

Beispielsammlung zum Kurs Statistik 2 (WS 2008)

Andreas Quatember

Ablauf des Kurses:

- In den einzelnen Lehrveranstaltungen wird der Lehrstoff des Kurses Statistik erarbeitet.
- Die Hausübung, die jeweils bis zum nächsten Kurstermin vorzubereiten ist, besteht immer aus den nächsten 3 HÜ-Beispielen (somit insgesamt $3 \times 11 = 33$), die *in der Art vorzubereiten sind, dass die Studierenden im Kurs den Lösungsweg vortragen können*. Vor jedem Kurs dürfen nur die in dieser Weise vorbereiteten Beispiele von den Studierenden in einer beim Hörsaalzugang ausgehängten „Kreuzelliste“ angekreuzt werden. Das Ankreuzen von Beispielen bei Nichtanwesenheit bzw. wenn kein Lösungsweg vorgetragen werden kann ist *selbstverständlich nicht erlaubt* und hat eine *negative Bewertung* der Mitarbeit und somit eine negative Gesamtbeurteilung zur Folge. Aufgrund der hohen Studierendenzahlen muss um Verständnis dafür ersucht werden, dass das ausnahmsweise **Abgeben oder Zusenden** von Hausübungsbeispielen aus organisatorischen Gründen **völlig unmöglich** ist.
- Sollte es Ihnen (z.B. wegen Elternschaft oder Berufstätigkeit) voraussichtlich nicht möglich sein, an der für die positive Beurteilung der Hausübungen notwendigen Anzahl an Lehrveranstaltungen teilzunehmen, dann setzen Sie sich schon vorab (bzw. so schnell es geht) mit dem Lehrveranstaltungsleiter in Verbindung.
- Die zusätzlichen Ü-Beispiele in diesem Skriptum sind Beispiele, die zur selbstständigen Vorbereitung auf die Klausur dienen.

Literatur für den Kurs:

Quatember, A. (2008; NEUE 2. Auflage). *Statistik ohne Angst vor Formeln*. Pearson Studium, München.

Weg zur Benotung:

Für eine positive Benotung müssen

- eine Klausur positiv absolviert werden und
- mindestens 17 Hausübungsbeispiele (= positive Mitarbeit) angekreuzt und gewertet werden. Somit ist diese Hausübungszahl Mindestvoraussetzung für den Klausurantritt. Eine negative Mitarbeitsbewertung (unter 17 gewertete Hausübungsbeispiele) ergibt auch eine negative Gesamtbeurteilung.

Die Klausur besteht aus 7 Beispielen zu je einem möglichen Punkt. Ab vier Punkten ist die Klausur positiv. Für die Gesamtbeurteilung werden – allerdings nur wenn die Klausur selbst positiv beurteilt wurde – zum Klausurergebnis ab 24 gewerteten Hausübungsbeispielen ein halber und ab 29 ein ganzer Punkt als Mitarbeitsbonus addiert.

Punkte und Noten:

Punkte:	Note:
7 – 8	sehr gut
6 – 6,5	gut
5 – 5,5	befriedigend
4 – 4,5	genügend
unter 4	nicht genügend



Denise Rudel
ÖH Shop-Referentin



Julia Sageder
ÖH Vorsitz-Team



Susi Aichinger
ÖH Vorsitz-Team

Liebe Kollegin, lieber Kollege!

Vor dir siehst du ein Skript des Open Courseware Projekts der ÖH Linz, welches allen Studierenden und Interessierten frei und kostenlos zur Verfügung steht.

Das OCW- Projekt der ÖH Linz

Im Jahr 2007 haben der Vorsitz der österreichischen HochschülerInnenschaft Linz und das Referat für Skripten, Lernbehelfe und OCW mit der Umsetzung von Open Courseware an der Johannes Kepler Universität begonnen. Alle Skripten sollten den Studierenden und Interessierten kostenlos zugänglich sein, zudem sollten die Unterlagen frei verändert und vervielfältigt werden dürfen um die Qualität und Aktualität der Unterlagen zu verbessern.

Zu diesem Zweck wurden alle Unterlagen, deren Lizenz bei der ÖH liegt, digitalisiert, mit einer Struktur und Suchfunktion versehen und über eine Homepage allen InternetnutzerInnen zugänglich gemacht. Darüber hinaus wurde den Lehrenden an der JKU die Möglichkeit gegeben jederzeit Verbesserungen und Ergänzungen bei den Unterlagen vorzunehmen.

Lizenz

Um die freie Verbreitung rechtlich zu gewährleisten steht dieses Werk unter einer Creative Commons Lizenz 3.0 Österreich.

Du darfst das Werk vervielfältigen, verbreiten und öffentlich zugänglich machen sowie Bearbeitungen des Werkes anfertigen.

Jedoch musst du dich dabei an gewisse Bedingungen halten:

- Du musst den Namen der/des Autorin/Autors / Rechteinhabers/Rechteinhaberin in der von ihm festgelegten Weise nennen.
- Das Werk darf nicht kommerziell genutzt werden.
- Die Weitergabe ist nur unter gleichen Bedingungen erlaubt, also unter der gleichen Lizenz.

Weitere und genauere Informationen über Creative Commons findest du unter

<http://www.creativecommons.at>.

Solltest du noch weitere Fragen zum OCW Projekt haben, oder dich beteiligen wollen, erreichst du uns unter oeh@oeh.jku.at oder **+43 732 2468 8535**.

Wir wünschen dir viel Spaß mit den OCW Skripten und viel Erfolg bei deinen Kursen!

Erlaubte Unterlagen bei der Klausur:

Alles mit Ausnahme der ausgerechneten Hausübungs- und Übungsbeispiele, ausgerechneter alter Klausuren, dieser Beispielsammlung und eines Laptops.

Termin der Klausur:

Dienstag, 16. Juni 2009, 13.45 bis 15.15 Uhr

(Nachklausurtermin: Dienstag, 29. September 2009, 18.15 bis 19.45 Uhr)

Da dafür gesorgt werden muss, dass genügend Klausurbögen kopiert, Hörsäle reserviert und Aufsichtspersonen rekrutiert werden, ist es absolut notwendig, sich für die Klausuren gesondert anzumelden:

Anmeldung für die Klausur:

Von Dienstag, 2., bis Dienstag, 9. Juni im KUSSS; ab Freitag, 12. Juni, wird im KUSSS auf der Lehrveranstaltungsseite eine Liste veröffentlicht, aus der jeder Studierende seine Klausurnummer und den Hörsaal erfährt.

Anmeldung für die Nachklausur:

Von Montag, 14., bis Montag, 21. September, im KUSSS; ab Montag, 28. September, wird auf der Lehrveranstaltungsseite im KUSSS eine Liste veröffentlicht, aus der jeder Studierende seine Klausurnummer und den Hörsaal für die Klausur erfährt. Für die Teilnahme zu diesem Nachklausurtermin ist die Teilnahme zum Juni-Termin natürlich *nicht* Voraussetzung.

Begleitendes Tutorium zum Kurs Statistik II mit Besprechung der Hausübungen (die nachfolgenden Termine sind provisorisch und werden in der 1. Lehrveranstaltung endgültig fixiert):

Jeden Dienstag vom 10. März bis 9. Juni in der Zeit von 12.45 bis 13.30 Uhr im K009D (Keplergebäude).

Bei allen Beispielen sind Ihre Ergebnisse von Ihnen auch zu interpretieren. Die Überschriften dieser Beispielsammlung stimmen mit jenen in Quatember (2008) überein.



Dieses Zeichen zeigt an, dass Sie sich für das Lösen des betreffenden Beispiels der Excel-Datei zum Kurs (Lehrveranstaltungsseite im KUSSS) bedienen müssen.

3 Schließende Statistik

3.1 Grundbegriffe und Handlungslogik

HÜ1)

Lesen Sie sich die ersten beiden Seiten dieser Beispielsammlung durch („Ablauf des Kurses“ etc.), damit Sie die Regeln, unter denen der Kurs und schließlich auch die Beurteilung stattfindet, und die Termine zur Kenntnis nehmen.

Bearbeiten Sie danach folgende Aufgabe: Sie sollen mittels einer repräsentativen Stichprobe erheben, wie viel Prozent der in Linz inskribierten Studierenden berufstätig sind. Überlegen Sie sich detailliert, wie Sie vorgehen würden, wenn Sie zu diesem Zweck 400 Studierende befragen sollen?

HÜ2)



Im Internet finden Sie in der für diesen Kurs nötigen Exceldatei (KUSSS) die Namen von 100 an einer Lehrveranstaltung teilnehmenden Studierenden und ihre Semesterzahlen. Sie sollen aus dieser Grundgesamtheit eine Zufallsstichprobe vom Umfang $n = 10$ ziehen, in die jeder dieser Studierenden mit gleicher Wahrscheinlichkeit gelangen soll.

- Überlegen Sie sich eine Vorgehensweise, die ohne Computereinsatz auskommt.
- Ziehen Sie die Stichprobe unter Zuhilfenahme der Excel-Funktion ZUFALLSZAHL.

HÜ3)

Berechnen Sie mit Ihrer in HÜ2b) gezogenen Stichprobe den Punktschätzer für

- die relative Häufigkeit derer, die in der Grundgesamtheit aller teilnehmenden Studierenden eine Semesterzahl von 2 aufweisen (d.h. für diejenigen, die im 2. Semester studieren)
- den Mittelwert der Semesteranzahl.

Ü1)

Besitzen alle Personen der gegenwärtigen österreichischen Wohnbevölkerung gleiche Auswahlchancen für eine Stichprobe, wenn diese aus

- der Volkszählung 2001
- dem Telefonverzeichnis erstellt wird?

Begründen Sie Ihre Antworten!

Ü2)

Warum ist eine Stichprobe, die

- aus in einer bestimmten Zeitspanne auf dem Linzer Hauptplatz Vorbeikommenden und Antwortwilligen besteht, eine hinsichtlich des Merkmals Musiktheaterstandort nichtrepräsentative Stichprobe für die Linzer Bevölkerung?
- aus den Anwesenden in dieser LV besteht, eine nichtrepräsentative Stichprobe für die Grundgesamtheit aller Linzer Studierenden hinsichtlich der Einstellung zum Studienfach Statistik?

Ü3)

Überlegen Sie sich eine mögliche Vorgehensweise zur Ziehung einer repräsentativen Stichprobe aus der Linzer Bevölkerung zur Erhebung der Einstellung zum Musiktheaterstandort UKH.

3.2 Schätzen und Testen einer relativen Häufigkeit

3.2.1 Schätzen einer relativen Häufigkeit

HÜ4)

Bei der Abstimmung über den EU-Beitritt Österreichs stimmten im Jahr 1994 66,6 % der Bevölkerung dafür. In welchem Bereich wäre an diesem Tag mit einer Wahrscheinlichkeit von 0,95 ein diesbezüglicher Anteil in einer zufällig aus dieser Grundgesamtheit gezogenen uneingeschränkten Zufallsauswahl gelegen, wenn man dazu

- 100
- 400
- 800

Personen ausgewählt hätte?

HÜ5)



Eine Woche vor der letzten Nationalratswahl wurden zufällig ausgewählte wahlberechtigte Österreicherinnen und Österreicher gefragt, welche Partei sie wählen werden. Die kodierten Antworten finden Sie in der im Internet bereit stehenden Excel-Datei.

Bestimmen Sie die relative Häufigkeit der Partei A in der Stichprobe mit Hilfe der Excel-Funktion HÄUFIGKEIT (siehe Statistik 1) und berechnen Sie anschließend mit dem Taschenrechner das Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ für die relative Häufigkeit an Wählerinnen und Wählern dieser Partei zu diesem Zeitpunkt in der Gesamtbevölkerung.

HÜ6)

Unter 800 zufällig ausgewählten Befragten ist der Prozentsatz derer, die auf die Frage: „Sind Sie mit den Auswirkungen des EU-Beitritts auf Österreich unzufrieden?“ mit „ja“ geantwortet haben, 70 %.

Berechnen Sie damit das Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ für die relative Häufigkeit an „Unzufriedenen“ in der Gesamtbevölkerung und berechnen Sie dieses auch für den Fall, dass der Stichprobenumfang nicht 800, sondern 400 bzw. 1.200 war.

HÜ7)

In einer Meinungsumfrage soll festgestellt werden, wie hoch der derzeitige Stimmenanteil einer bestimmten Partei wäre. Wie viele Wahlberechtigte sind in einer uneingeschränkten Zufallsauswahl aus der Grundgesamtheit mindestens zu befragen, wenn das Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ eine Schwankungsbreite ε von 0,03 besitzen soll und man davon ausgehen kann, dass diese relative Häufigkeit

- a) bei etwa 42 %
- b) bei höchstens 40 %
- c) zwischen 5 und 15 %
- d) bei höchstens 60 % liegen wird?

HÜ8)

Verwenden Sie die Angabe zu HÜ7 und bestimmen Sie die notwendigen Stichprobenumfänge für a) – d) nochmals bei erlaubten Schwankungsbreiten von 0,01 bzw. 0,05.

HÜ9)

Die relative Häufigkeit des Auftretens einer bestimmten Eigenschaft in der Gesamtbevölkerung soll in einer Stichprobe geschätzt werden. Welcher Stichprobenumfang ist zu wählen, wenn keinerlei Abschätzung des tatsächlichen Anteils existiert und das Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ eine von Ihnen festzulegende Schwankungsbreite aufweisen soll?

Ü4)

Der prozentuelle Anteil der Frauen sei in der über 18-jährigen Bevölkerung eines Landes 52 %. Wie groß wäre mit einer Wahrscheinlichkeit von 0,95 der diesbezügliche Anteil in einer zufälligen Stichprobe aus dieser Grundgesamtheit, wenn man dazu

- a) 100
- b) 1.000
- c) 10.000

Personen auswählt?

([0,422;0,618] [0,489;0,551] [0,510;0,530])

Ü5)

Eine Zufallsstichprobe von $n = 1.500$ Werkstücken aus einer Produktion ergab einen Ausschussanteil $p = 0,043$. Berechnen Sie daraus das Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ für den Ausschussanteil π in der gesamten Produktion. ([0,033; 0,053])

Ü6)

In einer Medienanalyse wurde bei einer Befragung von 1.400 zufällig ausgewählten Personen erhoben, dass 43,6 % der Oberösterreicher täglich zur *Kronen-Zeitung* greifen. Berechnen Sie das Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ für diese relative Häufigkeit in der Grundgesamtheit. ([0,410; 0,462])

Ü7)

In einer Meinungsumfrage soll mittels einer Zufallsstichprobe erhoben werden, wie hoch die derzeitige Ablehnung von Atomkraftwerken im Land ist. Wie viele Wahlberechtigte sind mindestens zu befragen, wenn das Konfidenzintervall zur Sicherheit

$1-\alpha = 0,95$ für diesen Anteil eine Schwankungsbreite von $\varepsilon = 0,025$ besitzen soll und man vermutet, dass dieser

- a) zwischen 0,8 und 0,9
- b) bei mindestens 0,4 liegen wird? (984 1.537)

Ü8)

Verwenden Sie die Angabe zu Ü7 und bestimmen Sie die notwendigen Stichprobenumfänge für a) – b) nochmals bei einer erlaubten Schwankungsbreite von 0,04. (385 601)

Ü9)

Verwenden Sie die Angabe zu Ü7 und bestimmen Sie den notwendigen Stichprobenumfang für den Fall, dass keinerlei Abschätzung des Anteils an Atomkraftgegnern existiert. (1.537)

3.2.2 Testen von Hypothesen über eine relative Häufigkeit

HÜ10)



Eine politische Partei will feststellen, ob sich die relative Häufigkeit an EU-Vertragsgegnern gegenüber dem Ergebnis einer Volksabstimmung verändert hat. Bei dieser Volksabstimmung waren 54,2 % dagegen gewesen.

- a) Formulieren Sie für dieses Problem geeignete statistische Hypothesen. In einer aktuellen Umfrage unter $n = 500$ Wahlberechtigten erhält man die Antworten, die sich in der im Internet bereit stehenden Excel-Datei kodiert widerspiegeln.
- b) Entscheiden Sie sich auf Basis dieses Stichprobenergebnisses auf einem Signifikanzniveau $\alpha = 0,05$ mit dem Taschenrechner für eine der beiden Hypothesen.
- c) Wie würde ihre Entscheidung bei $p = 0,662$ lauten?
- d) Wie würde ihre Entscheidung bei $p = 0,462$ lauten?

HÜ11)

In einer Zufallsstichprobe vom Umfang $n = 750$ aus der wahlberechtigten, oberösterreichischen Bevölkerung stellte man Ende des Jahres 2002 fest, dass 411 der Befragten der neuen Eurowährung positiv gegenüberstanden. Überprüfen Sie, ob man daraus auf einem Signifikanzniveau von $\alpha = 0,05$ folgern konnte, dass zu diesem Zeitpunkt eine Mehrheit in der oberösterreichischen Bevölkerung positiv eingestellt war?

HÜ12)

Entscheiden Sie sich bei den Hypothesen in HÜ10 und HÜ11 auf dem Signifikanzniveau $\alpha = 0,05$, wenn Sie erfahren, dass der zweiseitige p-Wert des Tests

- a) 0,225
 - b) 0,064
 - c) 0,021 betragen hat.
-

Ü10)

Von Werkstücken, die ein Jahr lang gelagert wurden, sind 40 % unbrauchbar. Nach einer Änderung der Lagerbedingungen soll überprüft werden, ob sich dieser Anteil an unbrauchbaren Werkstücken verringert hat.

- Formulieren Sie für dieses Problem geeignete statistische Hypothesen.
- Entscheiden Sie sich auf einem Signifikanzniveau $\alpha = 0,05$ für eine der Hypothesen, wenn sich unter $n = 100$ zufällig ausgewählten Stücken 36 % unbrauchbare befanden.
($H_0: \pi \geq 0,4$, $H_1: \pi < 0,4$; Entscheidung für H_0)

Ü11)

Entscheiden Sie sich mit den Daten aus Ü13 auf einem Signifikanzniveau $\alpha = 0,05$ für eine der beiden Hypothesen, wenn sich unter

- $n = 450$
- $n = 350$ zufällig ausgewählten Stücken 36 % unbrauchbare befanden.

(Entscheidung für: H_1 ; H_0)

Ü12)

Eine politische Partei will überprüfen, ob ihr Kandidat für ein Gemeindeamt der Stadt schon mehr als 50 % Zustimmung in der Grundgesamtheit der Wähler besitzt.

- Formulieren Sie für dieses Problem geeignete statistische Hypothesen. In einer dazu durchgeführten Umfrage unter $n = 200$ zufällig ausgewählten Wahlberechtigten betrug die diesbezügliche relative Häufigkeit $p = 0,525$.
- Entscheiden Sie sich auf Basis dieses Stichprobenergebnisses auf einem Signifikanzniveau von $\alpha = 0,05$ für eine der Hypothesen.

($H_0: \pi \leq 0,5$ $H_1: \pi > 0,5$; Entscheidung für H_0)

3.3 Schätzen und Testen eines Mittelwerts

3.3.1 Schätzen eines Mittelwerts

HÜ13)



Vor der Erprobung der Wirksamkeit eines Schlafmittels (siehe HÜ18) wird zuerst an 264 zufällig ausgewählten Testpersonen das Merkmal Schlafdauer gemessen. Die Messergebnisse finden Sie in der im Internet bereit stehenden Excel-Datei. Berechnen Sie mit diesen Daten ein Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ für die mittlere Schlafdauer in der Grundgesamtheit.

Ü13)

Bei der Abfüllung eines Mineralwassers in Literflaschen ist der Magnesiumgehalt je Liter normalverteilt. $n = 116$ Kontrollmessungen ergaben folgende Werte für den Gehalt x an Magnesium (in mg/l): $\bar{x} = 25,452$, $s^2 = 0,850$. Berechnen Sie das Konfidenzintervall zur Sicherheit $1-\alpha = 0,95$ für den mittleren Magnesiumgehalt in der Gesamtproduktion.
([25,284;25,620])

3.3.2 Testen von Hypothesen über einen Mittelwert

HÜ14)



Eine Zufallsstichprobe vom Umfang $n = 86$ ergab hinsichtlich eines normalverteilten Merkmals x Messergebnisse, die in der im Internet bereit gestellten Excel-Datei zu finden sind.

Testen Sie mit diesen Daten zum Signifikanzniveau $\alpha = 0,05$ folgende Hypothesen:

- a) $H_0: \mu = 4500$ gegen $H_1: \mu \neq 4500$
- b) $H_0: \mu \geq 4500$ gegen $H_1: \mu < 4500$.

HÜ15)

Die stetige Punktezahl bei einem Aufnahmetest sei annähernd normalverteilt mit $\mu = 75$. Nach Einführung verpflichtender vorbereitender Kurse soll überprüft werden, ob sich der Mittelwert erhöht hat.

- a) Formulieren Sie für dieses Problem geeignete statistische Hypothesen.
- b) Testen Sie diese Hypothesen auf einem Signifikanzniveau von $\alpha = 0,05$, wenn in einer Zufallsstichprobe vom Umfang $n = 120$ nach Einführung der Konzentrationsübungen ein Mittelwert von 78,4 Punkten erzielt wird und in dieser Stichprobe eine Standardabweichung von $s = 6$ gemessen wird.

Ü14)

Die Psychologen Stanford, Binet und Wechsler haben festgestellt, dass der Intelligenzquotient in der Bevölkerung nicht gleichverteilt, sondern normalverteilt ist. In der Bevölkerung besitzt der „IQ“ einen Mittelwert von 100.

Überprüfen Sie auf einem Signifikanzniveau von $\alpha = 0,05$, ob dieser Mittelwert unter Studierenden höher ist. Dazu steht eine Zufallsstichprobe vom Umfang $n = 100$ zur Verfügung, in der ein durchschnittlicher IQ von 108 bei einer Standardabweichung von $s = 19$ gemessen wurde. ($H_0: \mu \leq 100$ $H_1: \mu > 100$; H_0 wird verworfen, H_1 akzeptiert)

Ü15)

Entscheiden Sie sich bei den Hypothesen in HÜ14a), HÜ14b), HÜ15 und Ü14 jeweils auf dem Signifikanzniveau $\alpha = 0,05$, wenn Sie erfahren, dass der zweiseitige p-Wert des Tests

	HÜ14a)	HÜ14b)	HÜ15	Ü14
a) 0,225	a) H_0	H_0	H_0	H_0
b) 0,064	b) H_0	H_1	H_1	H_1
c) 0,021 betragen hat.	c) H_1	H_1	H_1	H_1

3.4 Testen von Hypothesen über zwei relative Häufigkeiten

HÜ16)



Zwei im Abstand zweier Monate gezogene unabhängige Stichproben von je 700 Wahlberechtigten erhoben für den Spitzenkandidaten einer Partei seine Sympathisanten. Sie finden die Daten in der im Internet bereit stehenden Excel-Datei.

Die Schlagzeile in einer Zeitung lautete: „Kandidat legt zu!“ Überprüfen Sie diese Behauptung mit den vorliegenden Daten auf einem Signifikanzniveau von $\alpha = 0,05$, nachdem Sie geeignete Hypothesen für diesen Test aufgestellt haben.

HÜ17)

Verwenden Sie die Daten aus HÜ16: Die Partei will nun überprüfen, ob sich die relative Häufigkeit an Sympathisanten für ihren Spitzenkandidaten innerhalb der zwei Monate auf einem Signifikanzniveau von $\alpha = 0,05$ verändert hat.

Ü16)

Es soll die positive Wirkung eines konzentrationssteigernden Mittels auf die Lernleistung getestet werden. Dazu werden unabhängig voneinander eine Gruppe von 200 Versuchspersonen ohne Einnahme des Mittels und eine gleich große nach Einnahme des Mittels geprüft. Es bestehen 41,5 % der ersten und 44,0 % der zweiten Gruppe.

Formulieren Sie geeignete Hypothesen und testen Sie diese auf einem Signifikanzniveau von $\alpha = 0,05$.

(A: Gruppe ohne, B: Gruppe mit dem Mittel: $H_0: \delta \geq 0$ $H_1: \delta < 0$;
 $p = 0,4275$ $d_u = -0,082$; H_0 wird beibehalten)

Ü17)

Eine Stichprobe vom Umfang $n = 350$ lieferte vor einem Jahr eine relative Häufigkeit von 64,0 % an Zustimmung zur Regierungsarbeit. Es soll nun auf einem Signifikanzniveau von $\alpha = 0,05$ überprüft werden, ob sich dieser Anteil im Jahresabstand verändert hat. Dazu wird eine neue Stichprobe vom Umfang $n = 620$ gezogen. In dieser ergibt sich ein diesbezüglicher Anteil von 62,1 %.

Formulieren Sie die diesbezüglichen Hypothesen und führen Sie den Test durch.

($H_0: \delta = 0$ $H_1: \delta \neq 0$ $p = 0,628$ $[-0,063; 0,063]$; H_0 wird beibehalten)

3.5 Testen von Hypothesen über zwei Mittelwerte

HÜ18)

In HÜ13 wurde an 264 zufällig ausgewählten Personen das Merkmal Schlafdauer gemessen. Nun wird eine zweite, von der ersten unabhängige Stichprobe vom Umfang 200 gezogen und abermals das Merkmal Schlafdauer gemessen, nachdem diesen neuen 200 Personen jedoch Schlafmittel verabreicht wurden. Als Kennzahlen dieser neuen Stichprobe ergeben sich: $\bar{x} = 7,1$, $s = 1,5$.

Überprüfen Sie auf einem Signifikanzniveau von $\alpha = 0,05$, ob sich die Einnahme des Schlafmittels positiv auf die Schlafdauer auswirkt.

Ü18)

Für die Lebensdauern von Taschenrechnerbatterien soll überprüft werden, ob sich die diesbezüglichen Mittelwerte zweier Hersteller in den Grundgesamtheiten unterscheiden. Eine Stichprobe vom Umfang 125 der einen Marke liefert einen Mittelwert \bar{x} von 5.996,5 h und eine Standardabweichung von $s = 65,3$ h. Eine Stichprobe von 122 Batterien der anderen Marke ergibt diesbezüglich die Werte: $\bar{x} = 6.125,6$ und $s = 57,0$.

Formulieren Sie geeignete Hypothesen und entscheiden Sie sich auf einem Signifikanzniveau von $\alpha = 0,05$ für eine der beiden.

($H_0: \delta = 0$, $H_1: \delta \neq 0$; $[-15,276;15,276]$; H_0 wird verworfen, H_1 akzeptiert)

3.6 Testen einer Hypothese über einen statistischen Zusammenhang zweier nominaler Merkmale

HÜ19)

In HÜ7 im Kurs Statistik (Ü11 in Quatember (2008), S.36) wurde das Ergebnis einer Befragung unter 794 Personen zum Thema Geschlecht und Euro-Einstellung präsentiert. In HÜ17 im Kurs Statistik (Ü28 in Quatember (2008), S.77) wurde mit Hilfe der Excel-Lerndatei zu diesem Kurs der statistische Zusammenhang dieser beiden Merkmale gemessen.

Verwenden Sie nun diese Ergebnisse, um zu überprüfen, ob die beiden Merkmale auch in der Grundgesamtheit aller Wahlberechtigten zusammenhängen.

- Formulieren Sie für dieses Problem geeignete statistische Hypothesen.
- Entscheiden Sie sich an Hand der Ergebnisse von HÜ17 auf einem Signifikanzniveau von $\alpha = 0,05$ für eine der Hypothesen.
- Verwenden Sie die Excel-Funktion CHIVERT: Wie groß ist der zur Teststatistik χ^2_{err} gehörende p-Wert?

HÜ20)

In einer Studie wird der Einfluss von Strategietraining bei $n = 235$ zufällig ausgewählten Anlageberatern auf den Verkaufserfolg untersucht:

		Verkaufserfolg	
		nein	ja
Training	nein	40	75
	ja	30	90

Prüfen Sie auf einem Signifikanzniveau $\alpha = 0,05$, ob in der Grundgesamtheit aller Anlageberater ein statistischer Zusammenhang zwischen den beiden Merkmalen besteht. Wie groß ist der zur Teststatistik gehörende p-Wert (Excel)?

HÜ21)

Zur Untersuchung der Wirksamkeit von Vorsorgeimpfungen gegen Grippe wurde eine Zufallsstichprobe aus der oberösterreichischen Bevölkerung vom Umfang $n = 1.000$ gezogen. Die Überprüfung des Zusammenhangs der beiden Merkmale Impfschutz (ja/nein) und Erkrankung (ja/nein) ergab einen p-Wert von 0,034.

Überprüfen Sie mit dieser Informationen über den Ausgang des statistischen Tests auf dem Signifikanzniveau $\alpha = 0,05$, ob die Merkmale Impfschutz und Erkrankung in der Grundgesamtheit einen statistischen Zusammenhang aufweisen.

Ü19)

In einer Studie zum Thema Museumsneubau wurde in einer Zufallsstichprobe unter 1.14 wahlberechtigten Bürgerinnen und Bürgern deren gemeinsame Verteilung auf den Merkmalen Geschlecht und Einstellung zum Neubau erhoben:

		Einstellung	
		positiv	negativ
Geschlecht	weiblich	149	450
	männlich	250	275

- a) Formulieren Sie für einen Test über einen statistischen Zusammenhang der Merkmale in der Grundgesamtheit geeignete statistische Hypothesen.
 b) Testen Sie diese Hypothesen anhand der Daten auf einem Signifikanzniveau $\alpha = 0,05$.
 ($H_0: \chi^2=0$, $H_1: \chi^2>0$; $\chi^2_{\text{err}} = 61,84$; Obere Schranke der schwachen Indizien gegen H_0 : 3,84; H_0 wird verworfen, H_1 akzeptiert)

Ü20)

In einer Zufallsstichprobe unter 350 Abonnenten einer Zeitung wurde erhoben, ob diese die Zeitung jeden Tag lesen und wie ihre Einstellung zu einer bestimmten Frage ist:

		Einstellung ...		
		positiv	neutral	negativ
Lese ...	täglich	10	20	120
	nicht täglich	100	60	40

- a) Formulieren Sie für die Überprüfung des Zusammenhangs von Leseverhalten und Einstellung unter den gesamten Abonnenten der Zeitung geeignete statistische Hypothesen und testen Sie diese mit den Daten auf einem Signifikanzniveau $\alpha = 0,05$.
 b) Wie groß ist der zur errechneten Teststatistik gehörende p-Wert (Excel)?
 ($H_0: \chi^2=0$, $H_1: \chi^2>0$; $\chi^2_{\text{err}} = 129,13$; Obere Schranke der schwachen Indizien gegen H_0 : 5,99; H_0 wird verworfen, H_1 akzeptiert; p-Wert α_1 : extrem klein)

Ü21)

Überprüfen Sie, ob es zwischen den nachfolgenden, in einer zufälligen Stichprobe erhobenen Merkmalen auf einem Signifikanzniveau von $\alpha = 0,05$ einen signifikanten Zusammenhang in der Grundgesamtheit gibt:

		Wähler einer ...		
		Regierungspartei	Oppositionspartei	
EU-Beitritt	positiv	184	96	280
	negativ	118	102	220
		302	198	500

Wie groß ist außerdem der p-Wert (Excel)?

($\chi_{\text{err}}^2 = 7,64$; Obere Schranke der

schwachen Indizien gegen H_0 : 3,84; H_0 wird verworfen, H_1 akzeptiert; p-Wert α_1 : 0,006)

3.7 Testen von Hypothesen über eine Verteilungsform

HÜ22)

Bei $n = 400$ zufällig ausgewählten Automobilen eines 1986 erstmals zugelassenen Typs wurde für einen elektronischen Bauteil das Merkmal Lebensdauer (in 1.000 km) erhoben:

Lebensdauer	p_i
unter 40	0,140
40 - 60	0,350
60 - 80	0,370
über 80	0,140

Überprüfen Sie auf einem Signifikanzniveau $\alpha = 0,05$, ob die Lebensdauer dieser Bauteile nicht normalverteilt ist. In der Stichprobe ergaben sich ein Mittelwert von 60.000 km und eine Standardabweichung von 20.000 km. Wie groß ist der zum errechneten Chiquadrat gehörende p-Wert (Excel)?

HÜ23)

Verwenden Sie die Daten aus HÜ22 und überprüfen Sie abermals, ob die Lebensdauer dieser Bauteile nicht normalverteilt ist. In der Stichprobe ergaben sich diesmal ein Mittelwert von 80.000 km und eine Standardabweichung von 30.000 km. Wie groß ist der zum errechneten Chiquadrat gehörende p-Wert (Excel)?

HÜ24)

Betrachten Sie folgende Ergebnisse eines χ^2 -Tests auf Nichtzutreffen der Verteilungsannahme Normalverteilung: Anzahl der Intervalle = 6, p-Wert = 0,01.

Wie groß sind die Freiheitsgrade? Welche Entscheidung hinsichtlich der Verteilungsannahme ist zu treffen? War der errechnete χ^2 -Wert größer, gleich oder kleiner als die obere Schranke aus der Tabelle?

Ü22)

Bei einer Untersuchung der Lebensdauer (in Jahren) von 100 zufällig ausgewählten Autobatterien wurde folgende empirische Häufigkeitsverteilung festgestellt:

Lebensdauer	Anzahl der Batterien
bis zu 2	14
2 - 3	38
3 - 4	28
über 4	20

In der Stichprobe ergibt sich ein Mittelwert von 3 Jahren mit einer Standardabweichung von 1. Überprüfen Sie auf einem Signifikanzniveau von $\alpha = 0,05$ die Hypothese, dass die Lebensdauer der Batterien in der betreffenden Grundgesamtheit nicht normalverteilt ist. Formulieren Sie dafür die geeigneten statistischen Hypothesen.

($H_0: \chi^2 = 0$, $H_1: \chi^2 > 0$; Freiheitsgrade = $4 - 3 = 1$; $\chi_{\text{err}}^2 = 2,82$;

Obere Schranke der schwachen Indizien gegen H_0 : 3,84; H_0 wird beibehalten)

Ü23)

Bei einem Test auf Nichtzutreffen der Normalverteilungsannahme wie in Ü28 werden die Messdaten der Stichprobe diesmal in 8 Intervalle zerlegt. Als Testergebnis liegt ein p-Wert von 0,025 vor.

War der errechnete χ^2 -Wert bei diesem Test im Vergleich zur oberen Schranke aus der Tabelle größer, gleich groß oder kleiner und für welche Hypothese ist auf einem Signifikanzniveau $\alpha = 0,05$ zu entscheiden? (größer; H_1)

Ü24)

Bei einem Test auf Nichtzutreffen der Normalverteilungsannahme wie in Ü28/29 werden die Messdaten der Stichprobe diesmal in 5 Intervalle zerlegt. Als Testergebnis errechnet sich: $\chi_{\text{err}}^2 = 4,70$.

Wie groß ist der p-Wert bei diesem Test (Excel) und für welche Hypothese ist auf einem Signifikanzniveau von $\alpha = 0,05$ zu entscheiden? (p-Wert $\alpha_1 = 0,095$; H_0)

3.8 Testen einer Hypothese über einen statistischen Zusammenhang zweier metrischer Merkmale



In den folgenden Beispielen setzen wir stillschweigend voraus, dass für alle Merkmale die Normalverteilungsannahme gültig ist, die diesbezüglichen Chiquadrattests (siehe Abschnitt 3.7) also keine signifikanten Ergebnisse erbrachten.

HÜ25)



In der im Internet bereit stehenden Excel-Datei finden Sie die Daten einer zufälligen Stichprobe aus den in den letzten zehn Jahren den „KS Statistik“ absolviert habenden Studierenden. Überprüfen Sie mit den Daten die Hypothese, dass die beiden erhobenen Merkmale in der Grundgesamtheit aller Absolventen und Absolventinnen gleichsinnig statistisch zusammenhängen. Verwenden Sie zur Berechnung des Korrelationskoeffizienten der Stichprobe die Excel-Funktion KORREL. Stellen Sie die beiden Hypothesen für diese Überprüfung auf und entscheiden Sie auf dem Signifikanzniveau $\alpha = 0,05$.

HÜ26)

Überprüfen Sie die Hypothese, dass die beiden metrischen Merkmale x (= Harnsäure) und y (= Cholesterin) in der Bevölkerung statistisch zusammenhängen. Stellen Sie dazu beide Hypothesen auf und entscheiden Sie sich auf dem Signifikanzniveau $\alpha = 0,05$, wenn für die beiden Zufallsvariablen unter 90 zufällig ausgewählten Personen eine Stichprobenkorrelation von +0,18 vorliegt.

HÜ27)

Zwei metrische Merkmale liefern in einer Stichprobe vom Umfang $n = 2.200$ eine Korrelation von $-0,06$. Formulieren Sie die Hypothesen für eine zweiseitige Überprüfung der Korrelation. Ist das Testergebnis zu einem von Ihnen festzulegendem Signifikanzniveau signifikant? Ist die Korrelation praktisch relevant? Muss der p-Wert des Tests größer, gleich oder kleiner als das von Ihnen gewählte Signifikanzniveau sein?

Ü25)

In einer empirischen Untersuchung zum Zusammenhang zweier metrischer Merkmale ergibt sich in einer Stichprobe vom Umfang 300 eine Korrelation r von $0,112$.

Formulieren Sie für eine Überprüfung, ob ein statistischer Zusammenhang zwischen den beiden Merkmalen in der Grundgesamtheit vorliegt, geeignete Hypothesen und entscheiden Sie sich für eine der beiden auf einem Signifikanzniveau von $\alpha = 0,05$.

($H_0: \rho = 0$ $H_1: \rho \neq 0$; Bereich der schwachen Indizien gegen H_0 : $[-1,96; 1,96]$; $u_{\text{err}} = 1,946$;
 H_0 wird beibehalten)

Ü26)

In einer empirischen Untersuchung des Zusammenhangs zweier metrischer Merkmale ergibt sich in einer Stichprobe vom Umfang 160 eine Korrelation von $0,142$. Formulieren Sie für eine Überprüfung, ob ein gleichsinniger Zusammenhang zwischen den beiden Merkmalen vorliegt, geeignete Hypothesen und entscheiden Sie sich für eine der beiden auf einem Signifikanzniveau von $\alpha = 0,05$.

($H_0: \rho \leq 0$ $H_1: \rho > 0$; Obere Schranke der schwachen Indizien gegen H_0 : $1,65$; $u_{\text{err}} = 1,803$;
 H_0 wird verworfen, H_1 akzeptiert)

Ü27)

Sie erfahren von einem Test der Korrelation in einer Stichprobe lediglich, dass der zweiseitige p-Wert $0,075$ betragen hat. Formulieren Sie die Hypothesen und entscheiden Sie sich für eine der beiden, wenn es sich um einen Test auf

- a) Zusammenhang
- b) gleichsinnigen
- c) gegensinnigen Zusammenhang gehandelt hat.

(a) $H_0: \rho = 0$ $H_1: \rho \neq 0$; H_0 wird beibehalten b) $H_0: \rho \leq 0$, $H_1: \rho > 0$; H_0 wird verworfen, H_1 akzeptiert;
d) $H_0: \rho \geq 0$, $H_1: \rho < 0$; H_0 wird verworfen, H_1 akzeptiert)

3.9 Testen von Hypothesen über Regressionskoeffizienten (Einfache Regressionsanalyse)



In den folgenden Beispielen setzen wir stillschweigend voraus, dass für alle Merkmale die Normalverteilungsannahme gültig ist, die diesbezüglichen Chiquadrattests (siehe Abschnitt 3.7) also keine signifikanten Ergebnisse erbrachten.

HÜ28)



Mit den Daten der Excel-Datei aus HÜ25 sollen auf einem Signifikanzniveau $\alpha = 0,05$ überprüft werden, ob die Regressionsgerade zu diesen beiden Merkmalen in der Grundgesamtheit nicht horizontal liegt.

HÜ29)



Mit den Daten der Excel-Datei aus HÜ25 sollen auf einem Signifikanzniveau $\alpha = 0,05$ überprüft werden, ob die Regressionsgerade zu diesen beiden Merkmalen in der Grundgesamtheit um mehr als 45 Grad ansteigt.

HÜ30)



Mit den Daten der Excel-Datei aus HÜ25 sollen auf einem Signifikanzniveau $\alpha = 0,05$ überprüft werden, ob die Regressionsgerade zu diesen beiden Merkmalen in der Grundgesamtheit einen Achsenabschnitt aufweist, der kleiner als 10 ist.

Ü28)

Überprüfen Sie mit den nachfolgenden Stichprobenergebnissen auf einem Signifikanzniveau $\alpha = 0,05$, ob die Steigung β_1 der Regressionsgeraden $y = \beta_1 \cdot x + \beta_2$ aus der interessierenden Grundgesamtheit nicht mit jener der vorgegebenen Regressionsgeraden $y = 2 \cdot x + 200$ übereinstimmt.

Die Stichprobenergebnisse einer Zufallsauswahl von $n = 252$ Personen lauten:

$\bar{x} = 15$, $\bar{y} = 200$ (beides nach (2) berechnet),

$s_x^2 = 16$, $s_y^2 = 2.500$ (beides nach (3) berechnet),

$s_{xy} = 31,2$ (nach (8) berechnet) und

nach (9): $r = 0,156$. ($H_0: \beta_1 = 2$; $H_1: \beta_1 \neq 2$; Bereich der schwachen Indizien gegen H_0 : [0,469; 3,531];
 $b_1 = 1,95$; $b_1 \in [0,469; 3,531] \rightarrow$ Beibehaltung von H_0 , nicht signifikantes Testergebnis)

Ü29)

Überprüfen Sie mit den nachfolgenden Stichprobenergebnissen auf einem Signifikanzniveau $\alpha = 0,05$, ob der Achsenabschnitt β_2 der Regressionsgeraden $y = \beta_1 \cdot x + \beta_2$ aus der interessierenden Grundgesamtheit nicht mit jenem der vorgegebenen Regressionsgeraden $y = 2 \cdot x + 200$ übereinstimmt.

Die Stichprobenergebnisse einer Zufallsauswahl von $n = 252$ Personen lauten:

$\bar{x} = 15$, $\bar{y} = 200$ (beides nach (2) berechnet),

$s_x^2 = 16$, $s_y^2 = 2.500$ (beides nach (3) berechnet),

$s_{xy} = 31,2$ (nach (8) berechnet) und

nach (9): $r = 0,156$.

($H_0: \beta_2 = 200$; $H_1: \beta_2 \neq 200$; Bereich der schwachen Indizien gegen H_0 : [176,240; 223,760];
 $b_2 = 170,75$; $b_2 \notin [176,240; 223,760] \rightarrow$ Entscheidung für H_1 , signifikantes Testergebnis)

Ü30)

Nehmen Sie an, es wurden die Tests von HÜ28 bis HÜ30 durchgeführt. Sie erfahren als Testergebnis bei jedem dieser Tests lediglich, dass der zweiseitige p-Wert 0,075 beträgt.

Für welche der beiden Hypothesen aus den jeweiligen Beispielen müssten Sie sich demnach beim herkömmlichen Signifikanzniveau entscheiden?

(HÜ28: H_0 , HÜ29: H_1 , HÜ30: H_1)

3.10 Testen von Hypothesen über mehr als zwei Mittelwerte (Einfache Varianzanalyse)



In den nachfolgenden Beispielen setzen wir still schweigend voraus, dass die Voraussetzungen für die Anwendung der Varianzanalyse gegeben sind.

HÜ31)

„*Emotionale Intelligenz* ist ein Sammelbegriff für Persönlichkeitseigenschaften und Fähigkeiten, welche den Umgang mit eigenen und fremden Gefühlen betreffen“ (http://de.wikipedia.org/wiki/Emotionale_Intelligenz). Zur Überprüfung der Behauptung, dass sich 3 Bevölkerungsgruppen A, B und C hinsichtlich der emotionalen Intelligenz nicht gleichen, wurden aus diesen Grundgesamtheiten jeweils uneingeschränkte Zufallsstichproben vom Umfang 3 gezogen. An den gezogenen Erhebungseinheiten wurde dazu der in jeder der 3 Grundgesamtheiten bei gleicher Varianz normalverteilte „Emotional-Intelligenzquotient“ berechnet. Es ergaben sich folgende Daten:

A:	103	104	99
B:	103	99	101
C:	102	104	103

Formulieren Sie die geeigneten statistischen Hypothesen und überprüfen Sie die Behauptung auf einem Signifikanzniveau von $\alpha = 0,05$.

HÜ32)



In der im Internet bereit stehenden Excel-Datei befinden sich die Daten von Messungen des Merkmals Ernteertrag (in kg) einer bestimmten Obstsorte (z.B. Weintrauben) pro Pflanze. Überprüft werden soll, ob sich die Ernteerträge in 4 Grundgesamtheiten (A-D), die sich durch die Bodenbeschaffenheit unterscheiden, nicht gleichen. Dazu wurden jeweils Stichproben von Pflanzen gezogen und der Ernteertrag pro ausgewählter Pflanze gemessen. Man kann davon ausgehen, dass sich das Merkmal in jeder der Grundgesamtheiten bei gleicher Varianz normalverteilt.

Formulieren Sie die für dieses Problem geeigneten statistischen Hypothesen und führen Sie den Test mit einem geeigneten statistischen Verfahren auf dem üblichen Signifikanzniveau durch.

HÜ33)

Der F-Test einer Varianzanalyse erbrachte einen einseitigen p-Wert α_1 von

- a) 0,46
- b) 0,06
- c) 0,01
- d) 0,0001.

Formulieren Sie die statistischen Hypothesen und entscheiden Sie sich auf Basis des p-Wertes auf dem herkömmlichen Signifikanzniveau für eine der beiden.

Ü31)

Es soll überprüft werden, ob ein metrisches Merkmal in 3 Grundgesamtheiten A-C unterschiedlich verteilt ist. In jeder der Grundgesamtheiten ist das Merkmal normalverteilt bei identischen Varianzen. In 3 kleinen Stichproben aus den Grundgesamtheiten wurde das Merkmal erhoben:

A:	106	107	111	108	108
B:	100	96	98		
C:	102	104	103		

Formulieren Sie die geeigneten statistischen Hypothesen und überprüfen Sie die Behauptung auf einem Signifikanzniveau von $\alpha = 0,05$.

($H_0: \mu_A = \mu_B = \mu_C$ H_1 : Mittelwerte $\mu_A = \mu_B = \mu_C$ nicht alle gleich;
 $F_{\text{crit}}=31,82$; Obergrenze der Schranken der schwachen Indizien gegen H_0 : 4,46; Entscheidung für H_1)

Ü32)

Die Wirkung dreier Medikamente auf den Blutdruck soll getestet werden. Es soll überprüft werden, ob die Wirkung der Medikamente verschieden ist. Man kann davon ausgehen, dass die Voraussetzungen für die Durchführung einer Varianzanalyse gegeben sind. An Stichproben aus der Gesamtbevölkerung wurden die Medikamente A-C verabreicht und der Blutdruck gemessen:

A:	125	124	123		
B:	120	114	122	123	121
C:	130	124	118		

Formulieren Sie die geeigneten statistischen Hypothesen und überprüfen Sie die Wirkung auf dem in der Medizin üblichen Signifikanzniveau von $\alpha = 0,01$.

($H_0: \mu_A = \mu_B = \mu_C$ H_1 : Mittelwerte $\mu_A = \mu_B = \mu_C$ nicht alle gleich;
 $F_{\text{crit}}=1,41$; Obergrenze der Schranken der schwachen Indizien gegen H_0 : 8,65; Beibehaltung von H_0)

Ü33)

Der F-Test einer Varianzanalyse erbrachte einen einseitigen p-Wert α_1 von

- a) 0,013
- b) 0,062
- c) 0,051
- d) 0,049.

Formulieren Sie die statistischen Hypothesen und entscheiden Sie sich auf Basis des p-Wertes auf dem Signifikanzniveau $\alpha = 0,05$ für eine der beiden. (H_1 H_0 H_0 H_1)
