Klausur
Einführung in die statistische Messdatenauswertung für Biotechnologen

WS 2007/2008

Kurzfragen (14 Punkte)

1. Über welche Naturkonstante sind die Einheit der Länge und die Einheit der Zeit miteinander verbunden?

2. Aus welchen Elementen besteht eine Messeinrichtung?

3. Die Wechselwirkung zwischen Messeinrichtung und Messobjekt führt stets zu Messabweichungen. Wir haben drei Möglichkeiten diskutiert, mit der Abweichung umzugehen. Welche sind das?

4. Welchen Wert kann der Parallaxenfehler bei der Ablesung eines Digitalmultimeters der Güteklasse 0,1 maximal annehmen?

5. Ein Messsystem verhalte sich wie ein lineares System 1. Ordnung. Um wieviel Prozent des Vollausschlags kann dieses System bei einer Sprungfunktion am Eingang überschwingen?

6. a) Ein Signal mit einer maximalen Frequenz von 47 kHz soll digitalisiert werden. Mit welcher Abtastrate muß man minimal arbeiten, um Aliasing zu vermeiden?

   b) Welche Abtastrate muss für ein Rechtecksignal mit einer Periodendauer von 1 ms gewählt werden, um das Abtasttheorem zu erfüllen?

7. Ist die folgende Aussage richtig oder falsch? „Die Verteilungsdichte für die Zahlen 1 bis 6 nimmt bei einem idealen Würfel die Form einer Gaußschen Glockenkurve an.“

8. Nennen Sie ein Beispiel für eine diskrete Verteilung.

9. Wird bei einem statistischen Test die Nullhypothese H_0 verworfen, ist dann die Alternativhypothese H_1 richtig?

10. Sie ermitteln für eine Anzahl von Wertepaaren (x_i, y_i) eine Regressionsgerade. Geben Sie einen Punkt an, durch den diese Gerade auf jeden Fall läuft.

Aufgabe 1. (13 Punkte)

Es soll untersucht werden, wie sich die Nahrungsaufnahme auf das Körpergewicht auswirkt. Die Körpergewichte der Testpersonen wurden vor dem Mittagessen (m_i) und eine Stunde nach dem Mittagessen (M_i) gemessen. Die folgende Tabelle enthält die Messwerte von 15 Testpersonen.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Vor dem Essen m_i / kg	36,2	49,0	98,5	79,0	85,0	58,0	72,1	105,5	63,3	72,1	86,0	55,0	86,4	91,0	77,0
Nach dem Essen M_i / kg	36,6	49,0	99,2	78,5	85,3	58,4	71,6	106,0	65,3	73,0	87,0	55,5	87,5	92,0	78,5

Es interessiert, ob das Mittagsessen auf einem Signifikanzniveau von $\alpha = 0,05$ eine Gewichtszunahme verursacht.

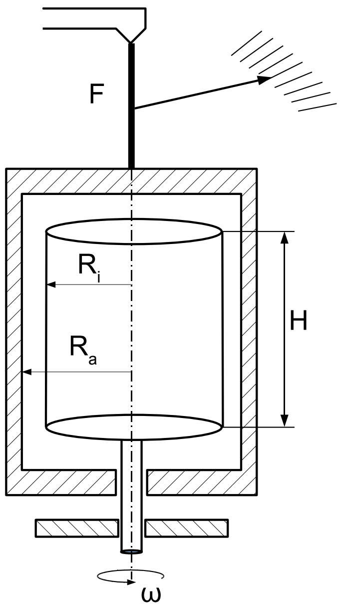
a) Welcher Test ist geeignet das Problem zu lösen?

b) Müssen wir eine einseitige oder zweiseitige Hypothese stellen? Warum?

c) Prüfen Sie die Hypothese durch Anwendung des Tests.

Aufgabe 2. (15 Punkte)

Zur Messung der dynamischen Viskosität η kann das im Aufbau sehr einfache Rotationsviskosimeter nach COUETTE benutzt werden. Die Versuchsflüssigkeit befindet sich dabei zwischen zwei koaxialen Zylindern, von denen sich der innere mit der konstanten Winkelgeschwindigkeit ω dreht und so durch die innere Reibung in der Flüssigkeit ein Moment M auf den äußeren, feststehenden Zylinder überträgt. Das Moment wird auf den Torsionsfaden F übertragen und dort zur Anzeige gebracht.



Die dynamische Viskosität berechnet sich hierbei zu

$$\eta = \frac{M}{4\pi \cdot H \cdot \omega} \left(\frac{1}{R_i^2} - \frac{1}{R_a^2} \right)$$

Hierbei sind:

M	Drehmoment
H	Höhe des inneren Zylinders
R_i	Radius des inneren Zylinders
R_a	Radius des äußeren Zylinders
ω	Winkelgeschwindigkeit des inneren Zylinders

Der Radius R_a des Rotationsviskosimeters wird mit einer Aussagewahrscheinlichkeit von $P = 99\,\%$ mit $R_a = (70 \pm 0,020)\,\mathrm{mm}$ angegeben.

Die Maße des inneren Zylinders seien mit $R_i = 60\,\mathrm{mm}$ und $H = 100\,\mathrm{mm}$ exakt bekannt.

Eine sechsmalige Messung des Drehmoments M ergab folgende Werte:

M / Ncm	2,12	2,10	2,14	2,12	2,11	2,13

Die Winkelgeschwindigkeit wird mit $\omega = 12 \frac{1}{\mathrm{s}}$ gemessen. Der Hersteller des Winkelgeschwindigkeitsmessgerätes gibt eine Unsicherheit (Vertrauensbereich) von 1 % des angezeigten Wertes bei einer Aussagewahrscheinlichkeit von $P = 95\,\%$ bei $n = 30$ Messungen an.

Berechnen Sie mit den angegebenen Werten das vollständige Messergebnis der Viskosität η mit einer Aussagewahrscheinlichkeit von $P = 99\,\%$.

Anmerkung:
Die Einheit der dynamischen Viskosität ist $Pa \cdot \mathrm{s} = N \cdot \mathrm{s/m^2} = \mathrm{kg/(s \cdot m)}$.
Für alle Größen kann eine Gaußsche Normalverteilung angenommen werden.

Elementare statistische Maßzahlen:

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

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

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

Konfidenzintervall:

Die Meßgröße *X* sei normalverteilt, σ sei bekannt:

$$\left[\bar{x} - \frac{c_{P\%}\sigma}{\sqrt{n}}, \quad \bar{x} + \frac{c_{P\%}\sigma}{\sqrt{n}} \right]$$

Die Meßgröße *X* sei normalverteilt, σ sei unbekannt.

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

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)

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

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

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

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 $x_1, \ldots, x_n$
3. Der Vertrauensbereich für den Regressionskoeffizienten b zur statistischen Sicherheit p=1-α beträgt:

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

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-α ist:

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

Fehlerfortpflanzung für statistische Fehler:

f sei $f(x_1, \ldots, x_n)$. Das Konfidenzintervall für ***f*** mit statistischer Sicherheit *P%=1-α*:

$$\left\lfloor f(\bar{x}_1, \ldots, \bar{x}_n) - c_f ; f(\bar{x}_1, \ldots, \bar{x}_n) + c_f \right\rfloor$$

$$c_f = \sqrt{\sum_{i=1}^{n} \left(\left. \frac{\partial f}{\partial x_i} \right|_{\bar{x}_1, \ldots, \bar{x}_n} c_{x_i} \right)^2}$$

$$c_{x_i} = \frac{t_{n_{x_i}-1;1-\alpha/2}}{\sqrt{n_{x_i}}} s_{x_i}$$

t-Test:

t-Test für Erwartungswert:

Die Testgröße:

$$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \qquad (df=n-1)$$

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 $|t_0| > t_{n-1;1-\alpha/2}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

t-Test für Vergleich zweier Erwartungswerte:

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

$$t_0 = \sqrt{n}\,\frac{\bar{x}-\bar{y}}{\sqrt{s_x^2+s_y^2}} \quad \textit{(df=2n-2)}$$

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 $|t_0| > t_{n_x+n_y-2;1-\alpha/2}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

t-Test für verbundene Stichproben:

Die Testgröße:

$$t_0 = \frac{\bar{d}}{s_d/\sqrt{n}} \qquad \textit{(df=n-1)}$$

$d_i=x_i-y_i$

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

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

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 $|t_0| > t_{n-1;1-\alpha/2}$, wird H_0 auf dem Signifikanzniveau α abgelehnt.

Der χ2-Test für Verteilungsfunktionen:

X sei eine Zufallsgröße mit unbekannter Verteilungsdichtefunktion. Aufgrund von Meßdaten oder Vorabinformationen wird vermutet, das X durch die Verteilungsdichtefunktion h(x) beschrieben wird. Um dies zu prüfen, kann ein χ2-Test durchgeführt werden.

Nullhypothese H0: X wird durch die Verteilungsdichtefunktion h(x) beschrieben.

Es wird eine Stichprobe von n Meßwerten x1,...xn 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.

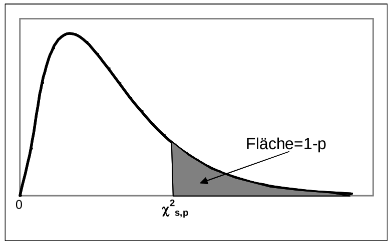
Vorgehensweise:

11. Aufteilen des Wertebereichs in r nicht überlappende Klassen Ti, so daß jede Klasse wenigstens 5 Werte der Stichprobe x1,...xn enthält. Die Intervalle können auch ungleich breit sein.
12. Bestimmen der Anzahl Bi von Meßwerten in der Klasse Ti
13. Falls die Verteilungsdichtefunktion h(x) Parameter enthält (z.B. μ und σ bei der Normalverteilung), so werden diese Parameter aus den Meßdaten x1,...xn abgeschätzt.
14. Berechnen der Wahrscheinlichkeit pi, mit der bei Annahme der hypothetischen Verteilungsdichte h(x) unter Annahme der unter 3. geschätzten Parameter ein Meßwert im Intervall Ti zu erwarten ist.
15. Berechnen der Produkte Ei=npi, die die theoretischen Besetzungszahlen der Klasse Ti bei Annahme der Verteilungsdichte h(x) darstellen.
16. Prüfen, ob für alle Klassen gilt: Ei $\geq$ 5. Klassen mit Ei < 5 werden mit benachbarten Klassen zusammengelegt. Nach diesem Schritt liegen r* Klassen vor mit r* $\leq$ r.
17. Berechnen der Testgröße
$$\chi_0^2 = \sum_{i=1}^{r^*}\frac{\left(B_i-E_i\right)^2}{E_i}$$
18. Bestimmung der Zahl der Freiheitsgrade
r* ist die Zahl der auswertbaren Klassen (Besetzungszahl >5)
s ist die Zahl der Parameter der Verteilungsdichtefunktion
Die Zahl der Freiheitsgrade ist r*-s-1
19. Festlegen der Irrtumswahrscheinlichkeit α

H0 ist abzulehnen mit Signifikanzniveau α, wenn $\chi_0^2 > \chi^2_{r^*-s-1;1-\alpha}$

p-Quantile $\chi^2_{s,p}$ der χ^2 - Verteilung mit s Freiheitsgraden

s \ p	0,900	0,950	0,975	0,990	0,995
1	2,71	3,84	5,02	6,63	7,88
2	4,61	5,99	7,38	9,21	10,60
3	6,25	7,81	9,35	11,34	12,84
4	7,78	9,49	11,14	13,28	14,86
5	9,24	11,07	12,83	15,09	16,75
6	10,64	12,59	14,45	16,81	18,55
7	12,02	14,07	16,01	18,48	20,28
8	13,36	15,51	17,53	20,09	21,95
9	14,68	16,92	19,02	21,67	23,59
10	15,99	18,31	20,48	23,21	25,19
11	17,28	19,68	21,92	24,72	26,76
12	18,55	21,03	23,34	26,22	28,30
13	19,81	22,36	24,74	27,69	29,82
14	21,06	23,68	26,12	29,14	31,32
15	22,31	25,00	27,49	30,58	32,80
16	23,54	26,30	28,85	32,00	34,27
17	24,77	27,59	30,19	33,41	35,72
18	25,99	28,87	31,53	34,81	37,16
19	27,20	30,14	32,85	36,19	38,58
20	28,41	31,41	34,17	37,57	40,00
21	29,62	32,67	35,48	38,93	41,40
22	30,81	33,92	36,78	40,29	42,80
23	32,01	35,17	38,08	41,64	44,18
24	33,20	36,42	39,36	42,98	45,56
25	34,38	37,65	40,65	44,31	46,93
26	35,56	38,89	41,92	45,64	48,29
27	36,74	40,11	43,19	46,96	49,64
28	37,92	41,34	44,46	48,28	50,99
29	39,09	42,56	45,72	49,59	52,34
30	40,26	43,77	46,98	50,89	53,67
40	51,81	55,76	59,34	63,69	66,77
50	63,17	67,50	71,42	76,15	79,49
60	74,40	79,08	83,30	88,38	91,95
70	85,53	90,53	95,02	100,43	104,21
80	96,58	101,88	106,63	112,33	116,32
90	107,57	113,15	118,14	124,12	128,30
100	118,50	124,34	129,56	135,81	140,17



p-Quantile $t_{s,p}$ der Student´schen t - Verteilung mit s Freiheitsgraden

s \ p	0,950	0,975	0,990	0,995
1	6,31	12,71	31,82	63,66
2	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60
5	2,02	2,57	3,36	4,03
6	1,94	2,45	3,14	3,71
7	1,89	2,36	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	3,05
13	1,77	2,16	2,65	3,01
14	1,76	2,14	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	1,72	2,09	2,53	2,85
21	1,72	2,08	2,52	2,83
22	1,72	2,07	2,51	2,82
23	1,71	2,07	2,50	2,81
24	1,71	2,06	2,49	2,80
25	1,71	2,06	2,49	2,79
26	1,71	2,06	2,48	2,78
27	1,70	2,05	2,47	2,77
28	1,70	2,05	2,47	2,76
29	1,70	2,05	2,46	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
60	1,67	2,00	2,39	2,66
80	1,67	1,99	2,38	2,65
100	1,66	1,99	2,37	2,64
200	1,66	1,99	2,37	2,63
∞	1,65	1,96	2,33	2,58