Statistik-Klausur vom 16.7.2002

Bearbeitungszeit: 120 Minuten

Aufgabe 1
Selten hat sich in der bundesrepublikanischen Öffentlichkeit ein Schlagwort so schnell verbreitet wie das vom „Euro, der ein Teuro“ sei; damit ist gemeint, dass die Einführung des Euro (€) als gemeinsame Währung in der Europäischen Union von den Unternehmern vorzugsweise dazu benutzt worden sei, die Verbraucherpreise drastisch zu erhöhen und die Konsumenten „abzuzocken“. Volkes Meinung äußert sich vorzugsweise in der Vermutung, die Unternehmer hätten die Währungsbezeichnung „DM“ durch „€ “ ersetzt und die numerischen Werte der Preise beibehalten („Brokkoli kostete vor dem Euro 1,99 DM und danach 1,99 € “, so eine Hausfrau beim Einkauf auf einem Wochenmarkt, zitiert in einem Filmbericht der Sendung „heute“ des ZDF).
Umgekehrt sagt das Statistische Bundesamt, das in der BRD zuständig ist für die Berechnung der Inflationsrate, dass sich die Preissteigerungsrate für die privaten Verbraucher im Januar 2002 auf 2,1 % belief (im Vergleich zum entsprechenden Vorjahresmonat 2001) und seitdem auf 1,2 % im Mai 2002 gesunken ist.
Die Empfindungen der Verbraucher und die Berechnungen der Statistiker passen offensichtlich nicht zusammen.

a) Erläutern Sie möglichst umfassend, wie in Deutschland die Preissteigerungsrate des privaten Verbrauchs von den amtlichen Statistikern berechnet wird!
(Anleitung zur Lösung: Arten der statistischen Erhebungen; Berechnungsformel; Periodizität; Aktualität; Repräsentativität; Fehlerquellen - Erwartet wird von Ihnen die verbale Beschreibung der Methoden der Preisstatistik ohne Berechnungen).

b) Versuchen Sie zu erklären, wie es zu dem enormen Unterschied zwischen der Meinung weiter Teile der Bevölkerung und der Journalisten einerseits und den Berechnungen des Statistischen Bundesamtes andererseits über die Höhe der Inflationsrate kommen kann!

Lösung:

a) Art der statistischen Erhebung: es werden die Preise p und die Verbrauchsmengen q von 750 Gütern erfasst
Berechnungsformel: Preisindex von Laspeyres mit den Ausgabeanteilen aus dem Basisjahr $\frac{p_i^0 \cdot q_i^0}{\sum p_j^0 \cdot q_j^0}$ und den Preisverhältnissen $\frac{p_i^1}{p_i^0}$ (Preismessziffern)
Periodizität: monatlich
Aktualität: alle fünf Jahre wird der Warenkorb neu zusammengestellt
Repräsentativität: befragt werden insgesamt 1 000 ausgewählte Haushalte von fünf verschiedenen Haushaltstypen

Fehlerquellen: Der Warenkorb ist schnell veraltet, insbesondere wenn neue Güter wie Handys am Markt auftauchen

b) Verschiedene Preissteigerungen sind nach der Euro-Einführung aufgetreten, und deshalb wird der Euro als Ursache vermutet.
Der Winter im Süden (Italien, Spanien) war im Januar 2002 vergleichsweise sehr kalt, so dass es hier zu enormen Preissteigerungen bei Obst und Gemüse kam. (Salatgurke ≈ 2 €) Zeitgleich wurde der Euro eingeführt, so dass er als Verursacher identifiziert wurde.
Ferner hat die BSE-Krise zu weniger Rindfleischverzehr geführt, so stand weniger Leder zur Verfügung, so dass Schuhe zeitgleich mit der Euro-Einführung teurer wurden.

Aufgabe 2

Am Samstag, dem 18. Mai 2002, wurden im „Kölner Stadt-Anzeiger" vier PKW der Marke VW Golf, Modell IV, von privaten Eigentümern (d.h. keine Händler) angeboten:

Lfd. Nr.	PS-Zahl	Baujahr Erstzulassung	Alter (in Monaten)	Laufleistung (in 1 000 km)	Verlangter Preis (in 1 000 €)
1	75	10/97	55	29	7,000
2	75	11/00	18	11	12,300
3	75	03/01	14	15	13,500
4	75	04/99	37	32	10,200

Statistisch untersucht wurde der lineare Zusammenhang zwischen der Laufleistung einerseits und dem verlangten Preis andererseits. Gemäß der Methode der kleinsten Quadrate ergaben sich folgende Gleichungen:

Verlangter Preis (in 1 000 €) $\approx$ 15,383 - 0,213 $\cdot$ Laufleistung (in 1 000 km)

Laufleistung (in 1 000 km) $\approx$ 51,818 - 2,797 $\cdot$ Verlangter Preis (in 1 000 €)

a) Jemand ist interessiert, im Kölner Raum einen VW Golf IV (75 PS) mit einer Laufleistung von rund 30 000 km von einem privaten Eigentümer zu erwerben. Mit welcher Kaufpreisforderung muss er rechnen?

b) Ein VW Golf IV mit 75 PS wird auf dem privaten Kölner Gebrauchtwagenmarkt für 11 000 € angeboten. Wie hoch dürfte die vermutete Laufleistung sein?

c) Wie gut kann man sich darauf verlassen, dass die Berechnungen unter a) und b) zutreffen?

d) Handelt es sich bei dem Zusammenhang zwischen Laufleistung und dem verlangten Kaufpreis um einen positiven oder um einen negativen Zusammenhang? Begründen Sie bitte Ihre Antwort!

e) Was halten Sie von einer Berechnung der Preisforderung für einen VW Golf IV mit einer Laufleistung von 100 000 km mittels der obigen Gleichung, die nach der Methode der kleinsten Quadrate berechnet wurde?

f) Die Methode der kleinsten Quadrate erlaubt Berechnungen wie unter a) und b). Worin liegt die größte Schwäche der Methode der kleinsten Quadrate?

Lösung:

a) $15{,}383 - 0{,}213 \cdot 30 = 8{,}993 \widehat{=} 9\,000$ €
d.h. es ist mit einem Kaufpreis von etwa 9 000 € zu rechnen.

b) $51{,}818 - 2{,}797 \cdot 11 = 21{,}05 \widehat{=} 21\,000$ km
d.h. es ist mit einer Laufleistung von 21 000 km zu rechnen.

c) Bestimmtheitsmaß $B = (-0{,}213) \cdot (-2{,}797) = 0{,}5958$
Korrelationskoeffizient $r_{xy} = -\sqrt{0{,}5958} = -0{,}77$
d.h. es liegt mittlerer linearer Zusammenhang vor.

d) Beide Steigungen der Regressiongeraden
$f(x) = a_1 - b_1 \cdot x = 15{,}383 - 0{,}213 \cdot x$
$g(y) = a_2 - b_2 \cdot y = 51{,}818 - 2{,}797 \cdot y$
sind negativ. Also liegt ein negativer Zusammenhang vor: je höher die Laufleistung, desto geringer der Verkaufspreis, und umgekehrt, je höher der Verkaufspreis, desto geringer die Laufleistung.

e) Da die obige Regressionsgerade auf Grund von Laufleistungen im Bereich von 11 000 km bis 32 000 km berechnet wurde, kann nicht ohne weiteres davon ausgegangen werden, dass diese Regressionsgerade auch für Laufleistungen gilt, die weit außerhalb dieses Bereichs liegen.

f) Die Methode der kleinsten Quadrate kann nur lineare Zusammenhänge erfassen.

Aufgabe 3

Im Rahmen einer Stellenausschreibung wird zur Vorauswahl der Bewerber ein schriftlicher Test durchgeführt, bei dem zehn Fragen beantwortet werden müssen. Zu jeder Frage stehen zwei Antworten zur Auswahl, von denen eine Antwort richtig und eine falsch ist. Ein Bewerber muss sich für genau eine Antwort entscheiden. Lässt er eine Frage unbeantwortet, so gilt sie als falsch beantwortet.
Alle Bewerber, die mindestens sieben der zehn Fragen richtig beantworten, kommen in die engere Auswahl.

a) Wie viele unterschiedliche Möglichkeiten gibt es, auf zehn Fragen zu antworten?

b) Wie wahrscheinlich ist es, dass ein Bewerber, der überhaupt nichts weiß und nur rät, d.h. zehn Antworten zufällig auswählt und ankreuzt, trotzdem alle zehn Fragen richtig beantwortet?

c) Wie viele Möglichkeiten aus a) führen dazu, dass ein Bewerber in die engere Auswahl kommt?

d) Wie groß ist die Wahrscheinlichkeit dafür, dass ein Bewerber in die engere Auswahl kommt?

e) Zwei Bewerber vereinbaren, bei allen zehn Fragen genau entgegengesetzt zu antworten. Wie groß ist die Wahrscheinlichkeit dafür, dass mindestens einer von den beiden in die engere Auswahl kommt?

Lösung:

a) 10 aus 2 mit Zurücklegen mit Berücksichtigung der Reihenfolge $= 2^{10} = 1\,024$
d.h. es gibt 1 024 Möglichkeiten.

b) X = Anzahl der richtig erratenen Antworten
$X \sim \mathsf{B}(n = 10; p = 0{,}5)$
$P(X = 10) = \binom{10}{10} \cdot 0{,}5^{10} \cdot 0{,}5^{0} = \frac{1}{1024} \approx 0{,}0010$
d.h. die Wahrscheinlichkeit beträgt 0,0010.

c)
$$\begin{array}{lrr}
7 \text{ aus } 10 \text{ ohne Zurücklegen ohne Berücksichtigung der Reihenfolge} = \binom{10}{7} = & & 120 \\
8 \text{ aus } 10 \text{ ohne Zurücklegen ohne Berücksichtigung der Reihenfolge} = \binom{10}{8} = & & 45 \\
9 \text{ aus } 10 \text{ ohne Zurücklegen ohne Berücksichtigung der Reihenfolge} = \binom{10}{9} = & & 10 \\
10 \text{ aus } 10 \text{ ohne Zurücklegen ohne Berücksichtigung der Reihenfolge} = \binom{10}{10} = & & 1 \\
\cline{3-3}
& \sum & 176
\end{array}$$

d.h. es gibt 176 Möglichkeiten.

d)
$$\begin{aligned}
P(X \geq 7) &= P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\
&= \binom{10}{7} \cdot 0{,}5^{7} \cdot 0{,}5^{3} + \binom{10}{8} \cdot 0{,}5^{8} \cdot 0{,}5^{2} + \binom{10}{9} \cdot 0{,}5^{9} \cdot 0{,}5^{1} + \binom{10}{10} \cdot 0{,}5^{10} \cdot 0{,}5^{0} \\
&= \frac{176}{1024} \\
&= 0{,}1719
\end{aligned}$$

d.h. die Wahrscheinlichkeit beträgt 0,1719.

e) Bezeichnet Y die Anzahl der richtig erratenen Antworten des zweiten Bewerbers, so interessieren wir uns für die Wahrscheinlichkeit:

$$P(X \geq 7 \cup Y \geq 7)$$

Da der zweite Bewerber immer entgegengesetzt antwortet, haben wir folgenden Zusammenhang zwischen X und Y:

Anzahl der richtigen Antworten

X	Y
0	10
1	9
2	8
3	7
4	6
5	5
6	4
7	3
8	2
9	1
10	0

Somit gilt:
$P(X \geq 7 \cup Y \geq 7) = P(X \leq 3) + P(X \geq 7)$

Jetzt ist es schneller, das Gegenereignis zu berechnen:
$P(X \leq 3) + P(X \geq 7) = 1 - [P(X = 4) + P(X = 5) + P(X = 6)] = 0{,}3438.$
d.h. die Wahrscheinlichkeit beträgt 0,3438.

Aufgabe 4

Für eine Verkehrsüberwachung sind insgesamt zehn Rechner eingesetzt. Jeder dieser Rechner hat pro Tag eine Ausfallwahrscheinlichkeit von 3%. Insgesamt müssen mindestens acht Rechner laufen, damit die Anlage arbeiten kann. Die Anzahl der Rechner, die pro Tag ausfallen, kann als binomialverteilt angesehen werden.

a) Wie groß ist die Wahrscheinlichkeit dafür, dass an einem Tag kein Rechner ausfällt?

b) Wie groß ist die Wahrscheinlichkeit dafür, dass an einem Tag genau zwei Rechner ausfallen?

c) Wie groß ist die Wahrscheinlichkeit dafür, dass an einem Tag die Anlage ausfällt?

d) An wie vielen Tagen im Jahr muss damit gerechnet werden, dass die Anlage ausfällt?

Lösung:
X = Anzahl der ausgefallenen Rechner
$X \sim \mathsf{B}(n = 10; p = 0{,}03)$

a) $P(X = 0) = 0{,}7374$
d.h. die Wahrscheinlichkeit beträgt 0,7374.

b) $P(X = 2) = 0{,}0317$
d.h. die Wahrscheinlichkeit beträgt 0,0317.

c)
$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - 0{,}7374 - 0{,}2281 - 0{,}0317 \\ &= 0{,}0028 \end{aligned}$$
d.h. die Wahrscheinlichkeit beträgt 0,0028.

d) $360 \cdot 0{,}0028 = 1{,}008$
d.h. im Mittel fällt die Anlage etwa einmal im Jahr aus.

Aufgabe 5

Die Bundestagswahl im September 2002 wirft ihre Schatten voraus. Im Auftrag des Fernsehsenders „n-tv“ führt das Meinungsforschungsinstitut „Emnid“ aus Bielefeld in regelmäßigen Abständen repräsentative Befragungen von rund 1 000 zufällig ausgewählten Wahlberechtigten in Deutschland durch und lässt sich dabei immer auch die berühmte „Sonntagsfrage“ beantworten: „Welcher Partei würden Sie Ihre Stimme geben, wenn am nächsten Sonntag Bundestagswahlen wären?“
Normalerweise liegen dabei die großen Volksparteien SPD und CDU/CSU ungefähr in einem Bereich von 30 bis 40 % derjenigen Wahlberechtigten, die sich bei der Umfrage für eine bestimmte Partei entscheiden.

a) Wie groß müsste eine Zufallsstichprobe (ohne Berücksichtigung der Antwortverweigerer, die auf die Sonntagsfrage nicht antworten bzw. angeben, nicht zur Wahl zu gehen bzw. sich noch nicht entschieden zu haben) sein, damit man mit einer Sicherheit von rund 95% davon ausgehen könnte, dass eine Partei bei allen Wahlberechtigten einen Stimmenanteil zwischen 34 und 36 % hat, wenn sich in der Zufallsstichprobe genau 35 % für diese Partei aussprechen?

b) Normalerweise sind rund 25 % der zufällig für dei Zufallsstichprobe ausgewählten Wahlberechtigten Antwortverweigerer [im Sinne der Darstellung unter a)], wobei es in der Natur der Sache liegt, dass dieser Anteil keine Konstante ist; gehen Sie indessen bei der Beantwortung von Frage b) der Einfachheit halber von einem Erfahrungssatz von 25 % Antwortverweigerern aus.
Wie müsste man die Größe der Zufallsstichprobe, die Sie unter a) berechnet haben, verändern, um bei sonst gleichen Voraussetzungen (Sicherheit: 95 %; Fehlerspielraum: 34 bis 36 %) die Antwortverweigerer zu berücksichtigen?

c) Nehmen Sie an, dass sich bei der Zufallsstichprobe [auf Basis der Teilaufgabe a); das heißt: verwenden Sie den dort berechneten Stichprobenumfang] 18% derjenigen, die eine Partei ihrer Wahl angeben, für die FDP votieren. Wie groß ist dann - wiederum bei einer Sicherheit von rund 95% - der Fehlerspielraum für den Stimmenanteil der FDP? Begründen Sie bitte Ihre Antwort durch Berechnung und anschließende Interpretation des Ergebnisses!

d) Wie beurteilen Sie entsprechende Berechnungen für eine Partei wie die PDS, deren Stimmenanteil normalerweise in einem Bereich von etwa 5 % dümpelt? Gibt es bei einer so kleinen Partei zusätzliche Probleme für die Umfrageforscher? Begründen Sie bitte Ihre Antwort aus der Sicht der statistischen Methodenlehre (und nicht aus der besonderen politischen Lage der PDS als Nachfolgepartei der SED)!

Lösung:

a) Auf Grund einer Stichprobe beträgt das 0,95-Konfidenzintervall für den Anteilswert laut Aufgabentext $[0{,}34; 0{,}36]$. Also beträgt die Breite des Konfidenzintervalls 0,02 und die halbe Breite $e = 0{,}01$. Ferner beträgt der Anteilswert in der Stichprobe laut Aufgabentext 0,35. Gesucht ist der Mindeststichprobenumfang n des 0,95-Konfidenzintervalls für den Anteilswert:
$$n \geq \frac{(1{,}96)^2 \cdot 0{,}35 \cdot 0{,}65}{(0{,}01)^2} = 8739{,}64$$
d.h. der Mindeststichprobenumfang beträgt etwa 8740.
Das in der Aufgabe gegebene Konfidenzintervall $[0{,}34; 0{,}36]$ ist ein Konfidenzintervall für den Anteilswert. Also müssen wir einerseits prüfen, ob wir die Binomialverteilung durch die Normalverteilung annähern dürfen. Die Faustregel $n \geq 100$ ist erfüllt. Und andererseits müssen wir überprüfen, ob überhaupt eine Binomialverteilung vorliegt, weil es sich bei einer Umfrage um eine Stichprobe ohne Zurücklegen handelt. Die Faustregel dazu $\frac{n}{N} = \frac{8740}{40\,000\,000} \leq 0{,}05$ stimmt. (Die Anzahl aller Wahlberechtigten in Deutschland beträgt etwa 40 Millionen.)

b) Dreisatz:

$$75\% \;\widehat{=}\; 8\,740 \text{ Befragte}$$

$$100\% \;\widehat{=}\; \frac{8\,740 \text{ Befragte}}{75\%} \cdot 100\% = 11\,653{,}3 \text{ Befragte}$$

d.h. der Stichprobenumfang müsste etwa 11 650 betragen.

c) Gesucht ist ein 0,95-Konfidenzintervall für den Stimmenanteil der FDP-Wähler bei der Wahl. Der Stimmenanteil der FDP-Wähler in der Stichprobe beträgt 0,18. Somit ergibt sich folgendes Konfidenzintervall:

$$\left[0{,}18 - 1{,}96 \cdot \sqrt{\frac{0{,}18 \cdot 0{,}82}{8740}}; 0{,}18 + 1{,}96 \cdot \sqrt{\frac{0{,}18 \cdot 0{,}82}{8740}}\right] = [0{,}1719; 0{,}1881]$$

d.h. das Konfidenzintervall beträgt etwa $[17{,}2\%; 18{,}8\%]$.

Interpretation: Das berechnete Konfidenzintervall ist ein geschätzter Bereich für das Intervall, in dem der FDP-Stimmenanteil in der Bevölkerung mit der Wahrscheinlichkeit 0,95 liegt.

e) Wenn ein Konfidenzintervall für den Stimmenanteil der PDS berechnet werden soll, so muss die Binomialverteilung durch die Normalverteilung angenähert werden. Die Normalverteilung ist eine symmetrische Verteilung. Die Binomialverteilung ist für kleine Anteile (so wie hier der PDS-Anteil 0,05) und für große Anteile eine schiefe Verteilung. Deshalb kann das Konfidenzintervall hier ungenau sein. Die Faustregel sagt aber aus, dass selbst bei schiefen Binomialverteilungen das Konfidenzintervall genau ist, sobald der Stichprobenumfang mindestens 100 beträgt.