Beispiele zum linearen Modell

Rang und Geschlechterverhältnis bei Hirschkühen

K. Pearsons Größen von Vätern und Söhnen

Prostata-Krebs-Daten von Stamey et al.

Ein Beispiel zur Varianzanalyse

Rang und Geschlechterverhältnis bei Hirschkühen

Frage: Hängt der Anteil männlicher Nachkommen einer Hirschkuh mit ihrem sozialen Rang zusammen?

Folgender Artikel berichtet über eine Langzeitstudie, bei der eine Gruppe Rothirsche auf der schottischen Insel Rùm über 15 Jahre beobachtet wurde, und die zu obiger Frage Daten gesammelt hat:

- Clutton-Brock, T. H. , Albon, S. D., Guinness, F. E. (1986) Great expectations: dominance, breeding success and offspring sex ratios in red deer. *Anim. Behav.* **34**, 460—471.

	Rang	Anteil männl. Nachkommen
1	0.01	0.41
2	0.02	0.15
3	0.06	0.12
4	0.08	0.04
5	0.08	0.33
6	0.09	0.37
.	.	.
.	.	.
.	.	.
52	0.96	0.81
53	0.99	0.47
54	1.00	0.67

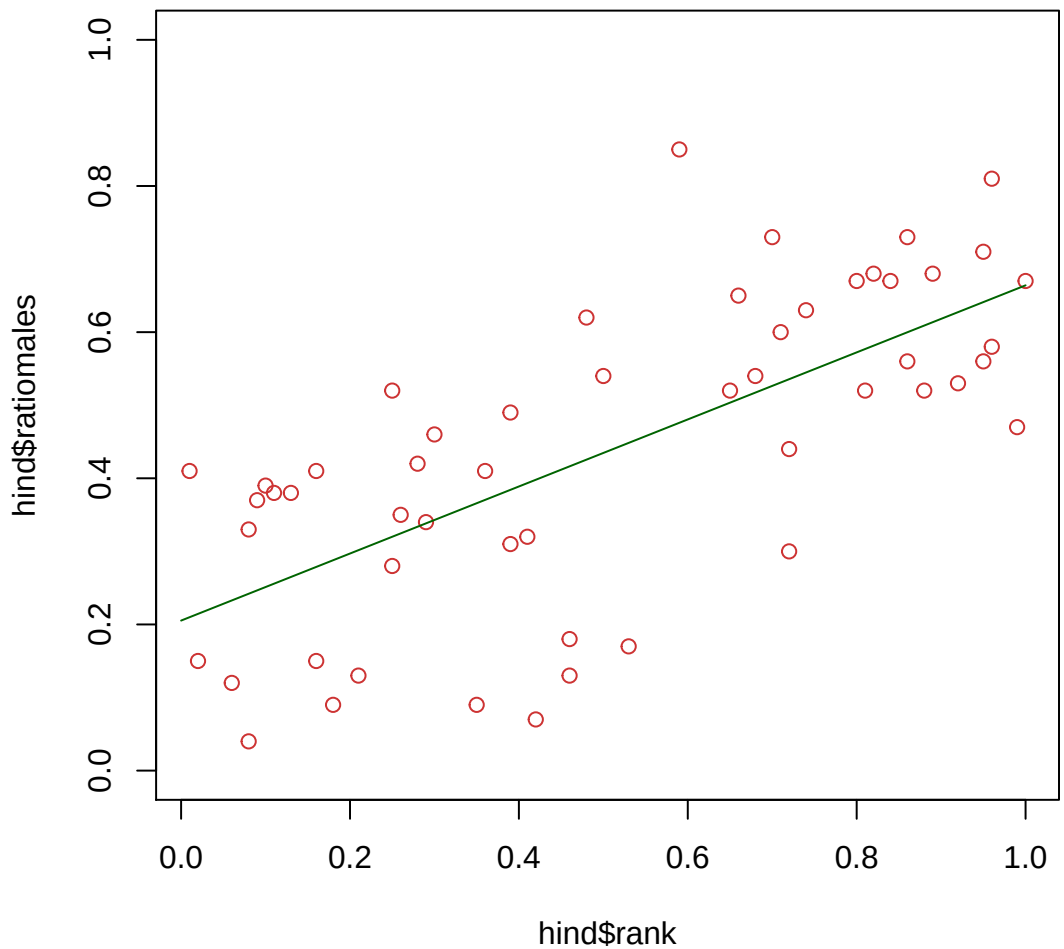
Die Beobachtungen:

Für 54 Hirschkühe wurde der Rang (normiert auf einen Wert in $[0, 1]$, grob gesprochen ein Schätzwert für die Wahrscheinlichkeit, dass die betreffende Kuh ein „Duell“ mit einer zufällig ausgewählten anderen Kuh „gewinnt“) und der Anteil m"annlicher Nachkommen beobachtet.

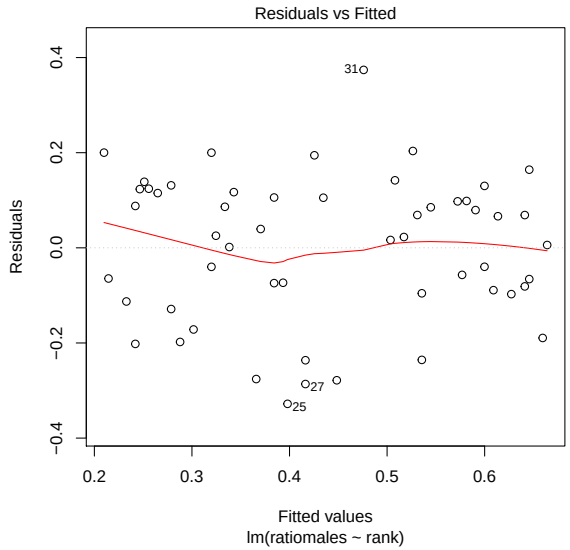
(Simulierte Daten, die sich an den Werten aus der Originalpublikation orientieren.)

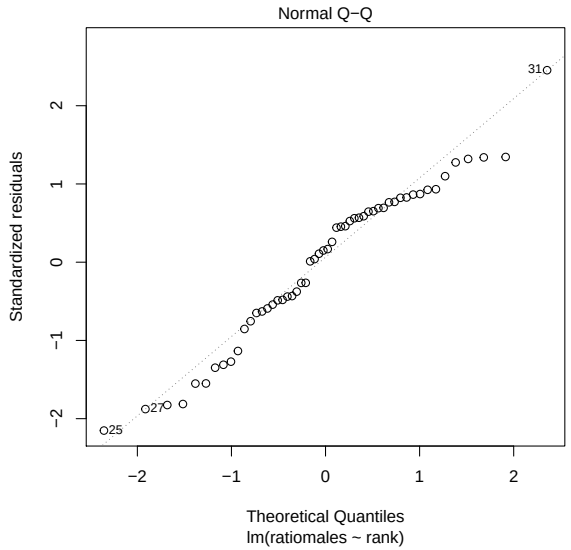
```
mod <- lm(ratiomales~rank,data=hind)
summary(mod)

Call:
lm(formula = ratiomales ~ rank, data = hind)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.32798 -0.09396  0.02408  0.11275  0.37403 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.20529    0.04011   5.119 4.54e-06 ***
rank         0.45877    0.06732   6.814 9.78e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 

Residual standard error: 0.154 on 52 degrees of freedom
Multiple R-squared:  0.4717, Adjusted R-squared:  0.4616 
F-statistic: 46.44 on 1 and 52 DF,  p-value: 9.781e-09
```

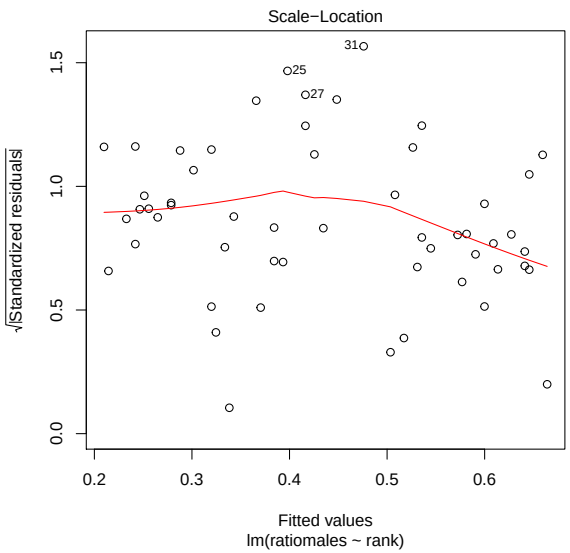
```
plot(hind$rank,hind$ratiomales,ylim=c(0,1))
abline(mod$coeff, col="darkgreen")
```

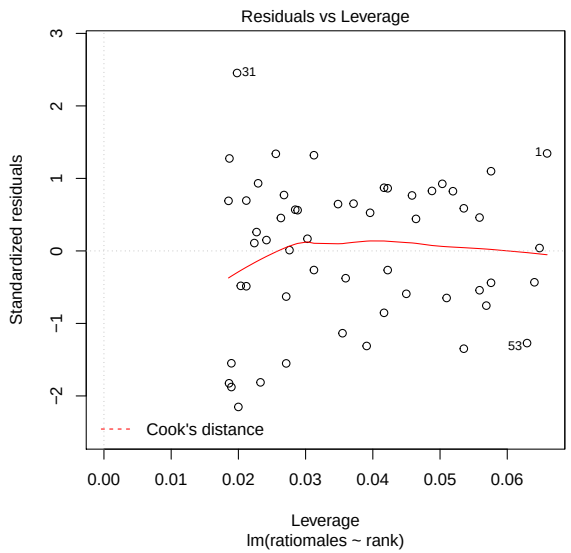


```
plot(mod)
```


Residuals vs Fitted
Residuals
Fitted values
lm(ratiomales ~ rank)
31
27
25


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(ratiomales ~ rank)
31
27
25


Scale−Location
√|Standardized residuals|
Fitted values
lm(ratiomales ~ rank)
31
25
27

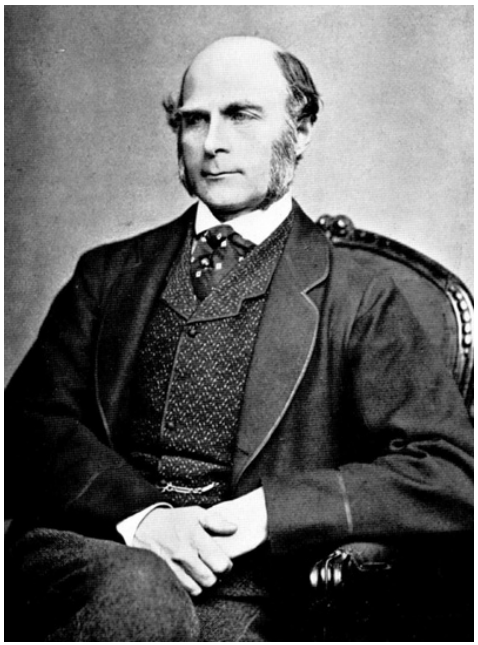

Residuals vs Leverage
Standardized residuals
Leverage
lm(ratiomales ~ rank)
Cook's distance
31
1
53

K. Pearsons Größen von Vätern und Söhnen

Woher kommt der Name „Regression“ (nach lat. regressio, Zurückkommen)?

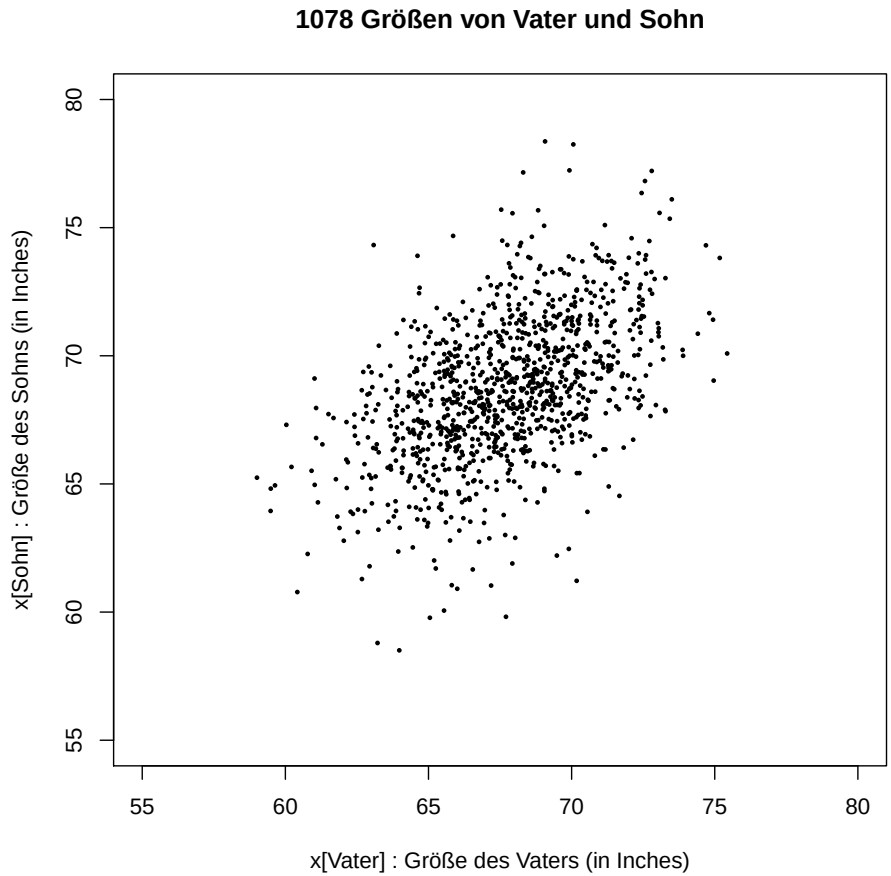
K. Pearsons Größen von Vätern und Söhnen

Woher kommt der Name „Regression“ (nach lat. regressio, Zurückkommen)?



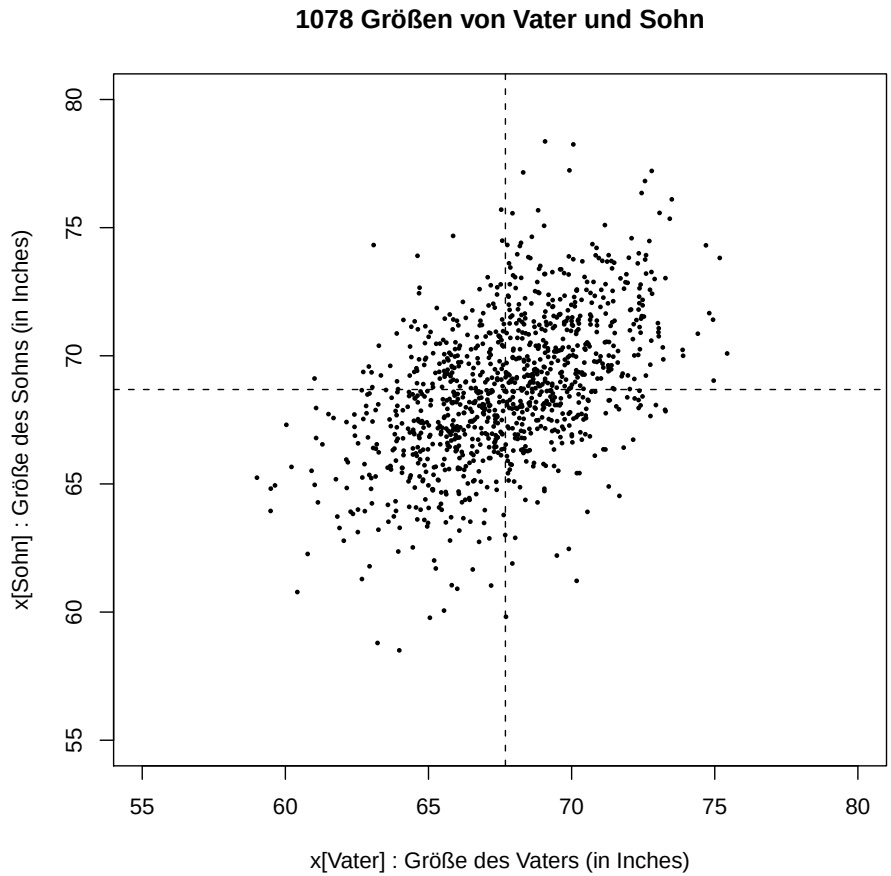
Francis Galton (1822–1911, engl. Wissenschaftler)
hat angesichts biometrischer Beobachtungen den Ausdruck “regression towards the mean” geprägt.

Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)



[1] siehe `data(father.son)` aus dem `R`-Paket `UsingR`

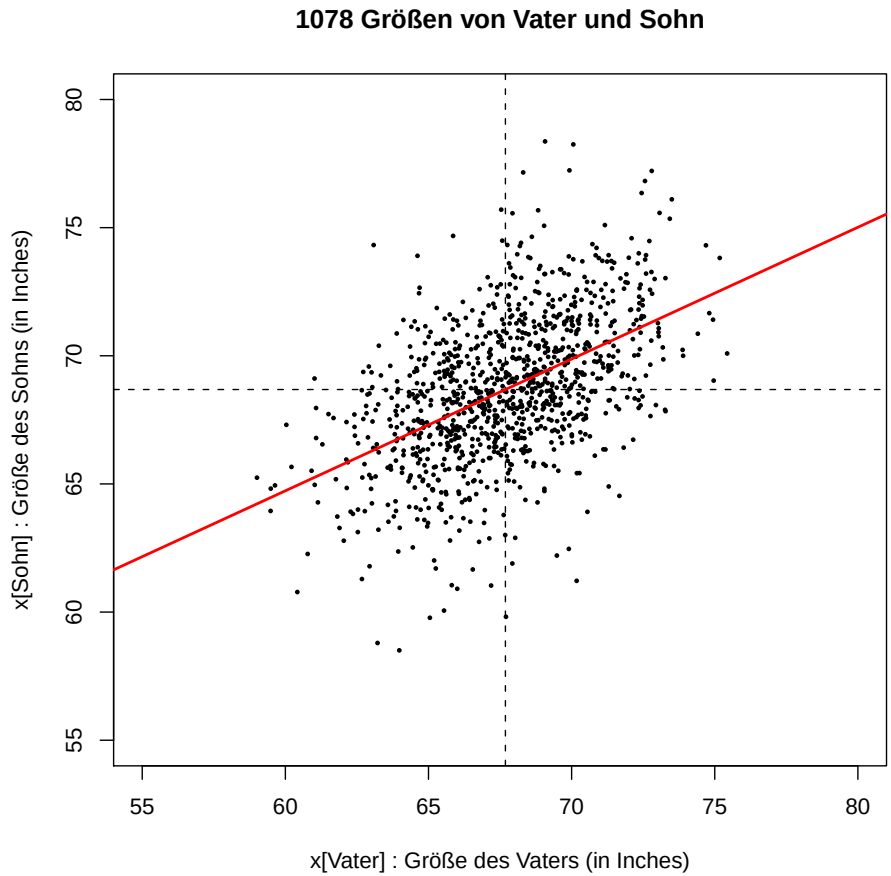
Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)



$\bar{x}_{\text{Vater}} = 67.7$, $\bar{x}_{\text{Sohn}} = 68.7$, $\sigma^2_{\text{Vater}} = 7.52$ ($\sigma_{\text{Vater}} = 2.74$, $\sigma_{\text{Sohn}} = 2.81$), $\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$
(Korrelationskoeffizient $\rho = \text{cov}(x_{\text{Vater}}, x_{\text{Sohn}})/(\sigma_{\text{Vater}}\,\sigma_{\text{Sohn}})) = 0.50$)

[1] siehe `data(father.son)` aus dem `R`-Paket `UsingR`

Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)

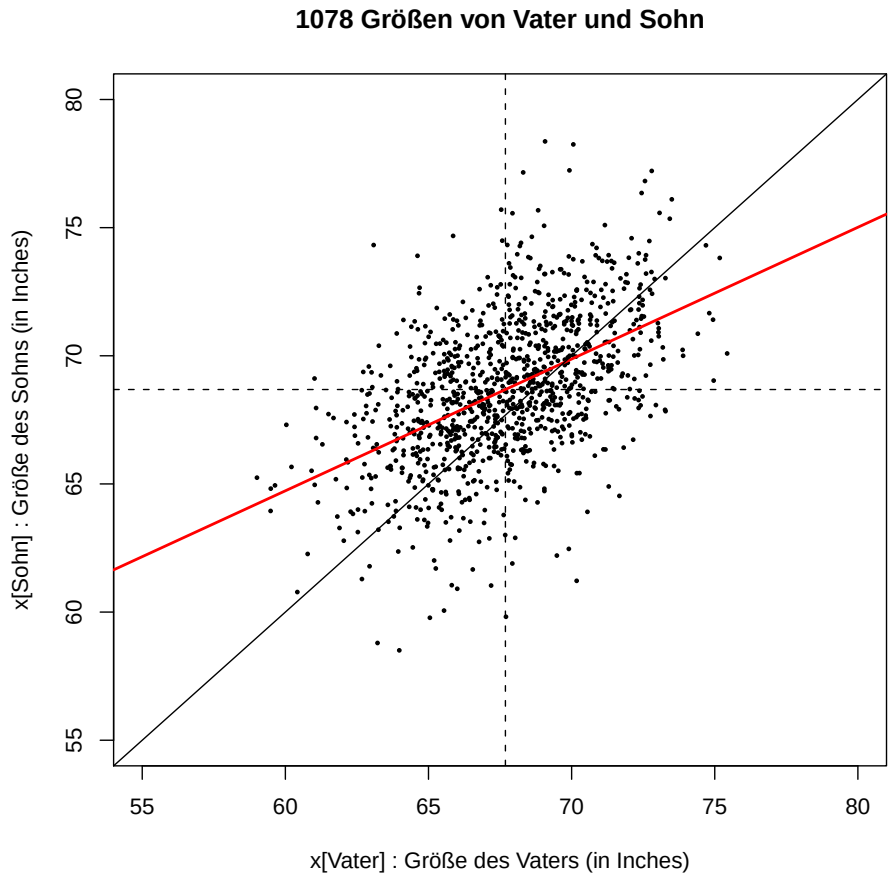


$\bar{x}_{\text{Vater}} = 67.7$, $\bar{x}_{\text{Sohn}} = 68.7$, $\sigma^2_{\text{Vater}} = 7.52$ ($\sigma_{\text{Vater}} = 2.74$, $\sigma_{\text{Sohn}} = 2.81$), $\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$
(Korrelationskoeffizient $\rho = \text{cov}(x_{\text{Vater}}, x_{\text{Sohn}})/(\sigma_{\text{Vater}}\,\sigma_{\text{Sohn}})) = 0.50$)

Regressionsgerade: $x_{\text{Sohn}} = 33.89 + 0.514 x_{\text{Vater}}$.

[1] siehe `data(father.son)` aus dem R-Paket `UsingR`

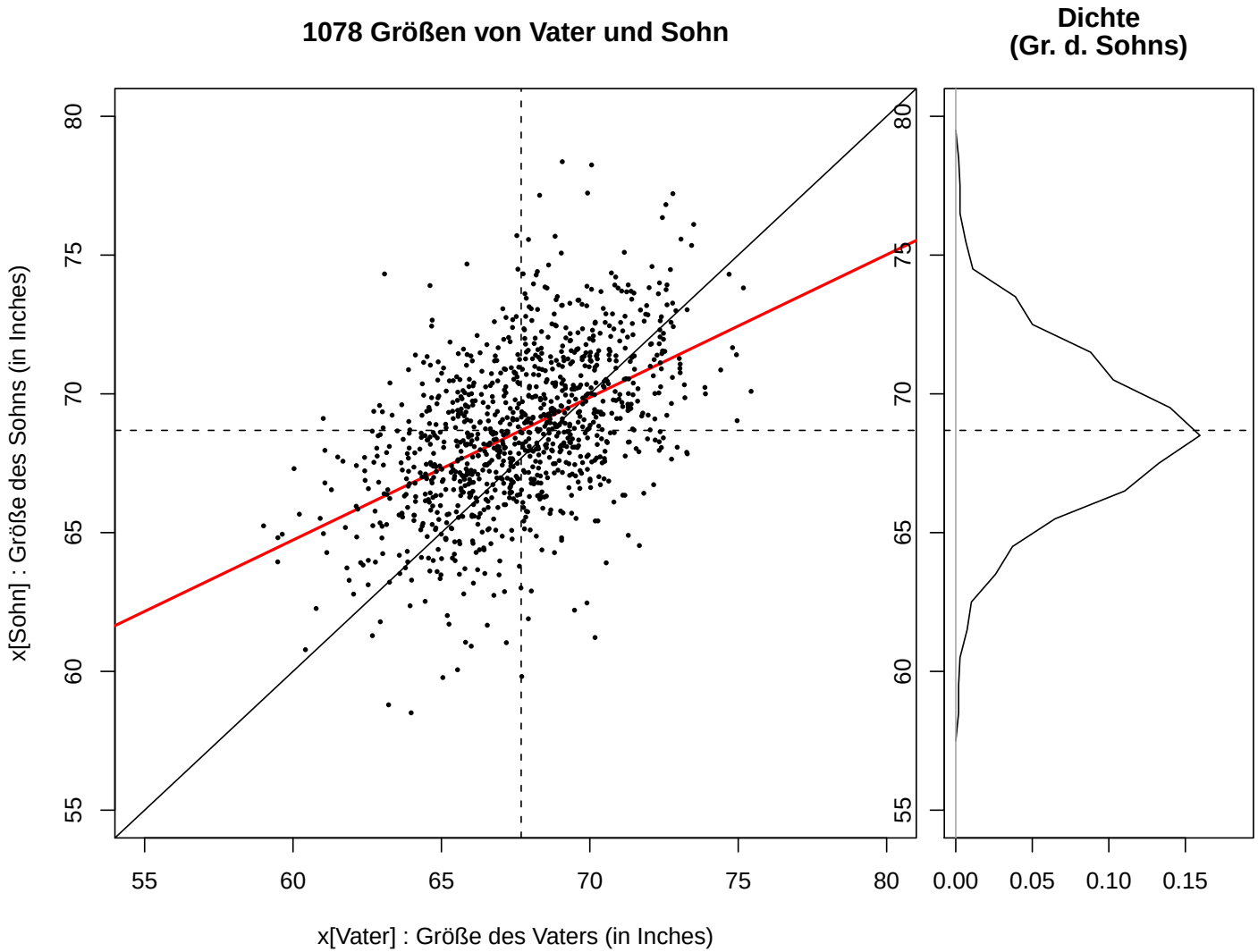
Ein (relativ berühmter) Datensatz[1] von Karl Pearson (1858-1936)

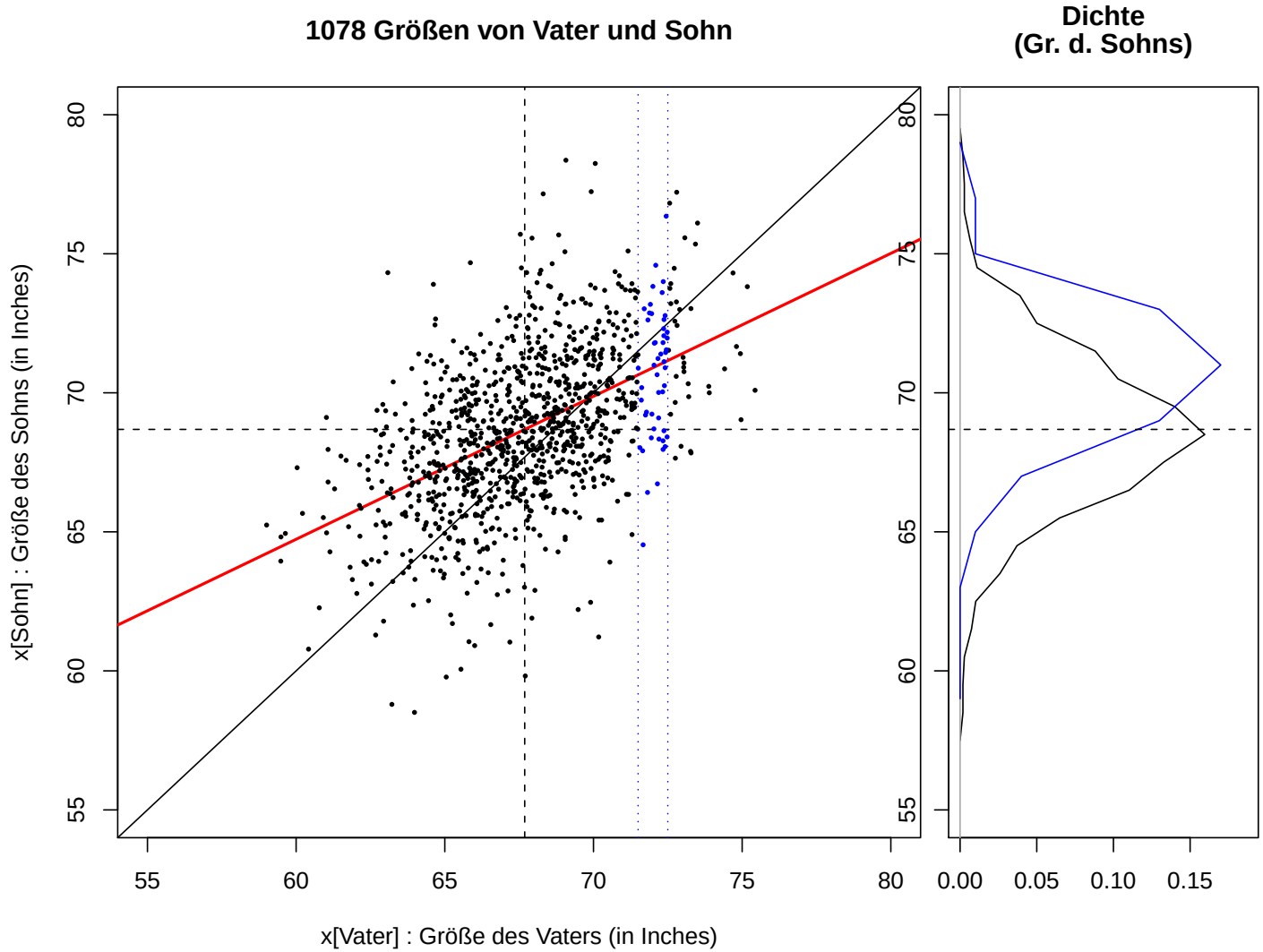


$\bar{x}_{\text{Vater}} = 67.7$, $\bar{x}_{\text{Sohn}} = 68.7$, $\sigma^2_{\text{Vater}} = 7.52$ ($\sigma_{\text{Vater}} = 2.74$, $\sigma_{\text{Sohn}} = 2.81$), $\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$
(Korrelationskoeffizient $\rho = \text{cov}(x_{\text{Vater}}, x_{\text{Sohn}})/(\sigma_{\text{Vater}}\,\sigma_{\text{Sohn}})) = 0.50$)

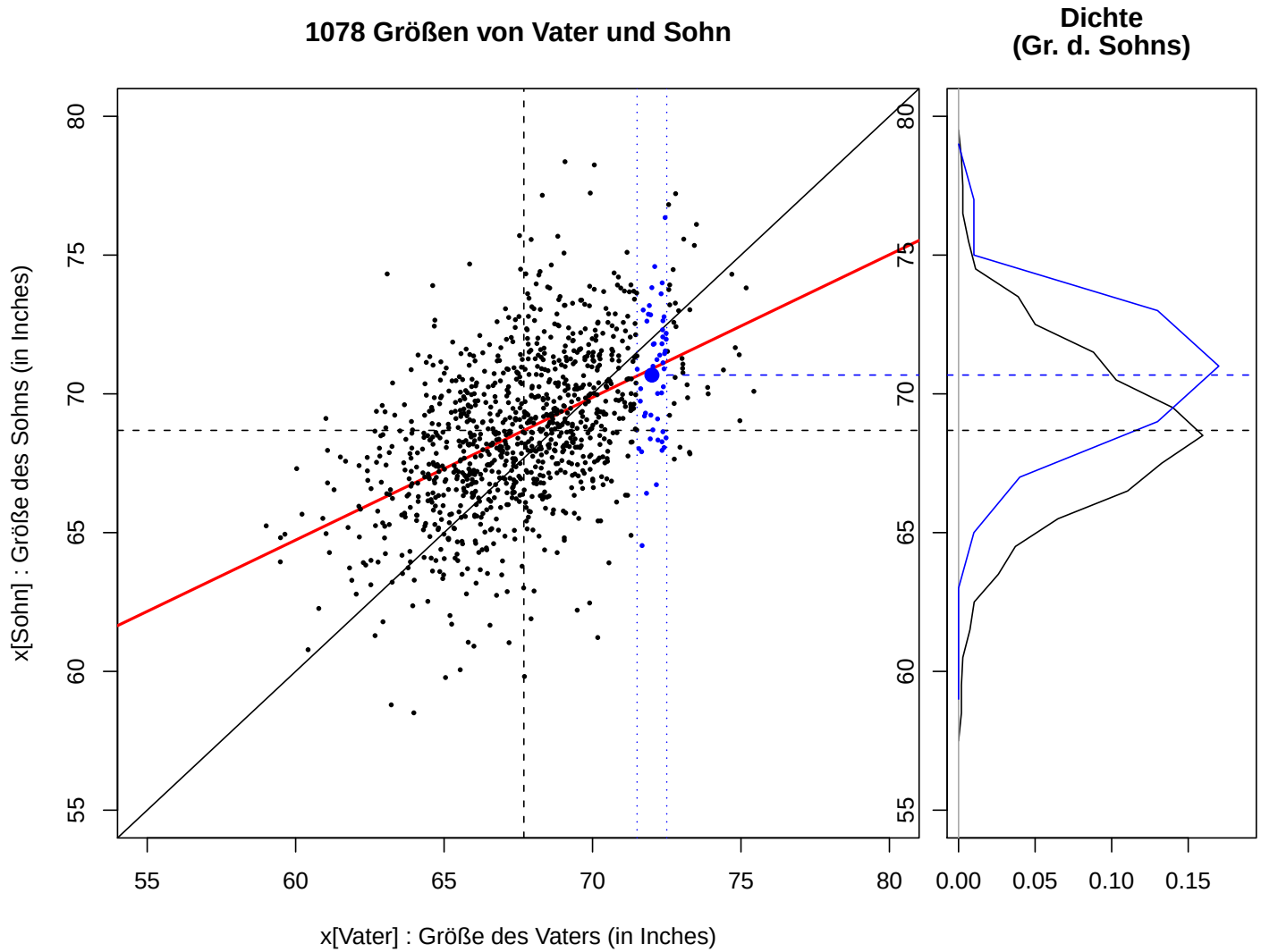
Regressionsgerade: $x_{\text{Sohn}} = 33.89 + 0.514 x_{\text{Vater}}$.

[1] siehe `data(father.son)` aus dem R-Paket `UsingR`

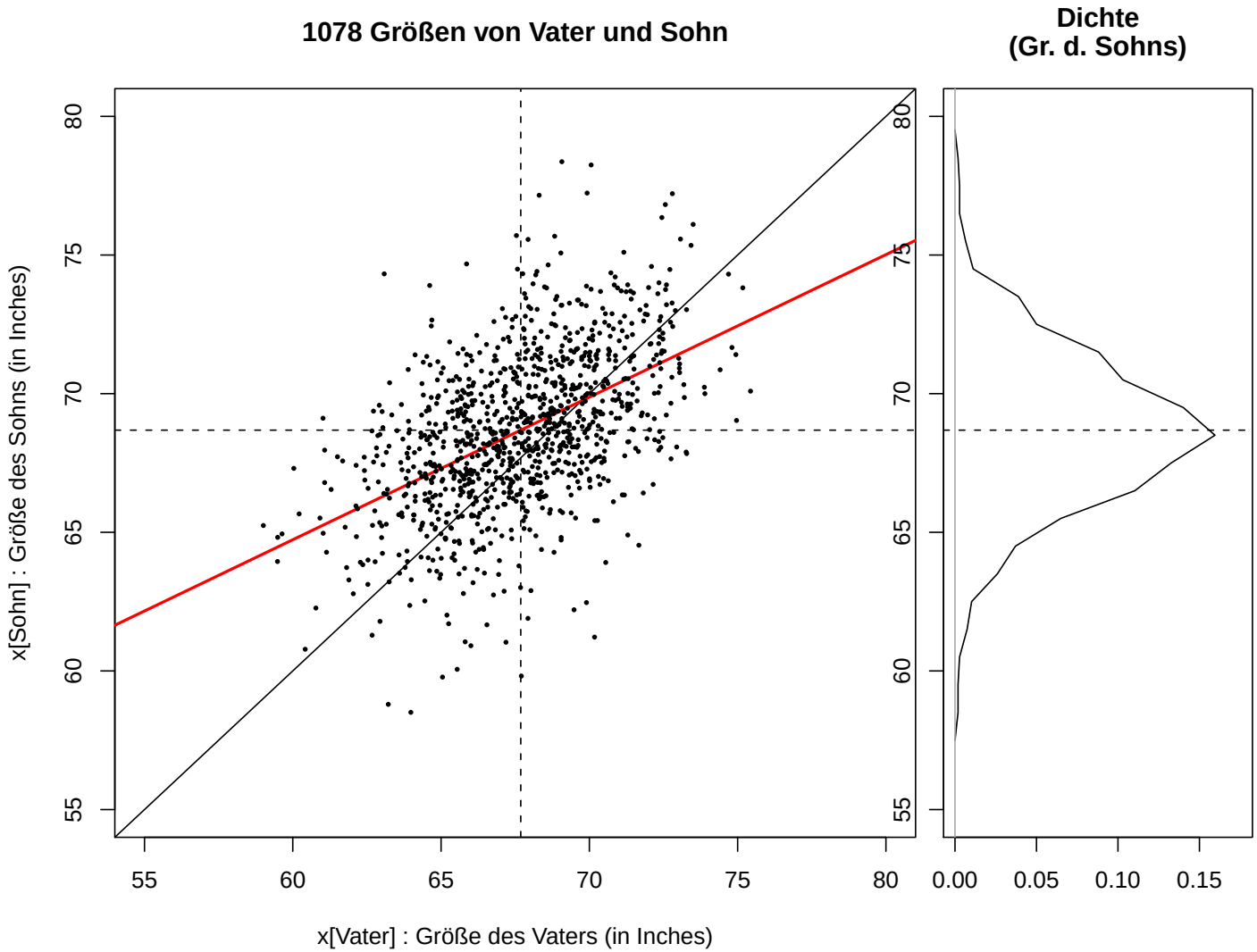

1078 Größen von Vater und Sohn
Dichte
(Gr. d. Sohns)
x[Sohn] : Größe des Sohns (in Inches)
x[Vater] : Größe des Vaters (in Inches)
55
60
65
70
75
80
0.00
0.05
0.10
0.15

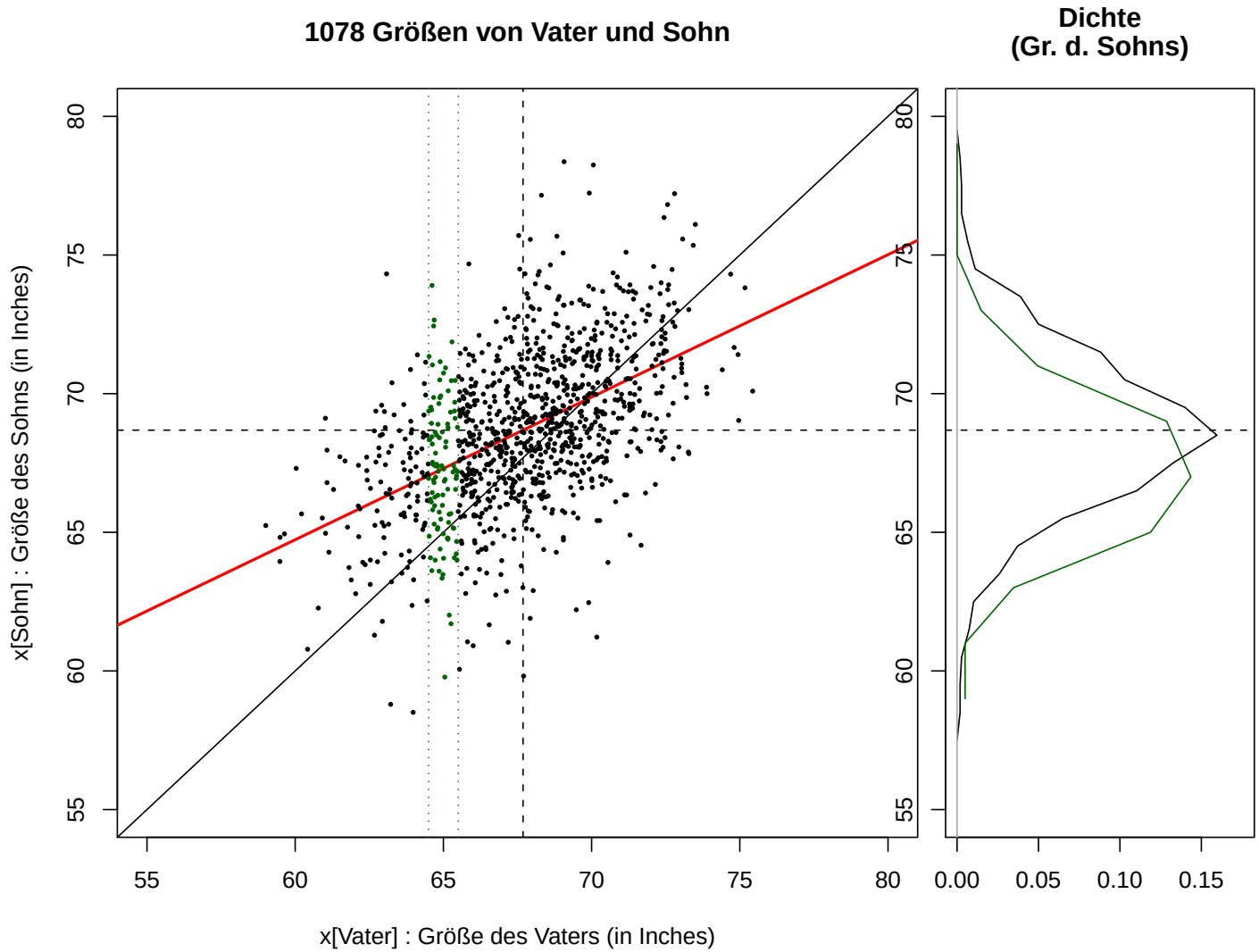


Betrachten wir die Söhne von überdurchschnittlich großen Vätern (z.B. Väter, die ca. 72 Inches groß sind):

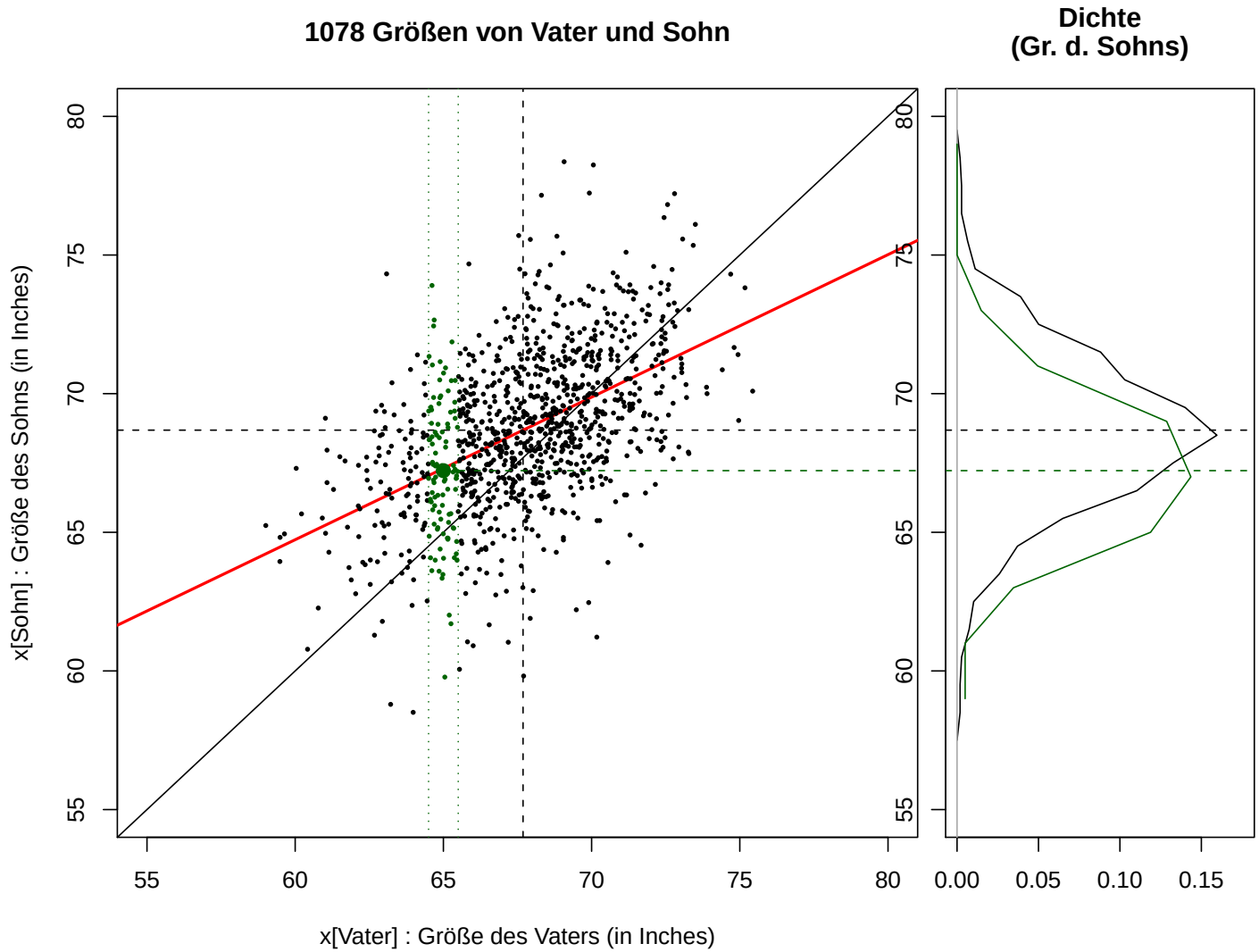


Betrachten wir die Söhne von überdurchschnittlich großen Vätern
(z.B. Väter, die ca. 72 Inches groß sind):
Diese Söhne sind überdurchschnittlich groß (im Vergleich zu allen Söhnen), aber im Mittel kleiner als ihr Vater.


1078 Größen von Vater und Sohn
Dichte
(Gr. d. Sohns)
x[Sohn] : Größe des Sohns (in Inches)
x[Vater] : Größe des Vaters (in Inches)
55
60
65
70
75
80
0.00
0.05
0.10
0.15

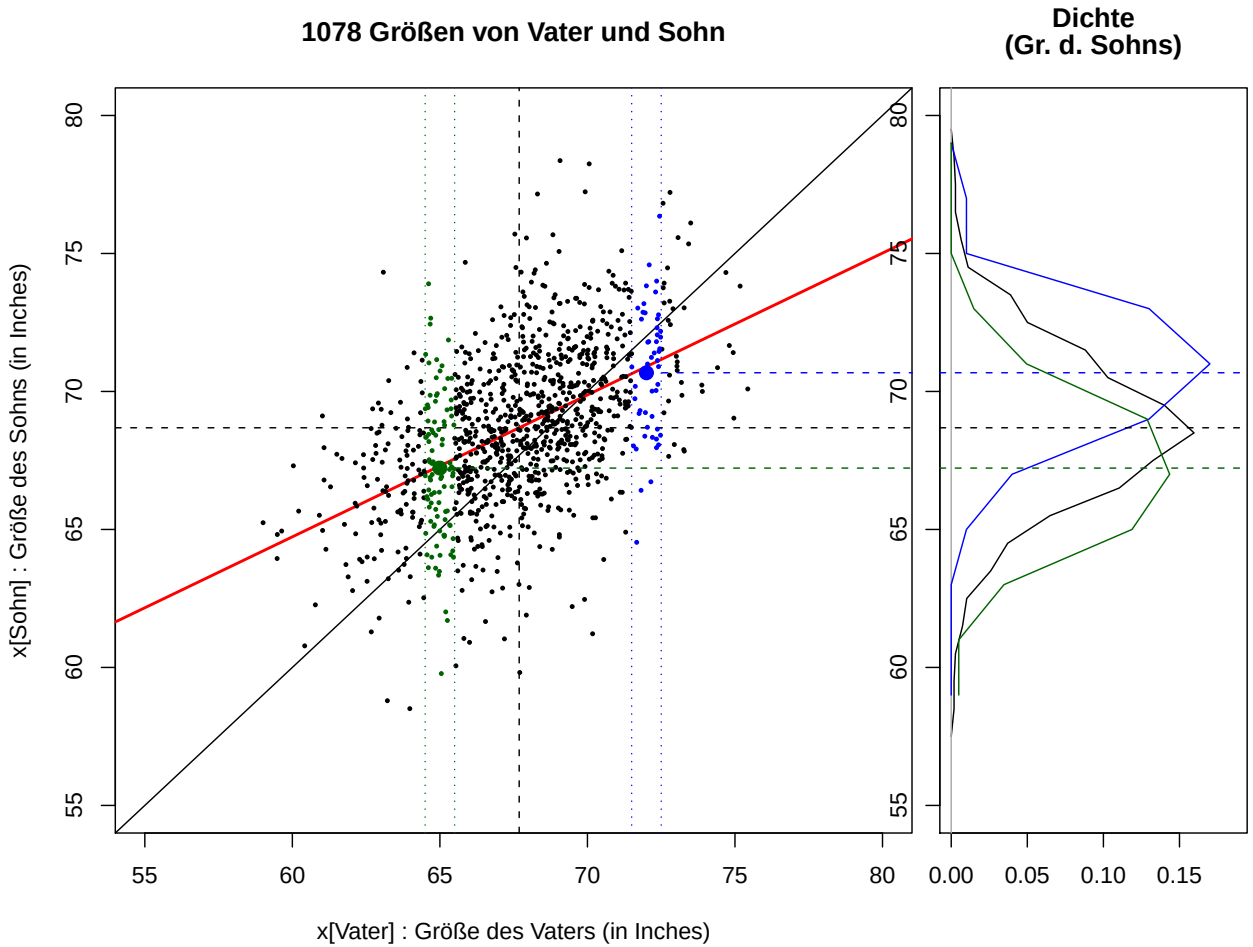


Betrachten wir andererseits die Söhne von unterdurchschnittlich großen Vätern (z.B. Väter, die ca. 65 Inches groß sind):



Betrachten wir andererseits die Söhne von unterdurchschnittlich großen Vätern (z.B. Väter, die ca. 65 Inches groß sind):
Diese Söhne sind unterdurchschnittlich groß (im Vergleich zu allen Söhnen), aber im Mittel größer als ihr Vater.

“Regression towards the mean”



Wir sehen: Söhne überdurchschnittlich großer Väter sind im Mittel kleiner als ihr Vater (aber immer noch größer als der Populationsdurchschnitt), für Söhne unterdurchschnittlich großer Väter ist es umgekehrt: „Rückkehr zum Mittelwert“.

Bemerkung: Das beobachtete Phänomen der „Rückkehr zum Mittelwert“ bedeutet nicht notwendigerweise einen tieferen kausalen Zusammenhang, es tritt stets im Zusammenhang mit natürlicher Variabilität auf (technisch gesehen stets, wenn für den Korrelationkoeffizient ρ gilt $|\rho| < 1$).

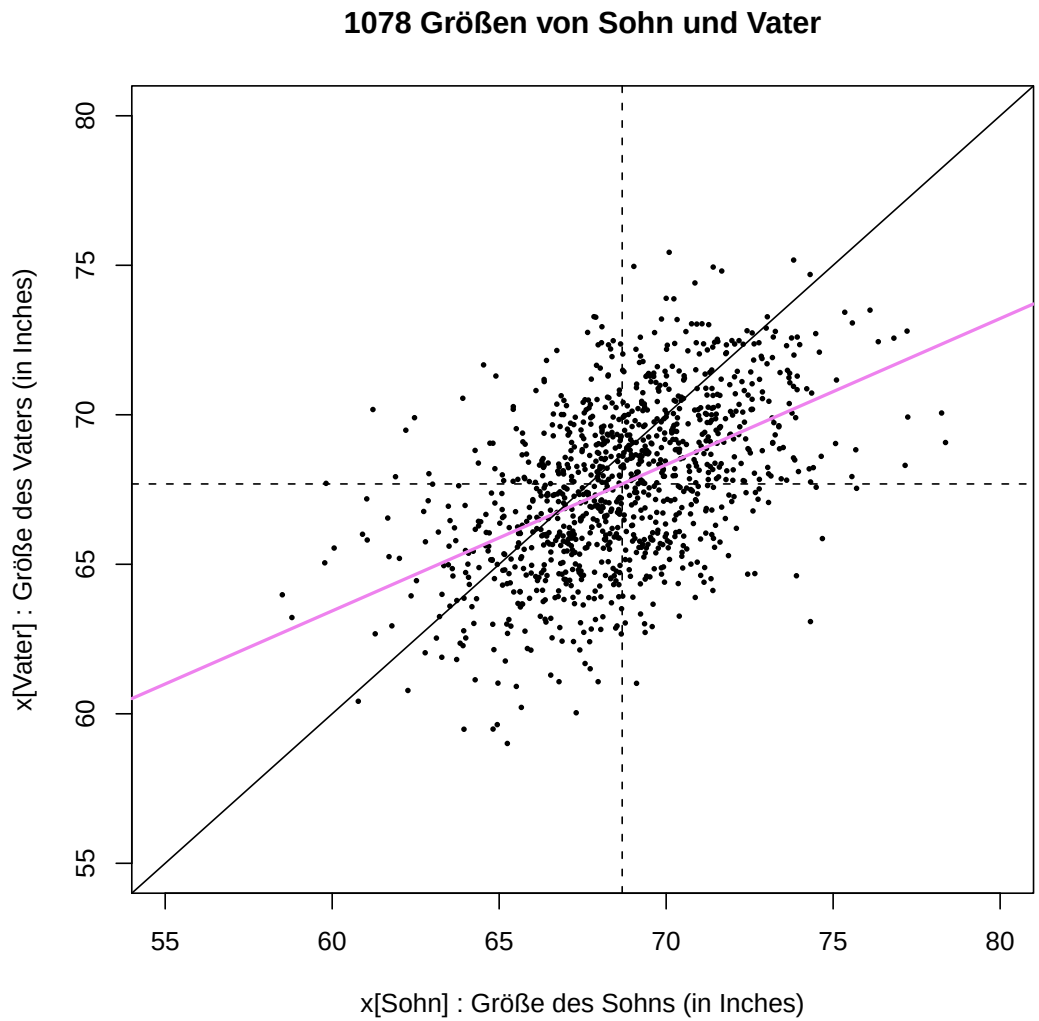
Bemerkung: Das beobachtete Phänomen der „Rückkehr zum Mittelwert“ bedeutet nicht notwendigerweise einen tieferen kausalen Zusammenhang, es tritt stets im Zusammenhang mit natürlicher Variabilität auf (technisch gesehen stets, wenn für den Korrelationkoeffizient ρ gilt $|\rho| < 1$).

Bestimmen wir (spaßeshalber) im Größen-Beispiel die Regressionsgerade für die Größe des Vaters als Funktion der Größe des Sohns:

Bemerkung: Das beobachtete Phänomen der „Rückkehr zum Mittelwert“ bedeutet nicht notwendigerweise einen tieferen kausalen Zusammenhang, es tritt stets im Zusammenhang mit natürlicher Variabilität auf (technisch gesehen stets, wenn für den Korrelationkoeffizient ρ gilt $|\rho| < 1$).

Bestimmen wir (spaßeshalber) im Größen-Beispiel die Regressionsgerade für die Größe des Vaters als Funktion der Größe des Sohns:

Wir hatten $\bar{x}_{\text{Vater}} = 67.7$, $\bar{x}_{\text{Sohn}} = 68.7$, $\text{cov}(x_{\text{Vater}}, x_{\text{Sohn}}) = 3.87$, $\sigma^2_{\text{Sohn}} = 7.92$
und finden die Regressionsgerade $x_{\text{Vater}} = 34.1 + 0.489 x_{\text{Sohn}}$



Regressionsgerade: $x_{\text{Vater}} = 34.1 + 0.489 x_{\text{Sohn}}$

Prostata-Krebs-Daten von Stamey et al

Datensatz beschrieben in T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning*, Example 3.2.1

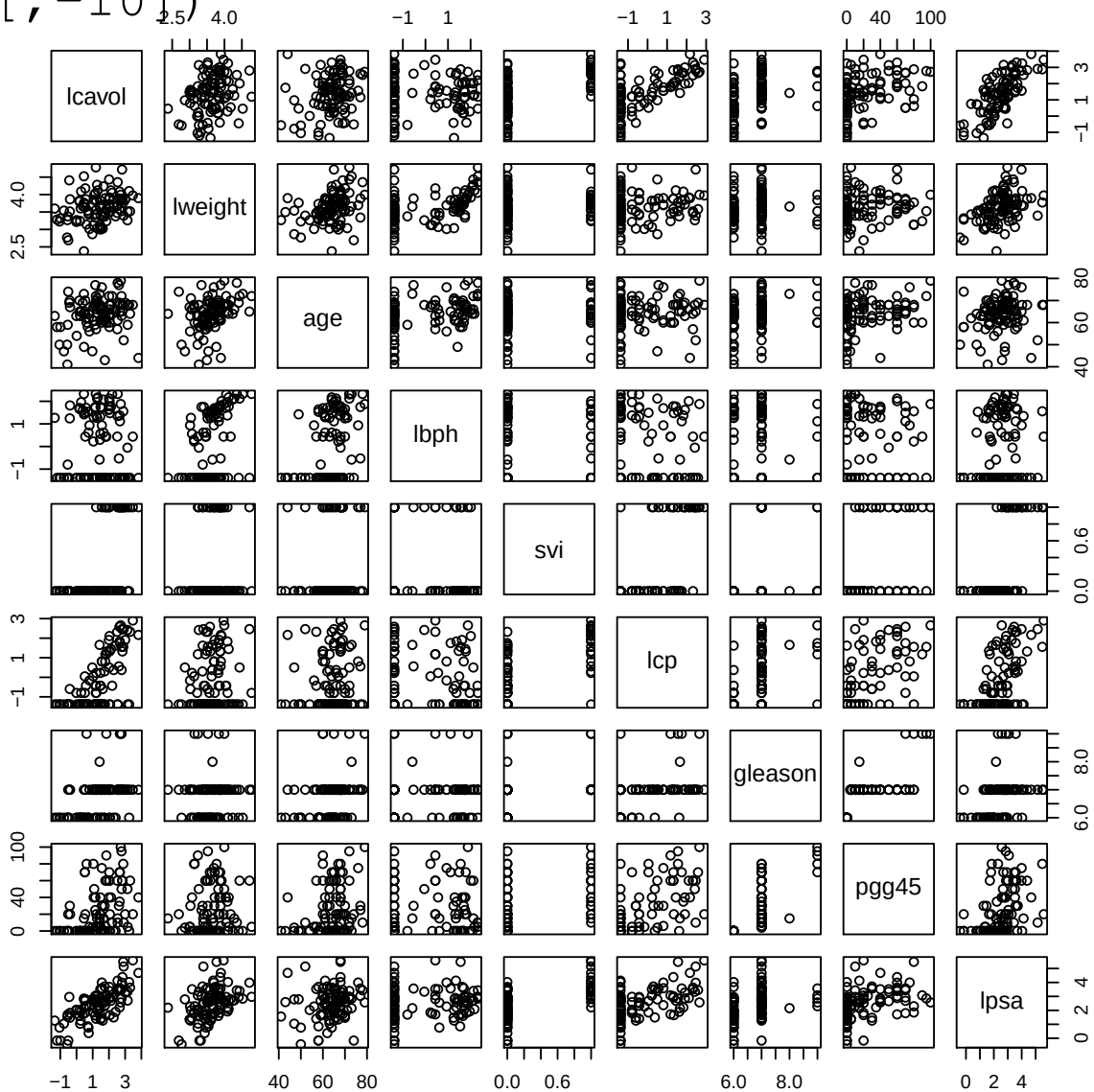
$n = 97$ Beobachtungen (Prostata-Krebs-Patienten),
$p = 8$ erklärende Variablen (plus "intercept")

log cancer volume (`lcavol`), *log prostate weight* (`lweight`), `age`, *log of the amount of benign prostatic hyperplasia* (`lbph`), *seminal vesicle invasion* (`svi`), *log of capsular penetration* (`lcp`), *Gleason score* (`gleason`), *percent of Gleason scores 4 or 5* (`pgg45`)

Zielgröße ("response") *log prostate specific antigen* `lpsa`

Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate II radical prostatectomy treated patients, Journal of Urology 16: 1076–1083

```
> dat <- read.table("prostate.data", header=T)
> pairs(dat[,-10])
```



Zentrierung/Standardisierung

Wir folgen Hastie, Tibshirani & Friedman (und „üblicher Praxis“):

Zentrierung der Spalten von $\boldsymbol{X}$:
ersetze $x_{i,j}$ durch $x_{i,j} - \bar{x}_j$ mit $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j}$

Standardisierung der Spalten von $\boldsymbol{X}$: Skaliere so, dass $\frac{1}{n}\sum_{i=1}^{n} x_{i,j}^2 = 1$ für $j = 1, \ldots, p$
(d.h. ersetze $x_{i,j}$ durch $x_{i,j}/(\frac{1}{n}\sum_{i=1}^{n} x_{i,j}^2)^{1/2}$, dann hat $\boldsymbol{X}^t\boldsymbol{X}$ auf der Diagonale 1en stehen)

```
> dat <- read.table("prostate.data", header=T)
> x <- dat[,1:8]
> xp <- scale(x,TRUE,TRUE)
> sdat <- data.frame(xp, lpsa=dat$lpsa)
```

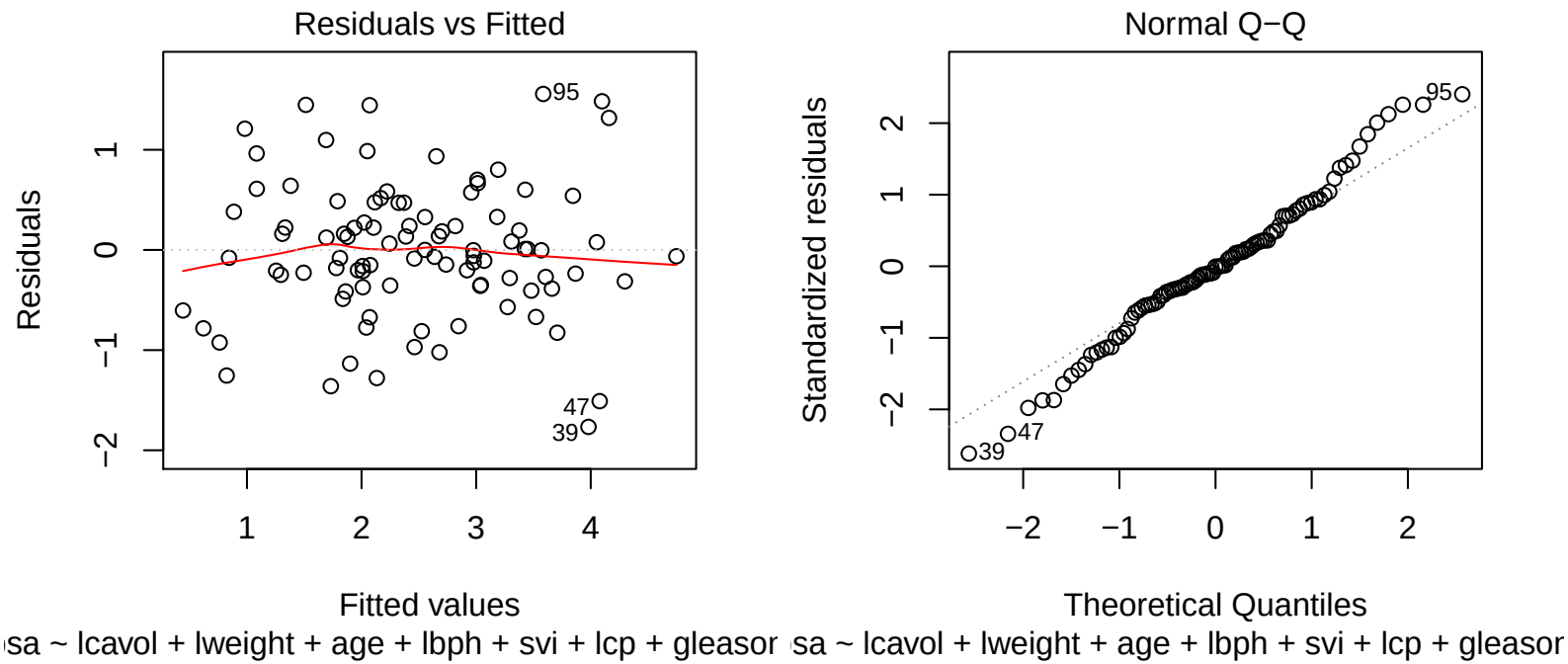
Es gibt nicht-triviale Korrelationen zwischen den Spalten von $\boldsymbol{X}$:

```
> round(cor(sdat[1:8]), digits=3)

        lcavol lweight   age   lbph    svi    lcp gleason pgg45
lcavol   1.000   0.281 0.225  0.027  0.539  0.675   0.432 0.434
lweight  0.281   1.000 0.348  0.442  0.155  0.165   0.057 0.107
age      0.225   0.348 1.000  0.350  0.118  0.128   0.269 0.276
lbph     0.027   0.442 0.350  1.000 -0.086 -0.007   0.078 0.078
svi      0.539   0.155 0.118 -0.086  1.000  0.673   0.320 0.458
lcp      0.675   0.165 0.128 -0.007  0.673  1.000   0.515 0.632
gleason  0.432   0.057 0.269  0.078  0.320  0.515   1.000 0.752
pgg45    0.434   0.107 0.276  0.078  0.458  0.632   0.752 1.000
```

```
> summary(modell)
Call:
lm(formula = lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
    gleason + pgg45, data = sdat)
Residuals:
     Min       1Q   Median       3Q      Max
-1.76644 -0.35510 -0.00328  0.38087  1.55770
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.47839    0.07102  34.895  < 2e-16 ***
lcavol       0.66515    0.10352   6.425 6.55e-09 ***
lweight      0.26648    0.08607   3.096  0.00263 **
age         -0.15820    0.08252  -1.917  0.05848 .
lbph         0.14031    0.08402   1.670  0.09848 .
svi          0.31533    0.09985   3.158  0.00218 **
lcp         -0.14829    0.12566  -1.180  0.24115
gleason      0.03555    0.11218   0.317  0.75207
pgg45        0.12572    0.12312   1.021  0.31000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6995 on 88 degrees of freedom
Multiple R-squared: 0.6634, Adjusted R-squared: 0.6328
F-statistic: 21.68 on 8 and 88 DF,  p-value: < 2.2e-16
```

Betrachten wir die Anpassung graphisch sowie einen Q-Q-Plot der Residuen

```
> plot(modell)
```



Wird der Fit schlechter, wenn wir `age`, `lcp`, `gleason` und `pgg45` weglassen (z.B. weil wir glauben, dass sie in Wirklichkeit keinen Einfluss haben, d.h. die entsprechenden $\beta_i =$ sind)?

```
> summary(modell2)
Call:
lm(formula = lpsa ~ lcavol + lweight + lbph + svi, data = sdat)

Residuals:
     Min       1Q   Median       3Q      Max
-1.85483 -0.35990 -0.01492  0.44467  1.53051
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.47839    0.07148  34.673  < 2e-16 ***
lcavol       0.62289    0.08781   7.094 2.63e-10 ***
lweight      0.22964    0.08415   2.729  0.00761 **
lbph         0.11403    0.08139   1.401  0.16454
svi          0.29207    0.08610   3.392  0.00102 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Wird der Fit schlechter, wenn wir `age`, `lcp`, `gleason` und `pgg45` weglassen (z.B. weil wir glauben, dass sie in Wirklichkeit keinen Einfluss haben, d.h. die entsprechenden $\beta_i =$ sind)?

```
> anova(modell2, modell)
Analysis of Variance Table

Model 1: lpsa ~ lcavol + lweight + lbph + svi
Model 2: lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
    pgg45
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     92 45.595
2     88 43.058  4     2.537 1.2963 0.2777
```

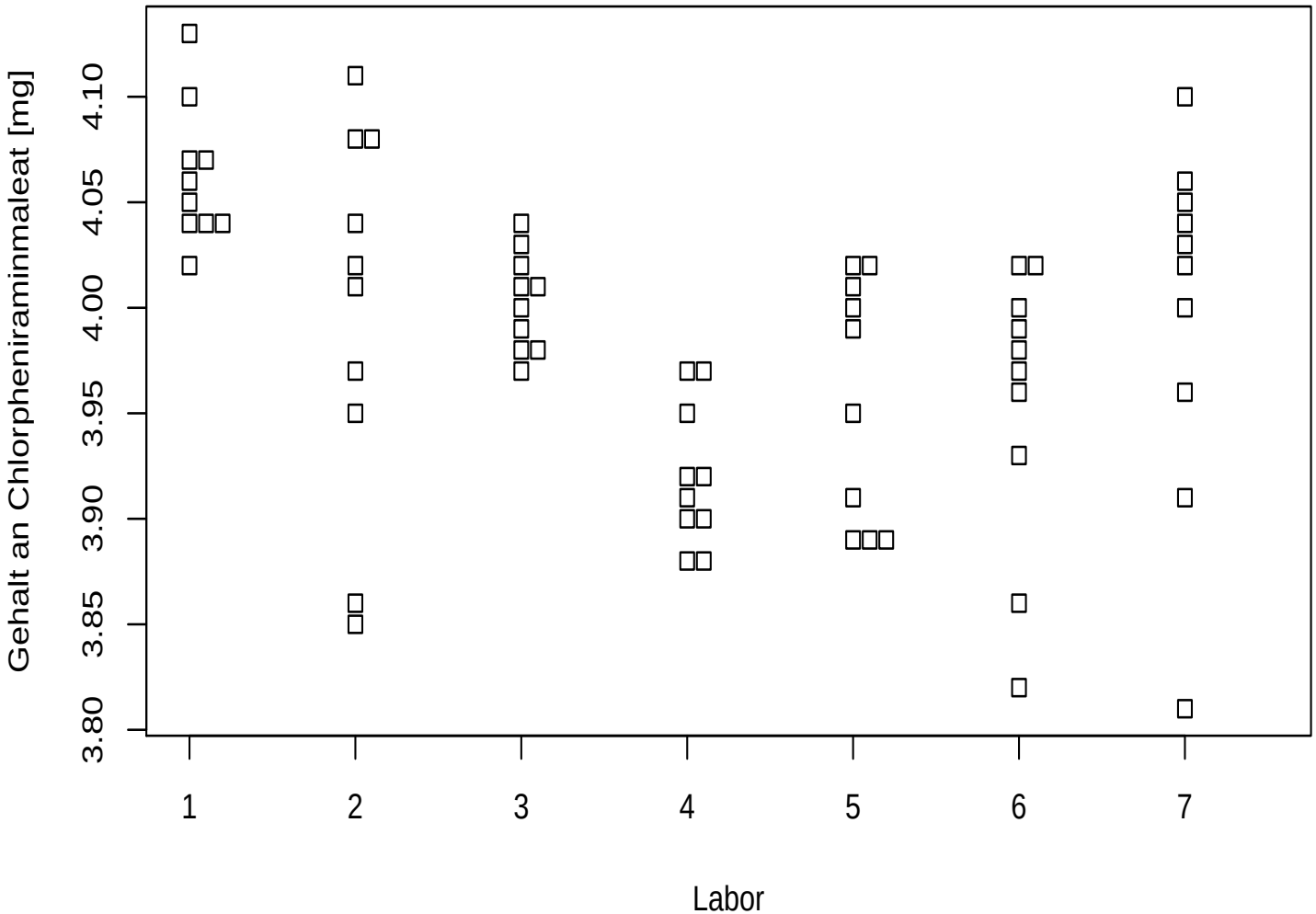
Ein Beispiel zur Varianzanalyse

7 verschiedene Labors haben jeweils 10 Messungen des Chlorpheniraminmaleat-Gehalts von Medikamentenproben vorgenommen.

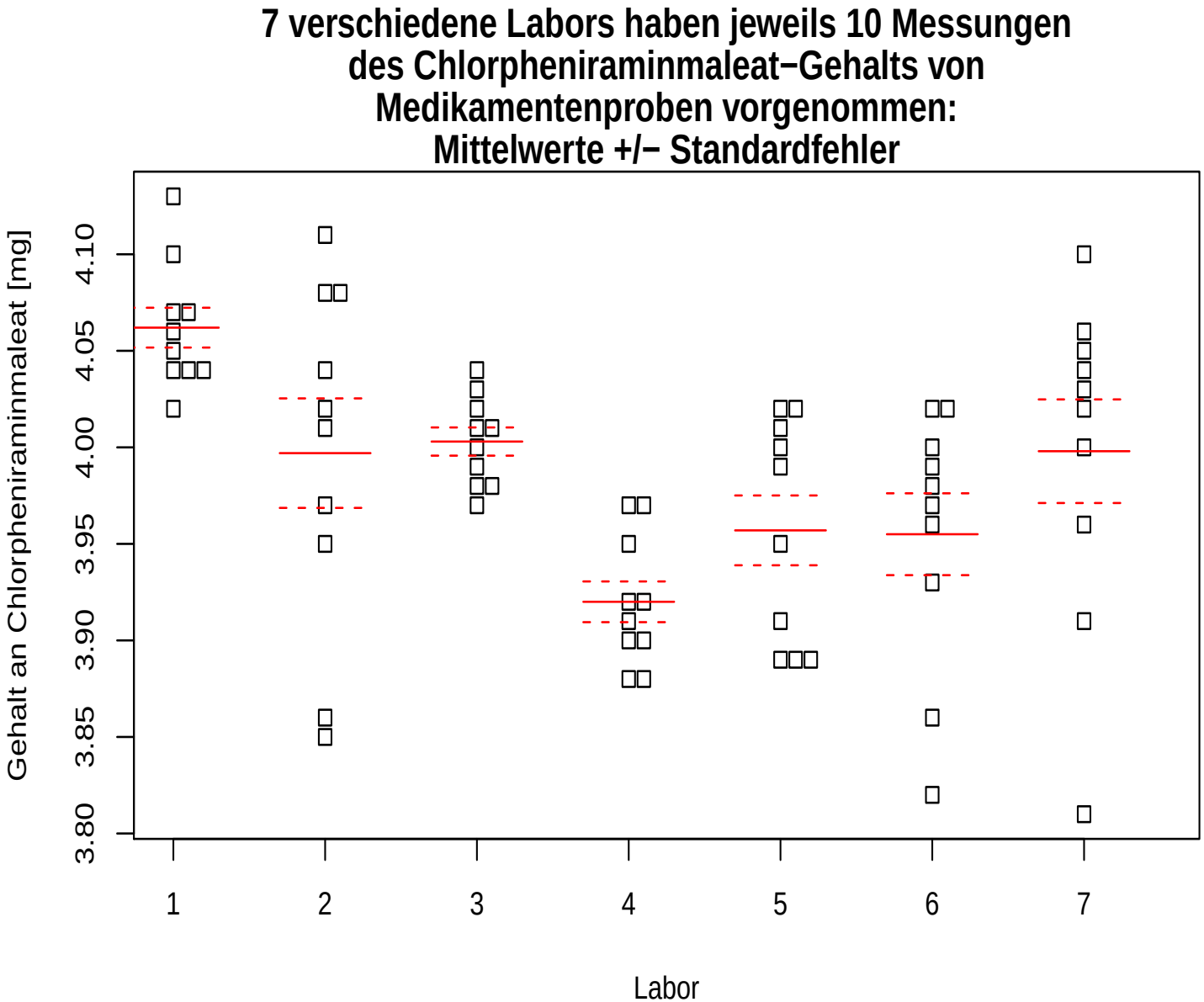
Die Daten liegen in der Datei `chlorpheniraminmaleat.txt` als Tabelle vor:

```
   Gehalt Labor
1  4.13 1
2  4.07 1
3  4.04 1
4  4.07 1
5  4.05 1
6  4.04 1
7  4.02 1
8  4.06 1
9  4.1 1
10 4.04 1
11 3.86 2
12 3.85 2
13 4.08 2
14 4.11 2
15 4.08 2
16 4.01 2
17 4.02 2
18 4.04 2
19 3.97 2
20 3.95 2
21 4 3
22 4.02 3
23 4.01 3
24 4.01 3
25 4.04 3
26 3.99 3
27 4.03 3
28 3.97 3
29 3.98 3
30 3.98 3
...
```

7 verschiedene Labors haben jeweils 10 Messungen des Chlorpheniraminmaleat−Gehalts von Medikamentenproben vorgenommen:



Daten aus R.D. Kirchhoefer, Semiautomated method for the analysis of chlorpheniramine maleate tablets: collaborative study, *J. Assoc. Off. Anal. Chem.* 62(6):1197-1201 (1979),
zitiert nach John A. Rice, *Mathematical statistics and data analysis*, 2nd ed., Wadsworth, 1995



Daten aus R.D. Kirchhoefer, Semiautomated method for the analysis of chlorpheniramine maleate tablets: collaborative study, *J. Assoc. Off. Anal. Chem.* 62(6):1197-1201 (1979),
zitiert nach John A. Rice, *Mathematical statistics and data analysis*, 2nd ed., Wadsworth, 1995

Beachte: Die Labore sind sind mit Zahlen nummeriert. Damit R das nicht als numerische Werte sondern als Nummern der Labore auffasst, müssen wir die Variable "Labor" in einen sog. Factor umwandeln:

```
> chlor <- read.table("chlorpheniraminmaleat.txt")
> str(chlor)
'data.frame': 70 obs. of  2 variables:
 $ Gehalt: num  4.13 4.07 4.04 4.07 4.05 4.04 4.02 4.06 4.1 4.04 .
 $ Labor : int  1 1 1 1 1 1 1 1 1 1 ...
> chlor$Labor <- as.factor(chlor$Labor)
> str(chlor)
'data.frame': 70 obs. of  2 variables:
 $ Gehalt: num  4.13 4.07 4.04 4.07 4.05 4.04 4.02 4.06 4.1 4.04 .
 $ Labor : Factor w/ 7 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1
```

Nun können wir die Varianzanalyse durchführen:

```
> chlor.aov <- aov(Gehalt~Labor,data=chlor)
> summary(chlor.aov)
            Df  Sum Sq  Mean Sq F value    Pr(>F)
Labor        6 0.12474 0.020789  5.6601 9.453e-05 ***
Residuals   63 0.23140 0.003673
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘
```

Über paarweise Vergleiche und multiples Testen

In vorigem Beispiel:

Sei μ_i der (uns unbekannte, wahre Populations-)Mittelwert
der Messungen aus Labor i, für $i = 1, \ldots, 7$.

Die Varianzanalyse zeigte, dass es signifikante Unterschiede zwischen den Laboren gibt.

Aber welche Labore unterscheiden sich signifikant?

Über paarweise Vergleiche und multiples Testen

In vorigem Beispiel:

Sei μ_i der (uns unbekannte, wahre Populations-)Mittelwert
der Messungen aus Labor i, für $i = 1, \ldots, 7$.

Die Varianzanalyse zeigte, dass es signifikante Unterschiede zwischen den Laboren gibt.

Aber welche Labore unterscheiden sich signifikant?

Wir könnten dazu für jedes Paar i, j von Labors jeweils einen
(zwei-Stichproben-)t-Test durchf"uhren, um die Nullhypothese

$$H_{0,(i,j)} : \mu_i = \mu_j$$

(zu einem vorgegebenen Signifikanzniveau α, sagen wir $\alpha = 5\%$) zu testen.

Welche Labore unterscheiden sich signifikant?

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Welche Labore unterscheiden sich signifikant?

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Das 97,5%-Quantil der t-Verteilung mit 18 Freiheitsgraden ist 2.101

Welche Labore unterscheiden sich signifikant?

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:
[(zweiseitiger) zwei-Stichproben t-Tests mit Annahme gleicher Varianzen]

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Das 97,5%-Quantil der t-Verteilung mit 18 Freiheitsgraden ist 2.101,
also würde für die rot markierten Paare (jeweils für sich betrachtet) ein t-Test $H_{0,(i,j)}$
zum Signifikanzniveau 5% ablehnen.

Welche Labore unterscheiden sich signifikant?

Alternative Darstellung:

p-Werte aus paarweisen Vergleichen mittels t-Tests:

Erinnerung:
p-Wert = W'keit (unter der Nullhypothese) einen mindestens so extremen Wert der t-Statistik wie den beobachteten zu erhalten

[Hier: $2(1 - F_{Student(18)}(|t|))$, mit $F_{Student(18)}$ Verteilungsfunktion der Student-Verteilung mit 18 Freiheitsgraden]

Welche Labore unterscheiden sich signifikant?

Alternative Darstellung:

p-Werte aus paarweisen Vergleichen mittels *t*-Tests:

[(zweiseitiger) zwei-Stichproben *t*-Tests mit Annahme gleicher Varianzen]

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	0.04506	0.00019	0.00000	0.00008	0.00025	0.03894
Lab2		0.84000	0.02033	0.24980	0.25103	0.97985
Lab3			0.00000	0.02982	0.04626	0.85929
Lab4				0.09398	0.15662	0.01446
Lab5					0.94356	0.22113
Lab6						0.22459

Erinnerung:

p-Wert = W'keit (unter der Nullhypothese) einen mindestens so extremen Wert der *t*-Statistik wie den beobachteten zu erhalten

[Hier: $2(1 - F_{Student(18)}(|t|))$, mit $F_{Student(18)}$ Verteilungsfunktion der Student-Verteilung mit 18 Freiheitsgraden]

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Wenn die Nullhypothese(n) („alles nur Zufallsschwankungen“) stimmt/en, verwirft man im Schnitt bei 5% der Tests die Nullhypothese zu Unrecht.

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Wenn die Nullhypothese(n) („alles nur Zufallsschwankungen“) stimmt/en, verwirft man im Schnitt bei 5% der Tests die Nullhypothese zu Unrecht.

Testet man mehr als 20 mal und gelten jeweils die Nullhypothesen, wird man also im Schnitt mehr als eine Nullhypothese zu Unrecht verwerfen.

Problem des multiplen Testens

Wir haben $7 \cdot 6 \cdot \frac{1}{2} = 21$ paarweise Vergleiche; auf dem 5%-Niveau zeigen einige davon Signifikanz an.

Wenn die Nullhypothese(n) („alles nur Zufallsschwankungen“) stimmt/en, verwirft man im Schnitt bei 5% der Tests die Nullhypothese zu Unrecht.

Testet man mehr als 20 mal und gelten jeweils die Nullhypothesen, wird man also im Schnitt mehr als eine Nullhypothese zu Unrecht verwerfen.

Dieses Phänomen müssen wir bei multiplen Tests berücksichtigen

(und mit entsprechend angepassten Tests bzw. mit korrigierten p-Werten arbeiten).

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Wert der *t*-Statistik aus paarweisen Vergleichen mittels *t*-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Betrachte $\alpha = 5\%$. Hier $m = 21$, das $(1 - \frac{1}{2}\frac{\alpha}{m})$-Quantil $(1 - \frac{1}{2}\frac{\alpha}{m} = 0.99881)$ der t-Verteilung mit 18 Freiheitsgraden ist 3.532

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Wert der t-Statistik aus paarweisen Vergleichen mittels t-Tests:

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	2.154	4.669	9.632	5.046	4.539	2.227
Lab2		-0.205	2.545	1.189	1.186	-0.026
Lab3			6.470	2.359	2.140	0.180
Lab4				-1.768	-1.478	-2.706
Lab5					0.072	-1.268
Lab6						-1.258

Betrachte $\alpha = 5\%$. Hier $m = 21$, das $(1 - \frac{1}{2}\frac{\alpha}{m})$-Quantil $(1 - \frac{1}{2}\frac{\alpha}{m} = 0.99881)$ der t-Verteilung mit 18 Freiheitsgraden ist 3.532,

also können wir für die rot markierten Paare $H_{0,(i,j)}$ zum multiplen Signifikanzniveau 5% ablehnen.

Labor-Vergleichs-Beispiel mit Bonferroni-Korrektur

Alternativ: $21 \times \left(p\text{-Wert aus paarweisem } t\text{-Test}\right)$

	Lab2	Lab3	Lab4	Lab5	Lab6	Lab7
Lab1	0.94626	0.00399	0.00000	0.00168	0.00525	0.8177
Lab2		17.64000	0.42693	5.24580	5.27163	20.576
Lab3			0.00000	0.62622	0.97146	18.045
Lab4				1.97358	3.28902	0.3036
Lab5					19.81476	4.6437
Lab6						4.7163

F"ur die rot markierten Paare ist der korrigierte p-Wert < 0.05.