

Klausur

Name, Vorname	Matrikelnummer
---------------	----------------

Bearbeiten Sie die folgenden Aufgaben direkt auf dem Aufgabenblatt oder auf einem separaten Zettel!

Aufgabe 1.

|10

Definieren Sie die folgenden statistischen Begriffe in einem Satz oder in einer Formel:

1. Histogramm
2. Empirische Varianz
3. Varianz einer Zufallsvariablen
4. Zentraler Grenzwertsatz
5. Binomial-Verteilung
6. Maximum-Likelihood-Methode
7. Overfitting

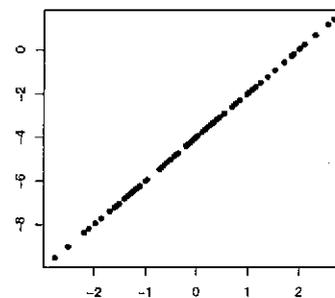
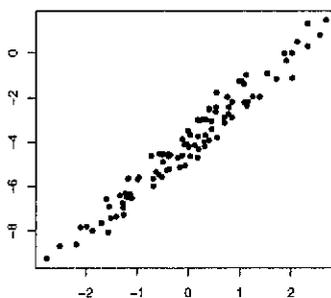
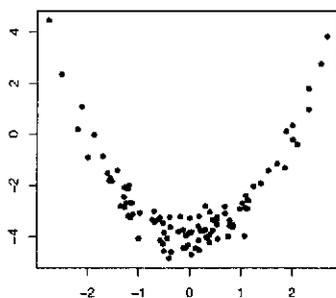
- | |
|---|
| 1 |
| 1 |
| 1 |
| 2 |
| 1 |
| 1 |
| 3 |

Aufgabe 2.

|10

1. Die Zufallsvariable X nimmt die Zustände -2, -1, 0, 1, und 2 mit den Wahrscheinlichkeiten 0.3, 0.1, 0.2, 0.1 und 0.3 an. Berechnen Sie Erwartungswert und Varianz von X .
2. $Y \sim N(0, 1)$ und $Z = 2 \cdot Y - 4$. Wie groß sind Erwartungswert, Median, Modalwert und Varianz von Z ?
3. Sind Y und Z unabhängige Zufallsvariablen? Sind Y und Z unkorreliert?
4. Welcher der folgenden Scatterplots stellt mögliche Realisierungen der Zufallsvariablen Y und Z dar? Kreuzen Sie die richtige Graphik an.

- | |
|---|
| 1 |
| 1 |
| 1 |
| 1 |



5. Die Zufallsvariablen S und T seien unabhängig und normalverteilt ($N(\mu, \sigma^2)$) mit $S \sim N(0, 4)$ und $T \sim N(0, 1)$. Wie groß ist dann der Loglikelihood-Ratio von S verglichen mit T für den Wert 4?

3

6. Die Zufallsvariablen V und W nehmen gleichverteilt Werte in $\{A, C, G, T\}$ an. Geben Sie **2 unterschiedliche Lösungen** für gemeinsame Verteilungen an, für die $P(V = W) = 0.5$ gilt.

3

Aufgabe 3.

10

1. Kommentieren Sie die folgenden Zeilen R-Code.

3

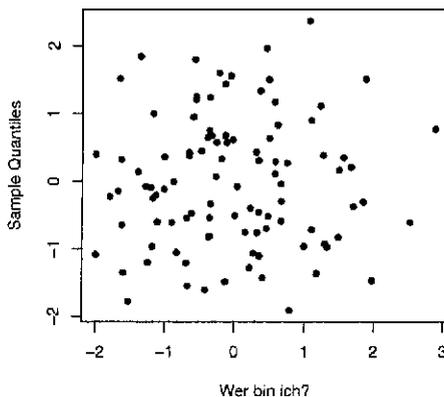
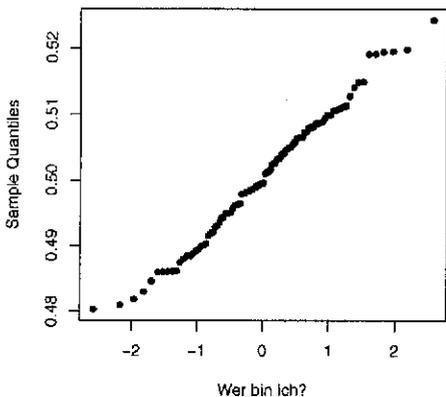
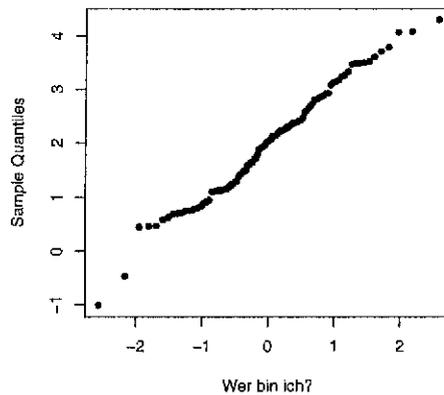
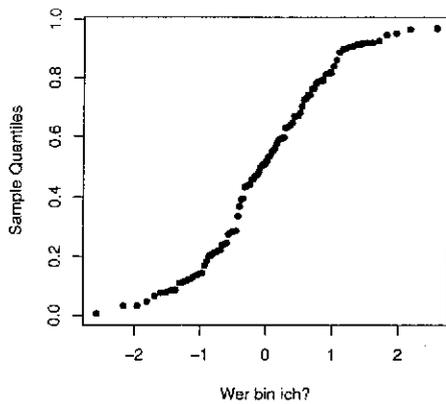
```
M <- matrix(runif(100*1000,min=0,max=1),100,1000)
a <- apply(M,1,function(x){sum(x>0.99)})
b <- apply(M,1,mean)
c <- apply(M,2,max)
```

2. Welche theoretische Verteilung hat a? Wie würden Sie den/die Parameter wählen?

3

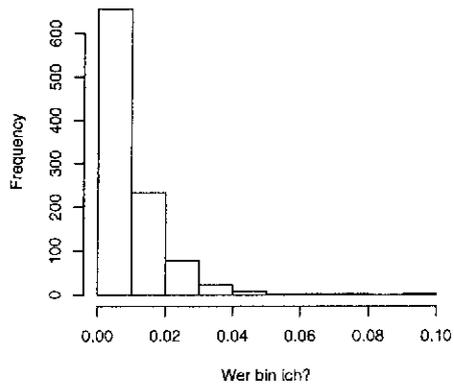
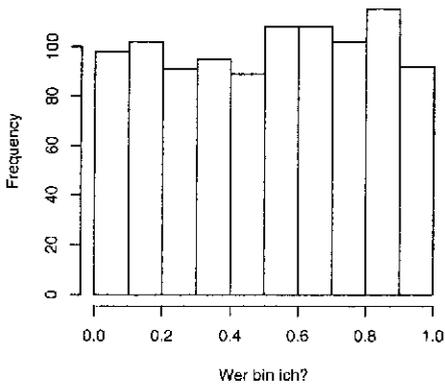
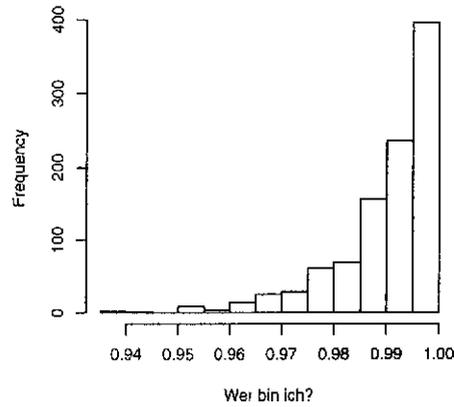
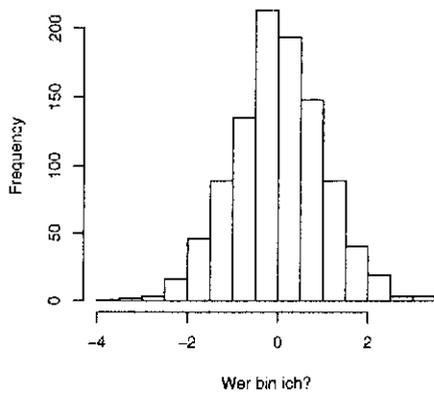
3. Welcher der folgenden Plots ist das Ergebnis von qqnorm(b)? Kreuzen Sie die richtige Graphik an.

2



4. Welcher der folgenden Plots ist das Ergebnis von `hist(c)`? Kreuzen Sie die richtige Graphik an.

2



Aufgabe 4.

10

In einer Studie wird die Expression eines bestimmten Genes bei kranken und gesunden Patienten untersucht. Von jedem Patient liegen sowohl der Genexpressionswert als auch die Konzentration des resultierenden Proteins vor. Wie stellen Sie fest, welches Verfahren sich besser zur Diagnose eignet?