

ÜBUNGEN ZUR WAHRSCHEINLICHKEITSTHEORIE UND STATISTIK FÜR BIOLOGEN

Blatt 8

1. Aufgabe Entgegen der landläufigen Meinung gibt es nicht nur grüne Marsmenschen, sondern auch rote und blaue. Grüne und blaue Marsmenschen scheinen sich zudem in ihrer Körpergröße zu unterscheiden. Belegen Sie dies mit einem statistischen Test; eine Vermutung kann man schließlich nicht publizieren! Der Datensatz `mars.txt` wurde bereits in Aufgabe 1 des vierten Übungsblattes mit folgenden Ergebnissen untersucht:

	grün	blau	rot
Anzahl	20	12	10
Mittelwert	60.87	71.35	62.31
Standardfehler	1.61	2.60	1.95

- Bestimmen Sie (ohne R, mit Taschenrechner) die 95%-Konfidenzintervalle der Schätzungen der durchschnittlichen Körpergrößen grüner, blauer und roter Marsbewohner. Die Tabelle zur t-Verteilung im Anhang liefert die nötigen Quantile.
- Ist der Stichprobenmittelwert der blauen Marsianer auf dem 95%-Niveau signifikant von 65 verschieden? Dies kann alleine mit den 95%-Konfidenzintervallen beantwortet werden.
- Sie glauben, dass sich die mittleren Körpergrößen von grünen und blauen Marsbewohnern unterscheiden. Mit welchem Test können Sie dies mit Hilfe der vorliegenden Daten untermauern? Stellen Sie zunächst die Nullhypothese auf. Benutzen Sie dann einen geeigneten R-Befehl, um den p-Wert zu berechnen.

2. Aufgabe Ein HIV-Test habe eine Sensitivität von 99,9% (d.h. er schlägt bei 99,9% der Infizierten an) und eine Spezifität von 95% (d.h. er schlägt bei 5% der nicht-Infizierten an). 8% einer Population seien infiziert. Wenn eine zufällig aus der Population gegriffene Person ein positives Testergebnis hat, wie wahrscheinlich ist es dann, dass sie tatsächlich infiziert ist?

3. Aufgabe Mukoviszidose ist eine menschliche Erbkrankheit, die von einem Gendefekt auf Chromosom 7 hervorgerufen wird. Das Allel für Mukoviszidose ist rezessiv, d.h. nur homozygote Individuen erkranken tatsächlich. Nehmen wir an, eine Population befindet sich (bezüglich dieses Gens) im Hardy-Weinberg-Gleichgewicht, und unter je 3000 Geburten findet sich ein an Mukoviszidose erkranktes Kind.

- Welcher Anteil der Population trägt genau eine Kopie des defekten Allels?
- Nehmen wir an, ein gesundes Paar habe ein an Mukoviszidose erkranktes Kind. Wie wahrscheinlich ist es dann, dass ein weiteres Kind ebenfalls krank sein wird?
- Nehmen wir an, ein gesundes Paar habe bereits ein gesundes Kind. Wie wahrscheinlich ist es dann, dass das zweite Kind krank sein wird?
- Eine gesunde Person habe gesunde Eltern, aber einen an Mukoviszidose leidenden Onkel. Wie wahrscheinlich ist es dann für ihr Kind (unter der Hypothese der Zufallspaarung in der Population), an Mukoviszidose zu leiden?

4. Aufgabe Um den Ertrag dreier verschiedener Düngersorten zu testen, wurde ein Acker in 18 gleichartige Parzellen eingeteilt, die Parzellen zufällig in drei Gruppen eingeteilt und dann jede Gruppe mit einem anderen Düngemittel behandelt. Folgende Tabelle zeigt die logarithmierten Ernteerträge:

Dünger	Ertrag
A	2.89 2.81 2.78 2.89 2.77
B	2.73 2.88 2.89 2.82 2.90 2.85
C	2.84 2.81 2.80 2.66 2.83 2.58 2.80

Stellen Sie von Hand die zugehörige ANOVA-Tafel auf, bestimmen Sie den Wert der F -Statistik und testen Sie die Hypothese, dass die Dünger sich nicht bezüglich des Ertrags unterscheiden. (Im Anhang finden Sie eine Tabelle der 95%-Quantile der Fisher-Verteilung für verschiedene Freiheitsgrade.)

5. Aufgabe Sie besuchen eine fremde Stadt, von der Sie wissen, dass die Taxis dort fortlaufend mit $1, 2, \dots, N$ nummeriert sind, Sie kennen aber nicht die Gesamtzahl N . Als Sie das Flughafengebäude verlassen, sehen Sie 12 Taxis, mit den Nummern 455, 190, 39, 542, 370, 289, 274, 237, 374, 899, 741, 706.

- Wie groß ist die Wahrscheinlichkeit, genau diese Taxinummern in dieser Reihenfolge zu sehen, wenn $N = 912$ ist? (Modellieren Sie das Erblicken eines Taxis durch "Ziehen mit Zurücklegen", d.h. ein Taxi kann prinzipiell auch zweimal gesehen werden.)
- Wie groß ist die Wahrscheinlichkeit, genau diese Taxinummern in dieser Reihenfolge zu sehen für beliebiges N ? (Insbesondere: Was passiert mit der Wahrscheinlichkeit der Beobachtung für $N < 899$?)
- Berechnen Sie auf der Basis Ihrer Beobachtung den Maximum-Likelihood-Schätzer für N .
- Wie sähe der ML-Schätzer im allgemeinen Fall aus: Sie sehen n Taxis mit Nummern x_1, x_2, \dots, x_n ?

6. Aufgabe Auf einem Abschnitt nicht-kodierender DNS der Länge 1000 Basenpaare werden zwischen Mensch und Schimpanse $N = 23$ Mutationen gezählt. Ein Wissenschaftler argumentiert: „Die Anzahl Mutationen auf einem Stück der Länge 1000 ist Binomial-verteilt mit Erwartungswert $np \approx n\hat{p} = 23$ und Standardabweichung $\sqrt{np(1-p)} \approx \sqrt{n\hat{p}(1-\hat{p})} \approx \sqrt{n\hat{p}} = \sqrt{23}$. Mit der asymptotischen Normalität folgt, dass $[13.6, 32.4]$ ein 95%-Konfidenzintervall für die erwartete Anzahl Unterschiede auf einem Stück DNS der Länge 1000 aus derselben genomischen Region ist.“

- Erklären Sie, wie dieses Intervall berechnet wurde.
- Erzeugen Sie unter der Annahme, dass das wahre $p = 0.023$ ist, zufällige Anzahlen von Mutationen und berechnen Sie aus den simulierten Daten Konfidenzintervalle nach dieser Methode (Hinweis: der R-Befehl `rbinom` simuliert binomialverteilte Zufallsgrößen). Mit welcher Wahrscheinlichkeit enthält das so konstruierte Konfidenzintervall den wahren Wert?
- Führen Sie Teil b) auch für $p = 0.007$ durch.

Quantile der Student- t - und der F -Verteilungen

95%-Quantil der Student- t -Verteilung in Abhängigkeit der Anzahl Freiheitsgrade

Freiheitsgrade	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Quantil	6.31	2.92	2.35	2.13	2.02	1.94	1.89	1.86	1.83	1.81	1.80	1.78	1.77	1.76
Freiheitsgrade	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Quantil	1.75	1.75	1.74	1.73	1.73	1.72	1.72	1.72	1.71	1.71	1.71	1.71	1.7	1.7

97.5%-Quantil der Student- t -Verteilung in Abhängigkeit der Anzahl Freiheitsgrade

Freiheitsgrade	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Quantil	12.71	4.3	3.18	2.78	2.57	2.45	2.36	2.31	2.26	2.23	2.20	2.18	2.16	2.14
Freiheitsgrade	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Quantil	2.13	2.12	2.11	2.10	2.09	2.09	2.08	2.07	2.07	2.06	2.06	2.06	2.05	2.05

99%-Quantil der Student- t -Verteilung in Abhängigkeit der Anzahl Freiheitsgrade

Freiheitsgrade	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Quantil	31.82	6.96	4.54	3.75	3.36	3.14	3	2.9	2.82	2.76	2.72	2.68	2.65	2.62
Freiheitsgrade	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Quantil	2.6	2.58	2.57	2.55	2.54	2.53	2.52	2.51	2.50	2.49	2.49	2.48	2.47	2.47

99.5%-Quantil der Student- t -Verteilung in Abhängigkeit der Anzahl Freiheitsgrade

Freiheitsgrade	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Quantil	63.66	9.92	5.84	4.60	4.03	3.71	3.5	3.36	3.25	3.17	3.11	3.05	3.01	2.98
Freiheitsgrade	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Quantil	2.95	2.92	2.90	2.88	2.86	2.85	2.83	2.82	2.81	2.80	2.79	2.78	2.77	2.76

95%-Quantile der Fisher-Verteilung Die folgende Tabelle zeigt (auf 2 Nachkommastellen gerundet) das 95%-Quantil der Fisher-Verteilung mit k_1 und k_2 Freiheitsgraden (k_1 Zähler- und k_2 Nennerfreiheitsgrade)

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	11
1	161.45	199.5	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	242.98
2	18.51	19	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.4	19.4
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6	5.96	5.94
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.7
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.1	4.06	4.03
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.6
8	5.32	4.46	4.07	3.84	3.69	3.58	3.5	3.44	3.39	3.35	3.31
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.1
10	4.96	4.1	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94
11	4.84	3.98	3.59	3.36	3.2	3.09	3.01	2.95	2.9	2.85	2.82
12	4.75	3.89	3.49	3.26	3.11	3	2.91	2.85	2.8	2.75	2.72
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63
14	4.6	3.74	3.34	3.11	2.96	2.85	2.76	2.7	2.65	2.6	2.57
15	4.54	3.68	3.29	3.06	2.9	2.79	2.71	2.64	2.59	2.54	2.51
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46
17	4.45	3.59	3.2	2.96	2.81	2.7	2.61	2.55	2.49	2.45	2.41
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37
19	4.38	3.52	3.13	2.9	2.74	2.63	2.54	2.48	2.42	2.38	2.34
20	4.35	3.49	3.1	2.87	2.71	2.6	2.51	2.45	2.39	2.35	2.31