

Klausur

Einführung in die statistische Messdatenauswertung für Biotechnologen

29.6.2012

Name:

Matrikel-Nr.:

Aufgabe	Kurzfragen	1	2	Gesamt
Punkte				

Kurzfragen

1. Aus welchen Elementen besteht ein vollständiges Messergebnis?
2. Welche Fehler können bei einem statistischen Test auftreten? Nennen Sie diese und erläutern Sie deren Bedeutung.
3. Was versteht man unter verbundenen Stichproben?
4. Was bedeutet es, dass es sich beim Shannon-Theorem um eine hinreichende aber nicht notwendige Bedingung handelt?
5. Welche Arten von Störungen können von außen auf ein Messsystem einwirken? Erläutern Sie deren Wirkungsweise.
6. Was versteht man unter Rückwirkung eines Messsystems? Was ist die Ursache für die Rückwirkung?
7. Nennen Sie drei Skalenniveaus und deren Eigenschaften (Welche Operationen können auf entsprechend skalierten Merkmalen durchgeführt werden?! Verzichten Sie auf die Nennung von Beispielen!
8. Existiert zu jeder Häufigkeitsverteilung ein arithmetisches Mittel? Begründen Sie!

Aufgabe 1. Ein Ornithologe möchte herausfinden, ob sich die Körpermasse durchziehender Kraniche während ihrer Rast in einem Feuchtgebiet verändert. Dazu wurden mehrere Vögel kurz nach ihrer Ankunft gefangen, beringt, gewogen und wieder freigelassen. Nach einigen Wochen konnten acht dieser Tiere wieder eingefangen und ihre Masse erneut bestimmen werden. Aus beiden Messungen ergaben sich folgende Körpermassen in Kilogramm für die acht Vögel vor und nach der Rast:

Vogel	1	2	3	4	5	6	7	8
Masse vor Rast / kg	5,204	5,377	5,194	5,290	5,721	5,221	5,420	5,664
Masse nach Rast/ kg	5,308	5,492	5,286	5,409	5,588	5,331	5,662	5,686

Testen Sie, ob sich die Körpermasse der Vögel während ihres Aufenthalts signifikant ($\alpha = 0,05$) verändert hat oder nicht! (Für die Körpermasse kann eine Normalverteilung angenommen werden.)

Aufgabe 2.

Eine häufig verwendete Methode zur Bestimmung von Massenströmen besteht in der Messung des Wirkdrucks an einer Drossel (genormt in ISO 5167). Hierbei wird der Strömungsquerschnitt künstlich verengt und der Druck vor und nach dieser Drosselstelle gemessen (Bild 1). Die Druckdifferenz ist ein Maß für den Massenstrom Q_m .

Es gilt:

$$Q_m = \frac{\alpha \cdot \pi d^2}{4} \sqrt{2\rho \cdot \Delta p}$$

d : Querschnittsdurchmesser der Drossel

Δp : Druckdifferenz

ρ : Dichte des Mediums

α : Durchflusszahl (dimensionslos)

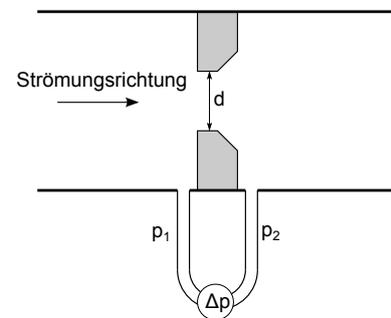


Bild 1

In einer Apparatur wird mit diesem Verfahren der Volumenstrom eines Mediums gemessen. Der Hersteller der verwendeten Drossel gibt folgende Kennwerte an:

Querschnittsdurchmesser $d = 50 \text{ mm} \pm 0,07 \text{ mm}$ (P=95%)

Durchflusszahl $\alpha = 0,6267 \pm 0,0009$ (P=95%)

Die Dichte ρ des Mediums beträgt laut zuvor durchgeführten Untersuchungen $0,84 \text{ kg/Liter} \pm 0,008 \text{ kg/Liter}$ (P=99%, für sehr große n). Zur Bestimmung des Massenstroms wird nun der Differenzdruck Δp fünf mal gemessen:

i	1	2	3	4	5
$\Delta p / \text{Pa}$	23223	23073	23017	23104	23149

$$1 \text{ Pa} = 1 \frac{\text{N}}{\text{m}^2} = 1 \frac{\text{kg}}{\text{s}^2 \text{m}}$$

Bestimmen Sie das vollständige Messergebnis des Massenstroms Q_m für eine Aussagewahrscheinlichkeit von P=95%! Verwenden Sie mindestens drei signifikante Stellen für die Angabe des Ergebnisses!

(Für alle Größen kann eine Normalverteilung angenommen werden.)

Produktregel: $(uv)' = u'v + uv'$

Quotientenregel: $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$

Kettenregel: $(u(v))' = u'(v) \cdot v'$

Elementare statistische Maßzahlen

Arithmetisches Mittel: $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

Empirische Varianz: $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$

Streuung: $S = +\sqrt{S^2}$

Konfidenzintervall

Die Messgröße X sei normalverteilt, σ sei bekannt:

$$\left[\bar{x} - \frac{k \cdot \sigma}{\sqrt{n}}, \bar{x} + \frac{k \cdot \sigma}{\sqrt{n}} \right]$$

Die Messgröße X sei normalverteilt, σ sei unbekannt:

$$\left[\bar{x} - \frac{S}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}}, \bar{x} + \frac{S}{\sqrt{n}} t_{n-1, 1-\frac{\alpha}{2}} \right]$$

Lineare Regression

Wenn durch eine Anzahl von Wertepaaren (x_i, y_i) nach der Methode der kleinsten quadratischen Abweichung eine Gerade gelegt wird, geht diese stets durch den Schwerpunkt (\bar{x}, \bar{y}) der Punkte:

$$(y - \bar{y}) = b(x - \bar{x})$$

(geschätzter) Regressionskoeffizient b (Steigung der Geraden)

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

Ein Schätzwert für σ^2 ist die Restvarianz $\hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{j=1}^n (y_j - \bar{y} + b(\bar{x} - x_j))^2$$

$$= \frac{n-1}{n-2} \cdot S_y^2 (1 - r_{xy}^2)$$

Bestimmung der Vertrauensgrenze für diese Schätzung des Steigungsmaßes:

1. Festlegen der geforderten statistischen Sicherheit P (z.B. 95%)
2. Berechnen der Streuung S_x aus den Messwerten x_1, \dots, x_n

3. Der Vertrauensbereich für den Regressionskoeffizienten b zur statistischen Sicherheit $P = 1 - \alpha$ beträgt:

$$\left[b - \frac{\hat{\sigma} t_{n-2, 1-\alpha/2}}{\sqrt{n} S_x}, b + \frac{\hat{\sigma} t_{n-2, 1-\alpha/2}}{\sqrt{n} S_x} \right]$$

4. Der Erwartungswert β für den Regressionskoeffizienten b liegt mit der statistischen Sicherheit P in diesem Intervall
5. Durch die berechnete Gerade wird einem beliebig gewählten x-Wert x^* der y-Wert

$$y^* = \bar{y} + b(x^* - \bar{x})$$

zugeordnet. Der Vertrauensbereich für y^* zur statistischen Sicherheit $P = 1 - \alpha$ beträgt:

$$\left[y^* - \frac{\hat{\sigma} t_{n-2, 1-\alpha/2}}{\sqrt{n}} \sqrt{1 + \frac{(x^* - \bar{x})^2}{S_x^2}}, y^* + \frac{\hat{\sigma} t_{n-2, 1-\alpha/2}}{\sqrt{n}} \sqrt{1 + \frac{(x^* - \bar{x})^2}{S_x^2}} \right]$$

Abweichungsfortpflanzung

f sei $f(x_1, \dots, x_n)$. Das Konfidenzintervall für f mit statistischer Sicherheit $P = 1 - \alpha$:

$$\left[f(\bar{x}_1, \dots, \bar{x}_n) - c_f, f(\bar{x}_1, \dots, \bar{x}_n) + c_f \right]$$

für den Fall zufälliger, normalverteilter Abweichungen mit:

$$c_f = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \Big|_{\bar{x}_1, \dots, \bar{x}_n} c_{x_i} \right)^2}, \quad c_{x_i} = \frac{S_{x_i}}{\sqrt{n_{x_i}}} t_{n_{x_i}-1, 1-\frac{\alpha}{2}}$$

t-Test

t-Test für Erwartungswert

Die Testgröße:

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{S}{\sqrt{n}}} \quad (\text{df} = n - 1)$$

Test der Nullhypothese bei vorgewähltem Signifikanzniveau α :

1. $H_0: \mu_x = \mu_0$ gegen $H_1: \mu_x < \mu_0$ (einseitige Hypothese)
Ist $t_0 < -t_{n-1, 1-\alpha}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.
2. $H_0: \mu_x = \mu_0$ gegen $H_1: \mu_x > \mu_0$ (einseitige Hypothese)
Ist $t_0 > t_{n-1, 1-\alpha}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.
3. $H_0: \mu_x = \mu_0$ gegen $H_1: \mu_x \neq \mu_0$ (zweiseitige Hypothese)
Ist $|t_0| > t_{n-1, 1-\frac{\alpha}{2}}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

t-Test für Vergleich zweier Erwartungswerte

Die Testgröße (einfachere Form, wenn $n_x = n_y = n$):

$$t_0 = \sqrt{n} \frac{\bar{x} - \bar{y}}{\sqrt{S_x^2 + S_y^2}} \quad (df = 2n - 2)$$

Test der Nullhypothese bei vorgewähltem Signifikanzniveau α :

1. $H_0: \mu_x = \mu_y$ gegen $H_1: \mu_x < \mu_y$ (einseitige Hypothese)

Ist

$$t_0 < -t_{n_x + n_y - 2; 1 - \alpha}$$

wird H_0 auf dem Signifikanzniveau α abgelehnt.

2. $H_0: \mu_x = \mu_y$ gegen $H_1: \mu_x > \mu_y$ (einseitige Hypothese)

Ist

$$t_0 > t_{n_x + n_y - 2; 1 - \alpha}$$

wird H_0 auf dem Signifikanzniveau α abgelehnt.

3. $H_0: \mu_x = \mu_y$ gegen $H_1: \mu_x \neq \mu_y$ (zweiseitige Hypothese)

Ist

$$|t_0| > t_{n_x + n_y - 2; 1 - \frac{\alpha}{2}}$$

wird H_0 auf dem Signifikanzniveau α abgelehnt.

t-Test für verbundene Stichproben

Die Testgröße:

$$t_0 = \frac{\bar{d}}{S_d / \sqrt{n}} \quad (df = n - 1)$$

mit:

$$d_i = x_i - y_i$$

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$$

$$S_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}}$$

Test der Nullhypothese bei vorgewähltem Signifikanzniveau α :

1. $H_0: \mu_d = 0$ gegen $H_1: \mu_d < 0$ (einseitige Hypothese)

Ist

$$t_0 < -t_{n-1; 1-\alpha}$$

wird H_0 auf dem Signifikanzniveau α abgelehnt.

2. $H_0: \mu_d = 0$ gegen $H_1: \mu_d > 0$ (einseitige Hypothese)

Ist

$$t_0 > t_{n-1; 1-\alpha}$$

wird H_0 auf dem Signifikanzniveau α abgelehnt.

3. $H_0: \mu_d = 0$ gegen $H_1: \mu_d \neq 0$ (zweiseitige Hypothese)

Ist

$$|t_0| > t_{n-1; 1 - \frac{\alpha}{2}}$$

wird H_0 auf dem Signifikanzniveau α abgelehnt.

Der χ^2 -Test für Verteilungsfunktionen

X sei eine Zufallsgröße mit unbekannter Verteilungsdichtefunktion. Aufgrund von Messdaten oder Vorabinformationen wird vermutet, dass X durch die Verteilungsdichtefunktion $h(x)$ beschrieben wird. Um dies zu prüfen, kann ein χ^2 -Test durchgeführt werden.

Nullhypothese H_0 : X wird durch die Verteilungsdichtefunktion $h(x)$ beschrieben.

Es wird eine Stichprobe von n Messwerten x_1, \dots, x_n aufgenommen.

Der Test erfolgt, indem zu dieser Messreihe ein empirisches Histogramm erstellt wird. Aus der Verteilungsdichtefunktion $h(x)$ wird ein theoretisches Histogramm berechnet. Als Testgröße wird eine normierte Differenz zwischen beiden Histogrammen berechnet. Wenn die Hypothese zutrifft, müsste diese Differenz hinreichend klein sein.

Vorgehensweise:

1. Aufteilen des Wertebereichs in r nicht überlappende Klassen T_i , so dass jede Klasse wenigstens 5 Werte der Stichprobe x_1, \dots, x_n enthält. Die Intervalle können auch ungleich breit sein.
2. Bestimmen der Anzahl B_i von Messwerten in der Klasse T_i
3. Falls die Verteilungsdichtefunktion $h(x)$ Parameter enthält (z.B. μ und σ bei der Normalverteilung), so werden diese Parameter aus den Messdaten x_1, \dots, x_n abgeschätzt.
4. Berechnen der Wahrscheinlichkeit p_i , mit der bei Annahme der hypothetischen Verteilungsdichte $h(x)$ unter Annahme der unter 3. geschätzten Parameter ein Messwert im Intervall T_i zu erwarten ist.
5. Berechnen der Produkte $E_i = np_i$, die die theoretischen Besetzungszahlen der Klasse T_i bei Annahme der Verteilungsdichte $h(x)$ darstellen.
6. Prüfen, ob für alle Klassen gilt: $E_i \geq 5$. Klassen mit $E_i < 5$ werden mit benachbarten Klassen zusammengelegt. Nach diesem Schritt liegen r^* Klassen vor mit $r^* \leq r$.
7. Berechnen der Testgröße:

$$\chi_0^2 = \sum_{i=1}^{r^*} \frac{(B_i - E_i)^2}{E_i}$$

8. Bestimmung der Zahl der Freiheitsgrade:

- r^* ist die Zahl der auswertbaren Klassen (Besetzungszahl ≥ 5)
- s ist die Zahl der Parameter der Verteilungsdichtefunktion, die aus der Stichprobe geschätzt werden
- Die Zahl der Freiheitsgrade ist $df = r^* - s - 1$

9. Festlegen der Irrtumswahrscheinlichkeit α

H_0 ist abzulehnen mit Signifikanzniveau α , wenn:

$$\chi_0^2 > \chi_{r^* - s - 1; 1 - \alpha}^2$$

Tabelle 1

Summenfunktion der standardisierten Normalverteilung

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt; \quad \Phi(z) = 1 - \Phi(-z)$$

Ablesebeispiel: $\Phi(0,76) = 0,776373$



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z
0,0	0,50000	0,503989	0,507978	0,511966	0,515953	0,519939	0,523922	0,527903	0,531881	0,535856	0,0
0,1	0,539828	0,543795	0,547758	0,551717	0,555670	0,559618	0,563559	0,567495	0,571424	0,575345	0,1
0,2	0,579260	0,583166	0,587064	0,590954	0,594835	0,598706	0,602568	0,606420	0,610261	0,614092	0,2
0,3	0,617911	0,621720	0,625516	0,629300	0,633072	0,636831	0,640576	0,644309	0,648027	0,651732	0,3
0,4	0,655422	0,659097	0,662757	0,666402	0,670031	0,673645	0,677242	0,680822	0,684386	0,687933	0,4
0,5	0,691462	0,694974	0,698468	0,701944	0,705401	0,708840	0,712260	0,715661	0,719043	0,722405	0,5
0,6	0,725747	0,729069	0,732371	0,735653	0,738914	0,742154	0,745373	0,748571	0,751748	0,754903	0,6
0,7	0,758036	0,761148	0,764238	0,767305	0,770350	0,773373	0,776373	0,779350	0,782305	0,785236	0,7
0,8	0,788145	0,791030	0,793892	0,796731	0,799546	0,802337	0,805105	0,807850	0,810570	0,813267	0,8
0,9	0,815940	0,818589	0,821214	0,823814	0,826391	0,828944	0,831472	0,833977	0,836457	0,838913	0,9
1,0	0,841345	0,843752	0,846136	0,848495	0,850830	0,853141	0,855428	0,857690	0,859929	0,862143	1,0
1,1	0,864334	0,866500	0,868643	0,870762	0,872857	0,874928	0,876976	0,879000	0,881000	0,882977	1,1
1,2	0,884930	0,886861	0,888768	0,890651	0,892512	0,894350	0,896165	0,897958	0,899727	0,901475	1,2
1,3	0,903200	0,904902	0,906582	0,908241	0,909877	0,911492	0,913085	0,914657	0,916207	0,917736	1,3
1,4	0,919243	0,920730	0,922196	0,923641	0,925066	0,926471	0,927855	0,929219	0,930563	0,931888	1,4
1,5	0,933193	0,934478	0,935745	0,936992	0,938220	0,939429	0,940620	0,941792	0,942947	0,944083	1,5
1,6	0,945201	0,946301	0,947384	0,948449	0,949497	0,950529	0,951543	0,952540	0,953521	0,954486	1,6
1,7	0,955435	0,956367	0,957284	0,958185	0,959070	0,959941	0,960796	0,961636	0,962462	0,963273	1,7
1,8	0,964070	0,964852	0,965620	0,966375	0,967116	0,967843	0,968557	0,969258	0,969946	0,970621	1,8
1,9	0,971283	0,971933	0,972571	0,973197	0,973810	0,974412	0,975002	0,975581	0,976148	0,976705	1,9
2,0	0,977250	0,977784	0,978308	0,978822	0,979325	0,979818	0,980301	0,980774	0,981237	0,981691	2,0
2,1	0,982136	0,982571	0,982997	0,983414	0,983823	0,984222	0,984614	0,984997	0,985371	0,985738	2,1
2,2	0,986097	0,986447	0,986791	0,987126	0,987455	0,987776	0,988089	0,988396	0,988696	0,988989	2,2
2,3	0,989276	0,989556	0,989830	0,990097	0,990358	0,990613	0,990863	0,991106	0,991344	0,991576	2,3
2,4	0,991802	0,992024	0,992240	0,992451	0,992656	0,992857	0,993053	0,993244	0,993431	0,993613	2,4
2,5	0,993790	0,993963	0,994132	0,994297	0,994457	0,994614	0,994766	0,994915	0,995060	0,995201	2,5
2,6	0,995339	0,995473	0,995604	0,995731	0,995855	0,995975	0,996093	0,996207	0,996319	0,996427	2,6
2,7	0,996533	0,996636	0,996736	0,996833	0,996928	0,997020	0,997110	0,997197	0,997282	0,997365	2,7
2,8	0,997445	0,997523	0,997599	0,997673	0,997744	0,997814	0,997882	0,997948	0,998012	0,998074	2,8
2,9	0,998134	0,998193	0,998250	0,998305	0,998359	0,998411	0,998462	0,998511	0,998559	0,998605	2,9

z	3,0	3,5	4,0	4,5	5,0	6,0	7,0	8,0	9,0	10,0	z
$\Phi(z)$	$1-1,350 \cdot 10^{-3}$	$1-2,326 \cdot 10^{-4}$	$1-3,167 \cdot 10^{-5}$	$1-3,998 \cdot 10^{-6}$	$1-2,867 \cdot 10^{-7}$	$1-9,866 \cdot 10^{-10}$	$1-1,280 \cdot 10^{-12}$	$1-6,221 \cdot 10^{-16}$	$1-1,129 \cdot 10^{-19}$	$1-7,620 \cdot 10^{-24}$	$\Phi(z)$

$\Phi(z)$	50%	60%	70%	80%	90%	95%	97,5%	99%	99,5%	99,75%	99,9%	99,95%	$\Phi(z)$
z	0	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	2,807	3,090	3,291	z

p-Quantile $t_{s,p}$ der Student'schen t-Verteilung mit s Freiheitsgraden

s	p	0,95	0,975	0,99	0,995
1		6,31	12,7	31,8	63,7
2		2,92	4,30	6,97	9,93
3		2,35	3,18	4,54	5,84
4		2,13	2,78	3,75	4,60
5		2,02	2,57	3,37	4,03
6		1,94	2,45	3,14	3,71
7		1,90	2,37	3,00	3,50
8		1,86	2,31	2,90	3,36
9		1,83	2,26	2,82	3,25
10		1,81	2,23	2,76	3,17
11		1,80	2,20	2,72	3,11
12		1,78	2,18	2,68	3,06
13		1,77	2,16	2,65	3,01
14		1,76	2,15	2,62	2,98
15		1,75	2,13	2,60	2,95
16		1,75	2,12	2,58	2,92
17		1,74	2,11	2,57	2,90
18		1,73	2,10	2,55	2,88
19		1,73	2,09	2,54	2,86
20		1,73	2,09	2,53	2,85
22		1,72	2,07	2,51	2,82
24		1,71	2,06	2,49	2,80
26		1,71	2,06	2,48	2,78
28		1,70	2,05	2,47	2,76
30		1,70	2,04	2,46	2,75
40		1,68	2,02	2,42	2,70
50		1,68	2,01	2,40	2,68
100		1,66	1,98	2,37	2,63
200		1,65	1,97	2,35	2,60
∞		1,64	1,96	2,33	2,58