Weiterbildungs-Lehrgang in Angewandter Statistik ETHZ 2009/2011 28.09.07

Musterlösungen — Serie 2 (Fakultativ)

1. Wir benützen

$$\gamma = \alpha + \beta x, \quad \sigma_\gamma^2 = \frac{\sigma^2}{n} \quad \text{und} \quad \sigma_\beta^2 = \frac{\sigma^2 1}{SSQ^{(X)}}.$$

a) Die Ellipsengleichung

$$\left(\frac{\gamma - \hat{\gamma}}{\sigma_\gamma}\right)^2 + \left(\frac{\beta - \hat{\beta}}{\sigma_\beta}\right)^2 = q^2$$

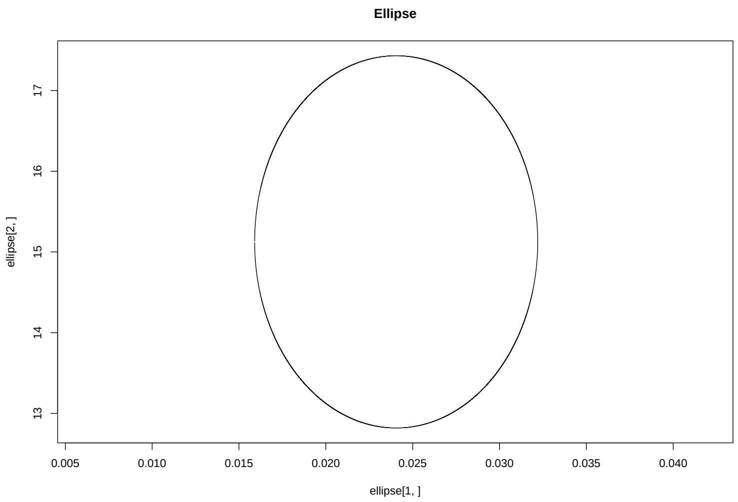
kann nach γ aufgelöst werden

$$\gamma_{1,2} = \hat{\gamma} \pm \sigma_\gamma \sqrt{q^2 - \left(\frac{\beta - \hat{\beta}}{\sigma_\beta}\right)^2}.$$

Somit können die untere und der obere Teil der Ellipse ploten.
R-Befehl:

```
d.buch <-
read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/buchpreis.dat",header=T)

r.lm <- lm(preis~seiten,data=d.buch)
r.lm.summary <- summary(r.lm)

t.n <- length(d.buch$seiten)
t.c <- sqrt(qf(0.95,df1=1,df2=t.n-2))
x.mean <- mean(d.buch$seiten)

est.gamma <- mean(d.buch$preis)
est.beta <- r.lm.summary$coefficients["seiten","Estimate"]

t.sigma.gamma <- r.lm.summary$sigma*sqrt(1/t.n)
t.sigma.beta <- r.lm.summary$coefficients["seiten","Std. Error"]

beta.range <- round(est.beta+c(-1,1)*t.c*t.sigma.beta,3)
beta.werte <- seq(beta.range[1]-0.01,beta.range[2]+0.01,by=0.0000001)
ellipse <-
est.gamma+sqrt(t.c**2-((beta.werte-est.beta)/t.sigma.beta)**2)
ellipse1 <- rbind(beta.werte,ellipse)
ellipse <-
est.gamma-sqrt(t.c**2-((beta.werte-est.beta)/t.sigma.beta)**2)
ellipse2 <- rbind(beta.werte,ellipse)
ellipse <- cbind(ellipse1,ellipse2)

plot(ellipse[2,]~ellipse[1,],main="Ellipse",type="l")
```


R-Output:



b) Zu einem x-Wert x_0 betrachten wir die Geraden

$$y_0 = \gamma + \beta(x_0 - \overline{x}),$$

mit den Parametern γ und β aus der Ellipse. Um für x_0 die Geraden, die den minimalen resp maximalen y-Wert representieren, zu finden, machen wir die folgende Bemerkung: da zu jedem β der Parameter γ des minimalen resp maximalen y-Wertes auf dem Ellipsenrand liegen muss, können wir die Gleichung für γ in die Geradengleichung substituieren. Dies ergibt:

$$y = \hat{\gamma} \pm \sigma_\gamma \sqrt{q^2 - \left(\frac{\beta - \hat{\beta}}{\sigma_\beta}\right)^2} + \beta(x_0 - \overline{x}).$$

Um die kritischen Werte (Minimum und Maxmium) für β zu erhalten, differenzieren wir die obige Gleichung nach β und setzen sie dann Null. Die Ableitung ist:

$$\frac{d}{d\beta} y = (x_0 - \overline{x}) \pm \frac{\frac{\beta - \hat{\beta}}{\sigma_\beta^2}}{\sqrt{q^2 - \left(\frac{\beta - \hat{\beta}}{\sigma_\beta}\right)^2}} = 0.$$

Als Zwischenresultat erhalten wir

$$\left(\frac{\beta - \hat{\beta}}{\sigma_\beta}\right)^2 = \frac{q^2 \sigma_\beta^2 (x_0 - \overline{x})^2}{\sigma_\gamma^2 + \sigma_\beta^2 (x_0 - \overline{x})^2}.$$

Schliesslich erhalten wir die Parameter-Werte:

$$\beta_{max,min} \quad = \quad \hat{\beta} \pm q \frac{\sigma_\beta^2 (x_0 - \overline{x})}{\sqrt{\sigma_\gamma^2 + \sigma_\beta^2 (x_0 - \overline{x})^2}} = \hat{\beta} \pm q \frac{\sigma_\beta^2 (x_0 - \overline{x})}{\mathrm{se}^{(\eta)}}$$

$$\gamma_{max,min} = \hat{\gamma} \pm q \frac{\sigma_\gamma^2}{\sqrt{\sigma_\gamma^2 + \sigma_\beta^2 (x_0 - \overline{x})^2}} = \hat{\gamma} \pm q \frac{\sigma_\gamma^2}{\mathrm{se}^{(\eta)}}$$

die zu jedem x_0 die (minimalen und maximalen) Geraden

$$\begin{aligned} y_{max} &= \left(\hat{\gamma} + q\frac{\sigma_\gamma^2}{\mathrm{se}^{(\eta)}}\right) + \left(\hat{\beta} + q\frac{\sigma_\beta^2 (x_0 - \overline{x})}{\mathrm{se}^{(\eta)}}\right)(x - \overline{x}) \\ y_{min} &= \left(\hat{\gamma} - q\frac{\sigma_\gamma^2}{\mathrm{se}^{(\eta)}}\right) + \left(\hat{\beta} - q\frac{\sigma_\beta^2 (x_0 - \overline{x})}{\mathrm{se}^{(\eta)}}\right)(x - \overline{x}) \end{aligned}$$

bilden. Setzen wir für x die Werte x_0 ein, so ergibt sich

$$y_{max,min} = \left(\hat{\gamma} \pm q\frac{\sigma_\gamma^2}{\mathrm{se}^{(\eta)}}\right) + \left(\hat{\beta} + q\frac{\sigma_\beta^2 (x_0 - \overline{x})^2}{\mathrm{se}^{(\eta)}}\right).$$

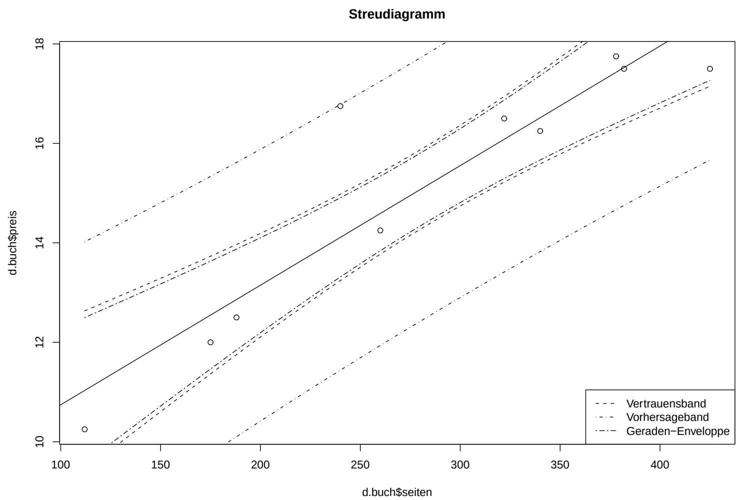
Diese Gleichungen können wir in R programmieren.
R-Befehle:

```
##Vertrauensband & Vorhersageband
t.range <- range(d.buch$seiten)
t.xwerte <- seq(t.range[1],t.range[2],by=1)
t.vert <- predict(r.lm,se.fit=T,newdata=data.frame(seiten=t.xwerte),
            interval ="confidence")$fit
t.vorh <- predict(r.lm,se.fit=T,newdata=data.frame(seiten=t.xwerte),
            interval ="prediction")$fit

# Geraden-Enveloppe
t.gamma.p <- function(x0){
est.gamma+t.c*t.sigma.gamma**2
   *sqrt(1/(t.sigma.gamma**2+t.sigma.beta**2*(x0-x.mean)**2))
}
t.gamma.m <- function(x0){
est.gamma-t.c*t.sigma.gamma**2
   *sqrt(1/(t.sigma.gamma**2+t.sigma.beta**2*(x0-x.mean)**2))
}

t.beta.p <- function(x0){
est.beta+(x0-x.mean)*t.c*t.sigma.beta**2
   *sqrt(1/(t.sigma.gamma**2+t.sigma.beta**2*(x0-x.mean)**2))
}
t.beta.m <- function(x0){
est.beta-(x0-x.mean)*t.c*t.sigma.beta**2
   *sqrt(1/(t.sigma.gamma**2+t.sigma.beta**2*(x0-x.mean)**2))
}

y.geraden.p <- function(x0,x){
   t.gamma.p(x0)+t.beta.p(x0)*(x-x.mean)
}
y.geraden.m <- function(x0,x){
   t.gamma.m(x0)+t.beta.m(x0)*(x-x.mean)
```



```
}

plot(d.buch$seiten,d.buch$preis,main="Streudiagramm")
abline(r.lm)
lines(x=t.xwerte,y=t.vert[,2],lty=2)
lines(x=t.xwerte,y=t.vert[,3],lty=2)
lines(x=t.xwerte,y=t.vorh[,2],lty=4)
lines(x=t.xwerte,y=t.vorh[,3],lty=4)
lines(x=t.xwerte,y=y.geraden.p(x0=t.xwerte,x=t.xwerte),lty=6)
lines(x=t.xwerte,y=y.geraden.m(x0=t.xwerte,x=t.xwerte),lty=6)
legend("bottomright", c("Vertrauensband",
"Vorhersageband","Geraden-Envelope"),
       lty=c(2,4,6), cex=1)
```

R-Output:



Bemerkung: Wir formen die die vorhergenden Gleichungen für y ein wenig um:

$$\begin{aligned} y &= \left(\hat{\gamma} + \hat{\beta}(x_0 - \overline{x})\right) \pm q\frac{\sigma_\gamma^2 + \sigma_\beta^2 (x_0 - \overline{x})^2}{\mathrm{se}^{(\eta)}} \\ &= \hat{\gamma} + \hat{\beta}(x_0 - \overline{x}) \pm q\mathrm{se}^{(\eta)} \\ &= \hat{\alpha} + \hat{\beta}x_0 \pm q\mathrm{se}^{(\eta)} \end{aligned}$$

Wir erhalten die gleiche Struktur wie das Vertrauensband, nur die q's haben verschiedene Bedeutungen respektive verschiedene Freiheitsgrade:
Vertrauensband: $q = t_{0.975,n-2} = \sqrt{F^{-1}_{1,n-2}(0.95)}$
Enveloppe: $q = \sqrt{F^{-1}_{2,n-2}(0.95)}$.