Universität Tübingen
Mathematisches Institut
Dr. Stefan Keppeler

Sommersemester 08

Mathematik II für Biologen

Nachklausur am 6.10.2008

Bitte schreiben Sie nicht mit Bleistift. Bitte beginnen Sie jede Aufgabe (nicht Teilaufgabe) auf einer neuen Seite. Zusätzliches Papier ist jederzeit verfügbar. **Zeigen Sie auch stets Ihren Rechenweg!**

Es sind 78 Punkte erreichbar, hinreichend zum Bestehen sind 33 Punkte.

Hilfsmittel: Zwei beidseitig handbeschriebene A4-Blätter, nicht internetfähiger Taschenrechner.
Bearbeitungszeit: 90 Minuten. **Viel Erfolg!**

Aufgabe 1 (2+4+1+4+4+4 = 19 Punkte)
Eine Herde von 1000 Rindern werde als gesund bezeichnet, wenn nicht mehr als 80 Rinder krank sind. Sie möchten mithilfe eines statistischen Tests beweisen, dass die Herde krank ist. Dazu lassen Sie 15 Rinder untersuchen und bezeichnen die Anzahl der kranken Rinder dieser Stichprobe mit X.

a) Formulieren Sie die Nullhypothese H_0 und die Alternativhypothese H_A.
HINWEIS: Bezeichnen Sie dazu mit k die (unbekannte) Anzahl kranker Rinder innerhalb der Herde.

b) Um zu bestimmen, wie die Teststatistik unter der Nullhypothese verteilt ist, simulieren Sie mit MATLAB 10^5mal eine gesunde Herde:

```
>> X=[];
for wdh=1:10^5
  N=1000;
  k=80;
  herde=[ones(1,k),zeros(1,N-k)];
  n=15;
  stichprobe=[];
  for j=1:n
    rind=unidrnd(length(herde));
    stichprobe=[stichprobe,herde(rind)];
    herde(rind)=[];
  end;
  X=[X,sum(stichprobe)];
end
verteilung=hist(X,0:15)./wdh;
cumsum(verteilung)

ans =
 Columns 1 through 8
  0.2847    0.6609    0.8902    0.9743    0.9956    0.9995    0.9999    1.0000
 Columns 9 through 16
  1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
```

Bestimmen Sie den Verwerfungsbereich K für auf dem Signifikanzniveau $\alpha = 1\%$.

c) Von den untersuchten 15 Rindern sind fünf krank, d.h. $X = 5$. Wie entscheidet der Test?

d) Geben Sie außerdem den p-Wert an (für die Beobachtung $X = 5$).
HINWEIS: Es hilft die MATLAB-Ausgabe aus Aufgabenteil b.

e) Um ein 99%-Vertrauensintervall für k zu bestimmen, wiederholen Sie die MATLAB-Simulation aus Aufgabenteil b) für unterschiedliche k-Werte. Für `k=100` erhalten Sie

```
>> cumsum(verteilung)
ans =
 Columns 1 through 8
  0.2059    0.5478    0.8182    0.9462    0.9881    0.9980    0.9998    1.0000
 Columns 9 through 16
  1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000    1.0000
```

Gehören die Werte 60, 80, 100 und 120 zum 99%-Vertrauensintervall für k (wenn $X = 5$ beobachtet wurde)?

f) Wie groß ist für eine Herde mit $k = 100$ kranken Rindern die Wahrscheinlichkeit β für einen Fehler zweiter Art?
HINWEIS: Es hilft die MATLAB-Ausgabe aus Aufgabenteil e.

Aufgabe 2 (2+4+4+2+1 = 13 Punkte)

Am 28.9.2008 wurde in Bayern ein neuer Landtag gewählt. Dabei erhielt die CSU mehr Zweitstimmen als Erststimmen, bei der SPD war es umgekehrt. Kontingenztafel der erzielten Stimmen* (gerundet auf ganze 1000):

	Erststimmen	Zweitstimmen	Gesamtstimmen
CSU	2 266 000	2 334 000	4 600 000
SPD	1 017 000	954 000	1 971 000
Summe	3 283 000	3 288 000	6 571 000

Testen Sie mit einem χ^2-Test auf dem Signifikanzniveau 5%, ob das Verhältnis von Erst- und Zweitstimmen für die beiden Parteien signifikant voneinander abweicht. Geben Sie dazu

a) die Nullhypothese H_0 und die Alternativhypothese H_A,
b) die Kontingenztafel für die unter H_0 erwarteten Werte,
c) die Teststatistik (Wert und Rechenweg),
d) das Verwerfungskriterium und
e) den Testentscheid an.

Aufgabe 3 (9 Punkte)

Für eine Stichprobe von Zahlenpaaren haben Sie die Korrelationskoeffizienten r_{xy} nach Pearson den Rangkorrelationskoeffizienten $r_{xy}^{(\mathrm{SP})}$ nach Spearman sowie die Steigung m der Regressionsgeraden bestimmt. Leider haben Sie anschließend Ihre Daten verloren, und auch zu den Zahlenwerten haben Sie sich nicht notiert, welche Größen sie beschreiben. Die Werte waren $-5{,}3$, -1 und $-0{,}8$. Ordnen Sie diese den drei Größen r_{xy}, $r_{xy}^{(\mathrm{SP})}$ und m zu.

Für jede richtige Zuordnung erhalten Sie 3 Punkte, für jede falsche werden 3 Punkte abgezogen. Die Mindestpunktzahl für diese Aufgabe ist Null.

*Quelle: `http://www.landtagswahl2008.bayern.de/taba2990.html`

Aufgabe 4 (1+1+7+7+2 = 18 Punkte)

Die Längen (in mm) von 14 verschiedenen Eiern einer anderen Kuckucksart (als in der Klausur vom 23.7.08) sind

22 23,9 20,9 23,8 25 24 21,7 23,8 22,8 23,1 23,1 23,5 23 23 .

Untersuchen Sie – mithilfe statistischer Tests auf dem Signifikanzniveau $\alpha = 5\%$ – ob die Annahme, dass diese Kuckuckseier im Mittel ungefähr 23,75mm lang sind, mit den Daten vereinbar ist.

Vielleicht helfen Ihnen dabei die folgenden MATLAB-Zeilen:

```
>> X = [22 23.9 20.9 23.8 25 24 21.7 23.8 22.8 23.1 23.1 23.5 23 23]
>> mean(X)  % Mittelwert
ans = 23.1143
>> var(X)   % Stichprobenvarianz
ans = 1.1013
>> sqrt(var(X))
ans = 1.0494
```

a) Wie lautet die Nullhypothese?

b) Wie lautet die Alternativhypothese?

c) Führen Sie dazu einen t-Test durch:
Wie lautet die Formel für die Teststatistik? Berechnen Sie ihren Wert.
Bestimmen Sie den Verwerfungsbereich bzw. das Verwerfungskriterium.†
Wie entscheidet der Test?

d) Führen Sie nun auch einen Vorzeichen-Test durch:
Wie lautet die Definition der Teststatistik? Berechnen Sie ihren Wert.
Bestimmen Sie den Verwerfungsbereich.
HINWEIS: Sie dürfen die Faustregel für den Binomialtest verwenden, obwohl die Stichprobe nicht besonders groß ist. Alternativ hilft auch der MATLAB-Befehl

```
>> binocdf(0:14,14,.5)
ans =
Columns 1 through 8
  0.0001    0.0009    0.0065    0.0287    0.0898    0.2120    0.3953    0.6047
Columns 9 through 15
  0.7880    0.9102    0.9713    0.9935    0.9991    0.9999    1.0000
```

Wie entscheidet der Test?

e) Sollten Sie nun dem t-Test oder dem Vorzeichen-Test glauben? Begründen Sie Ihre Antwort in *einem Satz.*

† Eine hilfreiche Tabelle finden Sie auf der letzten Seite.

Aufgabe 5 (6+4+4+5 = 19 Punkte)

Unter der Teilnehmern eines großen Sportwettbewerbs seien 10% der Sportler gedopt. Ein Dopingtest erkennt mit 98%iger Wahrscheinlichkeit einen gedopten Sportler als solchen, stuft aber auch 1% der ungedopten Teilnehmer als gedopt ein (vgl. Nature **454** (2008) 692–693). Für einen zufällig ausgewählten Teilnehmer definieren wir die folgenden Ereignisse,

$$D := \text{“Teilnehmer ist gedopt”}$$
$$T := \text{“Der Dopingtest fällt für diesen Teilnehmer positiv aus (d.h. zeigt an, dass der Teilnehmer gedopt ist).”}$$

a) Geben Sie die Werte für die folgenden (bedingten) Wahrscheinlichkeiten an, $P[D]$, $P[D^C]$, $P[T|D]$, $P[T|D^C]$, $P[T^C|D]$, $P[T^C|D^C]$.

b) Wie groß ist die Wahrscheinlichkeit, dass unter 7 zufällig ausgewählten Teilnehmern genau zwei gedopt sind?

c) Ein nicht gedopter Sportler werde im Verlauf des (mehrtägigen) Wettbewerbs 8 mal getestet. Wie groß ist die Wahrscheinlichkeit, dass der Test mindestens einmal positiv ausfällt (obwohl der Teilnehmer nicht gedopt ist)?

d) Für einen zufällig ausgewählten Teilnehmer falle der Dopingtest positiv aus. Wie groß ist die Wahrscheinlichkeit, dass dieser Teilnehmer tatsächlich gedopt ist?

Einige Quantile $t_{d_f,q}$ der Studentschen t-Verteilung für d_f Freiheitsgrade:

$d_f \setminus q$	0,90	0,95	0,975	0,99
12	1,356	1,782	2,179	2,681
13	1,350	1,771	2,160	2,650
14	1,345	1,761	2,145	2,624
15	1,341	1,753	2,131	2,602
16	1,337	1,746	2,120	2,583