Task p. 40-1 / Anwendung S. 40-1

FK

automatic

Working directory

```
> setwd("D:/kronthafranz/Documents/01Lehre/06Quantitative Forschungsmethoden
dt en")
```

Load data

```
> load("D:/kronthafranz/Documents/01Lehre/06Quantitative Forschungsmethoden
dt en/06Regression/reg_country.RData")
```

Regression function

$$\hat{y} = b_0 + b_1 \times urban + b_2 \times doc + b_3 \times bed + b_4 \times gdp + b_5 \times rad$$

or

$$y = \beta_0 + \beta_1 \times urban + \beta_2 \times doc + \beta_3 \times bed + \beta_4 \times gdp + \beta_5 \times rad + u$$

Regression function

Descriptive statistics

```
> numSummary(data_country_2[,c("bed", "doc", "expect", "gdp", "rad",
+    "urban")], statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,
+    .75,1))
```

	mean	sd	0%	25%	50%	75%
bed	34.87462	32.40236	2.5252525	11.887151	22.05074	50.31470
doc	10.52123	11.10770	0.1880017	1.185255	6.30517	16.66667
expect	66.31148	11.28524	41.0000000	56.000000	68.00000	76.00000
gdp	4157.71311	6113.61846	120.0000000	400.000000	1110.00000	4375.00000
rad	31.18610	30.01206	1.5625000	11.745690	21.27660	40.00000
urban	48.77623	24.62484	5.0000000	28.250000	48.00000	69.50000

	100%	n	NA
bed	135.13514	116	6
doc	42.91845	121	1
expect	83.00000	122	0
gdp	22470.00000	122	0
rad	200.00000	122	0
urban	100.00000	122	0

Scatterplot

```r
> scatterplotMatrix(~bed+doc+expect+gdp+rad+urban, reg.line=FALSE,
+    smooth=FALSE, spread=FALSE, span=0.5, ellipse=FALSE, levels=c(.5, .9),
+    id.n=0, diagonal = 'none', data=data_country_2)
```
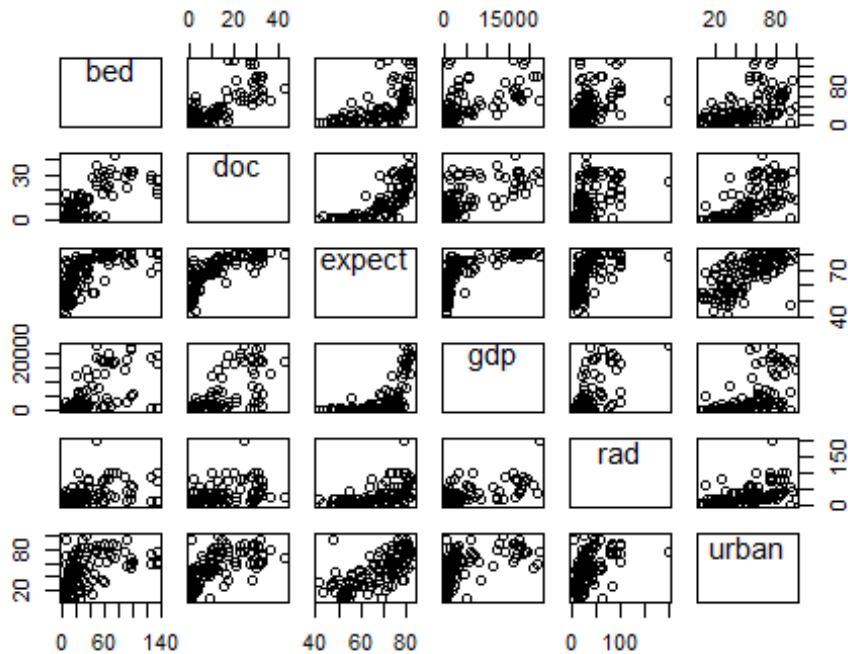

Correlation coefficients

```r
> cor(data_country_2[,c("bed","doc","expect","gdp","rad","urban")],
+    use="complete")
```

```
               bed       doc    expect       gdp       rad     urban
bed      1.0000000 0.7704197 0.6251971 0.6499957 0.4986648 0.5005619
doc      0.7704197 1.0000000 0.7818952 0.7142052 0.5280187 0.6842779
expect   0.6251971 0.7818952 1.0000000 0.6679829 0.5636072 0.6966961
gdp      0.6499957 0.7142052 0.6679829 1.0000000 0.6620972 0.6104549
rad      0.4986648 0.5280187 0.5636072 0.6620972 1.0000000 0.5339263
urban    0.5005619 0.6842779 0.6966961 0.6104549 0.5339263 1.0000000
```

Regressiona analysis

```r
> RegModel.1 <- lm(expect~bed+doc+gdp+rad+urban, data=data_country_2)

> summary(RegModel.1)


Call:
lm(formula = expect ~ bed + doc + gdp + rad + urban, data = data_country_2)
```

```
Residuals:
     Min       1Q   Median       3Q      Max
-17.5811  -3.6979  -0.3803   4.3983  15.6873

Coefficients:
             Estimate Std. Error t value  Pr(>|t|)
(Intercept) 53.9054353  1.4826881  36.357   < 2e-16 ***
bed          0.0085250  0.0302988   0.281   0.77896
doc          0.4646609  0.1039961   4.468 0.0000192 ***
gdp          0.0001668  0.0001616   1.032   0.30415
rad          0.0402463  0.0272514   1.477   0.14257
urban        0.1112802  0.0351610   3.165   0.00201 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.466 on 110 degrees of freedom
  (6 observations deleted due to missingness)
Multiple R-squared:  0.6788,    Adjusted R-squared:  0.6642
F-statistic: 46.49 on 5 and 110 DF,  p-value: < 2.2e-16
```

Interpretation:

F statistics: p-value is smaller than 0.01, we reject H0 for the model

R2 is 67.8%

We reject H0 for the variables doc and urban at the 1% significance level