Vorbereitung

Bitte führen Sie zur Vorbereitung folgende Schritte aus:

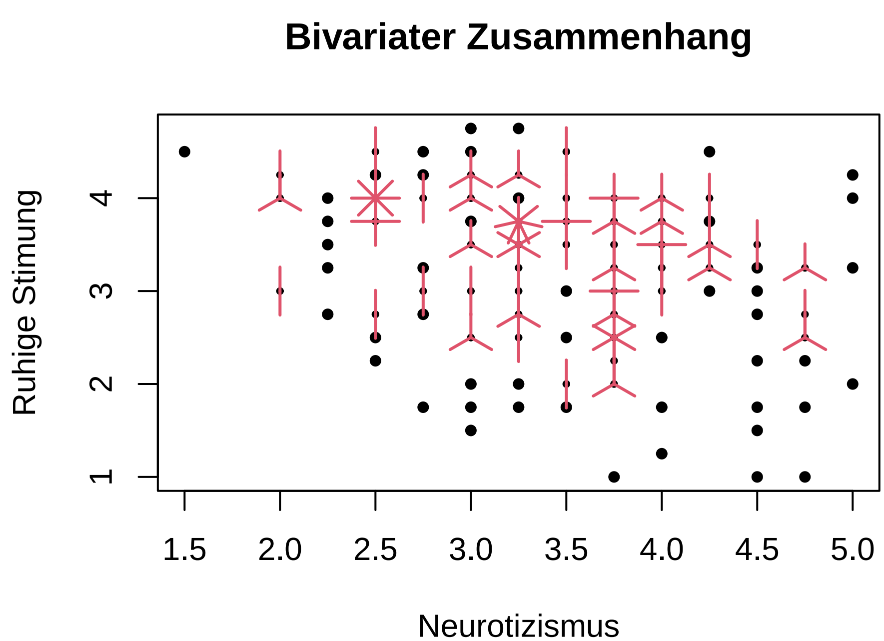
1. Starten Sie RStudio.
2. Löschen Sie den Workspace.
3. Setzen Sie das Arbeitsverzeichnis dort, wo Sie Ihre Daten abgelegt haben: Session > Set Working Directory > Choose Directory.
4. Öffnen Sie ein R-Skript.
5. Nachdem Sie die Aufgaben bearbeitet haben, speichern Sie das Skript unter einem geeigneten Namen ab.

Aufgabe 1

(i) Erstellen Sie ein Streudiagramm mit der ruhigen Stimmung (`ru.1`) auf der Y-Achse und Neurotizismus (`neuro`) auf der X-Achse. Beschriften Sie dabei die Achsen und geben Sie der Graphik einen Titel. Beschreiben Sie anhand des Punkteschwarms die Form und Art des Zusammenhangs.

Lösung

```
sunflowerplot(ru.1 ~ neuro, data = erstis,
              xlab = "Neurotizismus",
              ylab = "Ruhige Stimung",
              main = "Bivariater Zusammenhang")
```



- schwacher bis mittelstarker negativer linearer Zusammenhang

(ii) Erstellen Sie ein Regressionsmodell zur Vorhersage von der ruhigen Stimmung (`ru`) durch Neurotizismus (`neuro`). Lassen Sie sich eine Zusammenfassung des Modells ausgeben. Wie viele Personen wurden in die Berechnung des Modells mit eingeschlossen?

Lösung

```
m1 <- lm(ru.1 ~ neuro, data = erstis)
summary(m1)
```

```

Call:
lm(formula = ru.1 ~ neuro, data = erstis)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.1636 -0.5795  0.1630  0.5097  1.5197 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.46365    0.26694  16.722  < 2e-16 ***
neuro       -0.34668    0.07551  -4.591  8.1e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.776 on 186 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.1018,	Adjusted R-squared:  0.09696 
F-statistic: 21.08 on 1 and 186 DF,  p-value: 8.1e-06
```

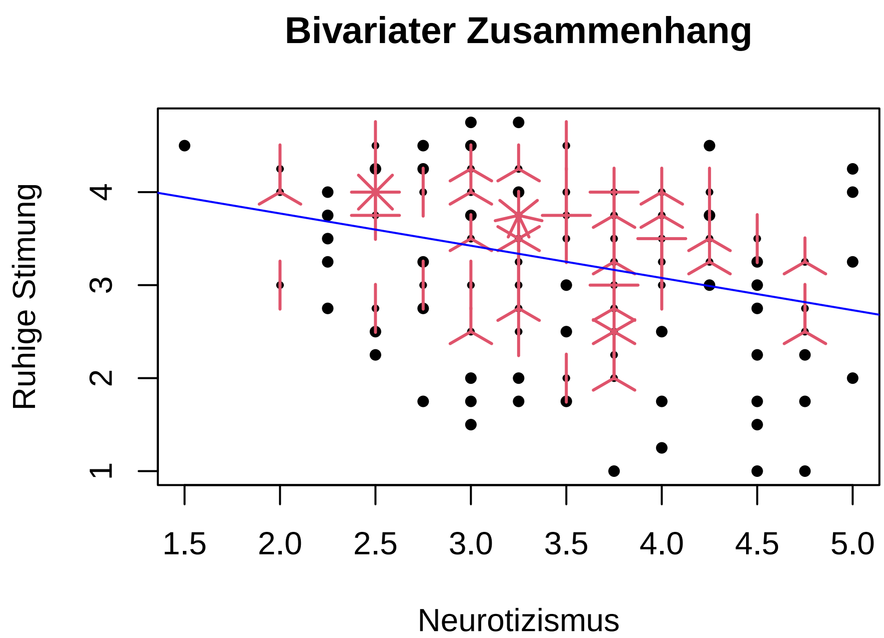
Es wurden 188 Personen in die Berechnung des Modells mit eingeschlossen (191 - 3 Beobachtungen, die als fehlend gelöscht wurden).

(iii) Fügen Sie die Regressionsgerade in ihr Streudiagramm ein und färben Sie die gerade blau.

Lösung

```
sunflowerplot(ru.1 ~ neuro, data = erstis,
              xlab = "Neurotizismus",
              ylab = "Ruhige Stimung",
              main = "Bivariater Zusammenhang")

abline(reg = m1, col = "blue")
```



(iv) Notieren Sie die unstandardisierte Regressionsgleichung und interpretieren Sie die Koeffizienten inhaltlich.

Lösung

```
coef(m1)
```

```
(Intercept)       neuro
  4.4636472  -0.3466754
```

```
# Alternativ: m1$coefficients
```

$$\text{ru.1} = 4.46 - 0.35 * \text{neuro}$$

- der vorhergesagte Wert der ruhigen Stimmung für Personen mit einem Neurotizismuswert von 0 beträgt $b_0 = 4.46$
- zwei Personen, die sich um eine Einheit im Neurotizismus unterscheiden, unterscheiden sich in ihren erwarteten Stimmungswerten um 0.35, wobei der Wert für die Person mit höherem Neurotizismus geringer (=unruhiger) ist
- auch hier wäre die Zentrierung des Prädiktors sinnvoll, da 0 außerhalb des Wertebereichs liegt

(v) Interpretieren Sie auch den Determinationskoeffizienten.

Lösung

```
summary(m1)$r.squared
```

```
[1] 0.1017889
```

- rund 10% der Unterschiede in der ruhigen Stimmung können auf Unterschiede im Neurotizismus zurückgeführt werden, $R^2 = 0.102$

(vi) Notieren Sie nun die standardisierte Regressionsgleichung.

Lösung

```
cor(erstis$ru.1, erstis$neuro, use = "complete")
```

```
[1] -0.3190438
```

$$\hat{z}_{ru.1} = -0.32 * z_{neuro}$$

Alternativ:

```
std2 <- data.frame(scale(m1$model))
m1_std <- update(m1, data = std2)
coef(m1_std)
```

```
  (Intercept)         neuro
-3.555797e-16 -3.190438e-01
```

Achtung: Die `scale()` Funktion kann nur auf numerische Vektoren angewendet werden, also nicht auf das gesamte `erstis` Objekt, da dieses auch Faktoren enthält. Bilden Sie einen Teildatensatz oder nutzen Sie Ihr Wissen, um die Gleichung zu formulieren.

Aufgabe 2

Angenommen Sie erhalten lediglich den Wert der Produkt-Moment-Korrelation (0.23) zwischen zwei Variablen, Neurotizismus und Mathematikfähigkeit. Kann man mithilfe dieser Informationen Aussagen über den Determinationskoeffizienten R^2 machen? Und wenn ja, welche?

Lösung

Bei der einfach linearen Regression entspricht das Bestimmtheitsmaß R^2 dem quadrierten Korrelationskoeffizienten (nach Pearson).

```
0.23^2
```

```
[1] 0.0529
```

Aufgabe 3

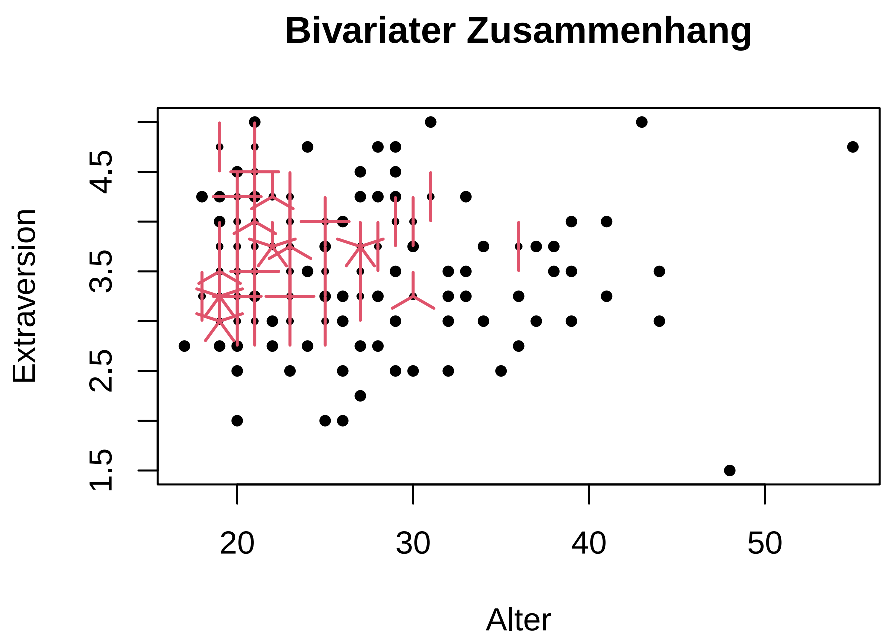
Sie interessieren sich für den Zusammenhang zwischen Extraversion (`extra`) und Alter (`alter`).

(i) Stellen Sie den Zusammenhang graphisch dar.

Lösung

```
# Erstellen eines Sub-Datensatzes ohne fehlende Werte
# nicht zwingend notwendig, aber sinnvoll für weitere Unteraufgaben
sub <- na.omit(erstis[,c("extra","alter")])

sunflowerplot(extra ~ alter, data = sub,
              xlab = "Alter",
              ylab = "Extraversion",
              main = "Bivariater Zusammenhang")
```



(ii) Bestimmen Sie, wie groß die Rangkorrelation nach Spearman ist. Lässt sich mithilfe dieses Maßes eine Aussage über den Determinationskoeffizienten R^2 machen?

Lösung

```
cor(sub$extra, sub$alter, method = "spearman", use= "complete")
```

```
[1] -0.01308037
```

Die Rangkorrelation nach Spearman erlaubt, anders als die Produkt-Moment-Korrelation nach Pearson, keine Aussagen über den Determinationskoeffizienten R^2 des bivariaten Regressionsmodells.

(iii) Erstellen Sie ein Regressionsmodell zur Vorhersage von der Extraversion durch das Alter, in dem Sie direkt ablesen können, wie groß der vorhergesagte Wert für eine Person mit durchschnittlichem Alter ist.

Lösung

```
sub$alter_c <- scale(sub$alter, scale = F, center = T)
m2 <- lm(extra ~ alter_c, data = sub)
summary(m2)
```

```

Call:
lm(formula = extra ~ alter_c, data = sub)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.97971 -0.35433 -0.04631  0.42145  1.49881 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.576149   0.049231  72.640   <2e-16 ***
alter_c     -0.004297   0.007395  -0.581    0.562    
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6494 on 172 degrees of freedom
Multiple R-squared:  0.001959,  Adjusted R-squared:  -0.003843
F-statistic: 0.3377 on 1 and 172 DF,  p-value: 0.5619
```

Die vorhergesagte Extraversion für eine Person mit durchschnittlichem Alter unseres Datensatzes beträgt 3.58 Punkte.

(iv) Wie groß ist der vorhergesagte Wert für die erste Person unseres Datensatzes?

Lösung

```
sub$yDach <- predict(m2)
sub[1,"yDach"]
```

```
[1] 3.52268
```

(v) In welcher Range bewegen sich die Residuen des Modells? Wie groß ist der Standardschätzfehler?

Lösung

```
summary(resid(m2))
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max. 
-1.97971 -0.35433 -0.04632  0.00000  0.42145  1.49881 
```

```
sd(resid(m2))
```

```
[1] 0.6475214
```

Die Residuen liegen im Intervall von -2 bis 1.5. Der Standardschätzfehler beträgt 0.65.

Aufgabe 4

Bestimmen Sie die Produkt-Moment-Korrelation zwischen Prokrastination (`prok`) und Alter (`alter`), nachdem Sie beide Variablen für die Gewissenhaftigkeit (`gewiss`) kontrolliert haben.

Lösung

```
#install.packages("ppcor")
library(ppcor)
sub2 <- na.omit(erstis[, c("prok", "gewiss", "alter")])
pcor(sub2[, c("prok", "gewiss", "alter")])$estimate
```

```
               prok      gewiss        alter
prok    1.000000000 -0.51291818 -0.007478243
gewiss -0.512918176  1.00000000  0.094318205
alter  -0.007478243  0.09431821  1.000000000
```

Die gesuchte Partialkorrelation beträgt -0.51.