

Klausur Statistik WS 03/04 Zweittermin

- Beginnen Sie jede Aufgabe auf einem neuen Blatt.
- Schreiben Sie zu jeder Lösung den Verweis in eckigen Klammern [X1], so daß die Zuordnung möglich ist. Fehlt die Zuordnung, können keine Punkte vergeben werden.
- Machen Sie zu jeder Lösung den Lösungsweg deutlich, ansonsten können keine Punkte vergeben werden.

Aufgabe 1: Univariate und bivariate Datensätze, Korrelation und Regression

In einer Marketing-Studie wurden drei Konsumententypen identifiziert und Daten zu deren Konsumverhalten erhoben.

Das Merkmal "Konsumententyp" (X) hat die drei Ausprägungen:

1=jung und dynamisch

2=Beständiger

3=sonstige

Das Merkmal "Häufigkeit des Konsums" (Y) hat die Ausprägungen

1= selten oder nie

2=häufig

Die gemeinsamen absoluten Häufigkeiten wurden in die folgende Kontingenztafel geschrieben

		Häufigkeit des Konsums (Y)		
		1	2	Σ
Konsumententyp (X)	1	5	25	30
	2	40	20	60
	3	5	5	10
	Σ	50	50	100

Tabelle 1: Die gemeinsame absolute Häufigkeitsverteilung der zwei Merkmale

[B1] Berechnen Sie die relativen Häufigkeiten des Merkmals "Konsumententyp", d.h. $h(X = x_i)$, und stellen Sie diese in einer geeigneten Grafik dar (Achsenbeschriftungen nicht vergessen!). Wie bezeichnet man die von Ihnen verwendete grafische Repräsentation? (5 Punkte)

[B2] Sind die beiden Merkmale "Konsumhäufigkeit" und "Konsumententyp" unabhängig? Argumentieren Sie warum (oder warum nicht). Welchen Wert für den Kontingenzkoeffizienten würden Sie erwarten (Vorzeichen, Größenordnung)? (5 Punkte)

[B3] Berechnen Sie die gemeinsame relative Häufigkeit der beiden Merkmale unter der Annahme der Unabhängigkeit der beiden Merkmale. (6 Punkte)

[B4] Berechnen Sie die bedingten relativen Häufigkeiten $h(Y = y_j | X = x_i)$, also die Häufigkeitsverteilung des Merkmals Konsumhäufigkeit gegeben den Konsu-

mententyp. Benutzen Sie die Tabellenform zur Präsentation Ihrer Resultate. (5 Punkte)

[B5] Im Rahmen derselben Studie wurden auch metrische Konsumentendaten erhoben, insbesondere Merkmal W , die monatlichen Ausgaben (in Euro) für das Produkt des Unternehmens, und Merkmal Z , das Alter (in Jahren) des Konsumenten. Die empirische Varianz der Variable "Alter" beträgt 25, die empirische Varianz der Konsumausgaben 2500. Die empirische Kovarianz der beiden Merkmale beträgt -150. Berechnen Sie den empirischen Korrelationskoeffizienten (Bravais/Pearson) und interpretieren Sie Vorzeichen und Größe des berechneten Wertes (5 Punkte)

[B6] Mit den erhobenen Daten zu Konsum und Konsumentenalter wurde auch eine Regressionsanalyse durchgeführt, mit den Konsumausgaben als abhängiger Variable und dem Konsumentenalter als erklärender Variable. Die Regressionsgleichung lautet:

$$W_v = b_0 + b_1 Z_v + e_v \quad (1)$$

wobei e_v eine unerklärte Restkomponente bezeichnet. Die unbekannt Parameter b_0 und b_1 sollen mit Hilfe der Kleinstquadratmethode geschätzt werden. Stellen Sie die Kleinstquadrat-Zielfunktion für das konkrete Beispiel auf (5 Punkte) und berechnen Sie unter Verwendung der Angaben unter [B5] den Kleinstquadrat-Schätzer für den Steigungsparameter, \hat{b}_1 (5 Punkte). Interpretieren Sie den von Ihnen berechneten Wert (3 Punkte). Das Bestimmtheitsmaß der Regression beträgt 0.36. Interpretieren Sie diesen Wert. (3 Punkte).

[B7] Als Praktikant/Praktikantin erhalten Sie den Auftrag aus einer Kundendatenbank den Durchschnittsumsatz pro Kunden im Geschäftsjahr 2003 zu berechnen. Die Controlling-Abteilung liefert die Information, daß die Kundendatenbank 10.000 Kunden umfaßt. Der Gesamtumsatz des Jahres 2003 dieser 10.000 Kunden beträgt eine Milliarde Euro. Nach der Berechnung ruft Sie ein Controller an und sagt Ihnen, er hätte vergessen, Ihnen die Informationen aus der Großkundendatenbank zu liefern. Für die zehn Großkunden in der separat geführten Datenbank hat der Controller den Durchschnittsumsatz pro Kunde mit 100 Mio Euro bereits berechnet. Berechnen Sie den Durchschnittsumsatz pro Kunden unter Verwendung aller Kundendaten (d.h. aller 10.010 Kunden). (5 Punkte).

Hätten Sie den Median-Umsatz anstelle des arithmetischen Mittels berechnet, wäre dieser Zentralwert größer oder kleiner als das arithmetische Mittel? Begründen Sie Ihre Antwort! (3 Punkte).

Aufgabe 2: Zeitreihenanalyse

Sie arbeiten als Praktikant/Praktikantin in der Controlling-Abteilung eines Unternehmens. Ihre Aufgabe ist es, mit Hilfe einer exponentiellen Glättung eine Prognose des Umsatzwachstums für die Produktgruppen des Unternehmens durchzuführen. Nach intensiver Forschungstätigkeit hat die Controlling-Abteilung die folgende Gleichung ermittelt:

$$y'_t = 0,3y'_{t-1} + 0,7y_t \quad (2)$$

wobei y_t das Umsatzwachstum von Periode $t - 1$ auf Periode t bezeichnet. y'_t

bezeichnet die Prognose zum Zeitpunkt t des Umsatzwachstums für die Periode $t + 1$.

[Z1] Das Umsatzwachstum der aktuellen Periode, y_t , beträgt 0,03. Der in der vergangenen Periode ermittelte Prognosewert, y'_{t-1} , beträgt 0,01. Berechnen Sie die Prognose für das Umsatzwachstum in der nächsten Periode. (3 Punkte)
Sie werden außerdem aufgefordert, mit Hilfe der obigen Gleichung das Umsatzwachstum in der übernächsten Periode, y'_{t+1} , zu prognostizieren. Wie gehen Sie hierbei vor? Berechnen Sie den Prognosewert und begründen Sie Ihr Vorgehen. (4 Punkte)

[Z2] Mit welchem Gewicht beeinflusst gemäß Gleichung (2) das zwei Perioden in der Vergangenheit zurückliegende Umsatzwachstum, y_{t-2} , die aktuelle Prognose y'_t ? (5 Punkte)

[Z3] Auf Nachfrage erklärt Ihnen ein Mitarbeiter, daß der Glättungsparameter, der für die exponentielle Glättung verwendet wurde, mit "praktischer Erfahrung" begründet wird. Beschreiben Sie ein nachvollziehbareres Verfahren zur Ermittlung des Glättungsparameters. (5 Punkte)

[Z4] Eines der Produkte des Unternehmens weist die in Abbildung 1 dargestellte Entwicklung des Umsatzwachstums auf: Ist die exponentielle Glättung mit dem

Abbildung 1: Umsatzwachstum eines Produkts

Glättungsparameter aus Gleichung (2) hier ein geeignetes Prognoseverfahren? Argumentieren Sie, warum, bzw. warum nicht! (5 Punkte). Welches alternative Verfahren würden Sie vorschlagen? (3 Punkte).

[Z5] Ein in der Controlling-Abteilung außerdem verwendetes Prognoseverfahren basiert auf Trendfunktionen. Unter anderem wird die folgende Trendfunktion verwendet:

$$y_t = a \cdot b^t$$

wobei t einen Zeitindex $t = 1, 2, ..$ bezeichnet. Wie würden Sie Werte für die unbekannt Parameter a und b berechnen/schätzen? Beschreiben Sie detailliert Ihr Vorgehen und die benötigten Daten. (5 Punkte)

Aufgabe 3: Frequentistische und Bayesianische Statistik

Ihre nächste Station in Ihrem Praktikum ist die Marketingabteilung. Deren Leiter ist ein Anhänger der Bayesianischen Statistik. Er erklärt Ihnen, daß er seine Einschätzungen für die Erfolgchancen eines neu einzuführenden Produkts mit einem Bayesianischen Modell bestimmt. Er weiß (oder glaubt?) daß Produkte, die später im Gesamtmarkt erfolgreich sind (gemessen an einem zu erreichenden Marktanteil), mit einer Wahrscheinlichkeit von 0,7 auch in einem Testmarkt erfolgreich sind. Produkte, die nicht im Gesamtmarkt erfolgreich sind, sind nur mit einer Wahrscheinlichkeit von 0,1 im Testmarkt ein Erfolg.

[W1] Einem der aktuellen Produkte (Produkt A), das in einem Testmarkt ein Erfolg war, weist der Marketingleiter eine (a-posteriori) Wahrscheinlichkeit für einen Produkterfolg von 0,9 zu. Einem anderen aktuellen Produkt (Produkt B), das in einem Testmarkt kein Erfolg war, weist er eine (a-posteriori) Wahrscheinlichkeit dafür, daß es ein Erfolg im Gesamtmarkt wird, von 0,7 zu. Berechnen Sie die a-priori Wahrscheinlichkeit für einen Produkterfolg im Gesamtmarkt, die der Marketingleiter Produkt A zuwies, bevor dieses im Testmarkt erprobt wurden. (10 Punkte)

[W2] Der Mißerfolg von Produkt B im Testmarkt hat den Marketingleiter so sehr verwundert, daß er es noch einmal in einem anderen Testmarkt erproben läßt. Sie führen die Studie durch und wieder wird das Produkt im Testmarkt kein Erfolg. Berechnen Sie die neue a-posteriori Wahrscheinlichkeit für den Gesamtmarkterfolg von Produkt B. (8 Punkte).

In Ihrem nächsten Praktikum ist die Leiterin Produktstrategie eher eine Anhängerin der frequentistischen Statistik. In einer Studie wurde eine Kundensegmentation vorgenommen, wonach es im Gesamtmarkt 40 % junge dynamische (Typ 1), 50 % Beständige (Typ 2) und 10 % sonstige (Typ 3) gibt. Aus Experimenten weiß die Leiterin Produktstrategie, daß die Wahrscheinlichkeit dafür, daß ein Konsument vom Typ 1 das Produkt kauft, gleich 0,6 ist. Die Wahrscheinlichkeit für das Ereignis, daß ein Konsument vom Typ 2 das Produkt kauft beträgt 0,3. Die Wahrscheinlichkeit, daß ein Konsument vom Typ 3 das Produkt kauft beträgt 0,1.

[W3] Aus der Grundgesamtheit werden drei Individuen zufällig ausgewählt. Die Auswahl erfolgt mit Zurücklegen, d.h. dieselbe Person kann mehrfach ausgewählt werden. Berechnen Sie die Wahrscheinlichkeit für das Ereignis, drei Konsumenten vom Typ 1 auszuwählen. (3 Punkte) und die Wahrscheinlichkeit dafür, daß keiner der drei ausgewählten Individuen vom Typ 3 ist. (3 Punkte) Wie groß ist die Wahrscheinlichkeit für das Ereignis, ein Individuum auszuwählen, das entweder vom Typ 1 oder Typ 3 ist? (3 Punkte)

[W4] Aus der wie oben zusammengesetzten Grundgesamtheit potentieller Konsumenten wird ein Individuum zufällig ausgewählt. Berechnen Sie die Wahrscheinlichkeit dafür, daß die ausgewählte Person das Produkt kauft. (6 Punkte).