Aufgaben

1. Gegeben seien folgende Daten einer statistischen Erhebung, bereits nach Größe sortiert (Rangliste):

10	13	14	14	15	16	17	17	18	18
18	19	19	20	20	20	20	20	21	21
21	21	21	21	21	21	22	22	22	22
22	22	23	23	23	23	24	24	25	25
25	25	25	26	26	26	27	27	28	30

Erstellen Sie eine Tabelle, in der die Merkmalsausprägungen übersichtlicher dargestellt werden, und die auch die Summe der Merkmalsausprägungen (Merkmalssumme) enthält.

2. Stellen Sie die in 1. entwickelte Tabelle grafisch dar.

3. Ihrem Auftraggeber ist sowohl die Tabelle als auch die Zeichnung zu groß und zu unübersichtlich. Fassen Sie deswegen die Merkmalsausprägungen in Klassen zusammen und erstellen Sie die Tabelle neu. Die Klassenbreite soll 2 betragen, die erste Klasse geht von der Merkmalsausprägung 9 bis unter 11.

4. Wieso stimmt die Merkmalssumme der Tabelle aus 3. nicht mit der Merkmalssumme der ursprünglichen Tabelle überein?

5. Wie lässt sich die Tabelle aus 3. grafisch darstellen?

6. Die grafische Darstellung der Tabelle aus 3. nennt man ein Histogramm. Lassen Sie dieses von Mathcad oder Excel zeichnen. Ein spezielles Statistikprogramm ist dafür nicht notwendig, jedoch folgender Hinweis:

   In Mathcad muss das *Data Analysis Extension Pack* installiert sein. Hier gibt es die Funktion *Histogramm(n,M)*, wobei n ein Vektor der Klassengrenzen ist und M eine Matrix der Merkmalsausprägungen. In Excel gehört die Histogrammfunktion nicht zur Standardinstallation, sondern zu den Analyse-Funktionen, die nachträglich zu installieren sind. In Excel 2003 wählt man den Befehl *Extras* > *Add-Ins* und aktiviert das Kontrollkästchen *Analyse-Funktionen*. In Excel 2007: *Office-Schaltfläche* > *Excel-Optionen* > *Add-Ins* > Im Listenfeld *Verwalten* den Eintrag *Excel-Add-Ins* wählen > Schaltfläche *Gehe zu…* > Im Fenster *Add-Ins* das Kontrollkästchen *Analyse-Funktionen* aktivieren > *OK*. In Excel 2010: *Datei* > *Optionen* > *Add-Ins* > Im Listenfeld *Verwalten* den Eintrag *Excel-Add-Ins* wählen > Schaltfläche *Gehe zu…* > Im Fenster *Add-Ins* das Kontrollkästchen *Analyse-Funktionen* aktivieren > *OK*.

7. Es gebe n Ausprägungen des Merkmals x. Die Merkmalsausprägungen werden mit i = 1…n indiziert. Wie hoch ist die Merkmalssumme?

8. Es gebe m unterschiedliche Ausprägungen des Merkmals x, die mit der Häufigkeit h vorkommen. Die unterschiedlichen Merkmalsausprägungen und die Häufigkeiten werden mit j = 1…m indiziert. Wie hoch ist die Merkmalssumme?

9. Die relative Häufigkeit ist die absolute Häufigkeit, geteilt durch die Anzahl *aller* Merkmalsausprägungen. Wie groß ist die Summe aller relativen Häufigkeiten?

10. Wie lässt sich der Ausdruck $\sum_{i=1}^{n} \overline{X}$ kürzer darstellen?

11. Gegeben ist folgende Häufigkeitsverteilung:

Merkmalsausprägung	Häufigkeit
1	2
2	7
3	15
4	10
5	6

Ermitteln Sie die relative Häufigkeit und die kumulierte relative Häufigkeit der Merkmalsausprägungen. Stellen Sie die relative Häufigkeit in Abhängigkeit von den Merkmalsausprägungen (Häufigkeitsfunktion) sowie die kumulierte relative Häufigkeit in Abhängigkeit von den Merkmalsausprägungen grafisch dar (Verteilungsfunktion).

12. Gegeben sind folgende Daten:

Merkmalsausprägung	Anzahl der Merkmalsträger
5	1
10	1
20	1
20	1
45	1

Ermitteln Sie den kumulierten Anteil der Merkmalsträger an der Gesamtzahl der Merkmalsträger und den kumulierten Anteil der Merkmalsausprägungen an der Merkmalssumme, jeweils beginnend bei null, und stellen sie dies in einem Koordinatensystem dar. Die kumulierten Anteile der Merkmalsträger werden dabei auf der Abszisse abgetragen und die kumulierten Anteile der Merkmalsausprägungen auf der Ordinate (Lorenzkurve).

Zeichnen Sie die Lorenzkurve (oder lassen Sie sie von Excel zeichnen) auch für folgende Daten:

Merkmalsausprägung	Anzahl der Merkmalsträger
20	1
20	1
20	1
20	1
20	1

Die Lorenzkurve bildet nun die sogenannte Gleichverteilungsgerade.

Verwenden Sie auch folgende Daten:

Merkmalsausprägung	Anzahl der Merkmalsträger
0	1
0	1
0	1
0	1
100	1

Die Lorenzkurve zeigt nun eine extreme Ungleichverteilung.

Wer sich daran stört, dass es „Merkmalsträger“ mit einer Merkmalsausprägung von null geben soll, lasse diese Kurve zeichnen:

Merkmalsausprägung	Anzahl der Merkmalsträger
1	1
1	1
1	1
1	1
96	1

Aufgaben

13. Die Beitragseinnahmen der 10 größten Versicherungsunternehmen Deutschlands betrugen im Jahre 2009[1]:

Beitragseinnahmen [Mio. Euro] 2009	Unternehmen
5.274	Signal Iduna Gruppe
6.144	Zurich Gruppe Deutschland
6.355	Versicherungskammer Bayern
8.142	Debeka Versicherungen
10.285	Axa Konzern AG
10.521	R+V Konzern
14.850	Generali Deutschland Holding
20.923	Talanx AG
41.423	Münchener-Rück-Gruppe
97.385	Allianz Group

Ermitteln Sie das 10%-Quantil, das 20%-Quantil, das untere Quartil, das 30%-Quantil, das 40%-Quantil, den Median, das 60%-Quantil, das 70%-Quantil, das obere Quartil, das 80%-Quantil und das 90%-Quantil.

14. Entwerfen Sie eine Grafik, in der die Zusammenhänge aus Aufgabe 13 sinnfällig dargestellt werden.

15. Ein Unternehmen hat folgende Daten ermittelt:

Monat	Geleistete Arbeitsstunden	Lohnkosten pro Arbeitsstunde
Januar	862.400	20,06
Februar	752.000	23,01
März	756.000	24,47
April	768.000	22,53
Mai	638.400	27,10
Juni	640.000	28,91
Juli	644.000	36,00
August	528.000	32,77
September	588.000	31,80
Oktober	792.000	21,84
November	784.000	22,07
Dezember	676.800	36,33

Wie hoch sind die Lohnkosten pro Arbeitsstunde im Jahresdurchschnitt?

16. Beweisen Sie: Wenn das gewogene arithmetische Mittel mithilfe von relativen Häufigkeiten ermittelt wird, muss nicht durch die Summe der Gewichte geteilt werden.

17. Beweisen Sie: Die Summe der Abweichungen vom arithmetischen Mittel ist gleich null.

18. Es sei:

n = Anzahl der Merkmalsausprägungen
i = 1…n Index der Merkmalsausprägungen
x_i = Merkmalsausprägung i

Das arithmetische Mittel ist $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$.

[1] Quelle: Institut der deutschen Wirtschaft Köln (Hrsg.), Deutschland in Zahlen - Ausgabe 2011 -, Köln 2011, S.52

Aufgaben

Beweisen Sie: Die Funktion $\sum_{i=1}^{n}\left(x_i - x\right)^2$ hat ein Minimum, wo $x = \overline{x}$.

19. Wie hoch ist für die Daten aus Aufgabe 13 die durchschnittliche Merkmalsausprägung?

20. Wie hoch ist für die Daten aus Aufgabe 13 die mittlere absolute Abweichung?

21. Wie hoch ist für die Daten aus Aufgabe 13 die Varianz und die Standardabweichung?

22. Gegeben sind folgende Daten:

$x = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \end{pmatrix}$ Ausprägungen des Merkmals x

$h = \begin{pmatrix} 2 \\ 7 \\ 15 \\ 10 \\ 6 \end{pmatrix}$ Häufigkeit der Merkmalsausprägungen

Wie hoch ist die Merkmalssumme?

23. Es gelte folgende Ungleichung:

$$\left(\sum_{i=1}^{n} a_i\right)^2 \le n \cdot \sum_{i=1}^{n} a_i^2$$

wobei

n = Anzahl der Merkmalsausprägungen
i = 1...n Index der Merkmalsausprägungen
a_i = Absolute Abweichung der Merkmalsausprägung i vom arithmetischen Mittel

Beweisen Sie folgende Behauptung:

Wenn alle Abweichungen einander gleich sind, wenn also gilt $a_i = a$, dann wird aus der Ungleichung eine Gleichung.

24. Beweisen Sie:

$$\sum_{i=1}^{n}\left(x_i - c\right)^2 = \sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 + n \cdot \left(\overline{x} - c\right)^2$$ Steinerscher Verschiebungssatz

25. Beweisen Sie:

Aus dem Steinerschen Verschiebungssatz folgt für c = 0:

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2$$

26. In einem bestimmten Zeitraum war der Durchschnittskurs von Aktie A 35,00 mit einer Standardabweichung von 5,00. Bei Aktie B war der Durchschnittskurs im gleichen Zeitraum 9,50 mit einer Standardabweichung von 2,60. Bei welcher Aktie waren die Kursschwankungen größer? Welche Aktie war – ohne Berücksichtigung der Dividenden – volatiler?[2]

[2] Quelle für das Beispiel: M. Piazolo, Statistik für Wirtschaftswissenschaftler – Daten sinnvoll aufbereiten, analysieren und interpretieren, Karlsruhe 2007, S. 95

Aufgaben

27. Beweisen Sie:

    Ein Rechteck mit der Breite B und der Höhe H, für welches gilt B + H = C = const, hat dann die größte Fläche, wenn B = H.

28. Stimmen die folgenden Gleichungen?

$$\frac{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\cdot\left(y_i-\overline{y}\right)}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\cdot\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}$$

$$\frac{\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\cdot\left(y_i-\overline{y}\right)}{\sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\cdot\sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}=\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)\cdot\left(y_i-\overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2}\cdot\sqrt{\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2}}$$

29. Wie hoch ist der Korrelationskoeffizient, wenn für jede Merkmalsausprägung gilt, dass $x_i-\overline{x}=y_i-\overline{y}$?

30. Wie hoch ist der Korrelationskoeffizient, wenn für jede Merkmalsausprägung gilt, dass $x_i-\overline{x}=-\left(y_i-\overline{y}\right)$?

31. Wie hoch ist der Korrelationskoeffizient, wenn für jede Merkmalsausprägung gilt, dass $x_i-\overline{x}=a\cdot\left(y_i-\overline{y}\right)$? Hierbei ist a eine Konstante.

32. Gegeben ist die Funktion $SQA\left(a,b\right)=\sum_{i=1}^{n}\left(y_i-a-b\cdot x_i\right)^2$. Wie lauten die ersten und zweiten Ableitungen der Funktion $\left(y_i-a-b\cdot x_i\right)^2$ nach a und b?

33. Gegeben ist folgende Aufstellung des Alters und des Einkommens von 5 Arbeitnehmern[3]:

Lfd. Nr.	Alter	Einkommen
1	21	1.850,00
2	28	1.800,00
3	35	2.230,00
4	46	2.500,00
5	55	2.560,00

    Vermutlich besteht ein Zusammenhang zwischen dem Alter und dem Einkommen. Wie stark ist die Korrelation?

34. Es hat sich bestätigt, dass bei den Daten aus Aufgabe 33 eine Korrelation besteht. Deswegen wird eine Regressionsfunktion ermittelt, welche diesen Zusammenhang beschreibt. Da der Zusammenhang linear zu sein scheint, wird eine lineare Regressionsfunktion vom Typ $y=a+b\cdot x$ gewählt. Welche konkreten Werte sind für a und b einzusetzen?

    Zur Ermittlung der Parameter a und b sind folgende Formeln erforderlich:

$$b=\frac{s_{xy}}{s_x^{\ 2}}$$

[3] Quelle für das Beispiel: A. Quatember, Statistik ohne Angst vor Formeln – Das Studienbuch für Wirtschafts- und Sozialwissenschaftler, 2. Aufl. München 2008, S. 67

Aufgaben

$$a = \bar{y} - b \cdot \bar{x}$$

wobei mit i = 1…n und x_i und y_i für die Merkmalsausprägungen:

$$s_x^{\;2} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$$

$$s_{xy} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right) \cdot \left(y_i - \bar{y}\right)$$

Alternativ kann b auch nach der Formel $b = r \cdot \frac{s_y}{s_x}$ errechnet werden, wobei

$$r = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right) \cdot \left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} \cdot \sqrt{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}}$$

$$s_x = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

$$s_y = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}$$

35. Mit welchem Einkommen kann ein Beschäftigter unter Zugrundelegung der Daten aus Aufgabe 33 im Alter von 60, 65 und 67 Jahren rechnen?

36. Gegeben sind folgende Merkmalsausprägungen

Lfd. Nr.	x	y
1	1	1,00
2	2	4,00
3	3	9,00
4	4	16,00
5	5	25,00

Wie hoch ist der Korrelationskoeffizient?

37. Durch welche Funktion wird die abhängige Variable in Aufgabe 36 vollständig erklärt? Wenn diese Funktion als Regressionsfunktion angenommen wird, wie hoch ist dann der Korrelationskoeffizient zwischen den Werten der Regressionsfunktion und y?

38. Gegeben sind folgende Merkmalsausprägungen

Lfd. Nr.	x	y
1	1	100,00
2	4	400,00
3	9	900,00
4	16	1.600,00
5	25	2.500,00

Wie hoch ist der Korrelationskoeffizient?

39. Durch welche Funktion wird die abhängige Variable in Aufgabe 38 vollständig erklärt? Wenn diese Funktion als Regressionsfunktion angenommen wird, wie hoch ist dann der Korrelationskoeffizient zwischen den Werten der Regressionsfunktion und y?

40. Eine Regressionsgerade muss die Bedingung $s_y^2 \cdot (1 - r^2) \geq 0$ erfüllen. Wieso folgt daraus, dass der Korrelationskoeffizient r nicht größer als |1| sein darf?

41. Die folgende Tabelle zeigt den durchschnittlichen Bruttomonatsverdienst der Arbeitnehmer im Kredit- und Versicherungsgewerbe in den Jahren 2001 bis 2010[4]. Durch welche lineare Funktion lässt sich diese Entwicklung annähern?

Jahr	Monatsgehalt
2001	3.321
2002	3.427
2003	3.525
2004	3.619
2005	3.692
2006	3.761
2007	3.865
2008	4.012
2009	4.097
2010	4.258

42. Der Verbraucherpreisindex insgesamt ist von 94,5 im Jahre 2001 auf 108,2 im Jahre 2010 gestiegen (Quelle: Statistisches Bundesamt). Wie hoch war die durchschnittliche jährliche Preissteigerungsrate?

43. Wie hoch ist die durchschnittliche jährliche Wachstumsrate der nominalen Gehälter aus Aufgabe 41 zwischen 2001 und 2010?

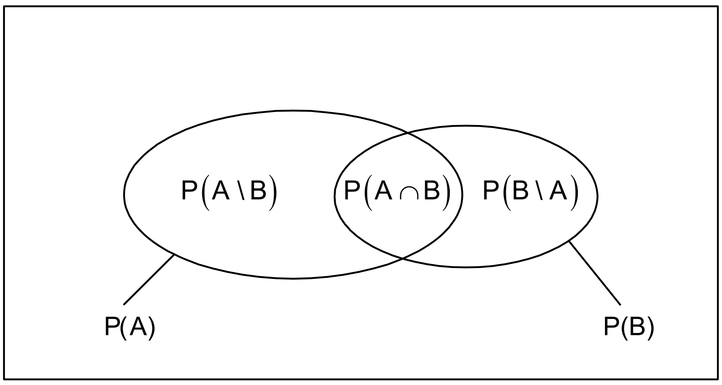
44. Wie hoch ist die reale Wachstumsrate der Gehälter aus Aufgabe 41 zwischen 2001 und 2010, wenn man die Preisentwicklung aus Aufgabe 42 zugrunde legt?

45. Ein Gehalt steigt von 3.321 im Jahre 2001 auf 4.258 im Jahre 2010. Im gleichen Zeitraum stieg der Index der Verbraucherpreise von 94,5 auf 108,2. Wie hoch ist das reale Einkommen des Jahres 2010 zu Preisen von 2001?

46. Eine Kapitalanlage verzinst sich mit 1,75 % pro Jahr. Die Inflationsrate beträgt 2,5 %. Wie hoch ist die reale Verzinsung?

47. Sind die Schadenseintritte bei Versicherungsnehmern disjunkte Ereignisse?

48. Wie lässt sich mithilfe der folgenden Zeichnung die Wahrscheinlichkeit $P(A \cup B)$ dafür ableiten, dass entweder ein beliebiges Ereignis A oder ein anderes beliebiges Ereignis B eintritt?



[4] Quelle: Institut der deutschen Wirtschaft Köln (Hrsg.), Deutschland in Zahlen - Ausgabe 2011 -, Köln 2011, S. 56

Aufgaben

49. Die Wahrscheinlichkeit für das Eintreten des Ereignisses A sei 0,1 und die Wahrscheinlichkeit für das Eintreten des Ereignisses B sei 0,2. Die Wahrscheinlichkeit für das gemeinsame Eintreten der Ereignisse sei 0. Wie hoch ist die Wahrscheinlichkeit, dass entweder A oder B eintritt?

50. Die Wahrscheinlichkeit für das Eintreten des Ereignisses A sei 0,1 und die Wahrscheinlichkeit für das Eintreten des Ereignisses B sei 0,2. Die Wahrscheinlichkeit für das gemeinsame Eintreten der Ereignisse sei 0,02. Wie hoch ist die Wahrscheinlichkeit, dass A oder B eintritt?

51. Die Wahrscheinlichkeit für das Eintreten des Ereignisses A sei 0,9 und die Wahrscheinlichkeit für das Eintreten des Ereignisses B sei 0,9. Die Wahrscheinlichkeit für das gemeinsame Eintreten der Ereignisse sei 0,81. Wie hoch ist die Wahrscheinlichkeit, dass A oder B eintritt?

52. Die Wahrscheinlichkeit für das Eintreten des Ereignisses A sei 0,9 und die Wahrscheinlichkeit für das Eintreten des Ereignisses B sei 0,9. Die Wahrscheinlichkeit für das gemeinsame Eintreten der Ereignisse sei 0,81. Wie hoch ist die Wahrscheinlichkeit, dass
nur A eintritt
nur B eintritt
nur A oder nur B eintritt
A und B gemeinsam eintreten
entweder nur A oder nur B oder beide Ereignisse gemeinsam eintreten?

53. Sind disjunkte Ereignisse abhängige oder unabhängige Ereignisse?

54. Die Wahrscheinlichkeit, dass ein Versicherungsnehmer einen Schaden erleidet, sei 0,1. Die Wahrscheinlichkeit, dass ein anderer Versicherungsnehmer einen Schaden erleidet, sei ebenfalls 0,1. Ein Kumulrisiko wird nicht in Betracht gezogen, d.h. die Wahrscheinlichkeit für den Schaden-eintritt bei einem Versicherungsnehmer wird als unabhängig vom Schadeneintritt bei einem anderen Versicherungsnehmer angesehen. Wie groß ist die Wahrscheinlichkeit, dass beide Versicherungsnehmer einen Schaden haben?

55. Die Wahrscheinlichkeit, dass bei einem Versicherungsnehmer der Schadenfall eintritt, sei w_s. Wie wahrscheinlich ist es, dass der Schadenfall bei i Versicherungsnehmern derselben Risikogruppe gemeinsam eintritt?

56. Die Wahrscheinlichkeit, dass bei einem Versicherungsnehmer der Schadenfall nicht eintritt, sei $1 – w_s$. Wie hoch ist die Wahrscheinlichkeit, dass bei $n – i$ Versicherungsnehmern kein Schaden eintritt?

57. Ein Kaufmann schätzt die Gewinnaussichten von zwei verschiedenen Geschäftsmöglichkeiten folgendermaßen ein:

	Gewinn	Wahrscheinlichkeit	
Geschäft 1	500,00	0,9	
	- 100,00	0,1	
Geschäft 2	600,00	0,9	
	- 1.000,00	0,1	

Wie hoch ist der Gewinnerwartungswert und die Standardabweichung beider Geschäfte?

58. Die Wahrscheinlichkeit für den Eintritt des Schadens s bei einem Versicherungsnehmer sei w_s. Die Wahrscheinlichkeit, dass der Schaden nicht eintritt, sei $1 – w_s$. Wie hoch ist der Erwartungs-wert, die Varianz und die Standardabweichung des Schadens?

59. Die Erwartungswerte der Schäden von Versicherungsnehmern können zum Erwartungswert des Gesamtschadens der Versicherungsunternehmung addiert werden. Die Varianzen der Schäden können zur Varianz des Gesamtschadens addiert werden, wenn die Schadenereignisse unab-hängig voneinander sind. Dies vorausgesetzt, wie hoch ist der Erwartungswert und die Standard-abweichung des Gesamtschadens einer Versicherungsunternehmung mit n gleichartigen Ver-sicherungsnehmern wie in Aufgabe 58 beschrieben?

60. Inwiefern ist der Value at Risk für den Statistiker nichts anderes als ein Quantil?