Arbeitsblatt HT 3 und 4: Binomial-Modell

1. Das Paradebeispiel für die Binomialverteilung ist der Wurf einer "fairen" Münze. Dabei ist die Wahrscheinlichkeit, Kopf (entspricht einem Erfolg) zu werfen, $\pi = 0.5$.

a) Zeichnen Sie die Wahrscheinlichkeitsverteilung für $\mathcal{B}\langle\pi = 0.5,\, m = 5\rangle$ und $\mathcal{B}\langle\pi = 0.5,\, m = 15\rangle$ auf.

R-Anleitung:

> `dbinom(0:5,size=5,prob=0.5)` # *Wahrscheinlichkeiten der Binomialverteilung mit* $m = 5$ *und* $\pi = 0.5$

b) Wie gross ist die Wahrscheinlichkeit, bei 5 Würfen genau 1x Kopf zu werfen?

c) Wie gross ist die Wahrscheinlichkeit, bei 15 Würfen genau 3x Kopf zu werfen?

d) Der eine Spieler hat bei 5 Würfen einmal Kopf geworfen. Der andere Spieler hat bei 15 Würfen 3 mal Kopf geworfen. Beide haben also "20% Erfolg" gehabt. Sind die Wahrscheinlichkeiten, dass jeweils diese Ereignisse (oder Situationen) eintreffen, deshalb auch gleich gross?

2. Die Fluggesellschaften wissen, dass in der Regel nicht alle Passagiere erscheinen, und überbuchen daher ihre Flugzeuge. Nach Angaben der KLM erscheinen die einzelnen Passagiere mit Wahrscheinlichkeit $\pi = 0.86$ (*SAS Communications, 4/1992*).

Wir nehmen an, dass die Passagiere unabhängig voneinander erscheinen. Wie gross ist dann die Wahrscheinlichkeit, dass ein Flugzeug mit 82 Plätzen und 85 verkauften Flugtickets alle wartenden Passagiere mitnehmen kann?

R-Anleitung:

> `pbinom(3,size=5,prob=0.5)` # *Kumulierte Wahrscheinlichkeiten der Binomialverteilung mit* $m = 5$ *und* $\pi = 0.5$ *von* $k = 0$ *bis* $k = 3$

3. Ein Unternehmen, das Wärmepumpen herstellt und verkauft, kann in etwa 50% der Fälle zusätzlich auch ein Service-Abonnement verkaufen. In der Filiale AA wurden im letzten Jahr bei 50 verkauften Wärmepumpen 15 Service-Abonnements unterzeichnet.

a) Mit welchem Modell lässt sich der Abschluss von Service-Abonnementen beschreiben?

b) Schätzen Sie die Erfolgswahrscheinlichkeit, dass Filiale AA im letzten Jahr einen Service-Vertrag abschliessen konnte.

c) Ist die beobachtete Erfolgswahrscheinlichkeit der Filiale AA plausible, wenn man von einer Erfolgswahrscheinlichkeit von 50% ausgeht? Testen Sie dazu auf dem 5% Niveau

(i) mit der R-Funktion `binom.test(...)`

(ii) (freiwillig) mit der Faustregel B1

d) Bestimmen Sie ein 95%-Vertrauensintervall für π mit R-Funktion `binom.test()` (und freiwillig mit Faustregel B2).

e) Erläutern Sie in wenigen Worten die Bedeutung des Vertrauensintervalls!

Bitte wenden!

4. Beispiel: Unfalldaten mit tötlichen Folgen auf Zürcher Strassen von 2010 und 2011 (siehe Folien). Auf den Strassen des Kantons Zürich starben im Jahre 2011 35 Personen gegenüber 37 im Vorjahr. Kann die Abnahme "rein zufällig" sein oder ist sie "signifikant"? - Beantworten Sie die Frage auf dem 5% Niveau.

5. Können unsere im Unterricht erhobenen "Kieselstein-"Daten mit einem Binomialmodell beschrieben werden? – Führen Sie zur Beantwortung dieser Frage einen Anpassungstest auf dem 5% Signifikanz-Niveau durch.

R-Anleitung:

```
> ## File "CAS-DA_ModulA2-HT3-4_chisq-bin.R" von Moodle, Unterordner R-Skript
> ## herunterladen und in den Unterordner ablegen, wo Sie die Übungen machen.
> ##File einlesen mit
> source("CAS-DA_ModulA2-HT3-4_chisq-bin.R")  # --> chisq.bin
> ## Anpassungstest mit der R-Funktion chisq.bin() vornehmen
```

6. Würfel-Experiment: Jede Teilnehmerin und jeder Teilnehmer des Kurses würfelt 10x mit dem orangen Würfel und zählt die Anzahl Würfe mit einer Augenzahl kleiner oder gleich 3.

a) Tragen Sie die Ergebnisse in eine gemeinsame Tabelle ein.

b) Vergleichen Sie die Resultate aus dem Experiment mit dem theoretischen Modell ($\mathcal{B}\langle m = 10, \pi = 0.5\rangle$) zuerst grafisch und dann mit einem statistischen Test.

7. Zeitungsartikel im Landboten vom 18. Mai 2013.
Im Kasten unten links wird auf ein Beispiel aus dem Buch von Christian Hesse verwiesen. Hier soll ein ähnliches Beispiel durchgerechnet werden:

Da ein neues medizinisches Test-Verfahren sich bei der Früherkennung einer Krankheit als wirksam gezeigt hat, wird eine Vorsorgeuntersuchung der Bevölkerung vorgeschlagen. Die Wahrscheinlichkeit, dass der Test korrekt jemanden mit der Krankheit als positiv identifiziert ist 0.99, und die Wahrscheinlichkeit, dass der Test korrekt jemanden ohne die Krankheit als negativ identifiziert, ist 0.95. Die Krankheit kommt in der allgemeinen Bevölkerung mit einer Wahrscheinlichkeit von 0.0001 vor.

Sie unterziehen sich ebenfalls dem Test, und das Ergebnis ist positiv. Wie gross ist die Wahrscheinlichkeit, dass Sie die Krankheit haben?$^{(*)}$

a) Bezeichne K das Ereignis, dass Sie die Krankheit haben, und S das Ereignis, dass der Test positiv anzeigt. Wie gross ist $P\langle K\rangle$? Wie lautet die Fragestellung $^{(*)}$ in der technischen Notation?

b) Bestimmen Sie folgende Wahrscheinlichkeiten: $P\langle S|K\rangle$, $P\left\langle \bar{K}\right\rangle$ und $P\left\langle S|\bar{K}\right\rangle$, wobei $\bar{K}$ das Ereignis ist, dass Sie die Krankheit nicht haben.

c) Wenden Sie das Theorem von Bayes auf $P\langle K|S\rangle$ an und setzen Sie zur Berechnung von $P\langle K|S\rangle$ die oben bestimmten Zahlen ein. Wie gross ist die Wahrscheinlichkeit, dass Sie die Krankheit haben?

d) Lesen Sie den Zeitungsartikel aus dem Landboten durch.
Was hat dieses Beispiel mit dem Haupttext, respektive mit den ersten Absatz im Kasten unten links zu tun?

8. Um die Wahrnehmung von industriellen Gerüchen zu untersuchen, wurde in drei Regionen, die in der Nähe von Industrieanlagen liegen, je eine Umfrage durchgeführt. Auf die Frage "Wie oft nehmen Sie Gerüche wahr, die von der Industrieanlage stammen?", hatten die befragten Personen folgende Antwortmöglichkeiten:

(1) jeden Tag
(2) mindestens einmal in der Woche
(3) mindestens einmal im Monat
(4) weniger als einmal im Monat
(5) überhaupt nicht

Die Resultate sind in der folgenden Tabelle festgehalten:

	Geruchswahrnehmung				
Beobachtet	(1)	(2)	(3)	(4)	(5)
Region I	20	28	23	14	12
Region II	14	34	21	14	12
Region III	4	12	10	20	53

Kann die Verteilung der Geruchswahrnehmung in allen drei Regionen die gleiche sein? (Testen Sie auf dem 5% Niveau.) Wo liegen die Unterschiede? (Machen Sie einen Mosaic-Plot.)

R-Anleitung:

```
> kt <- rbind(c(20,28,23,14,12), c(14,34,21,14,12), c( 4,12,10,20,53))
> chisq.test(kt)  # χ²-Test in Kontingenztafeln
```

9. Wir untersuchen die Erfolgswahrscheinlichkeiten bei einer Zulassungsprüfung an der Universität Berkeley im Jahr 1973. 8442 Männer und 4321 Frauen machten die Zulassungsprüfung. 44% der Männer und 35% der Frauen bestanden sie. Falls man davon ausgeht, dass Männer und Frauen gleich gut qualifiziert sind, so weisen diese Zahlen auf eine Benachteiligung der Frauen in der Zulassungsprüfung hin. (Die University of California, Berkeley wurde entsprechend verklagt.)

Die Departementsvorsteher/innen wurden deshalb aufgefordert, ihre Zahlen offenzulegen, um einer möglichen Benachteiligung der Frauen in der Zulassungsprüfung nachzugehen. Die Analyse der entsprechenden Zahlen konnte diesen Verdacht nicht bestätigen. In einigen Departementen werden Frauen bevorzugt und in anderen Männer. Insgesamt konnte man eher eine Benachteiligung der Männer herauslesen! Ein Auszug der Daten aus über 100 Departementen ist in folgender Tabelle gegeben:

Departement	Männer		Frauen	
	Bewerber	Zugelassene	Bewerberinnen	Zugelassene
A	825	511	108	89
B	560	353	25	17
C	325	120	593	202
D	417	138	375	131
E	191	53	393	94
F	373	22	341	24

Was steckt hinter diesem Phänomen, das auch als Simpson-Paradoxon bekannt ist? Um diese Frage zu beantworten, betrachten wir nur die Departemente A und F (zusammen bilden sie die "Universität") .

a) Erklären Sie in Worten das Phänomen, indem Sie die Anteile der Zugelassenen mit den entsprechenden Anteilen der weiblichen Kandidaten in den Departementen vergleichen.

Bitte wenden!

In diesem Beispiel können wir auch das Verständnis für bedingte Verteilungen verbessern. Dazu bilden wir relative Häufigkeiten und benutzen diese als entsprechende Wahrscheinlichkeiten.

b) Bestimmen Sie aus dem Zahlenmaterial die Wahrscheinlichkeiten, dass ein Kandidat weiblich bzw. männlich ist. (Formal: z.B. $P\langle\text{weiblich}\rangle = ?$) Bestimmen Sie aus dem Zahlenmaterial die Wahrscheinlichkeit, dass ein Kandidat weiblich, bzw. männlich, ist und zugelassen wird. (Formal: z.B. $P\langle\text{weiblich und zugelassen}\rangle = ?$)

c) Bestimmen Sie aus den Wahrscheinlichkeiten, die in Teilaufgabe (b) erarbeitet wurden, die Wahrscheinlichkeiten, dass eine Frau, bzw. ein Mann, zugelassen wird.
(Formal: z.B. $P\langle\text{zugelassen} \mid \text{weiblich}\rangle = ?$) Vergleichen Sie diese berechneten Wahrscheinlichkeiten mit den entsprechenden beobachteten relativen Häufigkeiten.

d) Führen Sie die Teilaufgaben (b) und (c) für die Departemente A und F jeweils separat durch. Kommentieren Sie!

e) (Freiwillig) Stellen Sie alle Daten in der Tabelle mit einem Mosaic-Plot dar und beschreiben Sie, was Sie feststellen.

10. (Freiwillig) Eine gewisse Konzentration einer Chemikalie wie z.B. Zyanid im Wasser führt bei 25% der Fische innerhalb von 24 Stunden zum Tod. Weil die Chemikalie im Wasser sehr schwierig nachzuweisen ist, verwendet man Fische als Bioindikatoren. Dazu werden 4 Fische in einen Wassertank gebracht, dessen Wasser mit der oben erwähnten Chemikalie und Konzentration verschmutzt ist (wir nehmen an, dass sich die Fische unabhängig voneinander verhalten).

a) Geben Sie die Wahrscheinlichkeit an, dass alle 4 Fische 24 Stunden überleben.

b) Geben Sie die Wahrscheinlichkeit an, dass höchstens 2 Fische 24 Stunden überleben.

c) In Teilaufgabe a) haben wir gesehen, dass zu oft alle Fische überleben werden. Deshalb ist diese Anordnung noch nicht als Bioindikator geeignet. Wie viele Fische müssten mindestens in den Wassertank gegeben werden, dass bei der Gegenwart der Chemikalie mit einer Wahrscheinlichkeit von höchstens 0.05 alle Fische 24 Stunden überleben?

11. (Freiwillig) Bei einer Untersuchung werden Wasserproben auf Verunreinigungen untersucht. Da nur 2 Prozent aller Proben verunreinigt sind, wird vorgeschlagen, von 10 Einzelproben jeweils die Hälfte der Probe zu einer Sammelprobe zusammenzufassen und zunächst nur die Sammelprobe zu untersuchen. Wird in der Sammelprobe keine Verunreinigung festgestellt, so ist die Untersuchung für die 10 Einzelproben beendet. Im anderen Fall müssen die anderen 10 übriggebliebenen Hälften in 10 Einzel-Untersuchungen geprüft werden.

a) Wie gross ist die Wahrscheinlichkeit, in der Sammelprobe keine Verunreinigung zu finden (unter der Annahme, dass die Einzelproben unabhängig voneinander sind)?

b) Sei die Zufallsvariable Y die Anzahl benötigter Untersuchungen für 10 Einzelproben. Welche Werte kann Y annehmen, und mit welchen Wahrscheinlichkeiten treten sie auf?

c) Wie viele Analysen werden „im Durchschnitt“ für jede Sammelprobe benötigt ?
Wie viele Analysen werden durch die Bildung von Sammelproben „im Durchschnitt“ eingespart?