

# Statistical Modeling

CH.2 - Multiple Linear Regression

SS 2021 || Prof. Dr. Buchwitz

Wir geben Impulse

**1** Organizational Information

**2** Multiple Linear Regression

| Session | Topic                               |
|---------|-------------------------------------|
| 1       | Simple Linear Regression            |
| 2       | Multiple Linear Regression          |
| 3       | Regression Diagnostics              |
| 4       | Qualitative Variables as Predictors |
| 5       | Transformation of Variables         |
| 6       | Weighted Least Squares              |
| 7       | Correlated Errors                   |
| 8       | Analysis of Collinear Data          |
| 9       | Working with Collinear Data         |
| 10      | Variable Selection Procedures       |
| 11      | Logistic Regression                 |
| 12      | Further Topics                      |

Grading is based on a portfolio examination with three parts:

- 1 One Lecture Recap Presentation (20%)
- 2 Hand-in Exercises (40%)
- 3 Final Case Study (40%)

**1** Organizational Information

**2** Multiple Linear Regression

$$Y = f(X_1, X_2, \dots, X_p) + \epsilon$$

- We now generalize the simple linear regression model so that the relation between the response  $Y$  and  $p$  predictor variables  $X_1, X_2, \dots, X_p$  can be studied.
- We still assume that **within the range** of the data the true relation between  $Y$  and the predictors can be approximated using a linear function.
- The previously discussed simple linear regression model can be seen as a special case of the general linear regression model where  $p = 1$ .

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

- Each regressor needs its own constant  $\beta$  so that the regression coefficients are now  $\beta_0, \beta_1, \dots, \beta_p$ .
- The random disturbance is noted using  $\epsilon$ . This term measures the discrepancy in the approximation and  $\epsilon$  contains **no systematic information for determining**  $Y$  that is not already captured by the  $X$ 's.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

$$i = 1, 2, \dots, n$$

- Following from the model equation, each observation can be written as above, where  $y_i$  represents the  $i$ -th value of the response variable  $Y$ .
- $x_{i1}, x_{i2}, \dots, x_{ip}$  are the values of the regressors that are associated with the  $i$ -th unit in the sample (usually  $i$ -th row of data).
- $\epsilon_i$  represents the individual error in the linear approximation for the  $i$ -th data unit.



## Example: Supervisor Performance

| Variable       | Description                                    |
|----------------|--|
| Y              | Overall rating of job being done by supervisor |
| X <sub>1</sub> | Handles employee complaints                    |
| X <sub>2</sub> | Does not allow special privileges              |
| X <sub>3</sub> | Opportunity to learn new things                |
| X <sub>4</sub> | Raises based on performance                    |
| X <sub>5</sub> | Too critical of poor performance               |
| X <sub>6</sub> | Rate of advancing to better jobs               |

- X<sub>1</sub>, X<sub>2</sub> and X<sub>5</sub> relate to direct interpersonal relationships between employee and supervisor
- X<sub>3</sub> and X<sub>4</sub> relate to the job as a whole
- X<sub>6</sub> is an indicator of the perceived progress of the employee

## Example: Supervisor Performance

*# Supervisor Performance Data*

P060

| ##    |  | Y  | X1 | X2 | X3 | X4 | X5 | X6 |
|-------|--|----|----|----|----|----|----|----|
| ## 1  |  | 43 | 51 | 30 | 39 | 61 | 92 | 45 |
| ## 2  |  | 63 | 64 | 51 | 54 | 63 | 73 | 47 |
| ## 3  |  | 71 | 70 | 68 | 69 | 76 | 86 | 48 |
| ## 4  |  | 61 | 63 | 45 | 47 | 54 | 84 | 35 |
| ## 5  |  | 81 | 78 | 56 | 66 | 71 | 83 | 47 |
| ## 6  |  | 43 | 55 | 49 | 44 | 54 | 49 | 34 |
| ## 7  |  | 58 | 67 | 42 | 56 | 66 | 68 | 35 |
| ## 8  |  | 71 | 75 | 50 | 55 | 70 | 66 | 41 |
| ## 9  |  | 72 | 82 | 72 | 67 | 71 | 83 | 31 |
| ## 10 |  | 67 | 61 | 45 | 47 | 62 | 80 | 41 |
| ## 11 |  | 64 | 53 | 53 | 58 | 58 | 67 | 34 |
| ## 12 |  | 67 | 60 | 47 | 39 | 59 | 74 | 41 |
| ## 13 |  | 69 | 62 | 57 | 42 | 55 | 63 | 25 |
| ## 14 |  | 68 | 83 | 83 | 45 | 59 | 77 | 35 |
| ## 15 |  | 77 | 77 | 54 | 72 | 79 | 77 | 46 |
| ## 16 |  | 81 | 90 | 50 | 72 | 60 | 54 | 36 |
| ## 17 |  | 74 | 85 | 64 | 69 | 79 | 79 | 63 |
| ## 18 |  | 65 | 60 | 65 | 75 | 55 | 80 | 60 |

## Example: Supervisor Performance

```
# Descriptive Statistics for all variables
```

```
dim(P060)
```

```
## [1] 30 7
```

```
t(sapply(P060, summary))
```

| ##    |  | Min. | 1st Qu. | Median | Mean     | 3rd Qu. | Max. |
|-------|--|------|---------|--------|----------|---------|------|
| ## Y  |  | 40   | 58.75   | 65.5   | 64.63333 | 71.75   | 85   |
| ## X1 |  | 37   | 58.50   | 65.0   | 66.60000 | 77.00   | 90   |
| ## X2 |  | 30   | 45.00   | 51.5   | 53.13333 | 62.50   | 83   |
| ## X3 |  | 34   | 47.00   | 56.5   | 56.36667 | 66.75   | 75   |
| ## X4 |  | 43   | 58.25   | 63.5   | 64.63333 | 71.00   | 88   |
| ## X5 |  | 49   | 69.25   | 77.5   | 74.76667 | 80.00   | 92   |
| ## X6 |  | 25   | 35.00   | 41.0   | 42.93333 | 47.75   | 72   |

## Example: Supervisor Performance

We now **assume** that  $Y$  is linearly related to the six explanatory variables:

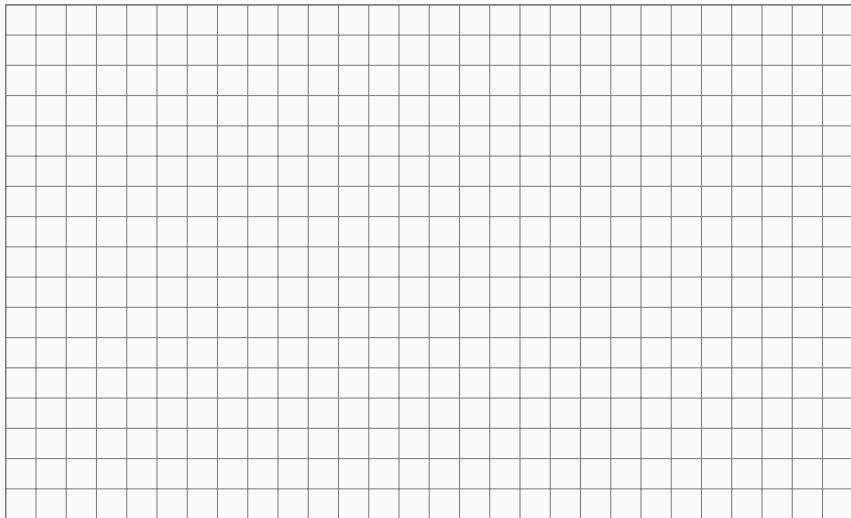
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon$$

Note that this assumption needs to be verified, which we ignore for now! It may however e.g. be possible that one or more variables have no relation with  $Y$  or that the model is incomplete and important variables are missing.

How do we estimate the parameters  $\beta_0, \beta_1, \dots, \beta_p$  for the multiple linear regression?

## Parameter Estimation

How do we estimate the regression parameters  $\beta_0, \beta_1, \dots, \beta_p$ ?



- **Solution:** Minimizing the sum of squares.
- The individual errors can be written as

$$\epsilon_i = y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip}$$

- The sum of squares therefore is

$$S(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip})^2$$

- $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$  are the constants that minimize  $S(\beta_0, \beta_1, \dots, \beta_p)$ .  
Formulating the minimization problem leads to a system of equations, which are called the *normal equations*. The solution is usually derived using matrix notation.

- **Solution:** Minimizing the sum of squares.
- The individual errors can be written as

$$\epsilon_i = y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip}$$

- The sum of squares therefore is

$$S(\beta_0, \beta_1, \dots, \beta_p) = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_p x_{ip})^2$$

- $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$  are the constants that minimize  $S(\beta_0, \beta_1, \dots, \beta_p)$ .  
Formulating the minimization problem leads to a system of equations, which are called the *normal equations*. The solution is usually derived using matrix notation.

We do not cover the solution in matrix notation in this course. If you are interested, please ask me about it or see the Appendix to Chapter 3 in RABE5.

- Using the fitted regression parameters, the **fitted least squares regression equation** can be noted

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p$$

- The fitted values follow analogously

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip} \quad \text{for } i = 1, 2, \dots, n$$

- The residuals are

$$e_i = y_i - \hat{y}_i \quad \text{for } i = 1, 2, \dots, n$$

- An unbiased estimate of the variance  $\sigma^2$  is

$$\hat{\sigma}^2 = \frac{SSE}{n - p - 1} \quad \text{with} \quad SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n e_i^2$$

- $n - p - 1$  is called the degrees of freedom and is equal to the number of observations minus the number of estimated regression coefficients.



# Interpretation

- The simple regression equation (only  $X_1$ ) represents a line.
- The multiple regression equation represents a plane ( $X_1$  and  $X_2$ ) or a hyperplane (more than two predictors).

```
library(car) # 3d Visualization of regression plane for two X variables
car::scatter3d(P060$Y, P060$X1, P060$X2)
```

- $\beta_0$  is the *constant coefficient* or *intercept* and results in the value of  $Y$ , when  $X_1 = X_2 = \dots = X_p = 0$ .
- The regression coefficient  $\beta_j$  has several interpretations:
  - $\beta_j$  is the change in  $Y$  corresponding to a unit change in  $X_j$ , when all other predictors are held constant (*ceteris paribus*).
  - $\beta_j$  is also called the *partial regression coefficient* as it represents the contribution of  $X_j$  to the response variable, after it has been adjusted for the other predictor variables.

## Example: Supervisor Performance

```
mod <- lm(Y ~ 1 + X1 + X2, data=P060)
summary(mod)
```

```
##
## Call:
## lm(formula = Y ~ 1 + X1 + X2, data = P060)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.7887  -5.6893  -0.0284   6.2745   9.9726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.32762     7.16023   2.141  0.0415 *
## X1           0.78034     0.11939   6.536 5.22e-07 ***
## X2          -0.05016     0.12992  -0.386  0.7025
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.102 on 27 degrees of freedom
## Multiple R-squared:  0.6831, Adjusted R-squared:  0.6596
## F-statistic: 29.1 on 2 and 27 DF, p-value: 1.833e-07
```

### Your turn

Write down the parameterized regression equation.

Plugging in the estimated regression coefficients from the output yields:

$$\hat{Y} = 15.3276 + 0.7803(X_1) - 0.0502(X_2)$$

- Lets explore the question what **adjusted for** actually means:
  - The coefficient for  $X_1$  suggests that each unit of  $X_1$  adds 0.7803 to  $Y$ , when the value of  $X_2$  is held fix.
  - The effect of  $X_2$  after adjusting for  $X_1$  is -0.0502, which means that for each unit increase of  $X_2$  the response variable  $Y$  decreases -0.0502 units and the value of  $X_1$  is held fix.
- These results can also be obtained iteratively by applying a series of simple regression models.

## 1 Relating $Y$ and $X_1$ to obtain the residuals $e_{Y \cdot X_1}$ .

```
mod1 <- lm(Y ~ 1 + X1, data=P060)
round(coef(mod1),4)
```

```
## (Intercept)      X1
##      14.3763      0.7546
```

## 2 Relating $X_2$ (temporary response) and $X_1$ to obtain residuals $e_{X_2 \cdot X_1}$ .

```
mod2 <- lm(X2 ~ 1 + X1, data=P060)
round(coef(mod2),4)
```

```
## (Intercept)      X1
##      18.9654      0.5130
```

## 3 Relating the residuals $e_{Y \cdot X_1}$ with $e_{X_2 \cdot X_1}$ .

```
e_y_x1 <- residuals(mod1)
e_x2_x1 <- residuals(mod2)
mod3 <- lm(e_y_x1 ~ 1 + e_x2_x1, data=P060)
round(coef(mod3),4)
```

```
## (Intercept)      e_x2_x1
##      0.0000     -0.0502
```

$e_{Y \cdot X_1}$  can be understood as  $Y$  with the linear influence from  $X_1$  removed. Analogously  $e_{X_2 \cdot X_1}$  is the regressor  $X_2$  with the partial effect of  $X_1$  removed.

- A Series of simple linear regressions can reproduce the coefficients (and standard errors) from a multivariate regression.
- The results from simple and multiple linear regression would only be equal if the regressors (the  $X$ 's) are not correlated, which is almost never the case in practice.
- Each regression coefficient  $\beta_j$  can be seen as **partial regression coefficient** because it represents the contribution of  $X_j$  to the response variable  $Y$  after both variables have been **adjusted for the linear dependence** of the other regressors.

- The value of the regression coefficient is dependent on the unit of measurement of the variables.
- To obtain unitless regression coefficients, we can center and scale (e.g. z-transform) our variables
- **Centering** can be achieved by subtracting the mean of the response  $Y - \bar{y}$  or predictor variables  $X - \bar{x}$ .
- **Scaling** can be done by *unit-length* scaling or standardizing

- **Unit Length Scaling** for  $Y$

$$Z_y = \frac{Y - \bar{y}}{L_y} \quad \text{with} \quad L_y = \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}$$

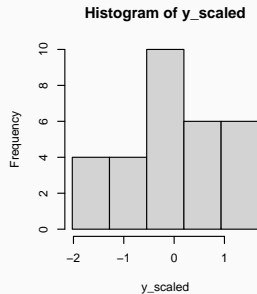
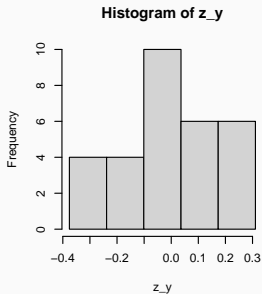
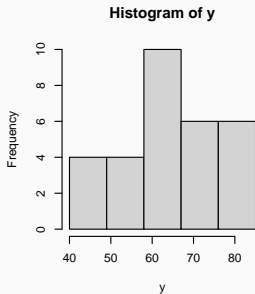
- **Standardizing** for  $Y$

$$Z_y = \frac{Y - \bar{y}}{s_y} \quad \text{with} \quad s_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}}$$

- Correlations are unaffected by these kind of transformations
- $\beta$ 's change as well as their interpretation, e.g. beta coefficients obtained after standardizing the variables represent marginal effects of the predictor variables in *standard deviation units*

# Centering and Scaling

```
y <- P060$Y  
L_y <- sqrt(sum((y-mean(y))^2))  
z_y <- (y - mean(y))/L_y           # Unit-length Scaling  
y_scaled <- scale(y)              # Standardizing by default
```





**BLUE**

# BLUE

Best Linear Unbiased Estimator

# Properties of the Least Squares Estimator

- 1 For all unbiased estimates that are linear in the observations the least squares estimators have the smallest variance.
- 2 The estimator  $\hat{\beta}_j$  is normally distributed with mean  $\beta_j$  and variance  $\sigma^2 c_{jj}$ , where  $c_{jj}$  is the  $j$ -th diagonal element of the inverse of a matrix known as the *corrected sums of squares and products* matrix.
- 3  $W = SSE/\sigma^2$  has a  $\chi^2$  distribution with  $n - p - 1$  degrees of freedom and  $\hat{\beta}_j$ 's and  $\hat{\sigma}^2$  are distributed independently of each other.
- 4 The vector  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p)$  has a  $p + 1$  dimensional normal distribution and with mean vector  $\beta = (\beta_0, \beta_1, \dots, \beta_p)$  and a variance-covariance matrix with elements  $\sigma^2 c_{ij}$ .

**It is due to these properties that we are able to test hypotheses about individual regression parameters and construct confidence intervals!**

## Multiple Correlation Coefficient

- The content discussed for testing the **goodness of fit** for the simple linear regression model can easily be extended for the larger regression models and the following holds also for the multiple regression model:

$$R^2 = [\text{Cor}(Y, \hat{Y})]^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

- Sometimes  $R = \sqrt{R^2}$  is referred to as the multiple correlation coefficient as it measures the correlation between the response and a set of predictors.
  - When the model provides a good fit for the data  $R^2$  is close to unity (and  $\sum (y_i - \hat{y}_i)$  will be small).
  - When the variables  $X_1, X_2, \dots, X_p$  do not have a linear relationship with the response, the best predicted value for each  $y_i$  is the mean  $\bar{y}$  (because it minimizes the sum of squares) and  $R^2$  will be near zero.

# Multiple Correlation Coefficient

```
mod <- lm(Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6, data=P060)
summary(mod)

##
## Call:
## lm(formula = Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6, data = P060)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9418  -4.3555   0.3158   5.5425  11.5990
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.78708    11.58926   0.931 0.361634
## X1           0.61319     0.16098   3.809 0.000903 ***
## X2          -0.07305     0.13572  -0.538 0.595594
## X3           0.32033     0.16852   1.901 0.069925 .
## X4           0.08173     0.22148   0.369 0.715480
## X5           0.03838     0.14700   0.261 0.796334
## X6          -0.21706     0.17821  -1.218 0.235577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.068 on 23 degrees of freedom
## Multiple R-squared:  0.7326, Adjusted R-squared:  0.6628
## F-statistic: 10.5 on 6 and 23 DF, p-value: 1.24e-05
```

## Your turn

What happens with  $R^2$  when we add one or many artificially generated variables  $z_j \sim N(0, 1)$  that are pure white noise to the model?

# Multiple Correlation Coefficient

```
mod <- lm(Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6, data=P060)
mod_summary <- summary(mod)
c(mod_summary$r.square, mod_summary$adj.r.squared)
```

```
## [1] 0.732602 0.662846
```

```
set.seed(1)
z1 <- rnorm(nrow(P060))
z2 <- rnorm(nrow(P060))
z3 <- rnorm(nrow(P060))
z4 <- rnorm(nrow(P060))
z5 <- rnorm(nrow(P060))

mod_large <- lm(Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6
               + z1 + z2 + z3 + z4 + z5, data=P060)
mod_large_summary <- summary(mod_large)
c(mod_large_summary$r.square, mod_large_summary$adj.r.squared)
```

```
## [1] 0.7738745 0.6356867
```

$$R_a^2 = 1 - \frac{SSE/(n - p - 1)}{SST/(n - 1)} = 1 - \frac{n - 1}{n - p - 1}(1 - R^2)$$

- The adjusted  $R^2$  tries to solve the observed issue by “adjusting” for the difference in the numbers of variables when judging the *goodness of fit*.
- The  $R_a^2$  cannot be interpreted as the proportion of total variation in  $Y$  accounted for by the predictors.

- Testing the **individual** regression coefficients for  $H_0 : \beta_j = \beta_j^0$  versus  $H_1 : \beta_j \neq \beta_j^0$  can be done using the t-test by comparing the observed value  $t_j$  with the appropriate critical value  $t_{n-p-1, \alpha/2}$ .

- For the two-sided test,  $H_0$  is rejected at the significance level  $\alpha$  if

$$|t_j| \geq t_{n-p-1, \alpha/2} \quad \text{with} \quad t_j = \frac{\hat{\beta}_j - \beta_j^0}{s.e.(\hat{\beta}_j)}$$

- An equivalent decision can be made using the p-value and reject  $H_0$  if  $p(|t_j|) \leq \alpha$ . The p-value is the probability that a random variable having a Student t-distribution with  $n-p-1$  degrees of freedom is greater than  $|t_j|$  (the observed value of the t-test).



# Inference for Regression Coefficients

```
mod <- lm(Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6, data=P060)
summary(mod)

##
## Call:
## lm(formula = Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6, data = P060)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9418  -4.3555   0.3158   5.5425  11.5990
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.78708    11.58926   0.931 0.361634
## X1           0.61319     0.16098   3.809 0.000903 ***
## X2          -0.07305     0.13572  -0.538 0.595594
## X3           0.32033     0.16852   1.901 0.069925 .
## X4           0.08173     0.22148   0.369 0.715480
## X5           0.03838     0.14700   0.261 0.796334
## X6          -0.21706     0.17821  -1.218 0.235577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.068 on 23 degrees of freedom
## Multiple R-squared:  0.7326, Adjusted R-squared:  0.6628
## F-statistic: 10.5 on 6 and 23 DF, p-value: 1.24e-05
```

## Your turn

Interpret the regression coefficients (assume  $\alpha = 0.05$ ).

$Y$  Overall rating of job being done by supervisor

$X_1$  Handles employee complaints

$X_2$  Does not allow special privileges

$X_3$  Opportunity to learn new things

$X_4$  Raises based on performance

$X_5$  Too critical of poor performance

$X_6$  Rate of advancing to better jobs

- 1 All the regression coefficients associated with the predictor variables are zero.
- 2 Some of the regression coefficients are zero.
- 3 Some of the regression coefficients are equal to each other.
- 4 The regression parameters satisfy certain specified constraints.

**We discuss a general approach for testing all stated hypotheses and illustrate that for testing if all regression coefficients are zero (option 1). Please see the recommended literature for the other options.**

- The model that contains all  $p + 1$  parameters will be referred to as the **full model** (FM).

$$\text{Full Model: } Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

- We are formulating a hypothesis for some coefficients  $\beta_0, \beta_1, \dots, \beta_p$  in the model. This incorporating the hypothesized values in the model we receive the so called **reduced model** (RM).
- The number of distinct parameters to estimate in the RM is smaller than the numbers of parameters to estimate in the FM.
- Subsequently we are going to test  
 $H_0$  : Reduced model is adequate.   vs.    $H_1$  : Full model is adequate.

- The main idea of the described approach is testing for **differences in goodness of fit** between the FM and RM.
- If the reduced model gives as good fit as the full model, the null hypothesis ( $H_0$  : Reduced model is adequate) is not rejected.
- The **lack of fit** can be measured by the sum of squared residuals (SSE), which we note using the predicted values for  $y_i$  of the full model  $\hat{y}_i$  and the reduced model  $\hat{y}_i^*$  respectively.

$$SSE(FM) = \sum (y_i - \hat{y}_i)^2 \quad SSE(RM) = \sum (y_i - \hat{y}_i^*)^2$$

- In the full model there are  $p + 1$  regression parameters  $(\beta_0, \beta_1, \dots, \beta_p)$  that need to be estimated. Lets suppose that the reduced model contains  $k$  distinct parameters.
- $SSE(RM) \geq SSE(FM)$  must hold, as the additional variables cannot increase the residual sum of squares.
- The difference  $SSE(RM) - SSE(FM)$  represents the increase in residual sum of squares due to fitting the reduced model. **If this difference is large, the RM is inadequate!**

$$F = \frac{[SSE(RM) - SSE(FM)]/(p + 1 - k)}{SSE(FM)/(n - p - 1)}$$

- The ratio above is called the statistic for the **F-Test**.
- We divide  $SSE(FM)$  and the discussed difference by their respective degrees of freedom to compensate for differences in number of parameters and ensure that the ratio follows a standard statistical distribution (F-distribution).
  - FM has  $p$  parameters, hence  $SSE(FM)$  has  $n - p - 1$  degrees of freedom
  - RM has  $k$  parameters, hence  $SSE(RM)$  has  $n - k$  degrees of freedom
  - $SSE(RM) - SSE(FM)$  has  $(n - k) - (n - p - 1) = p + 1 - k$  degrees of freedom
  - Required F-distribution has  $(p + 1 - k)$  and  $(n - p - 1)$  degrees of freedom

- If the observed F-value is large in comparison to the tabulated value of F with the respective degrees of freedom, the result is significant at level  $\alpha$ .
- This means that the reduced model is unsatisfactory and the null hypothesis (with the suggested values for the  $\beta$ 's) is rejected.
- $H_0$  is rejected if

$$F \geq F_{(p+1-k, n-p-1; 1-\alpha)} \quad \text{or} \quad p(F) \leq \alpha$$

- Here  $F$  is the observed value of the F-test,  $F_{(p+1-k, n-p-1; 1-\alpha)}$  is the appropriate critical value obtained from the F table and  $p(F)$  is the p-value for the F-Test (probability that a random variable having an F-distribution and the stated DoF is greater than the observed  $F$ ).

## Hypothesis Testing: All Regression Coefficients equal to Zero

$$RM : H_0 : Y = \beta_0 + \epsilon$$

$$FM : H_1 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

- The residual sum of squares resulting from the full model is  $SSE(FM) = SSE$ .
- Because the estimate for  $\beta_0$  in the reduced model is  $\hat{\beta}_0 = \bar{y}$ , the residual sum of squares from the reduced model is  $SSE(RM) = \sum (y_i - \bar{y})^2 = SST$ . Additionally,  $SST = SSR + SSE$  holds so that the F-test reduces to

$$F = \frac{[SST - SSE]/p}{SSE/(n - p - 1)} = \frac{SSR/p}{SSE/(n - p - 1)} = \frac{MSR}{MSE}$$



## Hypothesis Testing: All Regression Coefficients equal to Zero

- That F-test essentially tests  $H_0$  if the regression coefficients (excluding  $\beta_0$ ) contribute to the explanation of the response variable.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

- The hypothesis is equal to the multiple correlation coefficient  $R$  being zero. So the same hypothesis can be checked using the following where  $R_p$  denotes the multiple correlation coefficient, which is obtained from fitting a model to  $n$  observations with  $p$  predictor variables (+ intercept)

$$F = \frac{R_p^2/p}{(1 - R_p^2)/(n - p - 1)}$$

## Hypothesis Testing: All Regression Coefficients equal to Zero

- That F-test essentially tests  $H_0$  if the regression coefficients (excluding  $\beta_0$ ) contribute to the explanation of the response variable.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

- The hypothesis is equal to the multiple correlation coefficient  $R$  being zero. So the same hypothesis can be checked using the following where  $R_p$  denotes the multiple correlation coefficient, which is obtained from fitting a model to  $n$  observations with  $p$  predictor variables (+ intercept)

$$F = \frac{R_p^2/p}{(1 - R_p^2)/(n - p - 1)}$$

**Your turn**

What is  $H_1$ ?

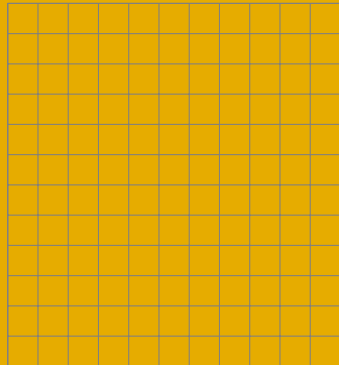
## Example

```
mod <- lm(Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6, data=P060)
summary(mod)

##
## Call:
## lm(formula = Y ~ 1 + X1 + X2 + X3 + X4 + X5 + X6, data = P060)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.9418  -4.3555   0.3158   5.5425  11.5990
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.78708    11.58926   0.931 0.361634
## X1           0.61319     0.16098   3.809 0.000903 ***
## X2          -0.07305     0.13572  -0.538 0.595594
## X3           0.32033     0.16852   1.901 0.069925 .
## X4           0.08173     0.22148   0.369 0.715480
## X5           0.03838     0.14700   0.261 0.796334
## X6          -0.21706     0.17821  -1.218 0.235577
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.068 on 23 degrees of freedom
## Multiple R-squared:  0.7326, Adjusted R-squared:  0.6628
## F-statistic: 10.5 on 6 and 23 DF, p-value: 1.24e-05
```

## Your turn

Calculate  $F$  and perform the F-test to test if all regression coefficients  $\beta_1, \dots, \beta_6$  are equal to zero. Use  $\alpha = 0.05$ .



As with simple linear regression multiple linear regression can be used to either predict the value of a response or the value of the mean response. The value  $x_0$  to be inserted for the predictor variables now consists of  $p$  values so that  $x_0 = (x_{01}, x_{02}, \dots, x_{0p})$ .

## Value of Response & Prediction Limits

The predicted value  $\hat{y}_0$  is given by  $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \dots + \hat{\beta}_p x_{0p}$ . Its confidence interval can be constructed by  $\hat{y}_0 \pm t_{(n-p-1, \alpha/2)} \cdot s.e.(\hat{y}_0)$ .

## Mean Response & Confidence Limits

The predicted value  $\hat{\mu}_0$  is given by  $\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \dots + \hat{\beta}_p x_{0p}$ . The confidence interval for that prediction is given by  $\hat{\mu}_0 \pm t_{(n-p-1, \alpha/2)} \cdot s.e.(\hat{\mu}_0)$ .

**Note:** The standard errors require matrix notation so that their derivation is not shown here. They can, however, be found in the recommended literature.