Klausur: Mathematik und Statistik

Lehrveranstaltung: Statistik

Fakultät für Wirtschaft

Studiengang: **BWL-Öffentliche Wirtschaft** **Datum: 28.06.2023**

Studierende(r) Matrikelnummer:

Dozent: Jürgen Meisel

Kurs: WOW A/B **Studienjahrgang: 2022** **Semester: 2**

Hilfsmittel: ***Wiss. TR (nicht programmierbar) / Formelsammlung***

Bearbeitungszeit: **60 Minuten**

Bewertung: **Maximale Punktzahl: 60**

Erreichte Punktzahl:

Datum, Unterschrift

……………………………………..

Anmerkungen: ***Von 7 gestellten Aufgaben müssen 5 ausgewählt und bearbeitet werden.***

Auf-gabennr.:	Thema / Bereich	maximale Punkte	erreichte Punkte	Bemerkungen
1	Mittelwerte & Streumaße (diskret)	12		
2	Mittelwerte & Streumaße (klassiert)	12		
3	Konzentration (Ginikoeff. & Lorenzkurve)	12		
4	Regression & Korrelation	12		
5	Preisindizes und Inflationsrate	12		
6	Stochastik: Verteilung (diskret)	12		
7	Stochastik: Verteilung (stetig)	12		
		5 aus 7		
Summe		60		

Klausur: Statistik – 28.06.2023

Jürgen Meisel

Hilfsmittel: **Wiss. nicht progr. Taschenrechner + Formelsammlung**
Bearbeitungszeit: **60 Minuten**

Aufgabe 1: Mittelwerte und Streumaße (diskret)

Dozent Knödelhuber unterrichtet an der Dualen Hochschule Mannheim im Schweiße seines Angesichts Mathematik und Statistik.
Im Rahmen der Klimadiskussion überlegt er sich, ob er zukünftig ganz auf die S-Bahn umsteigen sollte und hat hierzu einige Testdaten aus den bisherigen Fahrten gesammelt:

Benötigte Fahrzeit [in Minuten]	Anzahl der Fahrten
45	7
50	15
52	12
55	13
60	2
80	1
Summe	

a) Bestimmen Sie aus den Daten folgende Größen:
Arithmetisches Mittel, Modus und Median

b) Ermitteln Sie auch die zugehörigen Streumaße Standardabweichung und beide Quartile.

c) Beurteilen Sie aufgrund Ihrer Ergebnisse die Situation und geben Sie Dozent Knödelhuber eine Empfehlung.

Lösung:

Benötigte Fahrzeit [in Minuten]	Anzahl der Fahrten	Gesamminuten	rel. Hkeit	kum. rel. Hkeit	kum. abs. Hkeit	quadr. Fahrzeit	gewichtet
45	7	315	0,14	0,14	7	2025	283,5
50	15	750	0,3	0,44	22	2500	750
52	12	624	0,24	0,68	34	2704	648,96
55	13	715	0,26	0,94	47	3025	786,5
60	2	120	0,04	0,98	49	3600	144
80	1	80	0,02	1	50	6400	128
Summe	50	2604	1				2740,96

Arithm. MW	52,08	2604/50		Varianz	28,6336	
				Std.Abw.	5,35103728	
Modus	50	häufigster Wert				
Median	52	$0,5 * (x_{25} + x_{26})$		Quartil 1:	50	x_{13}
				Quartil 3:	55	x_{38}

Aufgabe 2: **Mittelwerte und Streumaße (klassiert)**

Dozent Knödelhuber unterrichtet (immer noch) an der Dualen Hochschule Mannheim Mathematik und Statistik.
Um eine endgültige klimafreundliche aber auch nervenschonende Entscheidung treffen zu können, hat er einige Testdaten aus den letzten Fahrten mit dem Pkw gesammelt:
Häufigkeitstabelle der Fahrten mit dem Pkw:

Fahrzeit [Min]	absolute Häufigkeit	relative Häufigkeit	Klassen-mitte	Klassen-breite	Häufigkeits-dichte	kum. rel. Häufigkeit
[32 ; 40[	20					
[40 ; 46[	40	0,4				
[46 ; 50[		0,25				
[50 ; 56[	10					
[56 ; 70[						
Summe						

a) Vervollständigen Sie die Tabelle.
b) Bestimmen Sie den arithmetischen Mittelwert, die modale Klasse und den Modalwert.
c) Berechnen Sie den Median, das untere Quartil und das obere Quartil.

Lösung:

Fahrzeit [Min]	absolute Häufigkeit	relative Häufigkeit	Klassen-mitte	Klassen-breite	Häufigkeits-dichte	kum. rel. Häufigkeit	quadriert	gewichtet
[32 ; 40[	20	0,2	36	8	2,50	0,2	1296	259,2
[40 ; 46[	40	0,4	43	6	6,67	0,6	1849	739,6
[46 ; 50[	25	0,25	48	4	6,25	0,85	2304	576
[50 ; 56[	10	0,1	53	6	1,67	0,95	2809	280,9
[56 ; 70[	5	0,05	63	14	0,36	1	3969	198,45
Summe	100	1						2054,15

Arith MW	**44,85**
Modus	**43**
Median	**44,5**
Varianz	**42,6275**
StdAbw.	**6,53**
Quartil 1	**40,75**
Quartil 3	**48,4**

$$\bar{x} = \sum_{i=1}^{k}\left[(x_i)_m \cdot \frac{n_i}{n}\right] = \sum_{i=1}^{k}\left[(x_i)_m \cdot p_i\right]$$

$mit\ (x_i)_m\ als\ Klassenmitte\ der\ Klasse\ i$

Modus: **[40 ; 46[** **HDI (maximal)**

$$\overline{x_M} \Rightarrow \overline{x_{0,5}} = [a;b] \rightarrow \overline{x_{0,5}} = a + \frac{\Delta_i \cdot [0,5 - F(a)]}{p_i}$$

$\Delta_i = Klassenbreite \quad p_i = rel.\,Häufigkeit$

$$V(X) \overset{Rechen}{\underset{formel}{=}} \left[\sum_{i=1}^{n}(x_i)_m^2 \cdot p_i\right] - \bar{x}^2 \Rightarrow S(X) = \sqrt{V(X)}$$

$mit\ (x_i)_m\ als\ Klassenmitte\ der\ Klasse\ i$

$$q_1 \Rightarrow \overline{x_{0,25}} = [a;b] \rightarrow \overline{x_{0,25}} = a + \frac{\Delta_i \cdot [0,25 - F(a)]}{p_i}$$

$$q_3 \Rightarrow \overline{x_{0,75}} = [a;b] \rightarrow \overline{x_{0,75}} = a + \frac{\Delta_i \cdot [0,75 - F(a)]}{p_i}$$

$\Delta_i = Klassenbreite \quad p_i = rel.\,Häufigkeit$

Aufgabe 3: **Gini-Koeffizient & Lorenzkurve**

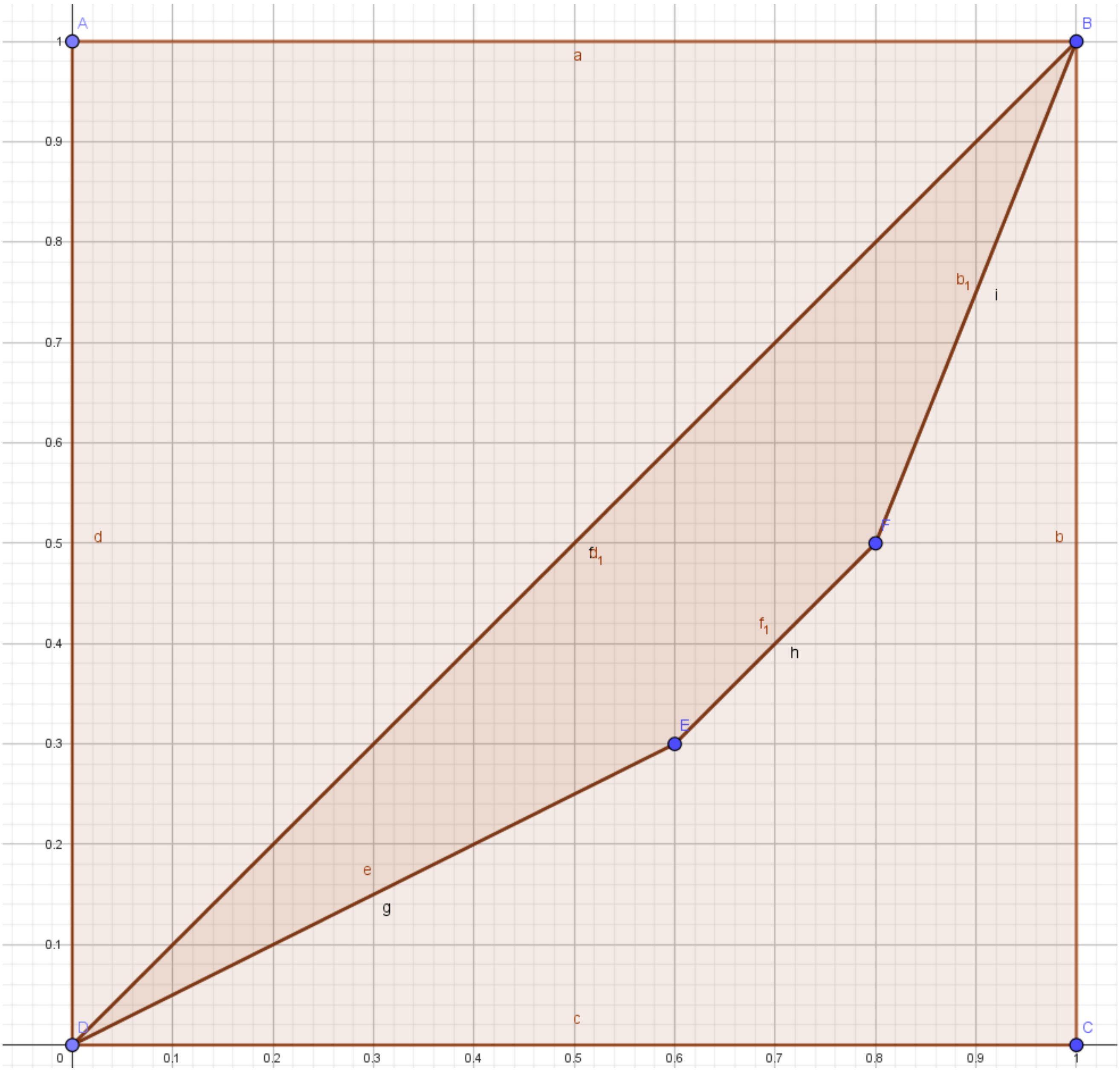
Im Landkreis Statistika gibt es 5 Krankenkassen (x-Achse), wobei sich die Gesamtzahl der **2 Mio. Mitglieder** (y-Achse) wie folgt aufteilt:

Krankenkasse	Absolute Mitgliederanzahl	Relative Mitgliederanzahl
KMK	**200.000**	
KDA	**400.000**	
MKD	**200.000**	
Zwerg		
Hightower		**0,5**

a) Zeichnen Sie die dazugehörige Lorenzkurve und berechnen Sie den Gini-Koeffizient.

b) Ermitteln Sie den normierten Gini-Koeffizienten und erläutern Sie den Unterschied.

Lösung:

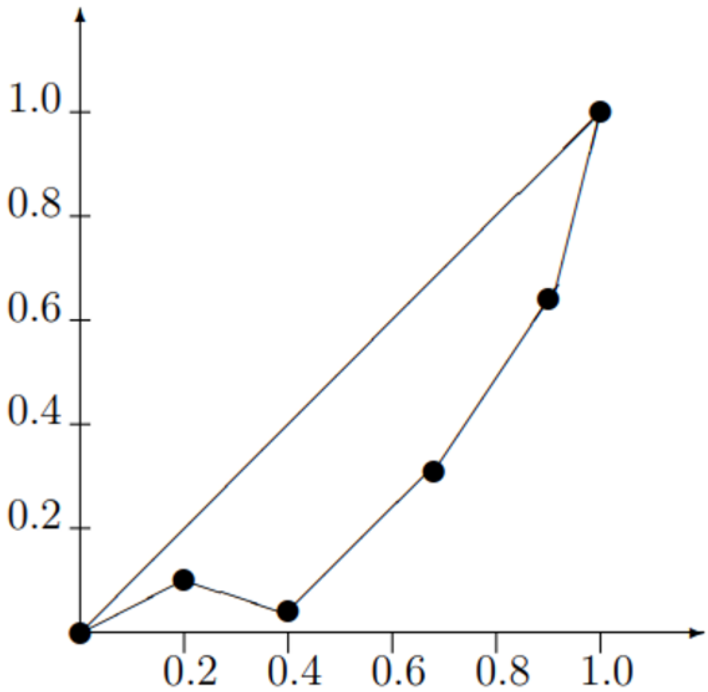


Krankenkasse	rel. Anteil	kum. rel. Anteil	Absolute Mitgliederanzahl	Relative Mitgliederanzahl	kum. Rel. Mit-Anzahl
KMK; MKD; Zwerg	0,6	0,6	600.000	0,3	0,3
KDA	0,2	0,8	400.000	0,2	0,5
Hightower	0,2	1	1.000.000	0,5	1
Summe	1	x-Achse	2.000.000	1	y-Achse

Fläche unter Lorenzkurve	
Fläche 1:	0,09
Fläche 2:	0,08
Fläche 3:	0,15
Gesamt:	0,32

Konz-Fläche:	0,18
GK:	0,36
GK normiert:	0,45

c) Oh je – hier ist ja wohl etwas schiefgelaufen.
Die nachfolgende Abbildung war ein erster Versuch einer Lorenzkurve.
Erläutern Sie den Fehler und machen Sie einen Korrekturvorschlag.



Lösung: **Problem der fehlerhaften Anordnung der Kundenanzahl bei den Apotheken.**

Aufgabe 4: **Regression & Korrelation**

Bei einer Bewerberaktion (15 Kandidaten) für die Stelle an einem Lehrstuhl wurden zwei Arten von Tests durchgeführt:

Aufgabe A: Erfassen von Informationen und Beantwortung von Fragen [Zeit in Min.]

Aufgabe B: Reaktionstest [Zeit in Sek.]

Die Ergebnisse sind in der folgenden Tabelle dargestellt – oh je, leider ist etwas Kaffee drübergelaufen 😉

Bew.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	4,5	5,2	6,0	5,5	4,2	[illegible]	5,5	4,7	4,9	7,2	6,8	4,0	5,8	5,1	6,2
B	11,0	10,3	9,2	11,0	13,5	11,4	[illegible]	9,5	12,0	8,4	10,2	11,4	10,2	10,0	9,6

a) Bestimmen Sie die lineare Regressionsfunktion aus den Ergebnissen von Aufgabe A (x-Wert) und Aufgabe B (y-Wert).

Ansatz:

$$y = b_0 + b_1 x$$

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i \rightarrow \mu_a = \frac{1}{15}\sum_{i=1}^{15} a_i = \frac{81,2}{15} = 5,413$$

$$\mu_y = \frac{1}{n}\sum_{i=1}^{n} y_i \rightarrow \mu_b = \frac{1}{15}\sum_{i=1}^{15} b_i = \frac{157,9}{15} = 10,5267$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i y_i) \rightarrow \frac{1}{15}\sum_{i=1}^{15}(a_i \cdot b_i) = \frac{844}{15} = 56,267$$

$$V(X) = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2\right) - \mu_x^2 \rightarrow V(A) = \sigma_A^2 = 0,875^2 = 0,765625$$

$$b_1 = \frac{\sigma_{xy}}{\sigma_{x^2}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i y_i) - \mu_x \cdot \mu_y}{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^2\right) - \mu_x^2} = \frac{Kov(X,Y)}{V(X)}$$

$$b_1 = \frac{\sigma_{AB}}{\sigma_{A^2}} = \frac{Kov(A,B)}{V(A)} = \frac{56,267 - 5,413 \cdot 10,5267}{0,765625} = -0,933$$

$$b_0 = \mu_y - b_1 \cdot \mu_x$$

$$b_0 = \mu_B - b_1 \cdot \mu_A \rightarrow b_0 = 10,53 - (-0,933) \cdot 5,413 = 15,58$$

b) Ermitteln Sie die ***Korrelation nach Pearson*** **und** ***die Kovarianz*** zwischen Aufgabe A (x-Wert) und Aufgabe B

Lösung:

$$r = \frac{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i y_i) - \mu_x \cdot \mu_y}{\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i^2) - \mu_x^2}\cdot\sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i^2) - \mu_y^2}} \rightarrow \frac{Kov(A,B)}{S(A)\cdot S(B)} = \frac{\sigma(A,B)}{\sigma(A)\cdot\sigma(B)}$$

$$r = \frac{56,267 - 5,413\cdot 10,5267}{0,875\cdot 1,206} = \frac{-0,7140271}{1,05525} = -0,677$$

Anmerkung:

(i) ***Geben Sie bei den Teilaufgaben a) und b) die notwendigen Formeln an;***

(ii) ***Sie können zudem die hier bereits berechneten Hilfsgrößen bei Ihrer eigenen Ermittlung der Lösungen verwenden.***

$$\sum_{i=1}^{15} a_i = 81,2 \qquad \sum_{i=1}^{15} b_i = 157,9 \qquad \sum_{i=1}^{15}(a_i \cdot b_i) = 844 \qquad \sigma_A = 0,875 \qquad \sigma_B = 1,206$$

c) **Sie möchten nun folgende Hypothese prüfen:**

„Die reaktionsschnellen Bewerber sind bei der Textarbeit eher langsamer."

Beurteilen Sie die Hypothese entsprechend des Korrelationswertes bzw. der Regressionsfunktion.

Lösung:

Die Hypothese kann aufgrund des hohen negativen Korrelationswertes bestätigt werden. Zudem besitzt die lineare Regressionsfunktion eine negative Steigung, was ebenfalls auf ein entgegengesetztes Zusammenwirken der beiden Eigenschaften schließen lässt.

d) Um die Korrelation auch zu ermitteln, kann man den Spearman-Rangkorrelationskoeffizienten berechnen – dieser liegt hier bei **r = -0,605**.
Erläutern Sie hierzu kurz die Vorgehensweise der Berechnung **(nicht berechnen!!!)**.

Lösung:

(1) Rangfolgen von A und B festlegen; bei gleicher Maßzahl den Mittelwert der betreffenden Ränge bilden
(2) Rangdifferenzen berechnen
(3) Quadrieren der Rangdifferenzen, um die positiven und negativen Differenzen auszugleichen
(4) Anwendung der Formel $$r_S = 1 - \frac{6\sum_{i=1}^{n}(d_i)^2}{n(n^2-1)}$$

Aufgabe 5: **Warenkorbmethode und Preisindexberechnung**

Ein Unternehmen hat eine Preis-Mengen-Übersicht für die bezogenen Güter A, B und C angefertigt.

Gut	Preise		Mengen	
	2020	2023	2020	2023
A	10	15	60	50
B	25	20	40	70
C	30	40	80	60

a) Ermitteln Sie hierzu die Preisindizes nach Laspeyres und Paasche.

b) Berechnen Sie den Preisindex nach Fisher.

c) Wie hoch ist die jährliche Inflationsrate auf der Grundlage der Daten nach Laspeyres?

Lösung:

$$Laspeyres: \quad L_P = \frac{\sum p_{23i} \cdot q_{20i}}{\sum p_{20i} \cdot q_{20i}} \rightarrow L_P = \frac{15 \cdot 60 + 20 \cdot 40 + 40 \cdot 80}{10 \cdot 60 + 25 \cdot 40 + 30 \cdot 80} = \frac{4900}{4000} = 1{,}225$$

$$\rightarrow \frac{\text{Ausgaben des Berichtsjahres mit Mengen des Basisjahres (Menge}_{\text{Periode I}} \text{ * Preis}_{\text{Periode II}})}{\text{Ausgaben/Umsatz des Basisjahres (Menge}_{\text{Periode I}} \text{ * Preis}_{\text{Periode I}})}$$

$$Paasche: \quad P_P = \frac{\sum p_{23i} \cdot q_{23i}}{\sum p_{20i} \cdot q_{23i}} \rightarrow P_P = \frac{15 \cdot 50 + 20 \cdot 70 + 40 \cdot 60}{10 \cdot 50 + 25 \cdot 70 + 30 \cdot 60} = \frac{4550}{4050} = 1{,}1234$$

$$\rightarrow \frac{\text{Ausgaben/Umsatz des Berichtsjahres (Menge}_{\text{Periode II}} \text{ * Preis}_{\text{Periode II}})}{\text{Ausgaben des Basisjahres mit Mengen des Berichtsjahres (Menge}_{\text{Periode II}} \text{ * Preis}_{\text{Periode I}})}$$

$$F_P = \sqrt{L_P \cdot P_P} \rightarrow \sqrt{1{,}225 \cdot 1{,}1234} = 1{,}1731$$

$$\text{Inflationsrate: } \sqrt[3]{1{,}225} = 1{,}069987 \rightarrow i_{eff} = 1{,}069987 - 1 = 0{,}069987 \xrightarrow{\cdot 100} 6{,}9987[\%]$$

d) Für einen aus 400 Gütern bestehenden Warenkorb wurden für die Jahre 2018, 2019 und 2020 folgende Umsatzsummen berechnet:

$$\sum_{i=1}^{400} q_{18;i}\, p_{18;i} = 870 \qquad \sum_{i=1}^{400} q_{18;i}\, p_{19;i} = 877 \qquad \sum_{i=1}^{400} q_{18;i}\, p_{20;i} = 898$$

$$\sum_{i=1}^{400} q_{19;i}\, p_{18;i} = 873 \qquad \sum_{i=1}^{400} q_{19;i}\, p_{19;i} = 879 \qquad \sum_{i=1}^{400} q_{19;i}\, p_{20;i} = 902$$

$$\sum_{i=1}^{400} q_{20;i}\, p_{18;i} = 878 \qquad \sum_{i=1}^{400} q_{20;i}\, p_{19;i} = 895 \qquad \sum_{i=1}^{400} q_{20;i}\, p_{20;i} = 905$$

Berechnen Sie hieraus die Preisindices nach Laspeyres und Paasche für 2020 zum Basisjahr 2018.

Lösung

$$Laspeyres: \quad L_P \quad = \frac{\sum p_{20i} \cdot q_{18i}}{\sum p_{18i} \cdot q_{18i}} \quad = \quad \frac{898}{870} \quad = \quad 1{,}032184$$

$$\rightarrow \quad \frac{\text{Ausgaben des Berichtsjahres mit Mengen des Basisjahres (Menge }_{\text{Periode I}} * \text{Preis }_{\text{Periode II}})}{\text{Ausgaben/Umsatz des Basisjahres (Menge }_{\text{Periode I}} * \text{Preis }_{\text{Periode I}})}$$

$$Paasche: \quad P_P \quad = \frac{\sum p_{20i} \cdot q_{20i}}{\sum p_{18i} \cdot q_{20i}} \quad = \quad \frac{905}{878} \quad = \quad 1{,}030752$$

$$\rightarrow \quad \frac{\text{Ausgaben/Umsatz des Berichtsjahres (Menge }_{\text{Periode II}} * \text{Preis }_{\text{Periode II}})}{\text{Ausgaben des Basisjahres mit Mengen des Berichtsjahres (Menge }_{\text{Periode II}} * \text{Preis }_{\text{Periode I}})}$$

Aufgabe 6: **Stochastik – Binomialverteilung – Flüssigdünger im Garten**

Bei einem Produzenten für Flüssigdünger werden zur Qualitätskontrolle 50 Stichproben aus der laufenden Produktion einer Produktlinie mit neuem Wirkungsgrad, von denen 95 % ein positiven Wachstumsverlauf generieren sollen, zur Kontrolle entnommen.

Mit welcher Wahrscheinlichkeit sind unter den entnommenen Proben

a) genau 45 mit hohem Wirkungsgrad?
b) zwischen 2 bis 5 ohne entsprechend intensive Wirkung?
c) höchstens 47 mit hohem Wirkungsgrad?

d) Wie viele Proben müssen der Produktion entnommen werden, um mit einer Wahrscheinlichkeit von 99 % mind. eine Probe ohne Wirkungsgrad zu erhalten?

Lösung

$$B_{50;0,95}(X=45) \quad = \quad \binom{50}{45} 0,95^{45} \cdot 0,05^{5} \quad = \quad 0,0658$$

$$B_{50;0,05}(2 \le Y \le 5) \quad = \quad \sum_{Y=2}^{5} \binom{50}{Y} 0,05^{Y} \cdot 0,95^{50-Y} \quad = \quad 0,6828$$

$$B_{50;0,95}(X \le 47) \quad = \quad \sum_{X=0}^{47} \binom{50}{X} 0,95^{X} \cdot 0,05^{50-X} \quad = \quad 0,4595$$

$$B_{n;0,05}(X \ge 1) \quad \ge \quad 0,99 \quad \to \quad 1 - B_{n;0,05}(X=0) \quad \ge \quad 0,99$$

$$\to \quad B_{n;0,05}(X=0) \quad \le \quad 0,01 \quad \to \quad \binom{n}{0} \cdot 0,05^{0} \cdot 0,95^{n} \quad \le \quad 0,01$$

$$\to \quad 1 \cdot 1 \cdot 0,95^{n} \quad \le \quad 0,01 \quad \xrightarrow{\ln} \quad n \cdot \ln 0,95 \quad \le \quad \ln 0,01$$

$$\xrightarrow{:\ln 0,95\,[<0]} \quad n \quad \ge \quad \frac{\ln 0,01}{\ln 0,95} \quad \approx \quad 89,78 \quad \to \quad n \quad \ge \quad 90$$

Aufgabe 7: **Stochastik – Normalverteilung - Parfümeriehandel**

Es wird eine Bestellung von 5.000 Flakons durchgeführt. Der Lieferant gibt die Zusage, dass die Duftqualität der Flakoninhalte zu 98 % einwandfrei sind.

a) Bestimmen Sie den Erwartungswert und die Werte für das σ-Intervall für einwandfreie Flakons.
b) Ermitteln Sie das 90 %-Intervall innerhalb dessen Flakons mit einwandfreier Duftqualität zu erwarten sind.
c) Wie hoch ist die Wahrscheinlichkeit, dass sich in der Lieferung höchstens 80 Flakons mit schlechter Duftqualität befinden?

ZUSATZAUFGABE: **(4 Punkte)**

Wie viele Flakons schlechter Qualität dürfen höchstens in der Lieferung enthalten sein, wenn die verbesserte Zusage des Lieferanten gelten soll:
Höchstens 1% der Lieferung ist nicht verwertbar!!!

Lösung

$$\mu(X) = n \cdot p \rightarrow \mu(X) = 5.000 \cdot 0{,}98 = 4.900$$

$$\sigma(X) = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{\mu(X) \cdot (1-p)} \rightarrow \sigma(X) = \sqrt{4.900 \cdot 0{,}02} = 9{,}9$$

$$\sigma(X) - Intervall: \left[\mu(X) \pm \sigma(X)\right] \rightarrow [4.890{,}1 \quad ; \quad 4.909{,}9]$$

$$P(k_1 \le X \le k_2) = 0{,}9 \rightarrow P(k_1 \le X \le k_2) = \Phi(z) - \Phi(-z) = 2 \cdot \Phi(z) - 1$$

$$z = \frac{k_2 - \mu}{\sigma} \xrightarrow{z \leftrightarrow \Phi(0{,}95)} 1{,}65 = \frac{k_2 - 4.900}{9{,}9} \rightarrow k_2 = 4.916{,}33 \rightarrow k_1 = 4.883{,}67$$

$$Intervall: [4.883 \quad ; \quad 4.916]$$

$$P(Y \le 80) = \Phi(-2{,}02) = 1 - \Phi(2{,}02) = 0{,}0217$$

$$z = \frac{k - \mu}{\sigma} \xrightarrow[\sigma=9{,}9]{\mu=100} z = \frac{80 - 100}{9{,}9} \rightarrow k = -2{,}02$$

ZUSATZAUFGABE:

$$P(Y \le k) \le 0{,}01 \rightarrow P(Y \le k) = \Phi(-2{,}33) = 1 - \Phi(2{,}33) = 0{,}0099$$

$$z = \frac{k - \mu}{\sigma} \xrightarrow[\sigma=9{,}9]{\mu=100} -2{,}33 = \frac{k - 100}{9{,}9} \rightarrow k = 76{,}9 \rightarrow k \le 76$$