Übung 5

1. a) In der Aufgabe 3, Serie 1 haben wir die Anzahlen der von weitem beobachteten Albatrosse durch ein Poissonmodell beschrieben. Der unbekannte Parameter λ wurde durch $\overline{X} = 0.9$ geschätzt.

Anzahl der Albatrosse, k	beobachtete Anzahl der Tage, n_k	erwartete Anzahl der Tage, $n\widehat{p}_k$	$(n_k - n\widehat{p}_k)^2/n\widehat{p}_k$
0	21	20.33	0.022
1	18	18.30	0.005
2	7	8.23	0.185
3	3	2.47	0.114
4	1	0.56	0.355
≥ 5	0	0.12	0.117

Beurteile diese Anpassung quantitativ mit einem Anpassungstest.

b) Der Kapitän hat in seiner letzten Expedition im Mittel 2 Albatrosse pro Tag beobachtet. Deshalb würde er zu den obigen Daten eine Poissonverteilung mit Parameter $\lambda = 2$ anpassen:

Anzahl der Albatrosse, k	beobachtete Anzahl der Tage, n_k	erwartete Anzahl der Tage, $n\widehat{p}_k$	$(n_k - n\widehat{p}_k)^2/n\widehat{p}_k$
0	21	6.77	29.94
1	18	13.53	1.47
2	7	13.53	3.15
3	3	9.02	4.02
4	1	4.51	2.73
≥ 5	0	2.63	2.63

Überprüfe diese Vermutung, indem Du den Anpassungstest mit *bekanntem* Parameter durchführst.

c) Mit `R` können wir den P-Wert in a) und b) berechnen: a) 0.8, und b) $< 10^{-7}$. Beurteile die Güte der Anpassungen mit Hilfe dieser Grössen.

2. In einem physikalischen Praktikum ist der Strahlungshintergrund gemessen worden. $k \mathrel{\widehat{=}}$ Anzahl Zerfälle pro Zeitintervall, $n_k \mathrel{\widehat{=}}$ Anzahl Zeitintervalle mit k Zerfällen.

k	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
n_k	1	2	4	7	11	16	18	20	19	16	13	9	6	3	2	1	1

Die Vermutung liegt nahe, dass die Daten einer Poissonverteilung folgen.

a) Schätze $\widehat{\lambda}$ ($n = \sum n_k = 149$ und $\sum k \cdot n_k = 1396$).

b) Wir wollen nun das Konfidenzintervall für λ berechnen.

i) Korrekt wird das Konfidenzintervall folgendermassen berechnet:
Wenn die X_i ($i = 1, \ldots, n$) unabhängig und poissonverteilt mit Parametern λ_i sind, so ist $X = \sum_{i=1}^{n} X_i$ poissonverteilt mit Parameter $\widetilde{\lambda} = \sum_{i=1}^{n} \lambda_i$. (In unserem Fall sind alle $\lambda_i = \lambda$, also ist $\widetilde{\lambda} = n \cdot \lambda$.) Mit x bezeichnen wir die Beobachtung $x = \sum_{i=1}^{n} x_i$ ($= \sum k \cdot n_k$ in unserem Beispiel). Dieses x setzen wir in die Faustregel für das Konfidenzintervall ein und erhalten so ein Konfidenzintervall für $\widetilde{\lambda}$. Indem wir nun die Grenzen durch n dividieren, erhalten wir das Konfidenzintervall für λ.

ii) Man könnte versucht sein, das Konfidenzintervall zu berechnen, indem man das arithmetische Mittel der Beobachtungen direkt in die Faustregel einsetzt. Dies ist allerdings **falsch**! Führe dies trotzdem durch, um in c) das Resultat mit dem korrekten zu vergleichen.

c) Vergleiche die beiden Intervalle die wir erhalten haben. Gib eine qualitative und eine statistische Begründung, warum das Vorgehen in ii) nicht richtig sein kann.

3. Um Messstandards für Asbestkonzentrationen zu entwickeln, löste das amerikanische National Institute of Science and Technology Asbest in Wasser auf und sprühte dieses gleichmässig über einen Filter. Dann wurden 23 Proben von Filterstückchen mit einem Durchmesser von 3 mm genommen und die sich darauf befindenden Asbestfasern unter dem Elektronenmikroskop ausgezählt. Aufgrund von Erfahrung mit ähnlichen Daten dürfen wir annehmen, dass die Anzahlen Asbestfasern X_i bei Probe i $(i = 1, \ldots, 23)$ unabhängig und identisch poissonverteilt sind. Die Zählungen sind im Daten Frame `asbest` gespeichert:

x_i: `31 29 19 18 31 28 34 27 34 30 16 18 26 27 27 18 24 22 28 24 21 17 24`

(Quelle: J. Rice 1995, „Mathematical Statistics and Data Analysis", S. 247)

a) Gib mit Hilfe der Faustregel ein 95%-Vertrauensintervall für die mittlere Anzahl Asbestfasern pro Probe an ($n = 23$, $\sum x_i = 573$, $\overline{x} = 24.91$).

b) Obige Messreihe wurde nach einer Umstrukturierung im Fabrikationsprozess des Asbests ermittelt. Vor der Umstrukturierung hatte man einen (recht genau bestimmten) Mittelwert von 28 Asbestfasern pro Probe. Hat sich der Mittelwert durch die Umstrukturierung signifikant geändert? Begründe mit Hilfe des Vertrauensintervalles.

c) Beantworte die Frage in b) mittels eines Tests.

d) Berechne mittels `R` (vgl. Übung 3, Aufgabe 1) den P-Wert in c) mit der Formel $2 \cdot P_{\tilde{\lambda}}[X \leq x]$ (Gilt nur für $x < \tilde{\lambda}$). Was kannst Du aus dem P-Wert schliessen?

4. Das *International Rice Research Institute* (IRRI) auf den Philippinen entwickelt neue Reissorten, die sowohl ertragreich als auch resistent gegen Fäulnis und Insektenbefall sein sollen.

In einem Projekt hoffte man, eine gegen den „Brown Plant Hopper" (eine Insektenart) resistente Sorte zu züchten. Dazu wurden 400 neue Sorten durch Kreuzungen gebildet. Von jeder neuen Sorte wurden mehrere Pflanzen dem Insekt ausgesetzt. Aufgrund der beobachteten Resultate wurden die neuen Sorten in drei Klassen eingeteilt:

Klasse	Anzahl Sorten
Alle Pflanzen resistent	97
Gemischt; einige Pflanzen resistent, einige anfällig	212
Alle Pflanzen anfällig	91

Gemäss einem Modell des IRRI sind die neuen Sorten unabhängig, und jede Sorte ist mit einer Wahrscheinlichkeit von 25% resistent, mit 50% Wahrscheinlichkeit gemischt und mit 25% Wahrscheinlichkeit anfällig.

a) Stelle eine Tabelle der beobachteten und der erwarteten Anzahl Sorten in allen drei Klassen auf.

b) Teste mit einer Faustregel für jede einzelne Klasse, ob die Abweichung der beobachteten Anzahl von der unter dem IRRI-Modell erwarteten Anzahl zufällig sein kann.

c) Beurteile mit einem globalen Test auf dem 5%-Niveau, ob das IRRI-Modell zu den beobachteten Anzahlen passt. (Verwende wiederum eine Faustregel.) Dokumentiere mit allen nötigen Angaben!

d) Ist der P-Wert hier grösser oder kleiner als 0.05? Begründe!

e) Berechne den exakten P-Wert.
`R`-Anleitung: `pchisq(7.1, df=5)` liefert $P[X \leq 7.1]$, falls $X \sim \chi^2_5$

Abgabe: Montag, 8. Mai, in der Pause der Vorlesungsstunde