Statistik-Klausur vom 10.2.2003

Bearbeitungszeit: 120 Minuten

Aufgabe 1

Ergebnisse der Klausuren „VWL-Teilprüfung I“ und „VWL-Teilprüfung II“
im Sommersemester 2001 am Fachbereich Wirtschaft der FH Köln

Note	Differenzierte Note	Anzahl der Prüfungsteilnehmer		Anteil der Prüfungsteilnehmer (in %)	
		Teilprüf.I	Teilprüf.II	Teilprüf.I	Teilprüf.II
Sehr gut	1,0	3	3	2	2
	1,3	8	0	5	0
Gut	1,7	9	2	6	1
	2,0	17	8	11	6
	2,3	14	11	9	8
Befriedigend	2,7	16	15	10	11
	3,0	14	20	9	14
	3,3	10	10	6	7
Ausreichend	3,7	16	20	10	14
	4,0	17	22	11	16
Nicht bestanden	5,0	31	30	20	21
$\sum$	·	155	141	≈ 100	100

Quelle: Prüfungsamt des Fachbereichs Wirtschaft der FH Köln

a) Welche statistischen Maßzahlen sind dazu geeignet, um das Leistungsniveau in beiden Klausuren miteinander zu vergleichen?
Nennen Sie eine solche Maßzahl.

b) In welcher der beiden Klausuren waren die Leistungsunterschiede, gemessen an den Noten(unterschieden), größer? Kennzeichnen und vergleichen Sie die Leistungsunterschiede in beiden Klausuren durch Berechnung und Interpretation einer geeigneten statistischen Maßzahl. Begründen Sie dabei dei Auswahl der Maßzahl.

c) Sie haben zur Lösung der Teil-Aufgabe a) für beide Klausuren eine statistische Maßzahl vorgeschlagen, die das Leistungsniveau kennzeichnet. Wovon hängt die Aussagekraft einer solchen Maßzahl zur Kennzeichnung des Leistungsniveaus ab? Begründen Sie bitte Ihre Antwort.

Lösung:
X = Noten der Teilprüfung I
Y = Noten der Teilprüfung II

a) Da es sich bei Noten um ordinal skalierte Daten handelt, können zum Vergleich des Niveaus sowohl der Modus als auch der Median berechnet werden.

Note	kum. rel. H.	
	X	Y
1,0	0,02	0,02
1,3	0,07	0,02
1,7	0,13	0,03
2,0	0,24	0,09
2,3	0,33	0,17
2,7	0,43	0,28
3,0	0,52	0,42
3,3	0,58	0,49
3,7	0,68	0,63
4,0	0,79	0,79
5,0	0,99	1,00

Bei nicht-klassierten Daten ist der 50-Prozentpunkt diejenige Merkmalsausprägung, bei der die kumulierte relative Häufigkeit (kum. rel. H.) erstmals über dem Wert 0,50 liegt. Also sind die Mediane folgende Werte:

$$x_{0,50} \approx 3{,}0$$

$$y_{0,50} \approx 3{,}7$$

d.h. Das Noten-Niveau war in der Teilprüfung I besser.

b) Da es sich bei Noten um ordinal skalierte Daten handelt, kann zum Vergleich der Unterschiede der Quartilsabstand berechnet werden.

$$x_{0,25} \approx 2{,}3$$

$$x_{0,75} \approx 4{,}0$$

Also beträgt der Quartilsabstand $x_{0,75} - x_{0,25} = 1{,}7$

$$y_{0,25} \approx 2{,}7$$

$$y_{0,75} \approx 4{,}0$$

Also beträgt der Quartilsabstand $y_{0,75} - y_{0,25} = 1{,}3$

d.h. gemessen am Quartilsabstand waren die Leistungsunterschiede in der Teilprüfung II etwas geringer.

c) Der Quartilsabstand der Werte im Datensatz Teilprüfung II ist geringer. Hier sind also die einzelnen Werte des Datensatzes konzentrierter um den Median.

Aufgabe 2

Sie sind Kostenrechner/in in einem mittelständigen Unternehmen und möchten Aufschluss über die Kostenfunktion Ihres Unternehmens in Abhängigkeit von der produzierten Menge haben.

Die folgenden Daten der letzten vier Quartale liegen Ihnen aus der Kostenrechnung vor:

	produzierte Menge (in Td Stück)	Gesamtkosten (in Td Euro)	
4. Quartal 2001	1	580	
1. Quartal 2002	1,2	690	
2. Quartal 2002	1,1	650	
3. Quartal 2002	0,9	550	

a) Bestimmen Sie anhand der vorliegenden Daten eine lineare Kostenfunktion Ihres Unternehmens. Verwenden Sie dazu die Methode der kleinsten Quadrate.

b) Ihr Geschäftsfüher bezweifelt, dass diese lineare Kostenfunktion die Kostensituation im Unternehmen richtig abbildet. Wie können Sie seine Zweifel, zumindest für die Ihnen vorliegenden Daten, ausräumen?

c) Im nächsten Quartal soll die Produktion auf 1 500 Stück erhöht werden. Welche Gesamtkosten sind Ihrer Meinung nach zu erwarten?

d) Die Produktion wurde tatsächlich auf 1 500 Stück erhöht. Die Gesamtkosten liegen aber viel höher, als Sie prognostiziert haben (Frage c)). Was könnte schief gegangen sein?

Lösung:
X = produzierte Menge (pro Quartal)
Y = Kosten (in Td Euro)

a)

i	unabhängige V. x_i produzierte Menge (in Td Stück)	abhängige V. y_i Gesamtkosten (in Td Euro)	$x_i \cdot y_i$	x_i^2	y_i^2
1	1	580	580	1	336 400
2	1,2	690	828	1,44	476 100
3	1,1	650	715	1,21	422 500
4	0,9	550	495	0,81	302 500
$\sum$	4,2	2 470	2 618	4,46	1 537 500

Regressionsgerade $f(x) = a_1 + b_1 \cdot x$

$$b_1 = \frac{4 \cdot 2\,618 - 4{,}2 \cdot 2\,470}{4 \cdot 4{,}46 - 4{,}2^2} = \frac{98}{0{,}2} = 490$$

$$a = \frac{2\,470 - 490 \cdot 4{,}2}{4} = 103$$

d.h. $f(x) = 103 + 490 \cdot x$ ist die lineare Kostenfunktion anhand der Methode der kleinsten Quadrate.

b) $$b_2 = \frac{98}{4 \cdot 1\,537\,500 - 2\,470^2} = \frac{98}{49\,100} = 0{,}002$$

Bestimmtheitsmaß
$B = b_1 \cdot b_2 = 490 \cdot 0{,}002 = 0{,}9780$
d.h. 97,80 % der Streuung der $y-$Werte wird erklärt durch die Streuung der Regressionsgeraden.

oder Korrelationskoeffizient
$r_{xy} = \sqrt{B} = \sqrt{0{,}9780} = 0{,}9889$
d.h. starker positiver linearer Zusammenhang

c) $f(1{,}5) = 103 + 490 \cdot 1{,}5 = 838$
d.h. es sind 838 000 Euro an Kosten zu erwarten.

d) Der Wert 1,5 liegt nicht im Bereich [0,9 ; 1,2], aus dem die x-Werte stammen. Bei der Berechnung unter c) handelt es sich somit um eine Extrapolation. Extrapolationen können schnell ungenau sein.
Außerdem sind vier Datenpunkte zur Festlegung einer Regressionsgeraden zu wenig.

Aufgabe 3

Die Gesellschaft für deutsche Sprache hat den „Teuro" mit der Begründung zum Wort des Jahres 2002 gekürt, diese „kreative und zugleich prägnante Wortschöpfung" bringe die Gefühle vieler Menschen nach der Euro-Einführung zum Ausdruck.
(Quelle: Kölner Stadt-Anzeiger, 21. Dezember 2002, Titelseite)

a) Erklären Sie - aus rationaler statistischer Sicht - den Unterschied zwischen der „gefühlten Inflationsrate" und der vom Statistischen Bundesamt gemessenen Inflationsrate von 1,3% im Jahr 2002 im Vergleich zum Vorjahr 2001.

b) „Das Statistische Bundesamt stellt zum 1. Januar 2003 den Preisindex für die Lebenshaltung auf das neue Basisjahr 2000 um und passt zugleich den Warenkorb an." (Quelle: Frankfurter Allgemeine Zeitung, 21. Dezember 2002, Seite 12)
Erläutern Sie den Sinn und den statistischen Gehalt dieser Kurzmeldung der FAZ.

Lösung:

a) Zeitgleich zur Einführung des Euros haben sich einige Lebensmittel verteuert. Durch den kalten Januar 2002 in Italien und Spanien wurde Gemüse schlagartig teurer. Ferner gibt es durch den Rückgang des Verzehrs von Rindfleisch auf Grund der BSE-Fälle immer weniger Leder, so dass auch Schuhe teurer wurden. Außerdem sind alkoholische Getränke, Benzin und Rohöl teurer geworden.
Negative Erlebnisse (wie Preiserhöhungen) sind für viele Menschen einprägsamer als positive (wie Preissenkungen). So kommt es zur Fehl-Einschätzung, dass alles teurer geworden sei, obwohl das statistische Bundesamt andere Daten liefert.

b) Alle fünf Jahre werden zum einen die Verbrauchsmengen der Güter im Warenkorb neu festgelegt und zum anderen neue Güter in den Warenkorb aufgenommen, um bei der Berechnung des Preisindex aktuell zu bleiben.

Aufgabe 4
Eine Unternehmung führt eine Kundenbefragung durch. Die Geschäftsführung möchte wissen, welche Gründe bei Gut A zur Kaufentscheidung beitragen (Mehrfachnennungen möglich). Drei von zehn genannten Gründen sind die Funktion, das Design und die Qualität von Gut A. Die Befragung ergibt folgendes Ergebnis:

- 60% der Kunden kaufen Gut A wegen dessen Funktion
- 30% der Kunden kaufen Gut A wegen dessen Designs
- 40% der Kunden kaufen Gut A wegen dessen Qualität
- 15% der Kunden kaufen Gut A wegen dessen Funktion und wegen dessen Designs
- 30% der Kunden kaufen Gut A wegen dessen Funktion und wegen dessen Qualität
- 40% der Kunden, die Gut A wegen dessen Qualität kaufen, treffen ihre Kaufentscheidung auch wegen des Designs von Gut A

a) Berechnen Sie die Wahrscheinlichkeit, dass ein zufällig ausgewählter Kunde Gut A wegen dessen Funktion oder wegen dessen Qualität kauft.

b) Berechnen Sie die Wahrscheinlichkeit, dass ein zufällig ausgewählter Kunde Gut A weder wegen dessen Funktion noch wegen dessen Qualität kauft.

c) Ein Kunde kauft Gut A wegen dessen Designs. Berechnen Sie die Wahrscheinlichkeit, dass er seine Kaufentscheidung auch wegen der Funktion von Gut A trifft.

d) Ein Kunde kauft Gut A wegen dessen Designs. Berechnen Sie die Wahrscheinlichkeit, dass er seine Kaufentscheidung auch wegen der Qualität von Gut A trifft.

Lösung:

$F =$ ein zufällig ausgewählter Kunde kauft Gut A wegen dessen Funktion
$D =$ ein zufällig ausgewählter Kunde kauft Gut A wegen dessen Designs
$Q =$ ein zufällig ausgewählter Kunde kauft Gut A wegen dessen Qualität

Laut Aufgabentext sind folgende Wahrscheinlichkeiten bekannt:

$$\begin{array}{ll} P(F) = 0{,}6 & P(F \cap D) = 0{,}15 \\ P(D) = 0{,}3 & P(F \cap Q) = 0{,}3 \\ P(Q) = 0{,}4 & P(D \mid Q) = 0{,}4 \end{array}$$

Wir tragen die Wahrscheinlichkeiten in Arbeitstabellen für F, Q und für F, D ein:

	F	$\overline{F}$	
Q	0,3		0,4
$\overline{Q}$			0,6
	0,6	0,4	1

	F	$\overline{F}$	
D	0,15		0,3
$\overline{D}$			0,7
	0,6	0,4	1

a) $P(F \cup Q) = P(F) + P(Q) - P(F \cap Q) = 0{,}6 + 0{,}4 - 0{,}3 = 0{,}7$
d.h. die Wahrscheinlichkeit beträgt 0,7.

b) $P(\overline{F} \cap \overline{Q}) = 1 - P(F \cup Q) = 1 - 0{,}7 = 0{,}3$
oder mit Hilfe der Arbeitstabelle:

	F	$\overline{F}$	
Q	0,3	0,1	0,4
$\overline{Q}$	0,3	0,3	0,6
	0,6	0,4	1

$P(\overline{F} \cap \overline{Q}) = 0{,}3$
d.h. die Wahrscheinlichkeit beträgt 0,3.

c) $P(F \mid D) = \dfrac{P(F \cap D)}{P(D)} = \dfrac{0{,}15}{0{,}3} = 0{,}5$
d.h. die Wahrscheinlichkeit beträgt 0,5.

d) $P(Q \mid D) = \dfrac{P(Q \cap D)}{P(D)} = \dfrac{P(D \mid Q) \cdot P(Q)}{P(D)} = \dfrac{0{,}4 \cdot 0{,}4}{0{,}3} = 0{,}53$
d.h. die Wahrscheinlichkeit beträgt 0,53.

Aufgabe 5

Eine Bank beabsichtigt einen neuen Aktienfonds aufzulegen. Die Geschäftsleitung möchte für diesen Fonds in sechs Anlage-Klassen investieren. Die Bank geht für das kommende Geschäftsjahr von folgenden Annahmen aus:

- Die Renditen der einzelnen Anlage-Klassen sind normalverteilt
- Die Renditen der einzelnen Anlage-Klassen sind stochastisch unabhängig

Anlage-Klasse	Verteilung der Rendite	Erwartete Rendite der	Standard-abweichung der Rendite	Wahrschein-lichkeit für eine positive Rendite (Gewinn)	Wahrschein-lichkeit für eine negative Rendite (Verlust)
A	normalverteilt	5%	3,90%	0,9	0,1
B	normalverteilt	5%	3,90%	0,9	0,1
C	normalverteilt	5%	3,90%	0,9	0,1
D	normalverteilt	5%	3,90%	0,9	0,1
E	normalverteilt	3%	3,56%	0,8	0,2
F	normalverteilt	2%	3,81%	0,7	0,3

a) Berechnen Sie die Wahrscheinlichkeit, dass die Rendite im kommenden Geschäftsjahr in Anlage-Klasse A unter 8,9 % liegt.

b) Berechnen Sie die Wahrscheinlichkeit, dass die Rendite im kommenden Geschäftsjahr in Anlage-Klasse F genau 2 % beträgt.

c) Welche Rendite wird in Anlage-Klasse E mit einer Wahrscheinlichkeit von 0,9 höchstens erzielt?

d) Berechnen Sie die Wahrscheinlichkeit, dass von den Anlage-Klassen A, B, C, D im kommenden Geschäftsjahr genau drei einen Gewinn erzielen.

e) Berechnen Sie die Wahrscheinlichkeit, dass in allen sechs Anlage-Klassen A, B, C, D, E, F im kommenden Geschäftsjahr ein Gewinn erzielt wird.

Lösung:
$X_A \sim \mathsf{N}(\mu = 5; \sigma = 3{,}90)$
$X_B \sim \mathsf{N}(\mu = 5; \sigma = 3{,}90)$
$X_C \sim \mathsf{N}(\mu = 5; \sigma = 3{,}90)$
$X_D \sim \mathsf{N}(\mu = 5; \sigma = 3{,}90)$
$X_E \sim \mathsf{N}(\mu = 3; \sigma = 3{,}56)$
$X_F \sim \mathsf{N}(\mu = 2; \sigma = 3{,}81)$

a) $P(X_A < 8{,}9) = F_U\left(\frac{8{,}9 - 5}{3{,}9}\right) = F_U(1) = 0{,}841$
d.h. die Wahrscheinlichkeit beträgt 0,841.

b) $P(X_F = 2) = 0$
d.h. die Wahrscheinlichkeit beträgt null.

c) $0{,}9 = P(X_E < x) = F_U\left(\frac{x - 3}{3{,}56}\right)$

$\Rightarrow 1{,}2816 = \frac{x - 3}{3{,}56} \Rightarrow x = 3 + 1{,}2816 \cdot 3{,}56 = 7{,}5625$
d.h. die mit einer Wahrscheinlichkeit von 0,90 maximal erzielte Rendite im kommenden Geschäftsjahr beträgt 7,56%.

d)

Gewinn	Verlust
A, B, C	D
A, B, D	C
A, C, D	B
B, C, D	A

$P(A \text{ mit Verlust}) = P(X_A < 0) = F_U\left(\frac{0 - 5}{3{,}9}\right) = F_U(-1{,}28) = 0{,}1$
$P(A \text{ mit Gewinn}) = P(X_A > 0) = 1 - P(X_A \leq 0) = 1 - 0{,}1 = 0{,}9$
Wkt. $= 4 \cdot 0{,}9^3 \cdot 0{,}1 = 0{,}2916$
d.h. die Wahrscheinlichkeit beträgt 0,2916.

e) $\underbrace{P(X_A > 0) \cdot P(X_B > 0) \cdot P(X_C > 0) \cdot P(X_D) > 0)}_{0{,}9^4} \cdot \underbrace{P(X_E > 0)}_{1-F_U(\frac{0-3}{3{,}56})=0{,}8} \cdot \underbrace{P(X_F > 0)}_{1-F_U(\frac{0-2}{3{,}81})=0{,}702} \approx$

$0{,}9^4 \cdot 0{,}8 \cdot 0{,}7 = 0{,}3685$
d.h. die Wahrscheinlichkeit beträgt 0,3685.

Aufgabe 6

Die Süddeutsche Zeitung (SZ) vom 14./15. September 2002 gab in einem Artikel mit der Überschrift „Die SPD überholt die Union“ einen Überblick über eine Wahlvorhersage zur Bundestagswahl.

> Massive Veränderungen haben sich knapp eine Woche vor der Bundestagswahl bei der politischen Stimmung ergeben: Die SPD legt deutlich auf 45%(+5%) zu, die CDU/CSU verliert hingegen kräftig und erreicht 35%(−4%). Konstant bleiben hingegen die kleinen Parteien: Grüne und FDP erreichen jeweils 8%, die PDS 3%. Die sonstigen Parteien kommen wie vorige Woche zusammen auf 1%.

Dieser Überblick basierte auf einer telefonischen Befragung von 1362 zufällig ausgewählten wahlberechtigten Deutschen in der zeit vom 9. bis 12. September 2002 durch die Forschungsgruppe Wahlen Mannheim für das ZDF-Politbarometer.

a) Auf welcher Grundgesamtheit basieren Wahlvorhersagen?

b) Ihr Freund, der leider noch keine Statistik-Vorlesung besucht hat, versteht nicht, wie man, basierend auf der Befragung von nur 1362 zufällig ausgewählten Personen, das Wahlverhalten von ca. 61,4 Mio. Wahlberechtigten vorhersagen und damit den Wahlausgang sinnvoll prognostizieren kann.
Was würden Sie ihm antworten? Wie würden Sie ihm den Schluss von den Werten einer Stichprobe auf den wahren, aber unbekannten, Populationsparameter erläutern?

Lösung.

a) Grundgesamtheit=Menge aller wahlberechtigten in der BRD

b) Mit Hilfe von Konfidenzintervallen lässt sich ein Bereich angeben, in dem der unbekannte Populaionsparameter mit der Wahrscheinlichkeit von z.B. 0,95 liegt. So dass hier auf Grund einer Stichprobe auf den wahren Anteilswert geschlossen werden kann.