# Lösungen zur Übung 7

**1.** b)
```
> mod1_lm(W~V+I(V^2)-1,data=bremsweg)
> summary(mod1, corr=F)

Call: lm(formula = W ~ V + V^2 - 1, data = bremsweg)
Residuals:
   Min     1Q Median    3Q   Max
 -22.3 -5.223 -0.259 4.198 27.77

Coefficients:
        Value Std. Error t value Pr(>|t|)
     V 0.5766 0.2008     2.8715  0.0056
I(V^2) 0.0621 0.0069     9.0011  0.0000

Residual standard error: 9.852 on 60 degrees of freedom
Multiple R-Squared: 0.9644
F-statistic: 813.5 on 2 and 60 degrees of freedom, the p-value is 0
```

$\hat{\beta}_1 = 0.5766,\ \hat{\beta}_2 = 0.0621,\ \hat{\sigma}^2 = 9.852^2 = 97.06$

$R^2 = 0.9644$, also sehr gross.

c) Modellannahmen sind verletzt: Residuen nehmen mit zunehmendem $\hat{y}$ zu.

d)
```
> mod2 <- lm(W ~ V + I(V^2) - 1,weights=1/V^2, data = bremsweg)
> summary(mod2, corr=F)

Call: lm(formula = W ~ V + I(V^2) - 1, data = bremsweg, weights = 1/V^2)

Residuals:
     Min      1Q   Median    3Q   Max
 -0.8124 -0.3596 -0.03149 0.314 0.939

Coefficients:
         Value Std. Error t value Pr(>|t|)
     V  0.6565  0.1234     5.3184  0.0000
I(V^2)  0.0590  0.0058    10.2055  0.0000

Residual standard error: 0.4497 on 60 degrees of freedom
Multiple R-Squared: 0.9468
F-statistic: 534.3 on 2 and 60 degrees of freedom, the p-value is 0
```

e)
```
> mod3 <- lm(W ~ V + I(V^2) - 1,weights=1/V, data = bremsweg)
> summary(mod3, corr=F)

Call: lm(formula = W ~ V + I(V^2) - 1, data = bremsweg, weights = 1/V)
Residuals:
```

```
   Min     1Q  Median    3Q   Max
 -4.07 -1.428 -0.1057 1.374 5.087

Coefficients:
        Value Std. Error t value Pr(>|t|)
     V 0.6145 0.1589     3.8669  0.0003
I(V^2) 0.0608 0.0061     9.8982  0.0000

Residual standard error: 1.997 on 60 degrees of freedom
Multiple R-Squared: 0.958
F-statistic: 683.6 on 2 and 60 degrees of freedom, the p-value is 0
```

f)

$$W_i = \beta_1 v_i + \beta_2 v_i^2 + \epsilon_i \Longleftrightarrow Z_i = \beta_1 + \beta_2 v_i + \epsilon_i' \text{ mit } Z_i = W_i/v_i \text{ und } \epsilon_i' = \epsilon_i/v_i$$

$$Var(\epsilon_i') = \frac{1}{v_i^2} Var\epsilon_i \approx \sigma^2$$

```
> mod4 <- lm(Z~ V , data = bremsweg)
> summary(mod4, corr=F)

Call: lm(formula = W/V ~ V, data = bremsweg)
Residuals:
     Min      1Q   Median    3Q   Max
 -0.8124 -0.3596 -0.03149 0.314 0.939

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  0.6565  0.1234     5.3184  0.0000
          V  0.0590  0.0058    10.2055  0.0000

Residual standard error: 0.4497 on 60 degrees of freedom
Multiple R-Squared: 0.6345
F-statistic: 104.2 on 1 and 60 degrees of freedom, the p-value is
9.881e-015
```

| | ungewichtet | $1/v_i^2$ | $1/v_i$ | Modell für $Z_i$ |
|---|---|---|---|---|
| $\hat{\beta}_1$ | 0.5766 | 0.6565 | 0.6145 | 0.6565 |
| $\hat{\beta}_2$ | 0.0621 | 0.0590 | 0.0608 | 0.0590 |
| $\hat{\sigma}$ | 9.852 | 0.4497 | 1.997 | 0.4497 |
| $R^2$ | 0.9644 | 0.9468 | 0.958 | 0.6345 |

Residuenanalyse: die Residuen der gewichteten Regressionen müssen korrigiert werden:

$r_i' = r_i \cdot \sqrt{\text{Gewicht}}$

Die Plots der mit $1/v_i^2$ gewichteten Regression und der Regression für $Z_i$ sehen gleich aus. $R^2$ ist in der Regression für $Z_i$ viel kleiner, weil die Gesamtvariabilität beim Übergang von $W_i$ zu $Z_i$ abgenommen hat.

**2.** a)
```
> mod1_lm(y~x,data=chicago)
> summary(mod1, corr=F)

Call: lm(formula = y ~ x, data = chicago)
```

```
Residuals:
    Min      1Q  Median    3Q   Max
 -10.79 -3.309 -0.7362 3.202 18.23

Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept) 6.6022 5.3335     1.2379  0.2251
          x 1.0056 0.1743     5.7677  0.0000

Residual standard error: 5.994 on 31 degrees of freedom
Multiple R-Squared: 0.5176
F-statistic: 33.27 on 1 and 31 degrees of freedom, the p-value is
2.38e-006
```

Beobachtung Nr. 33 hat ein grosses Residuum.

b) Nichts Besonderes.

```
c) > mod2_lm(y~x,weights=n,data=chicago)
   > summary(mod2, corr=F)

Call: lm(formula = y ~ x, data = chicago, weights = n)
Residuals:
    Min      1Q Median    3Q   Max
 -21.58 -9.117 -1.734 4.224 40.26

Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept) 3.3390 5.6339     0.5927  0.5577
          x 1.1219 0.1812     6.1899  0.0000

Residual standard error: 12.31 on 31 degrees of freedom
Multiple R-Squared: 0.5528
F-statistic: 38.31 on 1 and 31 degrees of freedom, the p-value is
7.157e-007
```

Beobachtung Nr. 29 ist einflussreich und ein Ausreisser ($n = 17$).

```
d) > mod3_lm(y~x,weights=n,data=chicago[-29,])
   > summary(mod3, corr=F)

Call: lm(formula = y ~ x, data = chicago[-29,  ], weights = n)
Residuals:
    Min      1Q  Median    3Q   Max
 -15.59 -5.928 -0.4955 5.032 42.97

Coefficients:
             Value Std. Error t value Pr(>|t|)
(Intercept) 8.0810 5.1341     1.5740  0.1260
          x 0.9235 0.1697     5.4418  0.0000

Residual standard error: 10.75 on 30 degrees of freedom
Multiple R-Squared: 0.4968
```

```
F-statistic: 29.61 on 1 and 30 degrees of freedom, the p-value is
6.691e-006
```

```
e) > chicago_cbind(chicago,rep(0,33))
   > names(chicago)[4]_"ch"
   > chicago$ch[c(2,23,29)]_1
   > mod4_lm(y~x+ch,weights=n,data=chicago[-33,])
   > summary(mod4, corr=F)

Call: lm(formula = y ~ x + ch, data = chicago[-33,  ], weights = n)
Residuals:
    Min     1Q Median    3Q   Max
 -12.26 -4.844 0.7599 5.362 12.33

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) 10.1691  3.5991     2.8254  0.0085
          x  0.8119  0.1212     6.7012  0.0000
         ch 10.4878  1.9368     5.4150  0.0000

Residual standard error: 7.179 on 29 degrees of freedom
Multiple R-Squared: 0.8323
F-statistic: 71.94 on 2 and 29 degrees of freedom, the p-value is
5.722e-012
```