

INHALT

2.2 Zeitreihenanalyse, Trend

Lineare Regressionsanalyse mit einem Einflussfaktor $X = \text{"Zeit"}$

Die trendbereinigten Saisonschwankungen $e_i = s_i = y_i - \hat{y}_i$

Die mittleren Saisonschwankungen $\bar{s}_j = \frac{1}{k} \sum_{i=1}^k s_j$

Die irregulären Komponenten $ir_i = s_i - \bar{s}_j$

2.5 Logistischer Trend, Sättigungsmodelle

Nichtlineare Regressionsanalyse, Bestimmtheitsmaß als Prüfmaß

Instrumente, die zeitliche Entwicklungen zusammenfassen:

2.6 Gleitende Mittelwerte

2.7 Indexzahlen

2.9 Wachstumsfaktoren und Zuwachsraten

2.12 Exponentielles Glätten

Die Beobachtungswerte werden mit den Werten einer geometrischen Folge gewichtet

THEIL's U als Prüfmaß

2.1 TRENDGERADE

Beispiel 2.1

Die folgende Tabelle enthält die Absatzmengen eines Betriebs für die Zeiträume 2000 bis 2008.

Zeitraum	Zeit X	Absatzmenge Y
x*(i)	x(i)	y(i)
Jahr		[1000 Stück]
2000	1	204
2001	2	225
2002	3	264
2003	4	288
2004	5	263
2005	6	290
2006	7	310
2007	8	328
2008	9	404

Ist ein linearer Zusammenhang zwischen der Zeitentwicklung X und einem Merkmal Y statistisch gesichert, dann liegt ein linearer Trend vor.

Die Regressionsfunktion $\hat{y} = m x + b$ nennt man **Trendgerade**; sie lässt sich wieder mit Hilfe der Methode der kleinsten Quadrate berechnen.

Trendgeraden sind Spezialfälle der Regressionsgeraden, wobei X das Merkmal **Zeit** darstellt.

Korrelationskoeffizient r und Bestimmtheitsmaß r^2 geben wieder Auskunft über die Stärke der Korrelation.

Wenn für die Prüfmaße aus der FISHER-Verteilung $x F_{\text{empirisch}} > x F_{\text{kritisch}}$ gilt, dann ist der Zusammenhang statistisch gesichert. \rightarrow Excel / Trend

Als Datenreihe X benutzt man die Folge der Zeiträume x(i), nicht die Jahresangaben x*(i).

Eine Folge von Daten Y, die von der zeitlichen Entwicklung abhängen, nennt man **Zeitreihe**. Die **Zeitreihenanalyse** beschäftigt sich mit der Untersuchung solcher Merkmale Y(X).

Hauptziel der Zeitreihenanalyse ist die **Prognose**, d.h. die Berechnung von Werten, die zeitlich nach dem letzten Beobachtungswert liegen, im obigen Beispiel: $\hat{y}(10)$, $\hat{y}(11)$ usw.

Daten einer Zeitreihe sind auf Zeitpunkte oder auf Zeiträume bezogen.

Zeitpunkte, z.B. Einwohner der Stadt Mannheim am 1.1.2006. Man behandelt die Daten als seien sie stetig, im Diagramm können die Punkte verbunden werden.

Zeiträume, z.B. Anzahl der Geburten in der Stadt Mannheim im Jahr 2005. Man behandelt die Daten als diskrete Werte, im Diagramm können die Punkte nicht verbunden werden. Keine Zwischenwerte!

Das Komponentenmodell

Die Beobachtungswerte y_i entstehen im Zusammenwirken vieler Einflussfaktoren. Im einfachsten Fall stellt man sich vor, dass sich die Faktoren **additiv** überlagern.

$$y_i = \hat{y}_i + s_i + ir_i$$

Beobachtungswerte = Trendwerte + Saisonkomponenten + irreguläre Restwerte

\hat{y}_i sind die **saisonbereinigten Trendwerte**, s_i sind die **trendbereinigten Saisonwerte**.

Die Saisonkomponenten sind zunächst einfache Differenzen: $s_i = y_i - \hat{y}_i$ (= e_i)

In den Saisonkomponenten sind also zunächst die Restwerte ir_i enthalten.

Diese Abweichungen s_i nennt man **Residuen, "errors", Saisonschwankungen, Zykluskomponenten oder trendbereinigte Werte**.

In Residuenanalysen untersucht man, ob diese Residuen s_i irgendeinem Zyklus folgen. (Autoregression)

Man bestimmt dazu den betragsmaximalen Auto-Korrelationskoeffizienten $R_h = \frac{\sum_{t=1}^{n-h} (y_t - \bar{y})(y_{t+h} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}$

Typischen Saisonschwankungen sind 12-Monate-Zyklen, die 4-Quartale-Zyklen, die 5-Arbeitstage-Zyklen, die 7-Tage-Zyklen, Tageszeitschwankungen. 1,5% yoy

2.2 SAISONSCHWANKUNGEN

Maximaler Auto-Korrelationskoeffizient $R_3 = 0,655 \Rightarrow$ Zyklus über 3 Zeiträume

Zeitraum	Zins-satz	Mittelwert $\bar{y}_m = 4,73$ Zähler = $[y(t) - \bar{y}_m] \cdot [y(t+1) - \bar{y}_m]$															
t [Jahre]	y [%]	$y(t) - \bar{y}_m$	$[y(t) - \bar{y}_m]^2$	$y(t+1)$	$Za-(t+1)$	$y(t+2)$	$Za-(t+2)$	$y(t+3)$	$Za-(t+3)$	$y(t+4)$	$Za-(t+4)$	$y(t+5)$	$Za-(t+5)$	$y(t+6)$	$Za-(t+6)$	$y(t+7)$	$Za-(t+7)$
1	5,10	0,375	0,141	4,100	-0,234	4,80	0,028	5,20	0,178	4,40	-0,122	4,90	0,066	5,20	0,178	4,30	-0,159
2	4,10	-0,625	0,391	4,800	-0,047	5,20	-0,297	4,40	0,203	4,90	-0,109	5,20	-0,297	4,30	0,266	4,90	-0,109
3	4,80	0,075	0,006	5,200	0,036	4,40	-0,024	4,90	0,013	5,20	0,036	4,30	-0,032	4,90	0,013	5,00	0,021
4	5,20	0,475	0,226	4,400	-0,154	4,90	0,083	5,20	0,226	4,30	-0,202	4,90	0,083	5,00	0,131	4,10	-0,297
5	4,40	-0,325	0,106	4,900	-0,057	5,20	-0,154	4,30	0,138	4,90	-0,057	5,00	-0,089	4,10	0,203	4,70	0,008
6	4,90	0,175	0,031	5,200	0,083	4,30	-0,074	4,90	0,031	5,00	0,048	4,10	-0,109	4,70	-0,004		-0,537
7	5,20	0,475	0,226	4,300	-0,202	4,90	0,083	5,00	0,131	4,10	-0,297	4,70	-0,012		0,786	R(7) =	-0,298
8	4,30	-0,425	0,181	4,900	-0,074	5,00	-0,117	4,10	0,266	4,70	0,011		-0,391	R(6) =	0,436		
9	4,90	0,175	0,031	5,000	0,048	4,10	-0,109	4,70	-0,004		-0,693	R(5) =	-0,217				
10	5,00	0,275	0,076	4,100	-0,172	4,70	-0,007		1,181	R(4) =	-0,384						
11	4,10	-0,625	0,391	4,700	0,016		-0,589	R(3) =	0,655								
12	4,70	-0,025	0,001		-0,758	R(2) =	-0,327										
13	4,94	0,212	1,803	R(1) =	-0,421												

Beispiel 2.2

Die Entwicklung einer betrieblichen Kennzahl Y über 13 Quartale soll untersucht werden.

Die Kennzahlen sind zeitpunktbezogen, die Absatzmengen in Beisp. 2.1 waren zeitraumbezogen.

Die Untersuchung umfasst **Trendanalyse und Residuenanalyse**:

Zeitraum	Betriebs-Kennzahl
x [Quartale]	y [%]
1	4,80
2	4,30
3	5,30
4	6,50
5	5,90
6	5,70
7	6,80
8	8,20
9	7,00
10	7,10
11	7,90
12	9,30
13	8,70

- a) Bestimmung der Trendgeraden, damit ergeben sich die saisonbereinigten Trendwerte $\hat{y}_i = \hat{y}(x_i)$
- b) Prüfung auf Stärke und Signifikanz des Trends: $r, r^2, xF_{empirisch}$
- c) Bestimmung der **Saisonkomponenten s_i** , das sind die trendbereinigten Saisonwerte $s_i = y_i - \hat{y}_i$
Ein 4-Quartale-Zyklus ist in der Grafik zu erkennen.
Die Maxima liegen im 4., 8., 12. Quartal (Weihnachtsgeschäft?).

- d) Bestimmung der mittleren Saisonkomponenten das ist je ein Mittelwert \bar{s}_{ij} für die I., II., III., IV. Quartale

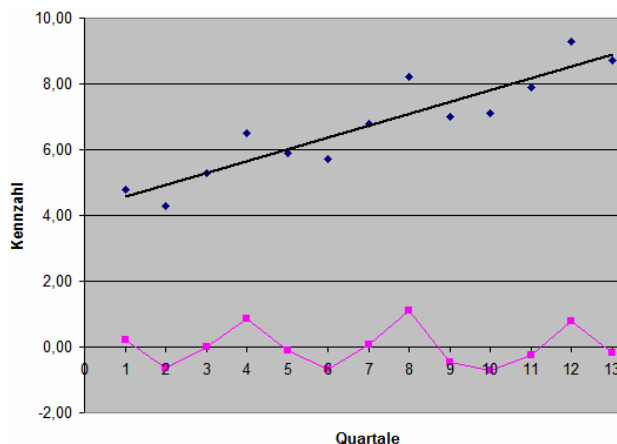
$$\bar{s}_j = \frac{1}{k} \sum_{i=1}^k s_{ij} \text{ wenn die Daten für } k \text{ Zyklen vorliegen.}$$

- e) Bestimmung des Prognosewertes $\hat{p}(x_{n+1})$ für den Zeitraum n+1 das ist einfach die Summe $\hat{p} = \hat{y}(x_{n+1}) + \bar{s}_{ij}$
d.h. die Summe aus Trendwert und Saison-Mittelwert.

- f) Die trendbereinigten Werte $s_i = y_i - \hat{y}(x_i)$ enthalten die eigentlichen Saisonkomponenten, die durch die mittleren Saisonkomponenten \bar{s}_j repräsentiert werden, und die irregulären Restwerte ir_i

$$s_i = \bar{s}_j + ir_i \Rightarrow ir_i = s_i - \bar{s}_j = y_i - \hat{y}_i - \bar{s}_j$$

→ Excel / Trend



2.3 AUFGABE TRENDAufgabe Trend

gegeben: Zwei Datenreihen: Zeitabschnitte und realisierte Daten der Zufallsvariablen Y .
Länge der Zyklen, Teile der Arbeitstabelle

gesucht / Schritte:

1. Funktionsgleichung der Trendgeraden

Arbeitstabelle ergänzen, Regressionskoeffizienten m und b berechnen.

Funktionsgleichung $\hat{y} = m x + b$ angeben.

2. Funktionswerte der Trendgeraden (Wertetabelle) \hat{y}_i .

Die Zeiten x_i in die Funktionsgleichung einsetzen.

3. Trendbereinigte Saisonkomponenten

$$s_i = y_i - \hat{y}_i$$

4. Mittlere Zykluswerte

$$\bar{s}_j = \frac{1}{k} \sum_{i=1}^k s_{ij}$$

5. Prognosewerte

$\hat{p} = \hat{y}(x_{n+z}) + \bar{s}_j$ für den z .ten Zeitabschnitt nach dem Zeitabschnitt n

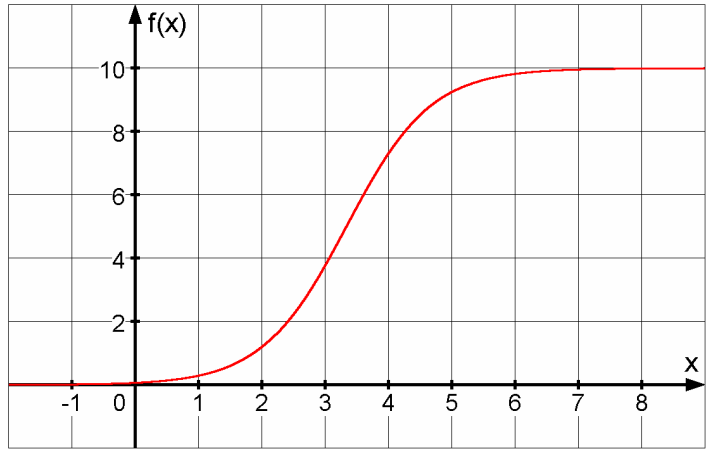
6. Irreguläre Restwerte

$$ir_i = s_i - \bar{s}_j = y_i - \hat{y}_i - \bar{s}_j$$

2.4 SÄTTIGUNGSNIVEAU

1. Bei langlebigen Konsumgütern wie Kühlschränken, Fernsehgeräte, Computer vermutet man eine Absatzentwicklung mit Sättigungsniveau S für den Bestand solcher Konsumgüter. Wachstumsprozesse mit Sättigungsniveau (saturation level) sind auch für Produktlebenszyklen typisch. Eine häufig verwendete Ansatzfunktion für solche zeitliche Entwicklungen ist die logistische Funktion mit der Gleichung

$$y = \frac{S}{1 + e^{mx+b}} = [Abb. hier] \frac{10}{1 + e^{-1,5x+5}}$$



2. Das Sättigungsniveau S schätzt man in der Regel mit Hilfe außermathematischer Überlegungen. Zum Beispiel wird man den Sättigungs-Bestand an Kühlschränken, Fernsehgeräten usw. auf $1 = 100\%$ der Haushalte schätzen: "100 % der Haushalte besitzen einen Kühlschrank". Für Markt-Prognosen müsste man eventuell nach zwei Kühlschränken pro Haushalt fragen, denn viele Haushalte besitzen mehr als *einen* Kühlschrank (in der Bar, im Wohnanhänger ...).

3. Die Ansatzfunktion $\hat{y} = \frac{S}{1 + e^{mx+b}}$ ist eine "echte" nichtlineare Funktion.

(Im Gegensatz zu den linearisierbaren $\hat{y} = a \varphi(x) + b$ in Abschnitt num4inpol, 3.20 oben) Das Logarithmieren der Funktionsgleichung führt zu einer linearen Ansatzfunktion. Dann kann man die Regressionskoeffizienten für die lineare Regression verwenden.

$$y = \frac{S}{1 + e^{mx+b}} \Rightarrow y \cdot (1 + e^{mx+b}) = S \Rightarrow \frac{S}{y} = 1 + e^{mx+b} \Rightarrow \frac{S}{y} - 1 = e^{mx+b}$$

$$\text{logarithmieren: } \ln\left(\frac{S}{y} - 1\right) = mx + b. \quad y^*_{\text{transformiert}} = \ln\left(\frac{S}{y} - 1\right) \Rightarrow y^* = mx + b.$$

4. Für die Regressionskoeffizienten für einfache lineare Regression gilt

$$m = \frac{n \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad b = \frac{1}{n} \sum y_i - \frac{m}{n} \sum x_i$$

und damit für die logistische Ansatzfunktion:

$$m = \frac{n \sum x_i \cdot y_i^* - \sum x_i \sum y_i^*}{n \sum x_i^2 - (\sum x_i)^2} \quad b = \frac{1}{n} \sum y_i^* - \frac{m}{n} \sum x_i$$

2.5 LOGISTISCHER TREND

Beispiel 2.3

Für die Ausstattung der Privathaushalte mit Mobiltelefonen rechnet man mit einer Sättigung von $100\%=S$. In der Tabelle sind für die Zeiträume x_i die Sättigungswerte $y(x_i)$ gelistet.

- a) Man bestimmt zuerst die Koeffizienten m und b und die Gleichung der Trendfunktion.
- b) Die Schätzwerte \hat{y}_i sind zu berechnen.
- c) Mit der Funktionsgleichung sollen Prognosewerte bis 2010 bestimmt werden.
- d) Das Bestimmtheitsmaß r^2 lässt sich berechnen und interpretieren.
- e) Die Prognosewerte lassen sich graphisch darstellen.

Jahr	Zeitraum x	y [%]
1990	1	1,80
1991	2	2,50
1992	3	3,50
1993	4	4,20
1994	5	8,80
1995	6	10,90
1996	7	15,00
1997	8	29,30
1998	9	42,20
1999	10	50,10
2000	11	60,90
2001	12	69,20
2002	13	74,10

→ Excel / Logist

Aufgabe Logist

gegeben: Zwei Datenreihen: n Zeitabschnitte und realisierte Daten der Zufallsvariablen Y .
Sättigungsniveau S , Zeiträume für Prognosen,
Teile der Arbeitstabelle, passende Koordinatenebene

gesucht / Schritte:

1. Herleitung der Formel für $y_{\text{transformiert}} = y^*$.
Herleitung wie in Abschnitt 2.4 Nr.3
2. Koeffizienten der logistischen Trendfunktion

Arbeitstabelle entwickeln bzw. vervollständigen, $y_{\text{transformiert}}^ = \ln\left(\frac{S}{y} - 1\right)$*

Koeffizienten m und b mit den entsprechenden Formeln berechnen

3. Gleichung der logistischen Trendfunktion $\hat{y} = \frac{S}{1 + e^{mx+b}}$
4. Funktionsgraph der Trendfunktion in die Koordinatenebene zeichnen
einige Wertepaare $(x_i | \hat{y}_i)$ bestimmen, Spalte \hat{y}_i
5. Prognosewerte für bestimmte (zukünftige) Zeiträume
 $\hat{y}(x_i)$ berechnen, wenn nicht schon in 4. geschehen
6. Bestimmtheitsmaß und seine Interpretation

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

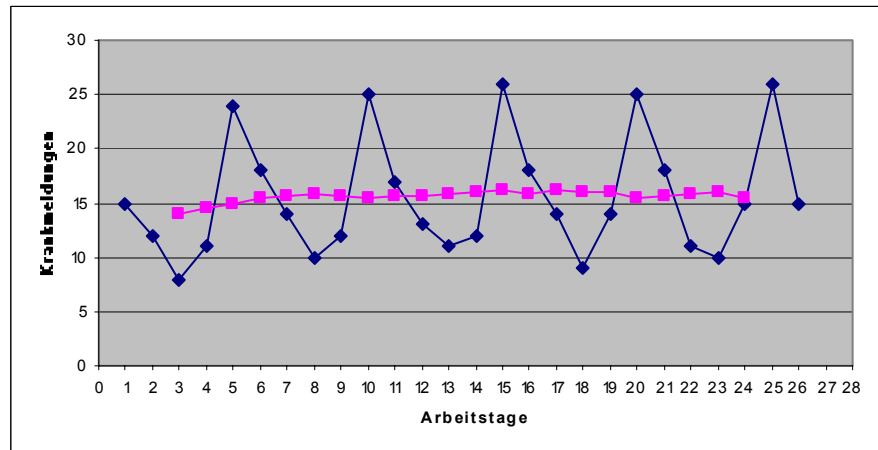
2.6 GLEITENDE MITTELWERTE

Die Methode der "gleitenden Mittelwerte" ist eine einfache Möglichkeit Zeitreihen zu glätten.
 Siehe Grafiken Abschnitt 2.7

Beispiel 2.4

Ein Dienstleistungsunternehmen notierte die Anzahl der Krankmeldungen Y an den Arbeitstagen x_i .
 Um die 5-Tage-Zyklen zu veranschaulichen, verbindet man (trotz Zeiträume!) die Punkte.

Tag	i	krank	\tilde{y}_i
Mo	1	15	
Di	2	12	
Mi	3	8	14
Do	4	11	14,6
Fr	5	24	15
Mo	6	18	15,4
Di	7	14	15,6
Mi	8	10	15,8
Do	9	12	15,6
Fr	10	25	15,4
Mo	11	17	15,6
Di	12	13	15,6
Mi	13	11	15,8
Do	14	12	16
Fr	15	26	16,2
Mo	16	18	15,8
Di	17	14	16,2
Mi	18	9	16,0
Do	19	14	16,0
Fr	20	25	15,4
Mo	21	18	15,6
Di	22	11	15,8
Mi	23	10	16,0
Do	24	15	15,4
Fr	25	26	
Mo	26	15	



Um die gleitenden Mittelwerte \tilde{y} zu erhalten, bildet man für jeweils 5 Tage einen Mittelwert und ordnet diesem dem mittleren Zeitraum zu.

$$\tilde{y}_3 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5} = \frac{15 + 12 + 8 + 11 + 24}{5} = \frac{70}{5} = 14$$

$$\begin{aligned} \tilde{y}_4 &= \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5} = \frac{12 + 8 + 11 + 24 + 18}{5} = 14,6 = \\ &= \frac{y_1 + y_2 + y_3 + y_4 + y_5 + y_6 - y_1}{5} = \frac{70 + 18 - 15}{5} = 14 + \frac{18 - 15}{5} = \end{aligned}$$

$$\tilde{y}_4 = \tilde{y}_3 + \frac{y_6 - y_1}{5} = \tilde{y}_3 + \frac{y_6 - y_1}{5} \dots$$

$$\tilde{y}_{11} = \frac{y_9 + y_{10} + y_{11} + y_{12} + y_{13}}{5} = \frac{12 + 25 + 17 + 13 + 11}{5} = \frac{78}{5} = 15,6$$

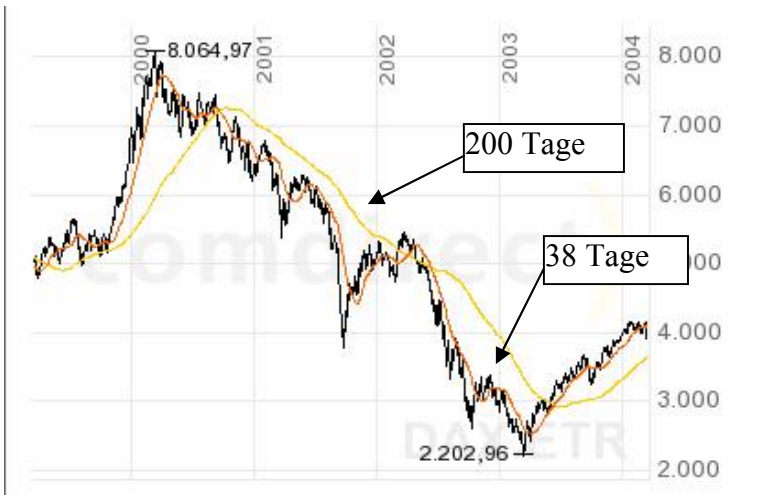
Mit den Werten der geglätteten Zeitreihe $\tilde{y}(i)$ kann man nun dieselbe Zeitreihenanalyse durchführen wie mit den durch die Methode der kleinsten Quadrate geglätteten Werten $\hat{y}(i)$.

Diese Methode hat drei Nachteile:

- a) Am Anfang und am Ende der Zeitreihe kommt es zu Informationsverlusten. Gerade die neuesten Werte am Ende der Zeitreihe fehlen; sie wären für Prognosen besonders interessant.
- b) Es gibt keine Funktionsgleichung wie $\hat{y} = m \cdot x + b$ mit der man saisonbereinigte Prognosewerte $\hat{y}(x_{n+z})$ berechnen könnte.
- c) Die gleitenden Mittelwerte $\tilde{y}(i)$ lassen sich nur dann eindeutig dem Zeitraum i zuordnen, wenn die **Zykluslänge k ungerade** ist. Bei 4-Quartale-Zyklen muss man den halben Wert der Vor-Periode und den halben Wert der Nachperiode einbeziehen:

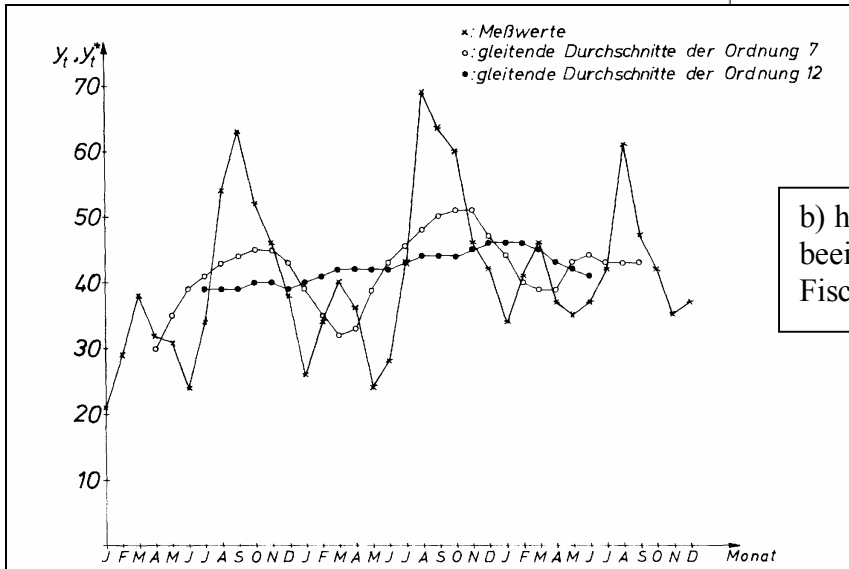
$$\tilde{y}_3 = \frac{0,5 \cdot y_1 + y_2 + y_3 + y_4 + 0,5 \cdot y_5}{4}$$

2.7 GLÄTTUNG UND ZYKLEN

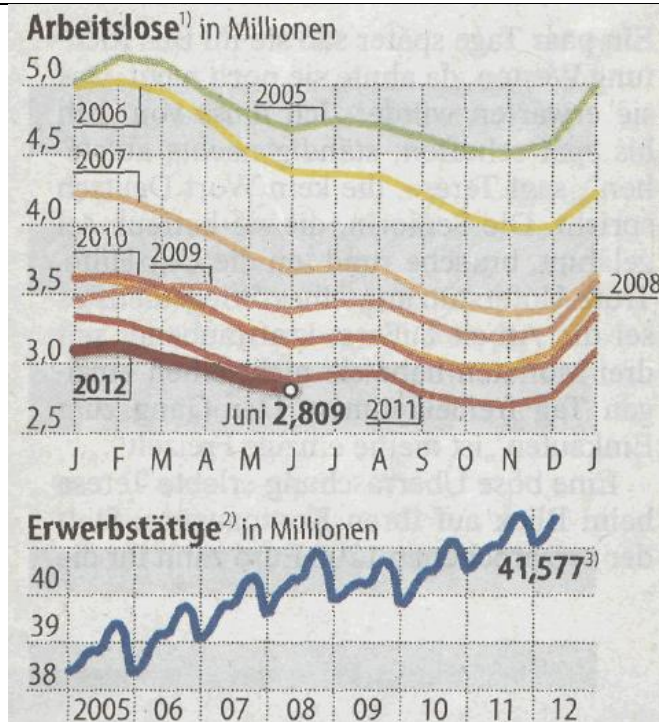


**a) DAX
über 5 Jahre**

Gleitende Mittelwerte für 38-Tage-Perioden und für 200-Tage-Perioden



b) hier geht es um jahreszeitlich beeinflusste Daten aus der Fischereiwirtschaft



c) Anzahl der Arbeitslosen in Deutschland, 12-Monate-Zyklusstruktur

2.8 INDEXZAHLEN

Mit Indexzahlen (Index, plur.: Indizes) lassen sich zeitliche Entwicklungen kompakt darstellen. Indizes fassen die durchschnittliche Veränderung einer Vielzahl von Faktoren in einer einzigen Prozentzahl zusammen.

Im Index wird ein Betrag B_1 aus der **Berichtszeit** mit dem Betrag B_0 der **Basiszeit** verglichen. Den Betrag B_0 der Basiszeit setzt man 100%.

Bei einem **Preisindex** sind die Beträge B_i Produkte aus Preisen p_i und Mengen q_i .

$$B_1 = \sum_i^n p_{1i} q_{0i} \quad B_0 = \sum_i^n p_{0i} q_{0i}$$

Die Mengenvektoren q_{0i} (bzw. q_{1i} für die Berichtszeit) nennt man auch "Warenkörbe". Zur Vereinfachung benutzt man zur Berichtszeit den **Warenkorb** q_{0i} .

Preisindex nach LASPEYRES $I_{0,1} = \frac{B_1}{B_0} = \frac{\sum p_{1i} q_{0i}}{\sum p_{0i} q_{0i}} = \frac{183}{160} = 1,144 = 114,4\%$

	Basiszeit	Basiszeit	Berichtszeit		
Produkt	Menge 0i	Preis 0i	Preis 1i		
i	q0(i)	p0(i)	p1(i)	q0(i) * p0(i)	q0(i) * p1(i)
1	5	4	5	20	25
2	10	6	7	60	70
3	8	10	11	80	88
				160	183

[LASPEYRES, Étienne, Tartu, Estland, 1871]

Tritt eine gravierende Änderung, z.B. Änderung des Warenkorbs ein, stellt man die Basiszeit um.

Beim amtlichen Verbraucherpreisindex für Deutschland (VPI) ist das **alle fünf Jahre**. Die Umbasierung führt man mit einfachen Proportionen (Dreisatzrechnungen) durch.

Beispiel 2.5 vgl. Bley Müller S. 185

Jahr i	Alter Index $I_{95,i}$ % (1995=100)	aus den Proportionen	aus den Proportionen	Neuer Index $I_{00,i}$ % (2000=100)
1995	100,0 = $I_{95,95}$			(93,5) = $I_{00,95}$
1996	101,3 = $I_{95,96}$			(94,8) = $I_{00,96}$
1997	103,2		→ 103,2 / 106,9 · 100 →	(96,5) = $I_{00,97}$
1998	104,1		→ 104,1 / 106,9 · 100 →	(97,4) = $I_{00,98}$
1999	104,9		→ 104,9 / 106,9 · 100 →	(98,1) = $I_{00,99}$
2000	106,9			100,0 = $I_{00,00}$
2001	(109,0) = $I_{95,01}$	← 106,9 · 102,0 / 100 ←		102,0
2002	(110,5)	← 106,9 · 103,4 / 100 ←		103,4
2003	(111,7)	← 106,9 · 104,5 / 100 ←		104,5

Eine Umbasierung kann erfolgen, wenn für einen Zeitraum i beide Indizes bekannt sind.

$$\frac{106,9}{100} = \frac{104,9}{I_{00,99}} \quad \text{oder} \quad 106,9 : 100 = 104,9 : I_{00,99} \Rightarrow 106,9 I_{00,99} = 100 \cdot 104,9 \Rightarrow I_{00,99} = 98,1$$

$$\frac{106,9}{100} = \frac{I_{95,01}}{102,0} \quad \text{oder} \quad 106,9 : 100 = I_{95,01} : 102,0 \Rightarrow I_{95,01} = 106,9 \cdot 102,0 / 100 \Rightarrow I_{95,01} = 109,0$$

Die fortgeführten und zurückgerechneten Werte werden eingeklammert hinzugefügt.

2.9 WACHSTUMSFAKTOREN

Statt Indizes kann man auch die Wachstumsfaktoren x_i oder Zuwachsraten r_i angeben.

Beispiel 2.7

Ein Unternehmen erzielte in den Jahren 2004 – 2008 folgende Umsätze y_i [Mio. €]:

i	Jahr	Umsatz y_i [Mio. €]	Wachstumsfaktor x_i	Zuwachsrate [%] gegenüber Vorjahresumsatz	Indizes 2004 = 100
0	2004	2,0	—	—	100,0
1	2005	2,4	1,200	20,0	120,0
2	2006	2,9	1,208	20,8	145,0
3	2007	2,7	0,931	- 6,9	135,0
4	2008	3,1	1,148	14,8	155,0

$$\text{Wachstumsfaktor } x_i = \frac{y_i}{y_{i-1}} = \frac{\text{Umsatz des aktuellen Monats}}{\text{Umsatz des Vormonats}}$$

Der Wachstumsfaktor des Jahres 2006 gegenüber 2005 ist $2,9 / 2,4 = 1,208$

Die Zuwachsrate (relativer Zuwachs) im Jahr 2006 gegenüber dem Vorjahr ist $r_{06} = 20,8\%$

$$\text{Zuwachsrate } r_i = (x_i - 1) \cdot 100$$

Der Index ist $I_{04,06} = 2,9 / 2,0 \cdot 100 = 145$.

Der Umsatz im obigen Beispiel wächst in den Jahren 2004 – 2008 um den Faktor

$$x_{\text{gesamt}} = x_1 \cdot x_2 \cdot \dots \cdot x_4 = 1,200 \cdot 1,208 \cdot 0,931 \cdot 1,148 = 1,550 \text{ (Gesamter Wachstumsfaktor)}$$

Dieser Faktor lässt sich auch berechnen mit $x_{\text{gesamt}} = y_4 / y_0 = 3,1 / 2,0 = 1,550$

Es sind **vier** Wachstumsperioden => "4. Wurzel aus ..."

Der mittlere Wachstumsfaktor ist **nicht** der Mittelwert $\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$

sondern das **geometrische Mittel** $GM(x_i) = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\frac{y_n}{y_0}}$

□ **mittlerer Wachstumsfaktor** $GM(x_i) = \sqrt[4]{\frac{y_4}{y_0}} = \sqrt[4]{\frac{3,1}{2,0}} = \sqrt[4]{1,55} = 1,116$

und die **mittlere Zuwachsrate** $r_{\text{mittel}} = (1,116 - 1) \cdot 100 = 11,6 \%$

□ Ähnliches ist aus der Zinseszinsrechnung bekannt: $K_n = K_0 \cdot q^n \Rightarrow \frac{K_n}{K_0} = q^n \Rightarrow \sqrt[n]{\frac{K_n}{K_0}} = q$

Endkapital $K_n = 60.000$ €, Anfangskapital $K_0 = 40.000$ €, Verzinsung über $n = 10$ Jahre.

Mittlerer Zinsfaktor $q = \sqrt[10]{\frac{60000}{40000}} = 1,0414$

Mittlerer Zinssatz $r = q - 1 = 0,0414 = 4,14 \%$

2.10 ZEITLICHE ENTWICKLUNGEN

Aufgabe Faktoren

gegeben: Zwei Datenreihen: n Zeitabschnitte und realisierte Daten der Zufallsvariablen Y .
Eventuell Teile der Arbeitstabelle

gesucht / Schritte:

1. Wachstumsfaktoren, Zuwachsraten

Formeln aus der Formelsammlung anwenden

2. Mittlerer Wachstumsfaktor, mittlere Zuwachsrate

$$GM(x_i) = \sqrt[n]{\frac{y_n}{y_0}}, \text{ Achtung: } n \text{ ist Anzahl der } y\text{-Werte } y_1 \text{ bis } y_n$$

3. Indizes mit einer Umbasierung

Alte und neue Indexreihe berechnen

Umbasierung mit Proportionen (oder Dreisatzrechnung) vornehmen

Fortgeführte und zurückgerechnete Werte in Klammern hinzufügen

4. Gleitende Mittelwerte

Formeln aus der Formelsammlung anwenden

beachten, dass am Anfang und Ende der Datenreihe keine Werte entstehen

Formel beachten, wenn Zykluslänge geradzahlig ist

Aus sta9loesung.xls

Faktoren 2						
Quartal	Umsatz	Wachstums- Faktoren	Zuwachs- raten	Index für die Basis	Index für die Basis	gleitende Mittelwerte
	y_i [1000 €]		[%]	II. Qu. 2007 = 100	I. Qu. 2008 = 100	$k = 4$ \bar{y}
II. Qu. 2007	17	---	---	100,0	94,4	---
III. Qu. 2007	23	1,353	35,3	135,3	127,8	---
IV. Qu. 2007	21	0,913	-8,7	123,5	116,7	20,875
I. Qu. 2008	18	0,857	-14,3	105,9	100,0	22,875
II. Qu. 2008	26	1,444	44,4	152,9	144,4	24,5
III. Qu. 2008	30	1,154	15,4	176,5	166,7	---
IV. Qu. 2008	27	0,900	-10,0	158,8	150,0	---

Gleitender Mittelwert wird dem IV.Quartal 2007 zugeordnet:

$$\frac{0,5 \cdot 17 + 23 + 21 + 18 + 0,5 \cdot 26}{4} = 20,875$$

$$\text{Mittlerer Wachstumsfaktor} = \sqrt[6]{\frac{27}{17}} = \sqrt[6]{1,588} = 1,08$$

$$\text{Mittlere Wachstumsrate} = (1,08 - 1) = 8 \%$$

2.11 GEOMETRISCHE FOLGEN

1. Filtert man eine Zeitreihe y_i mit gleitenden Mittelwerten, erfährt die Zeitreihe eine Glättung. Dabei gehen ältere y -Werte und neuere y -Werte mit gleichem Gewicht in die Rechnung ein.

$$\tilde{y}_4 = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5} \quad \text{In diesem Beispiel hat jeder Wert das Gewicht } w = \frac{1}{5} = 0,2.$$

Die neueren Werte sind sicher aussagekräftiger für die zukünftige Entwicklung und Prognosen. Es wäre also sinnvoller den neueren Werten höhere Gewichte w_i zuzumessen.

Eine höhere Gewichtung für neuere Werte könnte man für das obige Beispiel bei \tilde{y}_4 erreichen mit:

$$\tilde{y}_4 = \frac{0,2y_2 + 0,4y_3 + 0,6y_4 + 0,8y_5 + 1y_6}{3} \quad \text{oder} \quad \tilde{y}_4 = \frac{0,03y_2 + 0,07y_3 + 0,1y_4 + 0,3y_5 + 0,5y_6}{1}$$

Die zweite Variante ist einfacher anzuwenden, weil die Summe der Gewichte $\sum w_i = 1$ ist.

2. Stufungen bei technischen Maßen, bei Maßen des Alltags und bei wirtschaftlichen Prozessen benutzen Werte einer **geometrischen Folge**.

Beispiele sind DIN-Formate für Papier-Blätter $DIN A_i y_{\text{Länge Kante}} = 1,188 \cdot 0,707^i$ mit $i = 0, 1, 2, \dots$

Normzahlenreihe R5 $y = 100 \cdot 1,5849^i \Rightarrow 100, 160, 250, 400, 630, 1000, 1600, 2500, \dots$

Die harmonische Tonleiter (a' b' h' c' ...) $y = 440 \cdot 1,059^i$ (Kammerton bei $i = 0$)

Normmaße für Bauteile, Normmaße für Schrauben, Stahlprofile, Elektromotoren ...

Restwerte bei der degressiven Abschreibung, Diskontierung der Cash Flows (DCF-Methode)...

3. Eine geometrische Folge hat die Form **$w_i = \alpha q^i$** **$y = a q^n$**

Wenn die Summe der Gewichte $\sum w_i = 1$ werden soll,

setzt man $q = 1 - \alpha$ mit $0 \leq \alpha \leq 1$ Glättungsfaktor α **$\alpha(1 - \alpha)^i$**

denn es gilt: $\sum_{i=0}^{\infty} \alpha(1 - \alpha)^i = \alpha(1 - \alpha)^0 + \alpha(1 - \alpha)^1 + \alpha(1 - \alpha)^2 + \dots + \alpha(1 - \alpha)^i = 1$

□ Mit $\alpha = 0,6$ und $1 - \alpha = 0,4$ ergibt sich dann die Folge:

$$w_i = 0,6 \cdot 0,4^0 \quad 0,6 \cdot 0,4^1 \quad 0,6 \cdot 0,4^2 \quad 0,6 \cdot 0,4^3 \quad 0,6 \cdot 0,4^4$$

$$w_i = 0,6 \quad 0,24 \quad 0,096 \quad 0,0384 \quad 0,01536 \quad \sum_{i=0}^4 w_i = 0,99$$

und daraus: $\tilde{y}_4 = 0,6 \cdot y_i + 0,24 \cdot y_{i-1} + 0,096 \cdot y_{i-2} + 0,0384 \cdot y_{i-3} + 0,01536 \cdot y_{i-4}$

und allgemein für diese Summanden: $\alpha(1 - \alpha)^k \cdot y_{i-k}$

$$\frac{0,6 \cdot y_i + 0,24 \cdot y_{i-1} + 0,096 \cdot y_{i-2} + 0,0384 \cdot y_{i-3} + 0,01536 \cdot y_{i-4}}{1}$$

4. Exponentielles Glätten entsteht, wenn

jeder Beobachtungswert y_i absteigend mit dem Faktor $\alpha(1 - \alpha)^i$ gewichtet wird

Die Summe aller Gewichte konvergiert nach 1.

Die Abweichung gegenüber dem Grenzwert 1 kann man **vernachlässigen**, im obigen Beispiel

ergäbe der 10. Summand $= 0,6 \cdot 0,4^{10} y_{i-10} = 0,000063 y_{i-10}$

2.12 EXPONENTIELLES GLÄTTEN

Das Filtern einer Zeitreihe mit den Gewichten $\alpha(1-\alpha)^i$ nennt man exponentielles Glätten wegen der Exponenten i . Die Gewichtung der Beobachtungsdaten nimmt mit zunehmendem Alter der Beobachtungsdaten exponentiell ab. ("nehmen geometrisch ab")

Die Summe der so gewichteten Beobachtungswerte $\alpha(1-\alpha)^k \cdot y_{i-k}$ kann man zur Berechnung des Prognosewertes für den Zeitraum n benutzen:

$$\hat{y}_n = \alpha \sum_{i=0}^n (1-\alpha)^k \cdot y_{i-k} = \alpha \sum_{i=0}^{n-1} (1-\alpha)^k \cdot y_{i-k} + (1-\alpha)^n \cdot y_1 \approx \alpha \sum_{i=0}^{n-1} (1-\alpha)^k \cdot y_{i-k}$$

Dabei wird der älteste gewichtete Wert $(1-\alpha)^n \cdot y_1 \approx 0$ vernachlässigt.

Mit diesem Ausdruck lässt sich der Prognosewert \hat{y}_n direkt berechnen, der Prognosewert \hat{y}_{n+1} für den nächsten Zeitraum erfordert dann eine erneute Berechnung des Summenausdrucks.

Die Werte einer geometrischen Folge lassen sich **aber sehr einfach rekursiv** berechnen:

im obigen Beispiel: $0,6 \cdot 0,4^4 = 0,6 \cdot 0,4^3 \cdot 0,4$ oder $0,0384 \cdot 0,4 = 0,01536$

allgemein: $\alpha(1-\alpha)^3 \cdot (1-\alpha) = \alpha(1-\alpha)^4$,

d.h. wenn der geglättete Wert \hat{y}_i bekannt ist, lässt sich daraus der nächste \hat{y}_{i+1} berechnen.

$$\hat{y}_1 = \alpha y_1 + (1-\alpha) y_0$$

$$\hat{y}_2 = \alpha y_2 + (1-\alpha) \hat{y}_1 = \alpha y_2 + (1-\alpha)(\alpha y_1 + (1-\alpha) y_0) = \alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 y_0$$

$$\hat{y}_3 = \alpha y_3 + (1-\alpha) \hat{y}_2 = \alpha y_3 + (1-\alpha)(\alpha y_2 + \alpha(1-\alpha) y_1 + (1-\alpha)^2 y_0) =$$

$$= \alpha y_3 + \alpha(1-\alpha) y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3 y_0 = \alpha \sum_{i=0}^3 (1-\alpha)^k \cdot y_{i-k} + (1-\alpha)^3 y_0$$

$$\hat{y}_n = \alpha \sum_{i=0}^n (1-\alpha)^k \cdot y_{i-k} = \alpha \sum_{i=0}^{n-1} (1-\alpha)^k \cdot y_{i-k} + (1-\alpha)^n \cdot y_1 \approx \alpha \sum_{i=0}^{n-1} (1-\alpha)^k \cdot y_{i-k}$$

allgemein mit Rekursion: $\hat{y}_{i+1} = \alpha y_{i+1} + (1-\alpha) \cdot \hat{y}_i$ oder auch: $\hat{y}_i = \alpha y_i + (1-\alpha) \cdot \hat{y}_{i-1}$

Um den nächsten exponentiell geglätteten Wert \hat{y}_{n+1} zu berechnen benötigt man also nur den neuesten Beobachtungswert y_{n+1} und den letzten geglätteten Wert \hat{y}_n .

Den aktuell geglätteten Wert \hat{y}_{n+1} benutzt man als Prognosewert.

Um den Prognosewert \hat{y}_{n+2} zu bestimmen, wartet man üblicherweise den neuen Beobachtungswert, also hier den Wert y_{n+2} ab, und setzt diesen ein.

Als Glättungskonstanten α benutzt man Faktoren mit $0,1 \leq \alpha \leq 0,4$

Mit zunehmendem α werden die Ausdrücke $(1-\alpha)^k$ schnell kleiner.

Mit zunehmender Glättungskonstante α werden die Vergangenheitswerte niedriger gewichtet.

Bei $\alpha = 1$ ist $y_{n+1} = y_n$ und die Vergangenheitswerte werden "vergessen".

Bei $\alpha = 0$ würde der aktuellste Wert y_n überhaupt nicht berücksichtigt.

Die exponentielle Glättung benutzt man für Zeitreihen, die nicht trendbehaftet sind und für die Analyse von Residuen.

Die Residuen sind die trendbereinigten Saisonkomponenten, die sich im einfachsten Fall ergeben aus: Trendbereinigte Werte $y_i =$ Beobachtungswerte – Werte der Trendgeraden.

Es gibt auch ein zweistufiges Verfahren der exponentiellen Glättung, bei dem ein linearer Trend rechnerisch berücksichtigt wird.

→ Video: sta8expglätten.swf

2.13 THEIL'SCHES U

Als **Prüfmaß für die Güte der Prognose** benutzt man das "THEIL'sche U":

$$\text{THEIL'scher Ungleichheitskoeffizient } U = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{i-1})^2}} \quad [\text{THEIL, Henri, Chicago, USA, 1966}]$$

Dieses Prüfmaß vergleicht Schätzungen durch exponentielles Filtern mit der naiven Prognose.

Wenn $U < 1$, dann ist die Prognose signifikant besser als die autoregressive naive Prognose.

Bei einer naiven Prognose schließt man einfach von der Vergangenheit auf die Zukunft, etwa beim linearen Trend (Abschnitt 2.1) oder "morgen ist das Wetter zu 80% wie heute" ...

Wenn $U \geq 1$ lohnen sich die Aufwendungen für Prognoseverfahren nicht.

Dieses Prüfmaß wird zum Beispiel vom Sachverständigenrat der Deutschen Wirtschaft benutzt:

http://www.sachverstaendigenrat-wirtschaft.de/download/gutachten/ga06_ii.pdf

Beispiel 2.8

Über 12 Jahre wurde der Spirituosenverbrauch in einer bestimmten Stadt aufgezeichnet. Die Zeitreihe ist nicht trendbehaftet. (Die Werte y_i sind trendbereinigt.)

Das Beispiel stammt aus den ersten Veröffentlichungen über Filterungsverfahren wie exponentielles Glätten.

Man wird zunächst eine Zeitreihe mit geglätteten Werten \hat{y}_i bestimmen, dann damit eine Prognose wagen und mit dem THEIL'schen Ungleichheitskoeffizient U prüfen, ob die Prognose signifikant besser ist als das naive auto-regressive Schließen.

Zeitraum	Spir.Verbr.
t [Jahre]	y [liter]
1	6,80
2	7,90
3	6,60
4	7,20
5	7,20
6	7,60
7	8,40
8	6,80
9	7,90
10	7,90
11	8,00
12	7,70

→ Excel / Exp-Glätten

- (1) Als Startwert \hat{y}_1 benutzt man einfach y_1 .
- (2) Die Prognose ist signifikant besser als die naive Prognose, weil $0,522 < 1$
- (3) Um \hat{y}_{13} zu berechnen, benötigt man den Beobachtungswert y_{13} .

2.14 AUFGABE EXP. GLÄTTEN

Aufgabe Exp-Glätten

gegeben: Zwei Datenreihen: Zeitabschnitte und realisierte Daten der Zufallsvariablen Y .
Die y -Werte sind nicht trendbehaftet oder bereits trendbereinigt.
Glättungskonstanten α_1, α_2 , Teile der Arbeitstabelle, Dummy-Spalten

gesucht / Schritte:

- Exponentiell geglättete Werte \hat{y}_i für $i = 1, 2, \dots, n$
Startwert $\hat{y}_1 = y_1$ setzen
 $\hat{y}_{i+1} = \alpha y_{i+1} + (1 - \alpha) \cdot \hat{y}_i$ für die ganze Zeitreihe
- Eine weitere Reihe geglätteter Werte \hat{y}_i mit einer anderen Glättungskonstanten oder Funktionsgraph für die Funktion $\hat{y}_i(i)$
Startwert $\hat{y}_1 = y_1$ setzen
 $\hat{y}_{i+1} = \alpha_2 y_{i+1} + (1 - \alpha_2) \cdot \hat{y}_i$ für die ganze Zeitreihe
bzw. Funktionsgraph in die gegebene Koordinatenebene einzeichnen.
- Prognosewert für den Zeitraum $n+1$ mit Angabe der Einheit (Benennung)
 $\hat{y}_{i+1} = \alpha y_{i+1} + (1 - \alpha) \cdot \hat{y}_i$ ebenso für $i = n$
- Theil'scher Ungleichheitskoeffizient und dessen Interpretation
Die Datenreihen $(y_i - \hat{y}_i)^2$ und $(y_i - y_{i-1})^2$ bestimmen

$$U = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{i-1})^2}} \text{ berechnen}$$

$U < 1$ die durchgeführte Prognose ist signifikant besser als die naive
 $U \geq 1$ die durchgeführte Prognose ist überflüssig