

Klausur: Mathematik und Statistik

Lehrveranstaltung: Statistik

Fakultät für Wirtschaft

Studiengang: BWL-Öffentliche Wirtschaft

Datum: 18.06.2020

Studierende
Matrikelnummer:

Dozent/in: Jürgen Meisel

Kurs: WOW19B

Semester: 2.

Hilfsmittel:

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

Bearbeitungszeit: 60 Minuten

Bewertung:

Maximale Punktzahl: 60 Punkte

Erreichte Punktzahl:

Datum, Unterschrift

.....

Anmerkungen:

Von 7 gestellten Aufgaben müssen 5 ausgewählt und bearbeitet werden.

Aufgabennr.:	Thema / Bereich	maximale Punkte	erreichte Punkte	Bemerkungen
1	Mittelwerte & Streumaße (diskret)	12		
2	Mittelwerte & Streumaße (klassiert)	12		
3	Zeitreihenanalyse (Tertiale & Quartale)	12		
4	Konzentration (Ginikoeff. & Lorenzkurve)	12		
5	Regression & Korrelation	12		
6	Preisindizes und Inflationsrate	12		
7	Stochastik: Normalverteilung	12		
		5 aus 7		
Summe		60		

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

Aufgabe 1: Mittelwerte und Streumaße (diskret)

Die Schüler einer Klasse wurden in der aktuellen Corona-Phase befragt, wie viele WhatsApp-Nachrichten sie pro Tag schreiben.

Folgendes Ergebnis konnte festgestellt werden:

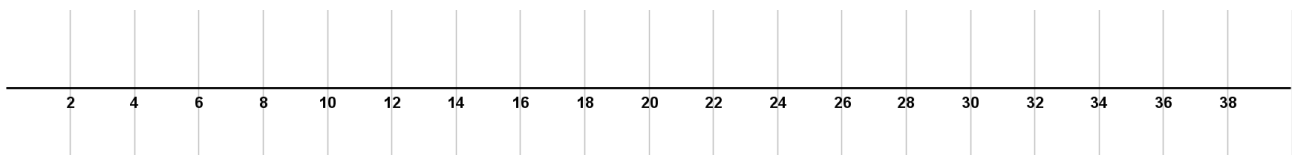
Anzahl Nachrichten pro Schüler	5	35	14	10	6	8	12
Anzahl Schüler	8	2	10	13	8	10	9

- Bestimmen Sie die Anzahl der Schüler.
- Ermitteln Sie folgende Werte bzgl. der Nachrichten:
 - ⇒ Arithmetisches Mittel und Standardabweichung
 - ⇒ Median und Quartil 1 und Quartil 3
 - ⇒ Modus/Modalwert
- Zeichnen Sie den zugehörigen BoxPlot

Anlage 1:

Anzahl Nachrichten pro Schüler							
Anzahl Schüler							

Anlage 2: Bitte zeichnen Sie hier den BoxPlot



Lösung:

a) $N = 60$

b) Arithmetisches Mittel: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \rightarrow \bar{x} = \frac{616}{60} = 10,27$

Standardabweichung:

$$V(X) = \left[\frac{1}{n} \sum_{i=1}^n (x_i)^2 \right] - \bar{x}^2 \rightarrow V(X) = \frac{8.134}{60} - 10,27^2 = 30,16$$

$$\Rightarrow S(X) = \sqrt{V(X)} \rightarrow S(X) = \sqrt{30,16} = 5,49$$

Median: n gerade: n = 60

$$\overline{x_M} = \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) \rightarrow \overline{x_M} = \frac{1}{2} (x_{30} + x_{31}) = \frac{1}{2} (10 + 10) = 10$$

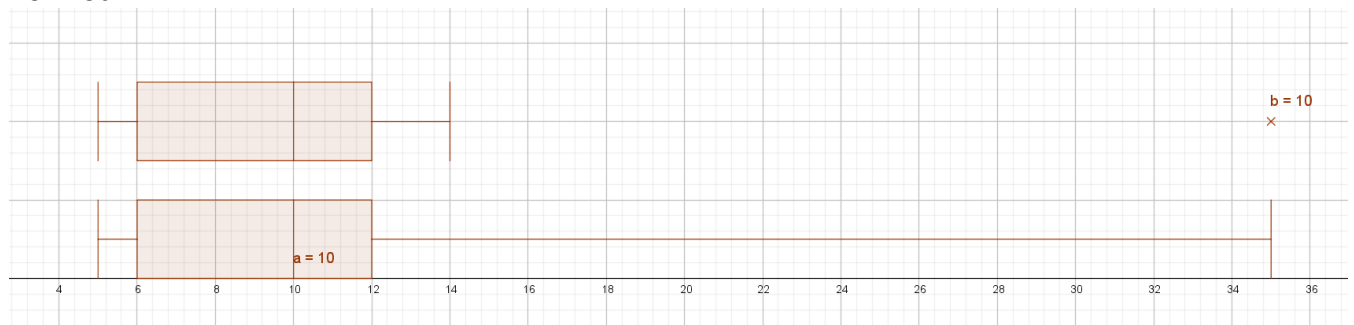
Quartile: (n · p) ganzzahlig → 60 · 0,25 = 15 und 60 · 0,75 = 45

$$\overline{x_{0,25}} = \frac{1}{2} (x_{n \cdot p} + x_{n \cdot p + 1}) \rightarrow \overline{x_{0,25}} = \frac{1}{2} (x_{15} + x_{16}) = \frac{1}{2} (6 + 6) = 6$$

$$\overline{x_{0,75}} = \frac{1}{2} (x_{n \cdot p} + x_{n \cdot p + 1}) \rightarrow \overline{x_{0,75}} = \frac{1}{2} (x_{45} + x_{46}) = \frac{1}{2} (12 + 12) = 12$$

Modus/Modalwert: Häufigster Wert x = 10 (13 Mal)

BoxPlot



Aufgabe 2: Mittelwerte und Streumaße (klassiert)

Aufgrund der Corona-Phase müssen die Trainingseinheiten im Handball überwiegend im Freien stattfinden.

Daher wird sehr intensiv auf das Lauftraining und die Ausdauer gesetzt.

In der Trainingsphase für die ersten vier Wochen (2 Läufe pro Woche) gibt eine Gruppe von 20 ambitionierten Spielerinnen, die regelmäßig eine Strecke von 10 km laufen, Rückmeldung über die Laufzeiten, die in der folgenden Tabelle dargestellt wurden:

Laufzeit (Grenzen in Min.)	Anzahl	p_i	Σp_i	Klassen- breite	Klassen- mitte	hdi	
[40:00 ; 45:00[	3						
[45:00 ; 50:00[	7						
[50:00 ; 55:00[	20						
[55:00 ; 57:00[	20						
[57:00 ; 60:00[	15						

[60:00 ; 70:00]	15						
Summe			---	---	---	---	

- a) Ermitteln Sie folgende Werte bzgl. der Laufzeit:
- ⇒ Arithmetisches Mittel und Standardabweichung
 - ⇒ Median und Quartil 1 und Quartil 3
 - ⇒ Modus/Modalwert
- b) Zeichnen Sie das zugehörige Histogramm.

Lösung:

Laufzeit (Grenzen in Min.)	Anzahl	p_i	Σ π_i	Klassen- breite	Klassen- mitte	Hdi
[40:00 ; 45:00[	3	0,0375	0,0375	5	42,5	0,6
[45:00 ; 50:00[	7	0,0875	0,125	5	47,5	1,4
[50:00 ; 55:00[	20	0,25	0,375	5	52,5	4
[55:00 ; 57:00[	20	0,25	0,625	2	56,0	10
[57:00 ; 60:00[	15	0,1875	0,8125	3	58,5	5
[60:00 ; 70:00]	15	0,1875	1	10	65,0	1,5
Summe	80	1	---	30	---	---

a) Arithmetisches Mittel:

$$\bar{x} = \sum_{i=1}^k \left[(x_i)_m \cdot \frac{n_i}{n} \right] = \sum_{i=1}^k \left[(x_i)_m \cdot p_i \right] \rightarrow \bar{x} = 56,03$$

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

Standardabweichung:

$$V(X) = \left[\sum_{i=1}^n (x_i)_m^2 \cdot p_i \right] - \bar{x}^2 \rightarrow V(X) = 32,58 \Rightarrow S(X) = \sqrt{V(X)} \rightarrow S(X) = \sqrt{32,58} = 5,71$$

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

Median:

$$\bar{x}_M \Rightarrow \bar{x}_{0,5} = [55:00 ; 57:00] \rightarrow \bar{x}_{0,5} = 55 + \frac{2 \cdot [0,5 - 0,375]}{0,25} = 56$$

$$q_1 \Rightarrow \bar{x}_{0,25} = [50:00 ; 55:00] \rightarrow \bar{x}_{0,25} = 50 + \frac{5 \cdot [0,25 - 0,125]}{0,25} = 52,5$$

Quartile:

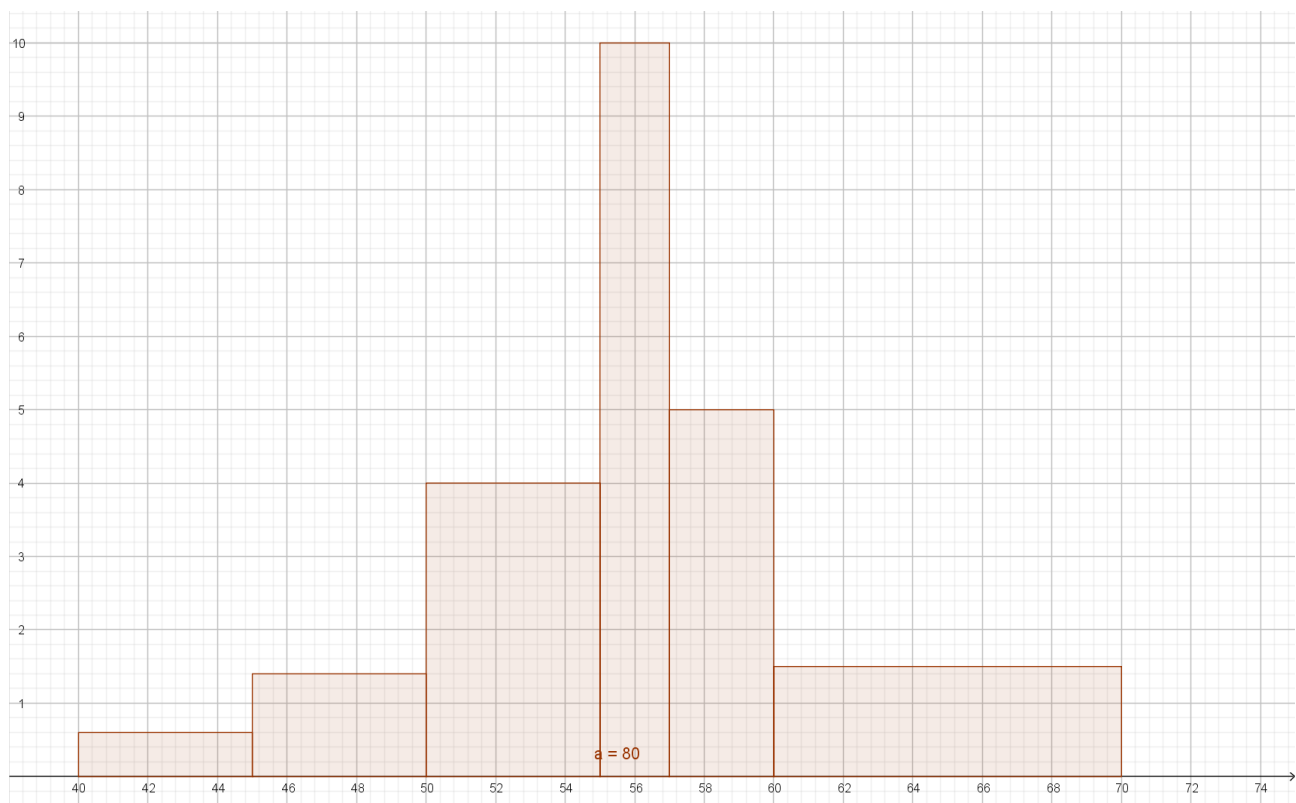
$$q_3 \Rightarrow \bar{x}_{0,75} = [57:00 ; 60:00] \rightarrow \bar{x}_{0,75} = 57 + \frac{3 \cdot [0,75 - 0,625]}{0,1875} = 59,0$$

Modus/Modalwert:

Schritt 1: Modale Klasse bestimmen $\Rightarrow$ Klasse mit max. Häufigkeitsdichte: [55:00 ; 57:00[

Schritt 2: Wert der Klassenmitte der modalen Klasse $\Rightarrow$ Klassenmitte: 56:00

b) Histogramm



Aufgabe 3: Zeitreihenanalyse

Bestimmen Sie die saisonbereinigten Zahlen der Anzahl der WhatsApp-Nachrichten
Bitte verwenden Sie hierfür die **Anlage – Zeitreihenanalyse**.

Aufgabe 4: Konzentration (Ginikoeffizient und Lorenzkurve)

- Bestimmen Sie den Gini-Koeffizient
- Zeichnen Sie die zugehörigen Lorenzkurven zu den drei Szenarien:

Unternehmen	A	B	C	D	E
Umsatzszenario 1	0	0	0	0	5.000
Umsatzszenario 2	1.000	1.000	1.000	1.000	1.000
Umsatzszenario 3	250	250	1.000	1.000	2.500

Bitte alle drei Lorenzkurven in ein Koordinatensystem zeichnen und entsprechend beschriften!

Lösung:

Unternehmen	Umsatz 1
A	0
B	0
C	0
D	0
E	5000
Summe	5 5000

kum. Unt	kum. Unt p	kum. Umsatz	kum. Umsatz p
4	0,8	0	0
5	1	5000	1

Unternehmen	Umsatz 2
A	1000
B	1000
C	1000
D	1000
E	1000
Summe	5 5000

kum. Unt	kum. Unt p	kum. Umsatz	kum. Umsatz p
5	1	5000	1

Unternehmen	Umsatz 3
A	250
B	250
C	1000
D	1000
E	2500
Summe	5 5000

kum. Unt	kum. Unt p	kum. Umsatz	kum. Umsatz p
2	0,4	500	0,1
4	0,8	2500	0,5
5	1	5000	1

Werte der Gini-Koeffizienten:

Unternehmensszenario 1:

$$Gini(GK) = \frac{0,5 - 0,5 \cdot 0,2 \cdot 1}{0,5}$$

$$Gini(GK) = \frac{0,4}{0,5} = 0,8$$

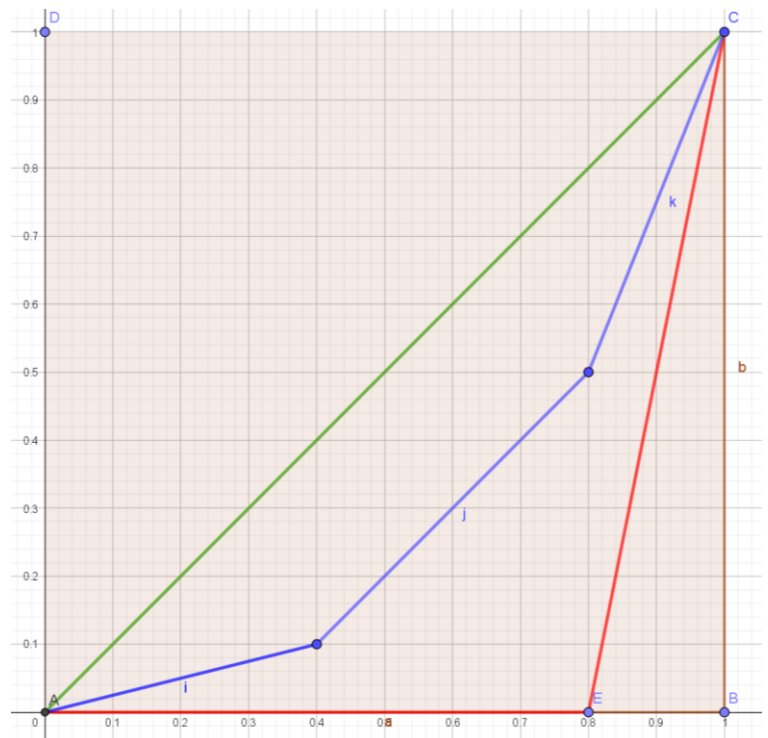
Unternehmensszenario 2:

$$Gini(GK) = \frac{0,5 - 0,5 \cdot 1}{0,5} = 0$$

Unternehmensszenario 3:

$$Gini(GK) = \frac{0,5 - 0,29}{0,5}$$

$$Gini(GK) = \frac{0,21}{0,5} = 0,42$$



$$\left. \begin{aligned} A_1 &= \frac{1}{2} \cdot 0,4 \cdot 0,1 = 0,02 \\ A_2 &= \frac{1}{2} \cdot (0,1+0,5) \cdot 0,4 = 0,12 \\ A_3 &= \frac{1}{2} \cdot (0,5+1,0) \cdot 0,2 = 0,15 \end{aligned} \right\} A_1 + A_2 + A_3 = 0,29$$

Aufgabe 5: Regression & Korrelation

Zur Vorbereitung einer Klausur in Statistik wurden 8 StudentInnen nach ihrer durchschnittlichen Lernzeit und den erreichten Punkten befragt.

StudentIn	A	B	C	D	E	F	G	H
Lernzeit [h]	10	14	5	8	9	10	10	6
Punktzahl	48	45	25	35	54	50	55	32

- Bestimmen Sie die daraus entstehende lineare Regressionsfunktion.
- Wie lange hätte demnach ein Student lernen müssen, um die Idealpunktzahl 60 zu erreichen?
- Ermitteln Sie die **beiden** Korrelationskoeffizienten nach **Pearson und Spearman**.

Lösung:

lineare Regressionsfunktion $f(x) = 18,5 + 2,72x$

$$f(x) = 18,5 + 2,72x = 60 \xrightarrow{-18,5} 2,72x = 41,5 \xrightarrow{:2,72} x = 15,25[h]$$

Korrelationskoeffizient nach Pearson: $r = 0,6853$

Rangkorrelationskoeffizient nach Spearman:

Student	Lernzeit	Rang Lernzeit	Punktzahl	Rang Punktzahl	Differenz Ränge	Quadrat Differenz
A	10	3	48	4	-1	1
B	14	1	45	5	-4	16
C	5	8	25	8	0	0
D	8	6	35	6	0	0
E	9	5	54	2	3	9
F	10	3	50	3	0	0
G	10	3	55	1	2	4
H	6	7	32	7	0	0
					Summe:	30

$$r_s = 1 - \frac{6 \sum_{i=1}^n (d_i)^2}{n(n^2 - 1)} \rightarrow r_s = 1 - \frac{6 \cdot 30}{8 \cdot 63} = \frac{9}{14} \approx 0,6429$$

Aufgabe 6: Preisindizes und Inflation

Der Warenkorb zur Ermittlung der Preisindizes wird auch länderspezifisch berechnet. Daher existiert natürlich auch der Pfälzische Warenkorb unter der Abkürzung WWW => **W(eck)W(orscht)W(oi)**.

Folgende Daten für Knut Knödel sind bekannt:

Produkte	Daten aus Juni 2015			Daten aus Juni 2019			
	Menge	Preis (pro ME)	Umsatz in €	Menge	Mengenänderung ggü. 2015	Preis (pro ME)	Umsatz
Brötchen	60	0,30	18,00	78	+ 30 %	0,40	31,20
Brot	10	3,00	30,00	8	- 20 %	4,00	32,00
Leberwurst	5	5,00	25,00	6	+ 20 %	4,50	27,00
Blutwurst	12	4,50	54,00	15	+ 25 %	6,00	90,00
Riesling	8	3,90	31,20	10	+ 25 %	4,80	48,00
Dornfelder	10	5,00	50,00	7	- 30 %	4,50	31,50

- Vervollständigen Sie die Tabelle
- Berechnen Sie die Preisindizes nach Laspeyres, Paasche und Fisher.
- Bestimmen Sie die jährliche Preisänderungsrate aufgrund des Index nach Laspeyres.

$$L_p = \frac{\sum p_{1i} \cdot q_{0i}}{\sum p_{0i} \cdot q_{0i}}$$

$$L_p = \frac{0,40 \cdot 60 + 4,00 \cdot 10 + 4,50 \cdot 5 + 6,00 \cdot 12 + 4,80 \cdot 8 + 4,50 \cdot 10}{18,00 + 30,00 + 25,00 + 54,00 + 31,20 + 50,00} = \frac{241,90}{208,20} = 1,1619$$

$$P_p = \frac{\sum p_{1i} \cdot q_{1i}}{\sum p_{0i} \cdot q_{1i}}$$

$$P_p = \frac{31,20 + 32,00 + 27,00 + 90,00 + 48,00 + 31,50}{78 \cdot 0,3 + 8 \cdot 3,00 + 6 \cdot 5,00 + 15 \cdot 4,50 + 10 \cdot 3,90 + 7 \cdot 5,00} = \frac{259,70}{218,90} = 1,1864$$

$$F_p = \sqrt{1,1619 \cdot 1,1864} = 1,1741$$

$$g_{MW} = \sqrt[4]{1,1619} = 1,03822$$

$$i_{eff} = 1,03822 - 1 = 0,03822 \rightarrow p = 3,822[\%]$$

Aufgabe 7: Stochastik - Normalverteilung

Das Gewicht von neugeborenen Kindern sei normalverteilt mit $\mu = 3.300$ g und $\sigma = 600$ g.

Wie groß ist die Wahrscheinlichkeit, dass ein Neugeborenes

- a) weniger als 3.500 g wiegt?
- b) zwischen 3.000 und 3.800 g wiegt?
- c) Wie schwer muss ein Neugeborenes mind. sein, damit es zu den 25 % schwersten Babys gehört?

$$P(X \leq 3.500) = \Phi_{\mu;\sigma}(0,33) = 0,6293 = 62,93[\%]$$

$$\text{mit } z = \frac{3.500 - 3.300}{600} = \frac{1}{3}$$

$$P(3.000 \leq X \leq 3.800) = \Phi_{\mu;\sigma}(0,83) - \Phi_{\mu;\sigma}(-0,5) = 0,7967 - [1 - 0,6915]$$

$$P(3.000 \leq X \leq 3.800) = 0,7967 - 0,3085 = 0,4882 = 48,82[\%]$$

$$\text{mit } z_1 = \frac{3.000 - 3.300}{600} = -\frac{1}{2} \quad \text{und} \quad z_2 = \frac{3.800 - 3.300}{600} = \frac{5}{6} = 0,83$$

$$P(X \geq k) = 0,25 \rightarrow P(X < k) = 0,75 \rightarrow z = 0,67$$

$$\rightarrow 0,67 = \frac{k - 3.300}{600} \rightarrow k = 3.702$$

Anlage zu Aufgabe 3: Zeitreihenanalyse

Jahr/ Tertial	Ursprungswert	gleitender Durchschnitt	Differenz Tertial I	Differenz Tertial II	Differenz Tertial III	Saisonkomponente	Saisonbereinigte Werte
2018/I	20						
2018/II	25						
2018/III	36						
2019/I	15						
2019/II	20						
2019/III	30						
2020/I	18						
2020/II	60						
2020/III	50						
2021/I	40						
2021/II	26						
2021/III	20						
	Werte in 1.000	Durchschnitt = Saisonkomponente					

Lösung Formelansicht

	E	F	G	H	I	J
3	gleitender Durchschnitt	Differenz Tertial I	Differenz Tertial II	Differenz Tertial III	Saison- komponente	Saisonbereinigte Werte
4						
5					=F17	=D5-I5
6	= (D5+D6+D7)/3		=D6-E6		=G17	=D6-I6
7	= (D6+D7+D8)/3			=D7-E7	=H17	=D7-I7
8	= (D7+D8+D9)/3	=D8-E8			=I5	=D8-I8
9	= (D8+D9+D10)/3		=D9-E9		=I6	=D9-I9
10	= (D9+D10+D11)/3			=D10-E10	=I7	=D10-I10
11	= (D10+D11+D12)/3	=D11-E11			=I8	=D11-I11
12	= (D11+D12+D13)/3		=D12-E12		=I6	=D12-I12
13	= (D12+D13+D14)/3			=D13-E13	=I10	=D13-I13
14	= (D13+D14+D15)/3	=D14-E14			=I11	=D14-I14
15	= (D14+D15+D16)/3		=D15-E15		=I6	=D15-I15
16					=I13	=D16-I16
17	Durchschnitt = Saisonkomponente	=MITTELWERT(F5:F16)	=MITTELWERT(G5:G16)	=MITTELWERT(H5:H16)		

Lösung:

	C	D	E	F	G	H	I	J
3	Jahr	Ursprungs- wert	gleitender Durchschnitt	Differenz Tertial I	Differenz Tertial II	Differenz Tertial III	Saison- komponente	Saisonbereinigte Werte
4	Tertial							
5	2018/I	20					-8,44	28,44
6	2018/II	25	27,00		-2,00		2,75	22,25
7	2018/III	36	25,33			10,67	6,00	30,00
8	2019/I	15	23,67	-8,67			-8,44	23,44
9	2019/II	20	21,67		-1,67		2,75	17,25
10	2019/III	30	22,67			7,33	6,00	24,00
11	2020/I	18	36,00	-18,00			-8,44	26,44
12	2020/II	60	42,67		17,33		2,75	57,25
13	2020/III	50	50,00			0,00	6,00	44,00
14	2021/I	40	38,67	1,33			-8,44	48,44
15	2021/II	26	28,67		-2,67		2,75	23,25
16	2021/III	20					6,00	14,00
17		Werte in 1.000	Durchschnitt = Saisonkomponente	-8,44	2,75	6,00		