>
</tr>
</tbody>
</table>

Using st.dev.

- Let $S_Y$ be the standard deviation of $Y$, and let $S_1, \ldots, S_p$ be the standard deviations of $X_1, \ldots, X_p$.
- We start with $Y_i = \hat{\alpha} + \hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_p X_{ip} + \hat{\epsilon}_i$.
- This can be rewritten as (derivation on board):
  $$\frac{Y_i - \bar{Y}}{S_Y} = \left( \hat{\beta}_1 \frac{S_1}{S_Y} \right) \frac{X_{i1} - \bar{X}_1}{S_1} + \ldots + \left( \hat{\beta}_p \frac{S_p}{S_Y} \right) \frac{X_{ip} - \bar{X}_p}{S_p} + \frac{\hat{\epsilon}_i}{S_Y}.$$
- Let $Z_{iY} = \frac{Y_i - \bar{Y}}{S_Y}$ and $Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}$, for $j = 1, \ldots, p$.
- Let $\hat{\beta}_j^* = \hat{\beta}_j \frac{S_j}{S_Y}$ and $\hat{\epsilon}_i^* = \frac{\hat{\epsilon}_i}{S_Y}$.
- Then we get $Z_{iY} = \hat{\beta}_1^* Z_{i1} + \ldots + \hat{\beta}_p^* Z_{ip} + \hat{\epsilon}_i^*$.
- $\hat{\beta}_j^* = \hat{\beta}_j \frac{S_j}{S_Y}$ is called the standardized regression coefficient.
Interpretation

- Interpretation: Increasing $Z_j$ by 1 and holding constant the other $Z_{\ell}$'s ($\ell \neq j$), is associated, on average, with an increase of $\hat{\beta}_j$ in $Z_Y$.
- Increasing $Z_j$ by 1, means that $X_j$ is increased by one SD of $X_j$.
- Increasing $Z_Y$ by 1 means that $Y$ is increased by one SD of $Y$.
- So increasing $X_j$ by one SD of $X_j$ and holding constant the other $X_{\ell}$'s ($\ell \neq j$), is associated, on average, with an increase of $\hat{\beta}_j$ SDs of $Y$ in $Y$.

Ozone example

- Ozone example:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff.</th>
<th>St.dev(variable)</th>
<th>Stand. coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>-0.196</td>
<td>3.99</td>
<td>-0.783</td>
</tr>
<tr>
<td>Rain</td>
<td>0.034</td>
<td>10.39</td>
<td>0.353</td>
</tr>
</tbody>
</table>

- Both methods (using hinge spread or standard deviations) only allow for a very limited comparison. They both assume that predictors with a large spread are more important, and that does not need to be the case.
### Added variable plots

- Suppose we start with $SF \sim YEAR$.
- We want to know whether it is helpful to add the variable RAIN.
- We want to model that part of SF that is not explained by YEAR (residuals of $lm(SF \sim YEAR)$) with the part of RAIN that is not explained by YEAR (residuals of $lm(RAIN \sim YEAR)$).
- Plotting these residuals against each other is called an *added variable plot* for the effect of RAIN on SF, controlling for YEAR.
- Regressing residuals of $lm(SF \sim YEAR)$ on the residuals of $lm(RAIN \sim YEAR)$ gives the coefficient for RAIN when controlling for YEAR.

### Summary

- Linear statistical model: $Y = \alpha + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$.
- We assume that the statistical errors $\epsilon$ have mean zero, constant standard deviation $\sigma$, and are uncorrelated.
- The *population parameters* $\alpha, \beta_1, \ldots, \beta_p$ and $\sigma$ cannot be observed. Also the statistical errors $\epsilon$ cannot be observed.
- We define the *fitted value* $\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_p X_{ip}$ and the residual $\hat{\epsilon}_i = Y_i - \hat{Y}_i$. We can use the residuals to check the assumptions about the statistical errors.
- We compute estimates $\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_p$ for $\alpha, \beta_1, \ldots, \beta_p$ by minimizing the *residual sum of squares* $SSE = \sum \hat{\epsilon}_i^2 = \sum (Y_i - (\hat{\alpha} + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_p X_{ip}))^2$.
- Interpretation of the coefficients?
Summary

- To measure how good the fit is, we can use:
  - the residual standard error $\hat{\sigma} = \sqrt{\frac{SSE}{n - p - 1}}$
  - the multiple correlation coefficient $R^2$
  - the adjusted multiple correlation coefficient $\tilde{R}^2$
  - the correlation coefficient $r$
- Analysis of variance (ANOVA): $TSS = SSE + RegSS$
- Standardized regression coefficients
- Added variable plots (partial regression plots)