Testat/Bachelorprüfung

zur Vorlesung Statistik für Geowissenschaftler (SS 05)

Name: Matrikelnummer:

Nehmen sie für dieses Testat grundsätzlich ein α-Niveau von 5% an. Die erreichbare Punktzahl ist bei allen Teilaufgaben in Klammern () angegeben.

Aufgabe 1: DDx in Brunnen

Früher wurde angenommen, daß DDT nicht ins Grundwasser gelangt, sondern in der oberen Bodenschicht gebunden bleibt. Neuere Überlegungen lassen allerdings vermuten, daß ein sehr giftiges DDT Abbauprodukt, das DDx, mit einiger Verzögerung doch ins Grundwasser gelangen könnte. Von jeweils 30 Brunnen stehen historische DDx Messwerte (DDx1990) aus dem Jahr 1990, als das DDT in Deutschland entgültig verboten wurde, zur Verfügung. Zu diesen 30 Brunnen wurden im Jahr 2002 nochmals Messungen gemacht (DDx2002). Jetzt soll nachgewiesen werden, daß – wie man nach dem neuen Abbaumodell vermuten würde – die DDx Verseuchung im Zeitraum von 1990-2002 angestiegen ist, obwohl kein zusätzlicher DDT-Eintrag stattgefunden hat. Dieser Nachweis des Anstieges aus empirischen Daten ist wichtigster Kernpunkt, um Geld für das von Ihnen betriebene Projekt zur DDx Erforschung zu erhalten.

(a) *Skalenniveau*

Welches Skalenniveau haben die gegebenen Daten (DDx2002)? (1)

Nennen Sie zwei statistische Graphiken, die sich zur Darstellung von Daten aus diesem Skalenniveau eignen? (2)

(b) *Nachweis der Änderung*

Sie können davon ausgehen, daß die Konzentrationsanstiege normalverteilt sind. Sie wollen nachweisen, daß ein DDx Anstieg stattgefunden hat. Welchen Test wählen Sie? (2)

```
....... .test(DDx1990,DDx2002,paired=...)
# t,wilcoxen,anova,kruskal; T = wahr, F=falsch
```

Beispiele:

```
>wilcox.test(CO[source=="BAAQMD"],CO[source=="refinery"],paired=F)
>t.test(CO[source=="BAAQMD"],CO[source=="refinery"],paired=T)
```

Alternativ können Sie den Test auch mit Namen bezeichnen:

(c) Sie finden in der Ausgabe dieses Testes den p-Wert $p = 1.453E - 12$. Was hat man damit nachgewiesen? (2)

(d) Es soll festgestellt werden, ob der pH-Wert des Boden die DDx Freisetzung beeinflußt. Mit dieser Erkenntnis könnte man eventuell Einfluß auf die DDx Freisetzung nehmen. Sie haben die durchschnittliche Boden pH-Wert im Einzugsbereich des Brunnens ermittelt.

```
> DDxAnstieg <- DDx2002-DDx1990
> modell <- lm(DDxAnstieg~pH)
> anova(modell)
Analysis of Variance Table

Response: DDxAnstieg
          Df  Sum Sq Mean Sq F value    Pr(>F)
pH         1 10.2221 10.2221  14.468 0.0007094 ***
Residuals 28 19.7823  0.7065
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> coef(modell)
(Intercept)          pH
  7.5              0.5
```

Welche Voraussetzung hat die lineare Regression?(3)

Welche dieser Voraussetzungen kann man mit den hier angeboteten R Ausgaben überprüfen? (1)

Nehmen wir an die Voraussetzungen der linearen Regression sind an diesem Datensatz erfüllt. Was haben wird dann mit dem hier gemachten Test nachgewiesen? (1)

Im Datensatz sind DDx in ppb und der pH-Wert in üblicher Weise angegeben. Welchen Anstieg der DDx Konzentration würde man seit 1990 bei Brunnen erwarten, deren pH-Wert des Bodens bei 7 liegt? (Benutzen Sie dazu die Regressiongerade der obigen Regressionsanalyse, die durch Achsenabschnitt (Intercept) und Steigung (pH) von R angegeben wurde) (2)

Aufgabe 2: Grundwasserspiegel bei Schaffhausen
Bei Schaffhausen wurde jeden Tag des Jahres 2000 die Höhe des Grundwasserspiegels (über NN) gemessen und mittels R ausgewertet:

```
> GW <- read.table("PegelSchaffhausen.txt",header=T)
> names(GW)
[1] "Grundwasserstand"
> GW$Grundwasserstand
  [1] 391.06 391.06 391.08 391.09 391.08 391.07 391.07 391.07 391.07 391.06
...
[361] 390.99 390.99 390.97 390.97 390.95 390.93
> shapiro.test(GW$Grundwasserstand)

         Shapiro-Wilk normality test

data:  GW$Grundwasserstand
W = 0.9574, p-value = 8.195e-09

> summary(GW$Grundwasserstand)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  390.9   391.0   391.1   391.1   391.2   391.2
> var(GW$Grundwasserstand)
[1] 0.00703621
> sd(GW$Grundwasserstand)
[1] 0.08388212
> plot(1:366,GW$Grundwasserstand,xlab="Tag im Jahr",pch=20)
> qqnorm(GW$Grundwasserstand)
```

```
> qqline(GW$Grundwasserstand)
> boxplot(GW$Grundwasserstand,ylab="Grundwasserstand",xlab="Schaffhausen")
> hist(GW$Grundwasserstand,nclass=14)
```

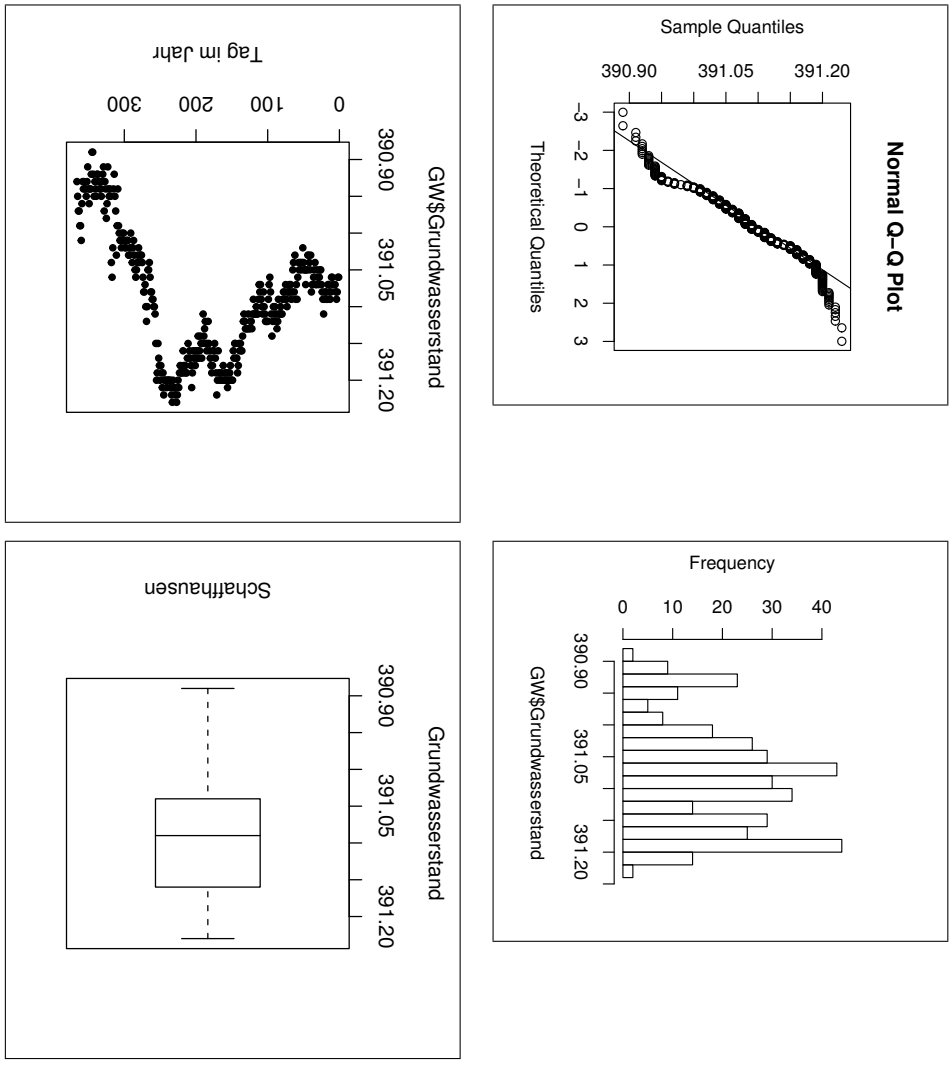


Abbildung 1: Graphiken zur Aufgabe 2

(a) Wie lauten die Kennzahlen des Datensatzes? (3)

- Mittelwert

- Median

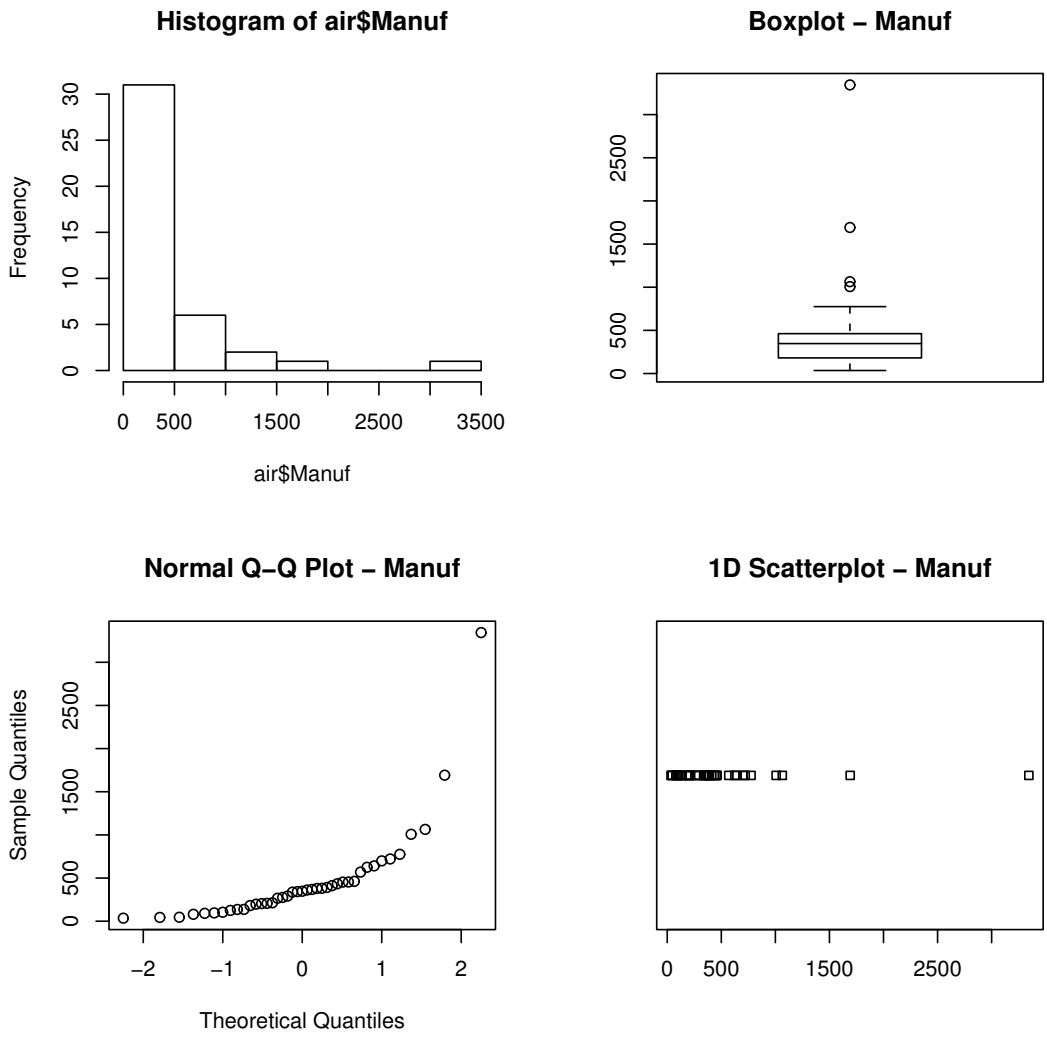
- Varianz

(b) Bestehen irgendwelche Zweifel an der stochastischen Unabhängigkeit der Beobachtungen? ....
Erläutern Sie: (1)

(c) Wie lautet das Ergebnis des Shapiro Wilk Test in allgemeinverständlichen Begriffen? (Unter der Voraussetzung, daß die Voraussetzungen erfüllt sind? (1)

Aufgabe 3: Luftverschmutzung in US-Städten
Quelle: Sokal and Rohlof, Air pollution in US cities
In 41 US Städten wurde die mittlere Schwefeldioxidkonzentration (`SO2`) und sechs erklärende Variablen gemessen. Diese sind die mittlere Jahrestemperatur (`T`), die Anzahl der produzierenden Unternehmen mit mehr als 20 Arbeitern (`Manuf`), die Bevölkerungszahl (`Pop`), mittlere Jahreswindgeschwindigkeit (`Wind`), mittlere Jahresniederschlag (`Prec`) und die mittlere Anzahl der Tage mit Niederschlag im Jahr (`DwP`).

(a) Univariate Statistik
Beschreiben Sie die Verteilung der Variable `Manuf` - Ausreißer, Schiefe, Verteilung? In welchen Graphiken sehen Sie das? (2)



(b) Im Folgenden wurden verschiedene lineare Modelle gerechnet. Welche der sechs erklärenden Variablen werden in irgendeinem Modell signifikant? (2)

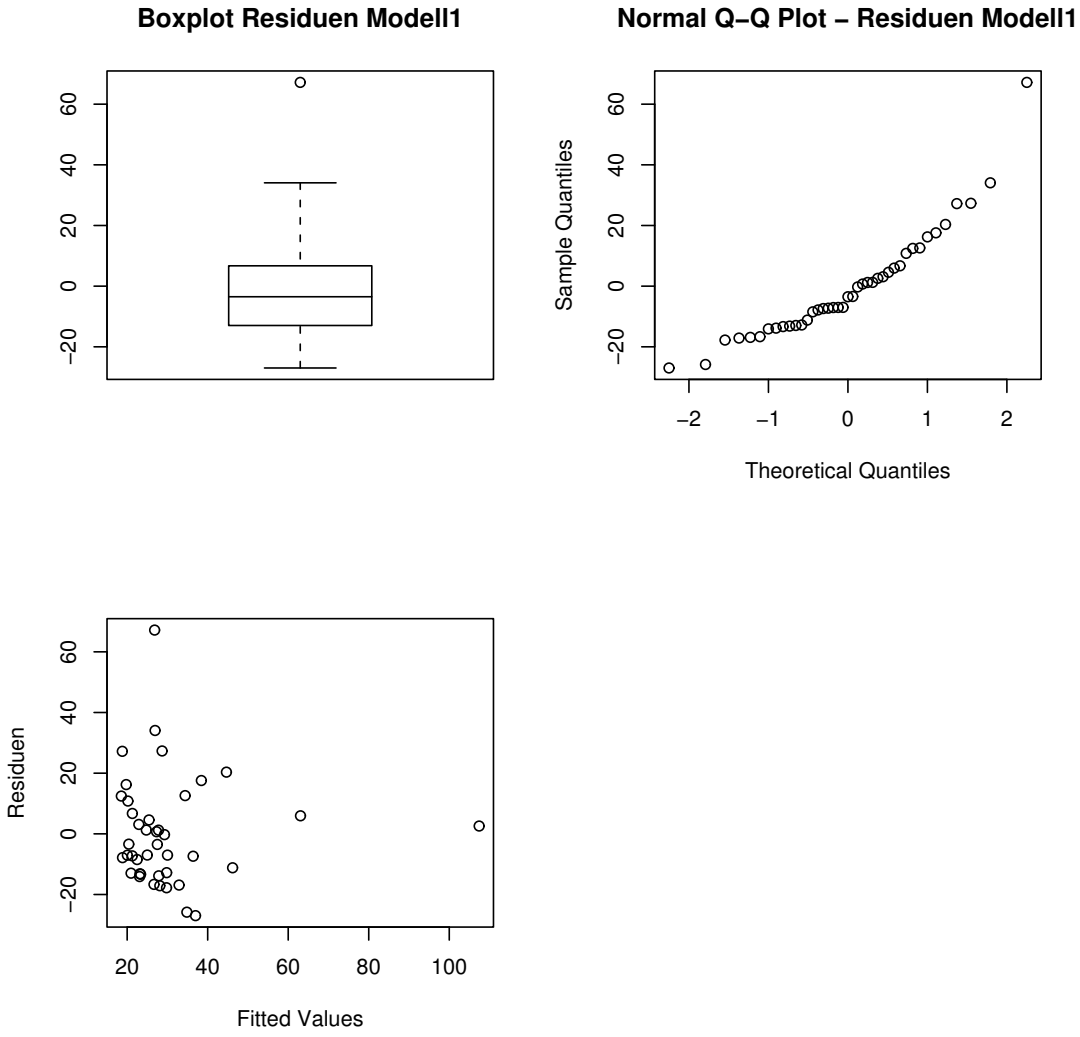


Abbildung 2: Graphiken zu Aufgabe 3

```
> Modell1 <- lm(SO2~Manuf,data=air)
> anova(Modell1)
Analysis of Variance Table
Response: SO2
          Df  Sum Sq Mean Sq F value    Pr(>F)
Manuf      1  9161.7  9161.7  27.750 5.363e-06 ***
Residuals 39 12876.2   330.2
---
> R2(Modell1)
[1] 0.4157267
```

```
==========================================================================
> Modell2 <- lm(SO2~Manuf+Pop,data=air)
> anova(Modell2)
Analysis of Variance Table
Response: SO2
          Df Sum Sq Mean Sq F value    Pr(>F)
Manuf      1 9161.7  9161.7  38.188 3.228e-07 ***
Pop        1 3759.5  3759.5  15.671 0.0003192 ***
Residuals 38 9116.6   239.9
---
> R2(Modell2)
[1] 0.5863202
==========================================================================
> Modell3 <- lm(SO2~Wind,data=air)
> anova(Modell3)
Analysis of Variance Table
Response: SO2
          Df  Sum Sq Mean Sq F value Pr(>F)
Wind       1   197.6   197.6  0.3528 0.5559
Residuals 39 21840.3   560.0
> R2(Modell3)
[1] 0.008966282
==========================================================================
> Modell4 <- lm(SO2~Prec,data=air)
> anova(Modell4)
Analysis of Variance Table
Response: SO2
          Df  Sum Sq Mean Sq F value Pr(>F)
Prec       1    65.0    65.0  0.1153  0.736
Residuals 39 21972.9   563.4
> R2(Modell4)
[1] 0.002947875
==========================================================================
> Modell5 <- lm(SO2~DwP,data=air)
> anova(Modell5)
Analysis of Variance Table
Response: SO2
          Df  Sum Sq Mean Sq F value  Pr(>F)
DwP        1  3009.9  3009.9  6.1691 0.01740 *
Residuals 39 19028.0   487.9
---
> R2(Modell5)
[1] 0.1365773
```

```
================================================================================
> Modell6 <- lm(SO2~T,data=air)
> anova(Modell6)
Analysis of Variance Table
Response: SO2
          Df  Sum Sq Mean Sq F value   Pr(>F)
T          1  4143.3  4143.3  9.0301 0.004624 **
Residuals 39 17894.6   458.8
---
> R2(Modell6)
[1] 0.1880091
================================================================================
> Modell7 <- lm(SO2~Manuf+Pop+T,data=air)
> anova(Modell7)
Analysis of Variance Table
Response: SO2
          Df Sum Sq Mean Sq F value    Pr(>F)
Manuf      1 9161.7  9161.7 39.7000 2.461e-07 ***
Pop        1 3759.5  3759.5 16.2909 0.0002621 ***
T          1  578.0   578.0  2.5045 0.1220313
Residuals 37 8538.7   230.8
---
> R2(Modell7)
[1] 0.6125468
>
```

(c) Welches der sieben Modelle ist das beste? .... Warum? (2)

(d) Welche der Voraussetzungen der linearen Regression (Aufgabe 1d) kann man für Modell1 mit den Graphiken in Abbildung 2 überprüfen und sind Sie erfüllt - woran sehen Sie das? (3)