Aufgabensammlung zur Übung „Statistische Messdatenauswertung für Biotechnologen“

Version 2012b

Inhaltsverzeichnis

1 Statistik-Grundlagen

A 1 Lage- und Streuungsparameter

Gegeben sei folgende Messreihe:

4, 5, 4, 6, 7, 4, 3, 5, 5, 4, 5, 6, 5, 12, 5, 4, 5, 3, 6, 5

Gesucht sind die statistischen Kennwerte Mittelwert, Varianz, Streuung, Modalwert, Medianwert, Spannweite, oberes und unteres Quartil sowie der Quartilsabstand.

A 2 Lage- und Streuungsparameter – Blattlängen

An 40 zufällig ausgewählten Blättern einer Rotbuche (*Fagus sylvatica*) wurde die Blattlänge (ohne Stiel) bestimmt. Dabei wurden folgende Messwerte erhalten (Längen in mm):

56,0	58,5	39,5	56,0	51,5	48,5	62,5	55,0	50,5	53,5
56,5	56,0	59,5	66,0	51,5	49,5	49,0	54,5	50,5	49,5
43,5	68,0	49,5	59,0	52,5	57,0	50,5	42,5	45,5	59,0
52,5	48,5	49,0	59,0	53,5	54,0	65,0	55,5	56,0	56,5

a) Berechnen Sie für die Originaldaten den Mittelwert $\overline{x}$, die Varianz S^2 bzw. die Streuung S, den Medianwert $\tilde{x}$, , den Modalwert M, die Spannweite R, das untere und das obere Quartil $Q_1, Q3$ sowie den Quartilsabstand Q!

b) Klassifizieren Sie die Originalwerte durch Einführung von Klassen der Breite $b = 5\,\mathrm{mm}$ mit den Klassenmitten 40,0 mm, 45,0 mm, 50,0 mm usw.! Stellen Sie diese in einem Histogramm graphisch dar!

c) Nehmen Sie an, dass die Längen der Blätter näherungsweise einer Normalverteilung genügen. In welchem Bereich sind, wenn Sie weitere Messungen durchführen, etwa 95 % aller Messwerte zu erwarten? In welchem Bereich wird der wahre Mittelwert der Grundgesamtheit aller Blätter liegen?

A 3 Gaußsche Normalverteilung – Vertrauensbereich

Bei einer Messung wurde folgende Messreihe aufgenommen:

79,81	80,51	80,06	79,57	80,36	79,13	80,14	80,25	79,41	79,92	80,00	79,98	80,57
79,00	81,00	80,48	80,02	80,36	79,74	79,53	80,09	79,29	80,74	80,90	79,69	79,88

a) Berechnen Sie die Schätzwerte des Erwartungswertes und der Standardabweichung der Stichprobe.

b) Wie groß ist der Vertrauensbereich des Mittelwertes bei einer statistischen Sicherheit von 99 %?

c) Mit welcher Wahrscheinlichkeit liegen die Messwerte im Bereich von 79,8 bis 80,2?

A 4 Fadenlänge von Seidenkokons

In einer Seidenspinnerei werden Rohfäden von Seidenkokons abgewickelt und zu Seidenfäden versponnen. Es wird angenommen, dass die verwertbare Fadenlänge pro Kokon (bei der betreffenden Seidenraupenart) durch eine normalverteilte Zufallsvariable mit Erwartungswert $\mu = 800\,\mathrm{m}$ und Varianz $\sigma^2 = 6400\,\mathrm{m}^2$ angemessen beschrieben werden kann.

a) Berechnen Sie die Wahrscheinlichkeit dafür, dass die verwertbare Fadenlänge eines beliebig herausgegriffenen Kokons mindestens 700 m beträgt und die Wahrscheinlichkeit dafür, dass sie 1000 m übersteigt.

b) Bestimmen Sie unter Zuhilfenahme geeigneter Annahmen eine Mindestgrenze x_u und eine Höchstgrenze x_o für die Gesamtlänge der von 100 000 Kokons abgewickelten verwertbaren Seidenfäden, die mit 95 % Wahrscheinlichkeit eingehalten werden. Man wähle diese Grenzen so, dass die Wahrscheinlichkeit für eine Unterschreitung von x_u und die Wahrscheinlichkeit für eine Überschreitung von x_o gleich groß sind.

c) Wieviele Kokons müssen abgewickelt werden, damit mit einer Wahrscheinlichkeit von mindestens 99 % die Gesamtlänge der verwertbaren Seidenfäden mindestens 100 000 km beträgt?

2 Abweichungsfortpflanzung

A 5 Zylindervolumen

Aus Stichprobenmessungen des Durchmessers d (Umfang $n_d = 10$) und der Länge l (Umfang $n_l = 5$) eines Zylinders wurden folgende Werte erfasst.

i	1	2	3	4	5	6	7	8	9	10
$d_i/\,\mathrm{mm}$	31,00	29,20	28,80	30,30	31,20	29,00	30,50	29,70	30,80	29,50
$l_i/\,\mathrm{mm}$	410,00	394,00	401,00	404,00	391,00					

Bestimmen Sie die Messunsicherheit (das Konfidenzintervall) c_v bei der Ermittlung des Zylindervolumens $V(d,l) = \frac{d^2\pi}{4}l$.
Die geforderte statistische Sicherheit beträgt $P\% = (1-\alpha) = 0,95 = 95\,\%$ $(\alpha = 0,05)$.

A 6 Unbekanntes Prüfteil

Ein unbekanntes, quaderförmiges Prüfteil mit den Parametern Länge l, Breite b und Höhe h wurde fünfmal gemessen (Volumen $V = lbh$):

i	1	2	3	4	5
Länge $x_1(l)/\,\mathrm{dm}$	1,1	1,0	0,9	1,0	1,0
Höhe $x_2(h)/\,\mathrm{dm}$	0,3	0,3	0,3	0,2	0,3
Breite $x_3(b)/\,\mathrm{dm}$	0,7	0,6	0,6	0,7	0,6

Die Masse des Prüfteils ist bekannt: $m = 1,402\,\mathrm{kg} \pm 0,041\,\mathrm{kg}$

Berechnen Sie nach den Gesetzen der Abweichungsfortpflanzung den Erwartungswert der Dichte des Prüfteils ($\rho = m/V$; $[\rho] = \mathrm{kg}/\,\mathrm{dm}^3$) und den dazugehörigen Vertrauensbereich mit $P\% = 1 - \alpha = 95\,\%$ statistischer Sicherheit.

A 7 Torsionsmodul Stahldraht

Für einen Stahldraht soll der Torsionsmodul G, eine der elastischen Konstanten eines Festkörpers, bestimmt werden. Hierzu wird an den Stahldraht eine Kreisscheibe gehängt und das System zu einer Drehschwingung angeregt. Die Periodendauer T dieser Drehschwingung wird mit einer Stoppuhr gemessen.

T/s	11,4	11,2	11,0	11,0	11,5	11,2	11,1

Der Torsionsmodul G lässt sich dann aus folgender Gleichung bestimmen:

$$G = \frac{8\pi L\Theta}{R^4 T^2}$$

mit

L : Länge des Drahtes
Θ : Trägheitsmoment der Kreisscheibe
R : Radius des Drahtes

Der Drahtdurchmesser d wurde mit einer Schiebelehre und die Drahtlänge L mit einem Maßstab in fünf Messungen bestimmt:

d/mm	0,490	0,493	0,491	0,493	0,495
L/cm	120,8	121,0	121,2	120,9	121,2

Für das Trägheitsmoment Θ der Kreisscheibe wird ein Wert von $\Theta = 1,3 \cdot 10^{-3}\,\mathrm{kg\,m^2}$ (aus acht Messungen bestimmt) mit einem Vertrauensbereich von $c_\Theta = 1,5\,\%$ bei einer Aussagesicherheit von 99 % angegeben.

Berechnen Sie den Torsionsmodul G für diesen Stahldraht inkl. Vertrauensbereich bei einer Aussagesicherheit von 95 %.

3 Student‘scher t-Test

3.1 t-Test für den Erwartungswert

A 8 Gewicht von Kohlköpfen

Eine Ladenkette fordert von den Erzeugern für Chinakohl ein mittleres Kopfgewicht von mindestens $1000\,\mathrm{g}$. Es wird eine Stichprobe (mit Umfang $n = 7$) aus einer Lieferung gezogen und folgende Kopfgewichte bestimmt:

Kopf i	1	2	3	4	5	6	7
Gewicht x_i/g	920	975	1030	910	955	925	1010

Überprüfen Sie, ob bei einem Signifikanzniveau von $\alpha = 0,05$ das mittlere Kopfgewicht der Forderung entspricht.

A 9 Ozongehalt der Luft

In München wurde an einem Sommertag der Ozongehalt der Luft an fünf verschiedenen Meßstellen bestimmt:

Meßstelle	1	2	3	4	5
Ozongehalt / $(\mu\mathrm{g}/\,\mathrm{m}^3)$	114	128	130	118	123

Überschreitet der mittlere Ozongehalt bei einem 5 %-Signifikanzniveau den Grenzwert von $120\,\frac{\mu\mathrm{g}}{\mathrm{m}^3}$, so muss eine Warnung der Bevölkerung erfolgen.

Formulieren Sie das Testproblem und überprüfen Sie es mit einem geeigneten Testverfahren.

3.2 t-Test für Vergleich zweier Erwartungswerte unabhängiger Stichproben

A 10 Futtereinfluss auf Gewichtszunahme bei Schweinen

In einem Versuch wurden zwei Gruppen von jeweils $n = 8$ Schweinen gleicher Rasse mit Futter unterschiedlichen Proteingehalts gemästet. Untersuchen Sie, ob ein hoher Proteingehalt auf einem Signifikanzniveau von $\alpha = 0,01$ eine höhere mittlere tägliche Gewichtszunahme der Tiere verursacht.

Futter	mittlere tägliche Gewichtszunahme/ g							
hoher Proteingehalt H_i	715	683	664	659	660	762	720	715
niedriger Proteingehalt N_i	684	655	657	531	638	601	611	651

A 11 Medikamenteneinfluss auf Tierwachstum

Bei der Untersuchung der Auswirkungen eines Medikaments auf das Wachstum wurde acht jungen Versuchstieren ein Medikament verabreicht. Einer Vergleichsgruppe mit ebenfalls acht Tieren wurde das Medikament nicht gegeben. Anschließend wurde die Zunahme der Körperlängen der einzelnen Tiere gemessen.

i	1	2	3	4	5	6	7	8
$x_i/\,\mathrm{cm}$	69	73	68	78	76	71	74	67
$y_i/\,\mathrm{cm}$	71	70	73	78	77	75	82	74

Untersuchen Sie anhand eines geeigneten Testverfahrens, ob sich die Zunahmen der Körperlängen der einzelnen Tiere signifikant unterscheiden.

A 12 Ertrag verschiedener Pilzarten

Austernpilze und Braunkappen sind Holzpilze, die auf Stroh kultiviert werden können. Es liegen zwei Stichproben über den Frischmasseertrag pro Strohballen von beiden Pilzarten vor.

	Frischmasse/(kg/Strohballen)											
Austernpilze	4,0	7,6	6,5	5,9	8,6	7,3	5,2	4,8	6,1	6,1		
Braunkappen	4,7	5,7	5,7	5,0	4,7	4,6	5,5	5,2	5,5	5,5	5,4	5,2
Stichprobe	1	2	3	4	5	6	7	8	9	10	11	12

Testen Sie auf einem Signifikanzniveau von 5 %, ob die mittleren Erträge der beiden Pilze unterschiedlich sind. Die Annahme der homogenen Varianzen ist hier gefährlich, da zwei Arten verglichen werden. Es wird daher der zweiseitige t-Test für verschiedene Streuungen verwendet.

3.3 t-Test für Vergleich zweier Erwartungswerte verbundener Stichproben

A 13 Einfluss der Benzinqualität auf Treibstoffverbrauch

Es soll geprüft werden, ob sich die Benzinqualität der Marken Eral und Asso auf dem Signifikanzniveau $\alpha = 0,05$ unterscheiden. Für beide Marken wurde der Verbrauch in $\mathrm{L}/100\,\mathrm{km}$ von $n = 5$ Autotypen gemessen. Die Daten tabellarisch:

Autotyp	Verbrauch [L/100 km]		d_i
i	Eral (E_i)	Asso (A_i)	
VW Passat	11,9	12,3	-0,4
Opel Astra	7,7	7,7	0
Mazda 323	8,2	8,5	-0,3
BMW 525	13,4	14	-0,6
Fiat Uno	8	8,3	-0,3

A 14 Wirksamkeit von Schlafmitteln

An einer Gruppe von 10 Personen wird die Schlaf verlängernde Wirkung (in Stunden) zweier Schlafmittel A und B in zwei aufeinanderfolgenden Nächten festgestellt. Die Versuchspersonen erhalten zunächst Medikament A. In der zweiten Nacht wird Medikament B verabreicht und jeweils die Wirksamkeit gemessen.

Patient i	1	2	3	4	5	6	7	8	9	10
A_i	1,9	0,8	1,1	0,1	-0,1	4,4	5,5	1,6	4,6	3,4
B_i	0,7	-1,6	-0,2	-1,2	-0,1	3,4	3,7	0,8	0,0	2,0

Untersuchen Sie mit Hilfe eines geeigneten Testverfahrens, ob sich die Wirksamkeit der beiden Medikamente signifikant unterscheidet.

4 Einfache Varianzanalyse (ANOVA)

A 15 Wirkung von Fungiziden

Ihre Aufgabe ist es, den Behandlungserfolg von Fungiziden gegen den pflanzenpathogenen Pilz *Septoria nodorum* zu testen. Der zur prüfende Faktor 'Behandlung' besteht aus folgenden Stufen:

Stufe	Behandlung
1	keine
2	Fungizid A
3	Fungizid B

Die folgende Tabelle zeigt die durchschnittliche Anzahl der Sporenlager pro cm^2 Blattfläche von jeweils fünf Versuchsparzellen. Es sind somit 3 Prüfglieder miteinander zu vergleichen.

		Behandlung (Gruppe) $j = 1 \ldots k$			
		1	2	3	
Wiederholungen i	1	5,2	3,9	2,2	
	2	5,2	3,4	2,6	
	3	2,5	2,5	3,4	
	4	4,4	2	4,4	
	5	6,2	1,2	3,4	Gesamt:
$\overline{x}_j$		**4,7**	**2,6**	**3,2**	$\overline{x} = 3,5$

5 χ^2-Test

A 16 Unkrautzählung

Ein Gebiet von 25 m² ist in $n = \sum n_i = 100$ Teilflächen zu je 0,25 m² aufgeteilt. In jeder dieser Zellen wird die Unkrautmenge u_i bestimmt. Alle 100 Werte $u_1, ..., u_{100}$ sind ganze Zahlen zwischen 0 und 8. Wir bilden $K = 9$ Klassen entsprechend der Zahl der gefundenen Unkrautpflanzen.

i	X_i	n_i
1	0	3
2	1	9
3	2	15
4	3	21
5	4	19
6	5	15
7	6	10
8	7	5
9	8	3
$K = 9$		$n = \sum n_i = 100$

Überprüfen sie auf einem Signifikanzniveau von $\alpha = 0,05$, ob die Unkrautmenge Poisson-verteilt ist.

A 17 Milchleistung von Kühen

Es wurde die Milchleistung von $n = 100$ $(u_1, \ldots, u_{100})$ Kühen untersucht. Wir vermuten, dass die empirische Häufigkeitsverteilung der Milchleistung normalverteilt ist.

Die Messergebnisse:

i	X_i	n_i
1	37,5	5
2	42,5	9
3	47,5	25
4	52,5	22
5	57,5	33
6	62,5	2
7	67,5	4
K=7		$n = \sum n_i = 100$

i	Nummer der Klasse
K	Anzahl der Klassen (hier : $K = 7$)
X_i	Klassenmitte der Milchleistung in dt/a
n_i	Beobachtete Klassenhäufigkeiten
Δx	Klassenbreite (hier: $\Delta x = X_{i+1} - X_i = 5\,\mathrm{dt/a}$)

Überprüfen sie diese Vermutung auf einem Signifikanzniveau von $\alpha = 0,05$.

A 18 Punktzahlen der Statistikklausur

Herr K. aus B. behauptet, dass die Punktzahl X von Biotechnologiestudenten in der Statistikklausur eine *kryptonormative Polydiskriminalverteilung* hat, die, wie allgemein nicht bekannt ist, folgendermaßen aussieht:

Klausurpunktzahl x	$0 < x \leq 30$	$30 < x \leq 50$	$50 < x \leq 70$	$70 < x \leq 100$
$H_t(x_j) = h_t(x_j) \cdot \Delta x$	0,1	0,2	0,3	0,4

In einer Klausur mit 100 Teilnehmern ergibt sich nun folgende Punkteverteilung:

Klausurpunktzahl x	$0 < x \leq 30$	$30 < x \leq 50$	$50 < x \leq 70$	$70 < x \leq 100$
$n(x_j)$	6	15	39	40

Kann Herr K.s Behauptung bei einem Signifikanzniveau von 5 % als statistisch gesichert gelten?

6 Lineare Regression

A 19 Ertrag von Sommerweizen

Eine Stichprobe im Umfang $n = 11$ über Erträge y_i von Sommerweizen (in dt/ ha) bei verschiedenen Stickstoffdüngergaben x_i (in $\mathrm{kg/\,ha}$) liefert folgende Messergebnisse:

$x_i/(\mathrm{kg/\,ha})$	80	90	110	90	110	130	90	110	130	70	150
$y_i/(\mathrm{dt/\,ha})$	56,2	55,7	75,5	55,7	68,4	67,7	63,3	58,3	80,8	52,1	87,3

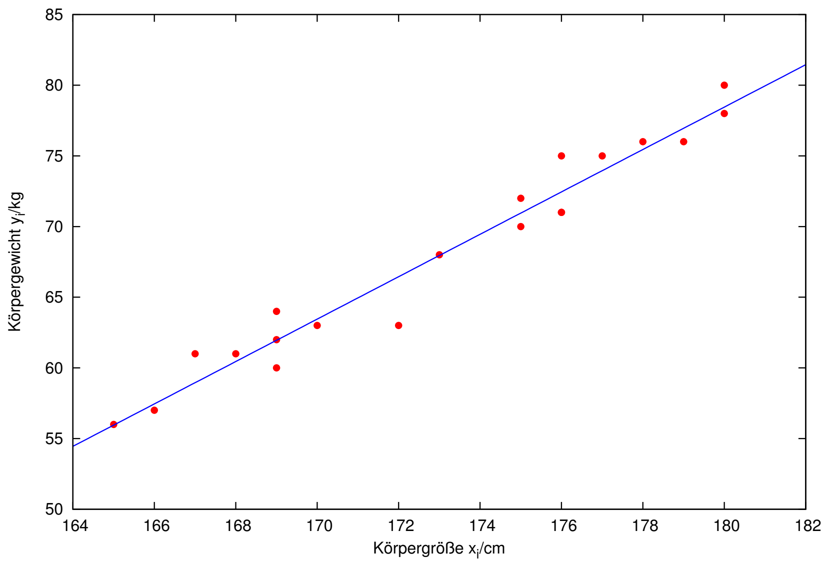
Es wird ein linearer Zusammenhang zwischen den Variablen X und Y unterstellt.

Überprüfen Sie den Zusammenhang zwischen den Erträgen an Sommerweizen bei verschiedenen Stickstoffdüngergaben bei einer statistischen Unsicherheit von 1 %. Geben Sie ferner den Vertrauensbereich für Y_i^* in den Punkten $x_1^* = 70\ \mathrm{kg/\,ha}$ und $x_2^* = 150\ \mathrm{kg/\,ha}$ an.

A 20 Zusammenhang von Körpergröße und Gewicht

Wir wollen den Zusammenhang zwischen Körpergröße (in cm) und Körpergewicht (in kg) untersuchen, daher haben wir die Daten von $n = 20$ erwachsenen Testpersonen aufgenommen:

i	**Größe** x_i	**Gewicht** y_i
1	165	56
2	176	75
3	175	70
4	168	61
5	167	61
6	172	63
7	175	72
8	180	80
9	179	76
10	173	68
11	166	57
12	178	76
13	169	60
14	169	64
15	170	63
16	176	71
17	180	78
18	169	62
19	177	75
20	176	71
$\sum$	**3460**	**1359**



Aufgrund des Diagramms wird ein linearer Zusammenhang vermutet.

a) Bestimmen Sie den Regressionskoeffizienten inkl. Vertrauensbereich für eine statistische Sicherheit von P % = 95 %.

b) Berechnen Sie für $x_1^* = 170,00\,\mathrm{cm}$ und $x_2^* = 172,00\,\mathrm{cm}$ die Funktionswerte mit zugehörigem Vertrauensbereich.