Klausur

Statistische Messdatenauswertung für Biotechnologen

im Modul

BP 14 Statistik und Programmieren

7. Februar 2025

Name: ____

Matrikel-Nr.: __________________________________

Prüfungsraum: SN 19.1

Mit meiner Unterschrift versichere ich, dass ich mich geistig und körperlich in der Lage befinde, die Prüfung abzulegen (d. h. prüffähig bin).

____
Unterschrift Studierende/r

AUFGABE	1	2	AWV A	AWV B	KF	Gesamt
PUNKTE	/18	/12	/18	/17	/5	/70

NOTE

Hinweise zur Prüfung

1. Bearbeitungsdauer: 120 Minuten
2. Als Hilfsmittel sind ausschließlich Taschenrechner ohne vorgefertigte Programme und ohne drahtlose Kommunikationsschnittstelle, einschließlich deren Bedienungsanleitung in gedruckter Form, zugelassen. Sonstige schriftliche Unterlagen sowie Bild-, Ton- und Videodokumente sind ausdrücklich nicht zugelassen. Die Verwendung elektronischer Geräte mit drahtloser Kommunikationsschnittstelle, gleich zu welchem Zweck, ist während der Klausur untersagt. Dieses Verbot gilt insbesondere auch für sogenannte Smartwatches. Nach allgemeinem Prüfungsrecht und aktueller APO stellt bereits das Mitführen eines nicht erlaubten Hilfsmittels im Prüfungsraum eine Täuschung dar. Verstöße dagegen bzw. andere Täuschungsversuche werden gemäß der Prüfungsordnung geahndet.
3. Auf das Deckblatt sind der Name, der Vorname, die Matrikelnummer und die Bezeichnung des Raumes, in welchem die Prüfung abgelegt wird einzutragen. Auf allen anderen abgegebenen Blättern ist zumindest der Name zu vermerken. Das Deckblatt ist als oberes Blatt der Klausur abzugeben. Der Rest der Aufgabestellung muss nicht abgegeben werden, sofern er keine für die Lösung relevanten Eintragungen enthält.
4. Zur Teilnahme an der Prüfung ist auf dem Deckblatt die Prüfungsfähigkeit durch Unterschrift zu bestätigen.
5. Der Studierendenausweis ist zusammen mit einem Lichtbildausweis und dem ausgefüllten Deckblatt der Aufgabenstellung sichtbar auszulegen.
6. Alle zur Lösung der gestellten Aufgaben benötigten nichttrivialen Gleichungen und Konstanten sowie alle notwendigen Tabellen und Diagramme sind der folgenden Formelsammlung, der Aufgabenstellung selbst oder dem Anhang auf den Seiten 13 bis 17 zu entnehmen.

Formelsammlung:

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

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

Kettenregel: $\frac{dy}{dx} = u'(v)v'(x) \quad \text{mit} \quad y = u\big(v(x)\big)$

Energie: 1 J (Joule) = 1 J = 1 N·m = 1 kg·m^2/s^2

<u>1. Aufgabe:</u>

Als spezifische Wärmekapazität bezeichnet man die auf die Masse bezogene Wärmekapazität. Sie ist eine Stoffeigenschaft der Thermodynamik und bemisst die Fähigkeit eines Stoffes, thermische Energie zu speichern.

Die spezifische Wärmekapazität c eines Stoffes ist gemäß folgender Gleichung die Wärme ΔQ, die einer Menge des Stoffes zugeführt oder entzogen wird, dividiert durch die zugehörige Erhöhung oder Absenkung der Temperatur ΔT und die Masse m des Stoffes:

$$c = \frac{\Delta Q}{m \cdot \Delta T}$$

Die spezifische Wärmekapazität ist somit ein Maß für diejenige Energie, die man benötigt, um 1 kg eines Stoffes um 1 K bzw. 1°C zu erwärmen.

Im Folgenden soll die spezifische Wärmekapazität c von flüssigem Glyzerin ($C_3H_8O_3$) auf der Grundlage von Messergebnissen für die Größen ΔQ, m und ΔT einschließlich der wahrscheinlichen Abweichungsgrenzen ermittelt werden.

Die der Flüssigkeit mittels eines Wärmetauschers zugeführte Wärmemenge ΔQ wird unter Verwendung eines elektronischen Wärmezählers ermittelt. Der Anzeigewert der zugeführten Wärmemenge beträgt $\Delta Q = 61{,}2\,\mathrm{J}$. Der Hersteller des Wärmezählers gibt eine relative Unsicherheit von $\pm 1\%$ des Anzeigewertes bei einer Aussagesicherheit von $P = 95\%$ und sehr großem Stichprobenumfang $n_{\Delta Q}$ an.

Die Masse m der untersuchten Flüssigkeitsmenge wurde im Vorfeld des Versuchs auf einer Laborwaage ermittelt. Die Masse beträgt demnach $m = 1264\ \mathrm{g}\ \pm 2\ \mathrm{g}$ bei $P = 99\%$.

Die Temperaturänderung wurde mittels Pt1000-Temperatursensoren in insgesamt $n_{\Delta T} = 8$ Einzelmessungen ermittelt. Die dabei erhaltenen Messwerte sind in Tabelle 1.1 zusammengefasst.

i	1	2	3	4	5	6	7	8
ΔT / K	20,04	19,95	19,89	19,83	20,12	20,01	20,18	19,74

Tabelle 1.1: Messwerte der Temperaturänderung ΔT

a) Berechnen Sie die gesuchte spezifische Wärmekapazität c und geben Sie das vollständige Messergebnis in der Einheit $\mathrm{J/(kg \cdot K)}$ mit einer Aussagewahrscheinlichkeit von $P = 99\%$ an!

Hinweis: Für alle Messgrößen kann eine Normalverteilung vorausgesetzt werden.

2. Aufgabe:

Im Rahmen Ihres Studiums absolvieren Sie ein biochemisches Praktikum, in welchem Sie auch mit Teclubrennern arbeiten. Zum Entzünden der Brenner steht allen teilnehmenden Studierenden ein gemeinsames Feuerzeug zur Verfügung, welches dadurch unangenehm auffällt, dass oftmals mehrere Versuche erforderlich sind, bis das Feuerzeug erfolgreich zündet.

Sie möchten die Gelegenheit nutzen, sich auf Ihre Statistikprüfung vorzubereiten und formulieren daher die Hypothese, dass die Anzahl der benötigten Versuche bis zur ersten erfolgreichen Zündung des Feuerzeugs einer geometrischen Verteilung genügt.

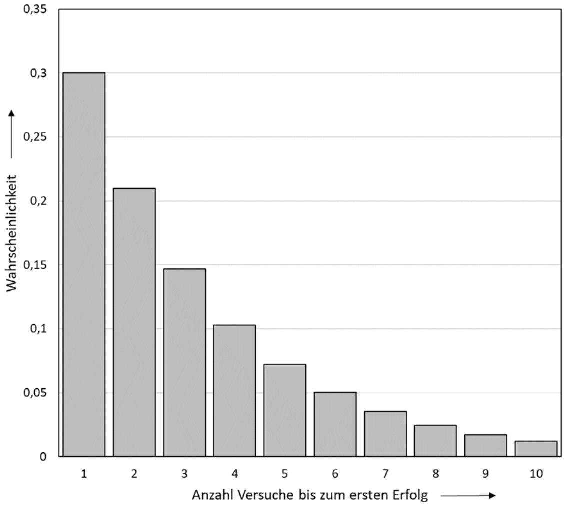


Abbildung 2.1: Geometrische Verteilung mit Parameter $p = 0{,}3$

Die geometrische Verteilung ist eine diskrete, univariate Wahrscheinlichkeitsverteilung, welche aus unabhängigen Bernoulli-Experimenten abgeleitet wird und in zwei Varianten A und B definiert ist. Die im vorliegenden Fall relevante Variante A gibt die Anzahl X der Versuche an, die notwendig sind, um einen ersten Erfolg zu erzielen. Diese Verteilung ist auf der Menge $\mathbb{N}$ der natürlichen Zahlen definiert. Für die Wahrscheinlichkeit, dass man genau m Versuche benötigt, um zum ersten Erfolg zu kommen, gilt:

$$P(X = m) = p(1 - p)^{m-1} = pq^{m-1}$$

Hierin steht m für die Anzahl der bis zum ersten Erfolg benötigten Versuche (einschließlich des erfolgreichen Versuchs), p gibt die Wahrscheinlichkeit an, mit welcher bei einem einzelnen Versuch ein Erfolg eintritt und q gibt die Wahrscheinlichkeit an, mit welcher bei einem einzelnen Versuch ein Misserfolg eintritt. Es gilt entsprechend $p + q = 1$.

Als empirische Daten aus dem Praktikum liegt Ihnen von allen Laborgruppen zusammengenommen die Anzahl m der bis zum jeweils ersten Zünden des Feuerzeugs benötigten Versuche vor. Insgesamt wurden bei $k = 480$ Versuchen $n = 144$ erfolgreiche Zündungen beobachtet. Für diese n erfolgreichen Versuche wurde jeweils die in Tabelle 2.1 angegebene Anzahl m an Versuchen benötigt.

Anzahl Versuche	1	2	3	4	5	6	7	8	9	10
Häufigkeit	46	25	18	16	8	9	10	8	2	2

Tabelle 2.1: Ermittelte Häufigkeiten für die jeweils benötigte Anzahl Versuche bis zum ersten Erfolg

Die Wahrscheinlichkeit p kann aus dem Verhältnis der erfolgreichen Versuche und der Anzahl aller Versuche abgeschätzt werden. Es gilt entsprechend $p = n/k$.

a) Untersuchen Sie mittels eines Chi-Quadrat-Tests, ob die in Tabelle 2.1 angegebene Verteilung auf einem Signifikanzniveau von $\alpha = 0{,}1$ einer geometrischen Verteilung genügt!

Erläuterungen zu Aufgaben nach dem Antwort-Wahl-Verfahren:

Bei jeder Fragestellung wird im Anschluss an die Antwortalternativen angegeben, um welchen Fragetyp es sich handelt. Die möglichen Fragetypen sind nachfolgend näher erläutert.

- *Fragetyp Einfachwahl*: Bei Fragen dieses Typs ist genau eine der angebotenen Antwortalternativen korrekt. Bei Fragen dieses Typs wird nur dann eine von null Punkten verschiedene Bewertung vergeben, wenn genau die eine korrekte Antwort markiert wurde.
- *Fragetyp Mehrfachwahl*: Bei Fragen dieses Typs ist mindestens eine der angebotenen Antwortalternativen korrekt. Entsprechend können auch mehrere oder alle Antwortalternativen korrekt sein. Bei Fragen dieses Typs werden auch dann anteilig Punkte vergeben, wenn einzelne Antworten unzutreffend sind (korrekte Antwort fälschlich nicht markiert oder unkorrekte Antwort fälschlich markiert). Hierbei gilt jedoch, dass eine Frage, bei welcher keine der Antworten markiert wurde als nicht bearbeitet gilt und mit null Punkten bewertet wird.

Für alle Fragetypen gilt, dass eine Frage nicht mit weniger als null Punkten bewertet werden kann. Es werden also keine negativen Punkte vergeben.

Antwort-Wahl-Verfahren, Teil A:

3. Bei einem Hersteller von Prallelementen für Volumendurchflussmesseinrichtungen wird im Rahmen der Qualitätssicherung der Durchmesser der zylinderförmigen Prallelemente mit einem Nenndurchmesser von $D_{nenn} = 25$ mm überwacht. Hierzu wird aus der laufenden Fertigung eine Stichprobe vom Umfang $n = 8$ entnommen und der Durchmesser D der Prallelemente ermittelt. Aus der Stichprobe ergibt sich ein Mittelwert des Durchmessers von $\overline{D} = 24{,}96$ mm und eine Streuung von $S_D = 0{,}032$ mm. Die Standardabweichung σ sei unbekannt.

3.1. Das Konfidenzintervall des Erwartungswertes des Prallelementdurchmessers D für eine Aussagewahrscheinlichkeit von $\mathrm{P} = 95\%$ beträgt für diesen Fall ungefähr:

a) $D = 24{,}96\ \mathrm{mm} \pm 0{,}0142\ \mathrm{mm}\,;\mathrm{P} = 95\%$ ☐
b) $D = 24{,}96\ \mathrm{mm} \pm 0{,}0186\ \mathrm{mm}\,;\mathrm{P} = 95\%$ ☐
c) $D = 24{,}96\ \mathrm{mm} \pm 0{,}0214\ \mathrm{mm}\,;\mathrm{P} = 95\%$ ☐
d) $D = 24{,}96\ \mathrm{mm} \pm 0{,}0222\ \mathrm{mm}\,;\mathrm{P} = 95\%$ ☐
e) $D = 24{,}96\ \mathrm{mm} \pm 0{,}0268\ \mathrm{mm}\,;\mathrm{P} = 95\%$ ☐

(Fragetyp Einfachwahl)

3.2. Angenommen, es sei bekannt, dass die Standardabweichung des Durchmessers der Prallelemente $D = 0{,}03$ mm betrage. Wie groß ist dann der minimal erforderliche Stichprobenumfang n, um bei einer Aussagewahrscheinlichkeit von $\mathrm{P} = 99\%$ das Konfidenzintervall des Erwartungswertes des Durchmessers auf maximal $\pm$ 0,01 mm abschätzen zu können?

a) $n = 15$ ☐
b) $n = 25$ ☐
c) $n = 49$ ☐
d) $n = 60$ ☐
e) $n = 64$ ☐

(Fragetyp Einfachwahl)

Fortsetzung Aufgabe 3 auf der nächsten Seite

3.3. Gehen Sie davon aus, dass Mittelwert und Streuung obiger Stichprobe mit dem Erwartungswert und der Standardabweichung der Grundgesamtheit übereinstimmen. Etwa wie viel Prozent aller Prallelemente weisen dann einen Durchmesser außerhalb des Intervalls $24{,}95\text{ mm} \leq D \leq 25{,}05\text{ mm}$ auf?

a) 12% ☐
b) 38% ☐
c) 49% ☐
d) 62% ☐
e) 87% ☐

(Fragetyp Einfachwahl)

3.4. Angenommen, der Erwartungswert des Durchmessers betrage $D = 25$ mm. Welchen (mathematisch gerundeten) Wert dürfte die Standardabweichung σ_D des Durchmessers dann maximal annehmen, damit 95% der Prallelemente innerhalb des Intervalls von $24{,}98\text{ mm} \leq D \leq 25{,}02\text{ mm}$ lägen?

a) 0,007 mm ☐
b) 0,008 mm ☐
c) 0,010 mm ☐
d) 0,012 mm ☐
e) 0,014 mm ☐

(Fragetyp Einfachwahl)

4. Sie möchten mittels einer Versuchsreihe in Erfahrung bringen, ob das von Ihnen entwickelte Herbizid „Auroator" dem Herbizid „Beruatrix" Ihres Wettbewerbers überlegen ist. Hierzu behandeln Sie in einem Feldversuch jeweils $n_A = 10$ Weizenparzellen mit „Auroator" und weitere $n_B = 10$ Weizenparzellen mit „Beruatrix". Nach einer definierten Wuchsdauer ermitteln Sie für jede Parzelle die mittlere Anzahl der Fremdpflanzen pro Quadratmeter. Anschließend möchten Sie mittels eines statistischen Tests untersuchen, ob die mittlere Anzahl der Fremdpflanzen bei Einsatz von „Auroator" niedriger ist als bei Einsatz von „Beruatrix".

4.1. Welcher statistische Test ist geeignet, die Frage zu beantworten?

a) lineare Regression ☐
b) t-Test für Erwartungswert ☐
c) t-Test für den Vergleich zweier Erwartungswerte bei unabhängigen Stichproben ☐
d) t-Test für den Vergleich zweier Erwartungswerte bei verbundenen Stichproben ☐
e) Chi-Quadrat-Test ☐

(Fragetyp Einfachwahl)

4.2. Welche Alternativhypothese ist für den Test zu wählen?

a) einseitige Alternativhypothese ☐
b) zweiseitige Alternativhypothese ☐

(Fragetyp Einfachwahl)

5. Anhand zweier verbundener Stichproben A und B möchten Sie einen t-Test für verbundene Stichproben durchführen. Aus den Stichproben, die jeweils einen Umfang von $n = 10$ aufweisen, haben Sie für die paarweisen Differenzen d_i einen Mittelwert von $\overline{d} = 0{,}52$ W und eine Streuung von $S_d = 0{,}14$ W errechnet.

5.1. Die Testgröße t_0 beträgt in diesem Fall gerundet:

- a) 0,85 ☐
- b) 1,17 ☐
- c) 11,75 ☐
- d) 16,61 ☐
- e) 37,14 ☐

(Fragetyp Einfachwahl)

5.2. Der für die Bestimmung des kritischen Wertes benötigte Freiheitsgrad s beträgt bei diesem Test:

- a) 9 ☐
- b) 10 ☐
- c) 18 ☐
- d) 19 ☐
- e) 20 ☐

(Fragetyp Einfachwahl)

6. Sie möchten mittels eines t-Tests für den Vergleich zweier Erwartungswerte bei unabhängigen Stichproben die Wirksamkeit zweier Pestizide A und B überprüfen. Der Stichprobenumfang beträgt $n = 15$. Ihre Nullhypothese lautet, dass die Wirksamkeit der beiden Pestizide sich nicht unterscheidet ($\mu_x = \mu_y$). Sie wählen eine zweiseitige Alternativhypothese ($\mu_x \neq \mu_y$). Sie wählen ein Signifikanzniveau von $\alpha = 0{,}05$. Die von Ihnen berechnete Testgröße beträgt $t_0 = -1{,}92$.

6.1. Geben Sie an, ob die Nullhypothese abgelehnt oder nicht abgelehnt werden muss!

- a) Nullhypothese wird nicht abgelehnt ☐
- b) Nullhypothese wird abgelehnt ☐

(Fragetyp Einfachwahl)

Ende von Antwort-Wahl-Verfahren, Teil A

Antwort-Wahl-Verfahren, Teil B:

7. Geben Sie an, bei welchen der folgenden Zustandsgrößen es sich um <u>intensive</u> Größen handelt!

a) dynamische Viskosität ☐
b) Temperatur ☐
c) Masse ☐
d) Brechungsindex ☐
e) Geschwindigkeit ☐
f) Impuls ☐
g) molare Masse ☐
h) Wärmekapazität ☐

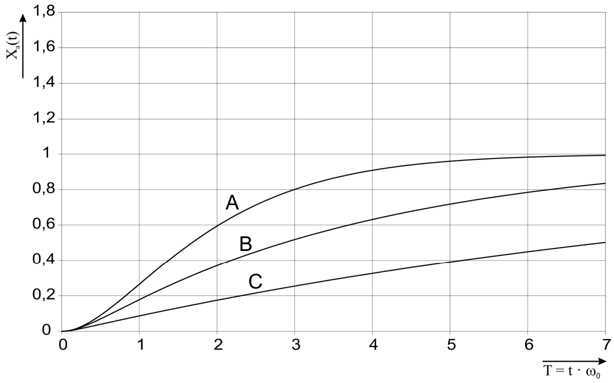
(Fragetyp Mehrfachwahl)

8. Geben Sie an, welche der folgenden Gleichungen korrekt sind!

a) $1\ \mathrm{TW} = 10^3\ \mathrm{GW}$ ☐
b) $10^3\ \mathrm{cm}^3 + 1\ \mathrm{dm}^3 = 2{\cdot}10^{-3}\ \mathrm{m}^3$ ☐
c) $100\ \mu\mathrm{s} + 2\ \mathrm{ms} = 2{,}1{\cdot}10^{-3}\ \mathrm{s}$ ☐
d) $100\ \mathrm{hPa} + 1\ \mathrm{kPa} = 1100\ \mathrm{Pa}$ ☐
e) $330\ \mathrm{pF} = 3{,}3{\cdot}10^{-10}\ \mathrm{F}$ ☐

(Fragetyp Mehrfachwahl)

9. In nachfolgender Abbildung sind die Sprungantworten dreier – mit A, B und C bezeichneter – linearer Systeme 2. Ordnung dargestellt, welche sich hinsichtlich ihrer Dämpfung D unterscheiden. Geben Sie an, welche Kombination von Dämpfungen D_A, D_B und D_C das Verhalten der dargestellten Systeme A, B und C qualitativ am besten beschreibt!



a) $D_A = 0{,}1;\quad D_B = 0{,}3;\quad D_C = 0{,}5$ ☐
b) $D_A = 0{,}3;\quad D_B = 1;\quad D_C = 3$ ☐
c) $D_A = 1;\quad D_B = 2;\quad D_C = 5$ ☐
d) $D_A = 3;\quad D_B = 2;\quad D_C = 1$ ☐

(Fragetyp Einfachwahl)

10. Ein lineares System 1. Ordnung mit der Zeitkonstanten T und dem Übertragungsfaktor $K = 3$ werde aus dem Beharrungszustand heraus zum Zeitpunkt $t = 0$ mit einer sprungförmigen Änderung der Eingangsspannung von -10 V auf $+10$ V beaufschlagt. Welche Spannung wird nach der Zeitdauer $t = T$ am Ausgang ungefähr anliegen?

a) $18{,}9$ V ☐
b) $7{,}8$ V ☐
c) $6{,}3$ V ☐
d) $-2{,}2$ V ☐
e) $-6{,}3$ V ☐

(Fragetyp Einfachwahl)

11. Sie beobachten einen radioaktiven Zerfallsprozess eines Uran-Isotops, bei welchem im Mittel $\lambda = 4{,}5$ Zerfälle pro Sekunde beobachtet werden. Geben Sie an, durch welche statistische Verteilung sich die jeweils pro Sekunde tatsächlich beobachtete Anzahl k der Zerfallsereignisse aller Wahrscheinlichkeit nach in guter Näherung beschreiben lässt!

a) Gleichverteilung ☐
b) Binomialverteilung ☐
c) Hypergeometrische Verteilung ☐
d) Normalverteilung ☐
e) Poissonverteilung ☐

(Fragetyp Einfachwahl)

12. Geben Sie an, wie viel Prozent der Elemente einer Verteilung unterhalb des dritten Dezils liegen!

a) 3% ☐
b) 30% ☐
c) $33{,}3\overline{3}$% ☐
d) 60% ☐
e) 75% ☐

(Fragetyp Einfachwahl)

13. Eine normalverteilte, dimensionslose Größe werde mit 20 Wiederholungen gemessen. Das Konfidenzintervall des Erwartungswertes wird zu $5 \leq \mu \leq 15$ bei $\mathrm{P} = 99\%$ bestimmt. Die Standardabweichung σ sei bekannt. Geben Sie an, wie viele Wiederholungsmessungen bei unveränderter Standardabweichung mindestens durchgeführt werden müssen, um das Konfidenzintervall bei unveränderter Aussagesicherheit auf $8 \leq \mu \leq 12$ zu reduzieren!

a) 50 ☐
b) 80 ☐
c) 90 ☐
d) 100 ☐
e) 125 ☐

(Fragetyp Einfachwahl)

14. Geben Sie an, welche der folgenden Aussagen über spezielle Verteilungsfunktionen zutreffend sind!

a) Erwartungswert und Varianz der Poissonverteilung sind wertgleich. ☐

b) Die Gaußsche Normalverteilung ist symmetrisch zum Erwartungswert μ und ihre Wendepunkte liegen bei $x = \mu \pm \sigma$. ☐

c) Die Gaußsche Normalverteilung beschreibt solche Prozesse gut, auf die eine große Zahl statistisch unabhängiger Einflussgrößen mit gleicher Größenordnung einwirkt. ☐

d) Für eine sehr große Zahl von Versuchen ($n \rightarrow \infty$) nähert sich die Student'sche t-Verteilung der Gaußsche Normalverteilung an. ☐

e) Die Binomialverteilung beschreibt den wahrscheinlichen Ausgang einer Folge gleichartiger Versuche, bei denen es nur zwei mögliche Ergebnisse gibt. ☐

(Fragetyp Mehrfachwahl)

15. Geben Sie an, welche der folgenden Aussagen über statistische Tests korrekt sind!

a) Als Fehlentscheidung 1. Art bezeichnet man den Fall, dass als Ergebnis eines statistischen Tests die Nullhypothese H_0 abgelehnt wird, obwohl H_0 tatsächlich zutrifft. ☐

b) Wird für einen statistischen Test ein Signifikanzniveau von 1% gewählt, bedeutet dies, dass mit einer Wahrscheinlichkeit von maximal 1% eine Fehlentscheidung auftritt. ☐

c) Die Güte eines statistischen Tests lässt sich durch Reduzierung des Signifikanzniveaus erhöhen. ☐

d) Eine Messreihe, die zur Bildung einer Hypothese verwendet wurde, darf <u>nicht</u> für einen Test dieser Hypothese genutzt werden. ☐

e) In experimentellen Wissenschaften können statistische Tests dazu genutzt werden, Hypothesen abzusichern oder begründet zu verwerfen. ☐

(Fragetyp Mehrfachwahl)

Ende von Antwort-Wahl-Verfahren, Teil B

Kurzfragen:

16. Erläutern Sie, wodurch sich *nominalskalierte* Daten und *ordinalskalierte* Daten unterscheiden! Nennen Sie für beide Datentypen je ein Beispiel!

17. Erläutern Sie, was unter der Hysterese eines Messgerätes zu verstehen ist!

18. Skizzieren Sie anhand eines Sinussignals exemplarisch, wie es durch Verletzung des Abtasttheorems nach Shannon zu einer fehlerhaften Rekonstruktion des Ursprungssignals kommen kann!

Ende der Kurzfragen

Leerseite

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)

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

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

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

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

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

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:

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

t-Test

t-Test für Erwartungswert

Die Testgröße:

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

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

t-Test für Vergleich zweier Erwartungswerte

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

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

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:

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

mit:

$$d_i = x_i - y_i$$

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

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

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
   $$|t_0| > t_{n-1,1-\frac{\alpha}{2}} \;,$$
   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 erforderlichenfalls 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 aus der Stichprobe abgeschätzten Parameter der Verteilungsdichtefunktion
   - Die Zahl der Freiheitsgrade ist $df = r^* - s - 1$
9. Festlegen der Irrtumswahrscheinlichkeit α

H_0 ist abzulehnen mit Signifikanzniveau α, wenn:

$$\chi_0^2 > \chi^2_{r^*-s-1,1-\alpha}$$

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

s \ p	0,9	0,95	0,975	0,98	0,99	0,995
1	3,078	6,314	12,706	15,895	31,821	63,657
2	1,886	2,920	4,303	4,849	6,965	9,925
3	1,638	2,353	3,182	3,482	4,541	5,841
4	1,533	2,132	2,776	2,999	3,747	4,604
5	1,476	2,015	2,571	2,757	3,365	4,032
6	1,440	1,943	2,447	2,612	3,143	3,707
7	1,415	1,895	2,365	2,517	2,998	3,499
8	1,397	1,860	2,306	2,449	2,896	3,355
9	1,383	1,833	2,262	2,398	2,821	3,250
10	1,372	1,812	2,228	2,359	2,764	3,169
11	1,363	1,796	2,201	2,328	2,718	3,106
12	1,356	1,782	2,179	2,303	2,681	3,055
13	1,350	1,771	2,160	2,282	2,650	3,012
14	1,345	1,761	2,145	2,264	2,624	2,977
15	1,341	1,753	2,131	2,249	2,602	2,947
16	1,337	1,746	2,120	2,235	2,583	2,921
17	1,333	1,740	2,110	2,224	2,567	2,898
18	1,330	1,734	2,101	2,214	2,552	2,878
19	1,328	1,729	2,093	2,205	2,539	2,861
20	1,325	1,725	2,086	2,197	2,528	2,845
21	1,323	1,721	2,080	2,189	2,518	2,831
22	1,321	1,717	2,074	2,183	2,508	2,819
23	1,319	1,714	2,069	2,177	2,500	2,807
24	1,318	1,711	2,064	2,172	2,492	2,797
25	1,316	1,708	2,060	2,167	2,485	2,787
26	1,315	1,706	2,056	2,162	2,479	2,779
27	1,314	1,703	2,052	2,158	2,473	2,771
28	1,313	1,701	2,048	2,154	2,467	2,763
29	1,311	1,699	2,045	2,150	2,462	2,756
30	1,310	1,697	2,042	2,147	2,457	2,750
40	1,303	1,684	2,021	2,123	2,423	2,704
50	1,299	1,676	2,009	2,109	2,403	2,678
60	1,296	1,671	2,000	2,099	2,390	2,660
70	1,294	1,667	1,994	2,093	2,381	2,648
80	1,292	1,664	1,990	2,088	2,374	2,639
90	1,291	1,662	1,987	2,084	2,368	2,632
100	1,290	1,660	1,984	2,081	2,364	2,626
200	1,286	1,653	1,972	2,067	2,345	2,601
∞	1,282	1,645	1,960	2,054	2,326	2,576

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

s \ p	0,90	0,95	0,975	0,99	0,995
1	2,71	3,84	5,02	6,63	7,88
2	4,61	5,99	7,38	9,21	10,6
3	6,25	7,81	9,35	11,3	12,8
4	7,78	9,49	11,1	13,3	14,9
5	9,24	11,1	12,8	15,1	16,8
6	10,6	12,6	14,5	16,8	18,6
7	12,0	14,1	16,0	18,5	20,3
8	13,4	15,5	17,5	20,1	22,0
9	14,7	16,9	19,0	21,7	23,6
10	16,0	18,3	20,5	23,2	25,2
11	17,3	19,7	21,9	24,7	26,8
12	18,6	21,0	23,3	26,2	28,3
13	19,8	22,4	24,7	27,7	29,8
14	21,2	23,7	26,1	29,1	31,3
15	22,3	25,0	27,5	30,6	32,8
16	23,5	26,3	28,9	32,0	34,3
17	24,8	27,6	30,2	33,4	35,7
18	26,0	28,9	31,5	34,8	37,2
19	27,2	30,1	32,9	36,2	38,6
20	28,4	31,4	34,2	37,6	40,0
21	29,6	32,7	35,5	38,9	41,4
22	30,8	33,9	36,8	40,3	42,8
23	32,0	35,2	38,1	41,6	44,2
24	33,2	36,4	39,4	43,0	45,6
25	34,4	37,7	40,6	44,3	46,9
26	35,6	38,9	41,9	45,6	48,3
27	36,7	40,1	43,2	47,0	49,6
28	37,9	41,3	44,5	48,3	51,0
29	39,1	42,6	45,7	49,6	52,3
30	40,3	43,8	47,0	50,9	53,7
40	51,8	55,8	59,3	63,7	66,8
50	63,2	67,5	71,4	76,2	79,5
60	74,4	79,1	83,3	88,4	92,0
70	85,5	90,5	95,0	100,4	104,2
80	96,6	101,9	106,6	112,3	116,3
90	107,6	113,1	118,1	124,1	128,3
100	118,5	124,3	129,6	135,8	140,2

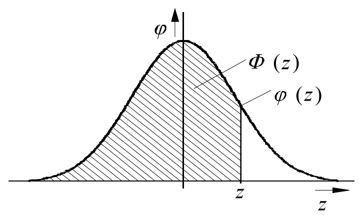
Tabelle 1

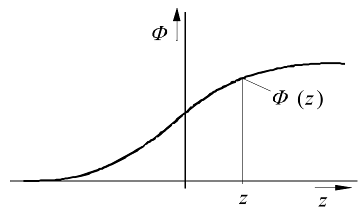
Summenfunktion der standardisierten Normalverteilung

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

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

z-Transformation: $z = \dfrac{x - \mu}{\sigma}$





z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z
0,0	0,500000	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
Φ(z)	$1-1{,}350\cdot10^{-3}$	$1-2{,}326\cdot10^{-4}$	$1-3{,}167\cdot10^{-5}$	$1-3{,}398\cdot10^{-6}$	$1-2{,}867\cdot10^{-7}$	$1-9{,}866\cdot10^{-10}$	$1-1{,}280\cdot10^{-12}$	$1-6{,}221\cdot10^{-16}$	$1-1{,}129\cdot10^{-19}$	$1-7{,}620\cdot10^{-24}$	Φ(z)

Φ(z)	50%	60%	70%	80%	90%	95%	97,5%	99%	99,5%	99,75%	99,9%	99,95%	Φ(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