Lineare Regression in R

```r
str(hauspreise)  # Quelle: wooldridge::hprice1 (transformiert)
```

```
'data.frame':	88 obs. of  4 variables:
 $ preis       : num  300000 370000 191000 195000 373000 ...
 $ wohnflaeche : num  226 193 128 135 234 ...
 $ schlafzimmer: int  4 3 3 3 4 5 3 3 3 3 ...
 $ grundflaeche: num  569 920 483 427 566 ...
```

Lineare Regression in R

Falls es dich interessiert, genau so habe ich den Datensatz `hauspreise` gebildet:

```
library(dplyr)
umrechnungsfaktor <- 0.09290304
hauspreise <- wooldridge::hprice1 %>%
  summarize(
    preis = price * 1000,
    wohnflaeche = sqrft * umrechnungsfaktor,
    schlafzimmer = bdrms,
    grundflaeche = lotsize * umrechnungsfaktor
  )
```

Lineare Regression in R

```r
str(hauspreise)  # Quelle: wooldridge::hprice1 (transformiert)
```

```
'data.frame':	88 obs. of  4 variables:
 $ preis       : num  300000 370000 191000 195000 373000 ...
 $ wohnflaeche : num  226 193 128 135 234 ...
 $ schlafzimmer: int  4 3 3 3 4 5 3 3 3 3 ...
 $ grundflaeche: num  569 920 483 427 566 ...
```

Lineare Regression in R

```r
str(hauspreise)  # Quelle: wooldridge::hprice1 (transformiert)
```

```
'data.frame':	88 obs. of  4 variables:
 $ preis       : num  300000 370000 191000 195000 373000 ...
 $ wohnflaeche : num  226 193 128 135 234 ...
 $ schlafzimmer: int  4 3 3 3 4 5 3 3 3 3 ...
 $ grundflaeche: num  569 920 483 427 566 ...
```

```r
model <- lm(
  formula = preis ~ wohnflaeche + I(schlafzimmer > 4) + grundflaeche,
  data = hauspreise
)
```

Lineare Regression in R

```r
str(hauspreise)  # Quelle: wooldridge::hprice1 (transformiert)
```

```
'data.frame':   88 obs. of  4 variables:
 $ preis       : num  300000 370000 191000 195000 373000 ...
 $ wohnflaeche : num  226 193 128 135 234 ...
 $ schlafzimmer: int  4 3 3 3 4 5 3 3 3 3 ...
 $ grundflaeche: num  569 920 483 427 566 ...
```

```r
model <- lm(
  formula = preis ~ wohnflaeche + I(schlafzimmer > 4) + grundflaeche,
  data = hauspreise
)
```

```r
summary(model)
```

```
Call:
lm(formula = preis ~ wohnflaeche + I(schlafzimmer > 4) + grundflaeche,
    data = hauspreise)

Residuals:
    Min      1Q  Median      3Q     Max
-149356  -27653   -3891   24175  202625

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 53990 on 84 degrees of freedom
Multiple R-squared:  0.7333,    Adjusted R-squared:  0.7237
F-statistic: 76.97 on 3 and 84 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = preis ~ wohnflaeche + I(schlafzimmer > 4) + grundflaeche,
    data = hauspreise)
```

Ganz oben steht der Funktionsaufruf (“`Call`”) der `lm()` Funktion:

1. `formula` ist unsere Modellformel
2. `data` ist der Name unseres Datensatzes

```
Call:
lm(formula = preis ~ wohnflaeche + I(schlafzimmer > 4) + grundflaeche,
    data = hauspreise)
```

Ganz oben steht der Funktionsaufruf (“`Call`”) der `lm()` Funktion:

1. `formula` ist unsere Modellformel
2. `data` ist der Name unseres Datensatzes

Die Modellformel übersetzen wir so:

- Allgemeine Formel: $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + u_i$, $i = 1, \ldots, n$
- R Formel: `preis ~ wohnflaeche + I(schlafzimmer > 4) + grundflaeche`
- Übersetzt:

$$\text{preis}_i = \beta_0 + \beta_1 \text{wohnflaeche}_i + \beta_2 \underbrace{1\{\text{schlafzimmer} > 4\}}_{\text{Indikator-Variable}} + \beta_3 \text{grundflaeche} + u_i$$

```
Residuals:
    Min      1Q  Median      3Q     Max
-149356  -27653   -3891   24175  202625
```

Als nächstes kommt eine Zusammenfassung der Residuen.

Was sind nochmal Residuen?

Für alle Beobachtungen $i = 1, \ldots, n$ ist das Residuum $\hat{u}_i$ die Differenz aus dem beobachteten Preis preis_i und dem durch das Modell erklärten Preis $\widehat{\text{preis}}_i$, also:

$$
\begin{aligned}
\hat{u}_i &= \text{preis}_i - \widehat{\text{preis}}_i \\
&= \text{preis}_i - (\hat{\beta}_0 + \hat{\beta}_1 \text{wohnflaeche}_i + \hat{\beta}_2 1\{\text{schlafzimmer} > 4\} + \hat{\beta}_3 \text{grundflaeche}).
\end{aligned}
$$

```
Residuals:
    Min      1Q  Median      3Q     Max 
-149356  -27653   -3891   24175  202625 
```

Die konkreten Residuenwerte für unser geschätztes Modell bekommen wir so:

```r
residuals <- model$residuals
head(residuals, n = 5)
```

```
        1         2         3         4         5 
-22070.79  78966.08 -17285.44 -19952.80  42983.91 
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-149356  -27653   -3891   24175  202625 
```

Die konkreten Residuenwerte für unser geschätztes Modell bekommen wir so:

```r
residuals <- model$residuals
head(residuals, n = 5)
```

```
        1         2         3         4         5 
-22070.79  78966.08 -17285.44 -19952.80  42983.91 
```

```r
summary(residuals)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
-149356  -27653   -3891       0   24175  202625 
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-149356  -27653   -3891   24175  202625
```

Was für Werte wünschen wir uns hier eigentlich?

```
Residuals:
    Min      1Q  Median      3Q     Max
-149356  -27653   -3891   24175  202625
```

Was für Werte wünschen wir uns hier eigentlich?

Wir möchten, dass die Residuen symmetrisch um die geschätzte Regressionslinie streuen, also sollte:

- $|\texttt{Min}| \approx \texttt{Max}$
- $|\texttt{1Q}| \approx \texttt{3Q}$
- $\texttt{Median} \approx 0$

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *  
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In dieser Tabelle stehen Informationen über die Kleinste-Quadrate-Schätzwerte.

Die Zeilennamen sind:

- `(Intercept)`, also β_0
- `wohnflaeche`, also der zur Variable `wohnflaeche` gehörige Koeffizient β_1
- `I(schlafzimmer > 4)TRUE`, also β_2
- `grundflaeche`, also β_3

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Spalte `Estimate` stehen die Schätzwerte, also:

- $\hat{\beta}_0 = \texttt{5.384e+04} = 5.384 \cdot 10^4 = 53840$
- $\hat{\beta}_1 = \texttt{1.134e+03} = 1134$
- $\hat{\beta}_2 = \texttt{1.049e+05} = 10490$
- $\hat{\beta}_3 = \texttt{2.019e+01} = 20.19$

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche              1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche             2.019e+01  6.253e+00   3.229  0.00177 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Spalte `Estimate` stehen die Schätzwerte, also:

- $\hat{\beta}_0 = \texttt{5.384e+04} = 5.384 \cdot 10^4 = 53840$
- $\hat{\beta}_1 = \texttt{1.134e+03} = 1134$
- $\hat{\beta}_2 = \texttt{1.049e+05} = 10490$
- $\hat{\beta}_3 = \texttt{2.019e+01} = 20.19$

Damit können wir jetzt interpretieren, dass ein zusätzlicher Quadratmeter Wohnfläche *ceteris paribus* (alles andere bleibt gleich) den Verkaufspreis um 1134 Dollar erhöht.

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *  
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Spalte `Std. Error` stehen die Standardfehler der geschätzten Koeffizienten (für deren Berechnung wir übrigens die Homoskedastie der Fehler u_i annehmen), also:

- $se(\hat{\beta}_0) = \texttt{2.339e+04} = 23390$
- $se(\hat{\beta}_1) = \texttt{1.272e+02} = 127.2$
- $se(\hat{\beta}_2) = \texttt{2.233e+04} = 22330$
- $se(\hat{\beta}_3) = \texttt{6.253e+00} = 6.253$

Hier gilt, dass ein kleinerer Standardfehler für einen präziseren Schätzwert steht.

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *  
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Spalte `t value` stehen die Teststatistiken mit denen wir testen, ob ein Regressor einen signifikanten Einfluss auf die zu erklärende Variable (den Preis) hat:

H_0: $\beta_k = 0$

H_1: $\beta_k \neq 0$

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)            5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche            1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche           2.019e+01  6.253e+00   3.229  0.00177 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In Spalte `t value` stehen die Teststatistiken mit denen wir testen, ob ein Regressor einen signifikanten Einfluss auf die zu erklärende Variable (den Preis) hat:

H_0: $\beta_k = 0$

H_1: $\beta_k \neq 0$

Für jeden Koeffizienten $k = 0, \ldots, 3$ wird dafür folgendes berechnet:

$$\texttt{t value} = t_k = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} = \frac{\texttt{Estimate}}{\texttt{Std. Error}}$$

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche              1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche             2.019e+01  6.253e+00   3.229  0.00177 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Unter der Annahme, dass die Fehler normalverteilt sind, gilt:

$$t_k = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \sim t(n - K - 1)$$

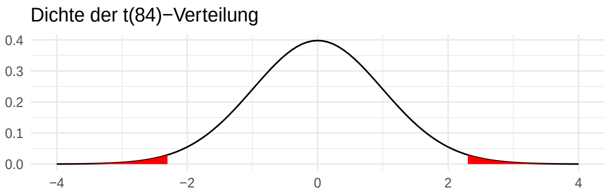
- $t(n-K-1)$ ist die t-Verteilung mit $n-K-1$ Freiheitsgraden
- $n = 88$ ist der Stichprobenumfang
- $K = 3$ ist die Anzahl an Regressoren (ohne Intercept!)

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche              1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche             2.019e+01  6.253e+00   3.229  0.00177 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In der Spalte `Pr(>|t|)` stehen die “p-values”. Das sind die Wahrscheinlichkeiten, mit denen wir unter der $t(n - K - 1)$-Verteilung einen betragsmäßig größeren Wert als t_k beobachten.

```
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)    
(Intercept)            5.384e+04  2.339e+04   2.302  0.02380 *  
wohnflaeche            1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche           2.019e+01  6.253e+00   3.229  0.00177 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In der Spalte `Pr(>|t|)` stehen die "p-values". Das sind die Wahrscheinlichkeiten, mit denen wir unter der $t(n-K-1)$-Verteilung einen betragsmäßig größeren Wert als t_k beobachten. Also zum Beispiel für den Intercept:



```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wie trifft man nochmal die Testentscheidung?

Wenn `p-value` $< \alpha$, verwerfe $H_0 : \beta_k = 0$.

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *  
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wie trifft man nochmal die Testentscheidung?

Wenn `p-value` $< \alpha$, verwerfe $H_0 : \beta_k = 0$.

Also ist der Intercept signifikant zu $\alpha = 5\%$, denn $0.02380 < 0.05$.

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              5.384e+04  2.339e+04   2.302  0.02380 *  
wohnflaeche              1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05  2.233e+04   4.699 1.01e-05 ***
grundflaeche             2.019e+01  6.253e+00   3.229  0.00177 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Wie trifft man nochmal die Testentscheidung?

Wenn `p-value` $< \alpha$, verwerfe $H_0 : \beta_k = 0$.

Also ist der Intercept signifikant zu $\alpha = 5\%$, denn $0.02380 < 0.05$.

Die Sterne in der rechten Spalte bieten einen schnellen Überblick:

- Koeffizienten mit `p-value` $< 10\%$ bekommen einen `.`
- `p-value` $< 5\%$ gibt `*`, `p-value` $< 1\%$ gibt `**`, `p-value` $< 0.1\%$ gibt `***`

```
Residual standard error: 53990 on 84 degrees of freedom
Multiple R-squared:  0.7333,    Adjusted R-squared:  0.7237
F-statistic: 76.97 on 3 and 84 DF,  p-value: < 2.2e-16
```

Am Ende stehen noch drei Zeilen mit viel Information:

1. ein Schätzwert für die Standardabweichung σ der Fehler u_i
2. die Gütekriterien R^2 und adjustiertes R^2
3. das Ergebnis des "Overall-F-Test"

```
Residual standard error: 53990 on 84 degrees of freedom
```

Der Schätzwert für die Standardabweichung σ der Fehler ist

$$\hat{\sigma} = \sqrt{SSR/(n - K - 1)}$$

```
Residual standard error: 53990 on 84 degrees of freedom
```

Der Schätzwert für die Standardabweichung σ der Fehler ist

$$\hat{\sigma} = \sqrt{SSR/(n - K - 1)}$$

wobei SSR (für "Sum of squared residuals") die Summe der quadrierten Residuen ist:

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i^2)$$

```
Residual standard error: 53990 on 84 degrees of freedom
```

Der Schätzwert für die Standardabweichung σ der Fehler ist

$$\hat{\sigma} = \sqrt{SSR/(n-K-1)}$$

wobei SSR (für "Sum of squared residuals") die Summe der quadrierten Residuen ist:

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i^2)$$

```
sqrt(sum(model$residuals^2)/84)
```

```
[1] 53986.52
```

(im `summary()` Output wurde gerundet, daher der kleine Unterschied)

```
Residual standard error: 53990 on 84 degrees of freedom
```

Zur Erinnerung: Neben SSR gibt es auch noch

- $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$ (für “Sum of squares total”) und
- $SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ (für “Sum of squares explained”).

```
Residual standard error: 53990 on 84 degrees of freedom
```

Zur Erinnerung: Neben SSR gibt es auch noch

- $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$ (für "Sum of squares total") und
- $SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ (für "Sum of squares explained").

Es gilt $SST = SSE + SSR$, das heißt, die Gesamtvariabilität des Datensatzes (SST) ist gleich der durch die Regression erklärten Variabilität (SSE) plus der unerklärten Variabilität (SSR).

```
Multiple R-squared:  0.7333,    Adjusted R-squared:  0.7237
```

`Multiple R-squared` ist das R^2 mit

$$R^2 = SSE/SST = 1 - SSR/SST,$$

das heißt die Größe R^2 gibt den Anteil der durch das Modell erklärten Variabilität an der Gesamtvariabilität im Datensatz an (hier $73,33\%$ - ganz gut!).

```
Multiple R-squared:  0.7333,    Adjusted R-squared:  0.7237
```

`Multiple R-squared` ist das R^2 mit

$$R^2 = SSE/SST = 1 - SSR/SST,$$

das heißt die Größe R^2 gibt den Anteil der durch das Modell erklärten Variabilität an der Gesamtvariabilität im Datensatz an (hier $73,33\%$ - ganz gut!).

Daneben steht das adjustierte R^2:

$$\bar{R}^2 = 1 - \frac{SSR/(n-K-1))}{SST/(n-1)}.$$

Was ist eigentlich der Unterschied?

Während R^2 zwingend mit jedem hinzugefügten Regressor steigt, tut das $\bar{R}^2$ nicht notwendigerweise. Also ist R^2 nicht zur Modellselektion geeignet, $\bar{R}^2$ aber schon.

```
F-statistic: 76.97 on 3 and 84 DF,  p-value: < 2.2e-16
```

Die letzte Zeile ist das Ergebnis des “Overall-F-Test” (Globaler F-Test). Er testet

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$

H_1: $\beta_1 \neq 0$ oder $\beta_2 \neq 0$ oder $\beta_3 \neq 0$

```
F-statistic: 76.97 on 3 and 84 DF,  p-value: < 2.2e-16
```

Die letzte Zeile ist das Ergebnis des “Overall-F-Test” (Globaler F-Test). Er testet

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$
H_1: $\beta_1 \neq 0$ oder $\beta_2 \neq 0$ oder $\beta_3 \neq 0$

Der Wert der Teststatistik wurde berechnet mit der Formel

$$\texttt{F-statistic} = F = \frac{R^2/K}{(1 - R^2)/(n - K - 1)}$$

```
r.squared <- summary(model)$r.squared
(r.squared/3)/((1 - r.squared)/84)
```

```
[1] 76.97402
```

```
F-statistic: 76.97 on 3 and 84 DF,  p-value: < 2.2e-16
```

Die letzte Zeile ist das Ergebnis des "Overall-F-Test" (Globaler F-Test). Er testet

H_0: $\beta_1 = \beta_2 = \beta_3 = 0$
H_1: $\beta_1 \neq 0$ oder $\beta_2 \neq 0$ oder $\beta_3 \neq 0$

Der Wert der Teststatistik wurde berechnet mit der Formel

$$\texttt{F-statistic} = F = \frac{R^2/K}{(1-R^2)/(n-K-1)}$$

```r
r.squared <- summary(model)$r.squared
(r.squared/3)/((1 - r.squared)/84)
```

```
[1] 76.97402
```

Die Teststatistik ist F-verteilt zu K und $n-K-1$ Freiheitsgraden (`DF`). Außerdem ist der `p-value` angegeben. Natürlich gilt auch hier: H_0 verwerfen, wenn `p-value` $< \alpha$.

```
Call:
lm(formula = preis ~ wohnflaeche + I(schlafzimmer > 4) + grundflaeche,
    data = hauspreise)

Residuals:
    Min      1Q  Median      3Q     Max
-149356  -27653   -3891   24175  202625

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)             5.384e+04  2.339e+04   2.302  0.02380 *
wohnflaeche             1.134e+03  1.272e+02   8.909 9.03e-14 ***
I(schlafzimmer > 4)TRUE 1.049e+05              4.699 1.01e-05 ***
grundflaeche            2.019e+01  6.253e+00   3.229  0.00177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 53990 on 84 degrees of freedom
Multiple R-squared:  0.7333,    Adjusted R-squared:  0.7237
F-statistic: 76.97 on    and    DF,  p-value: < 2.2e-16
```

Dein Selbsttest:

1. Was ist $se(\hat{\beta}_2)$?
2. Wie viele Signifikanzsterne hat $\hat{\beta}_3$?
3. Wie groß ist n?
4. Wie groß ist SSE?
5. Welche Freiheitsgrade hat die F-Statistik?