Math. II, Statistik, Uebung1, Termin 04.04.2013 (AI), 08.04.2013 (BI)
1-Korrelation, 2-Multiple Korr., 3-A'-Regression, 4-Zeitreihenanalyse

ALLGEMEINES

1. Interpretation von Korrelationskoeffizient und Bestimmtheitsmaß beachten.
   "entgegengesetzt", 48,7% des Anstiegs der Konzentration ist auf die Verringerung des Druckes zurückzuführen, 51,5% auf andere Ursachen.
   mindestens: …die Veränderung der Konzentration ist auf die Veränderung des Druckes…

2. Rundungsfehler: 85,196 ≈ 86 Verkehrsunfälle ?
   55,04 ≈ 56 Untersuchungen ?

2. Aufgabe 3 "Regressionsgerade" ist falsch, es müsste Regressionsfunktion heißen
   "= 0" gehört nicht zur Ableitung
   Normalgleichungen, direkt aus der Formelsammlung

3. Aufgabe 4 enthielt viele kleine Rechnungen, stumpfsinnig, tut mir leid.
   Bei Aufgabe 4 fehlte die Punktangabe, wird hier nachgereicht

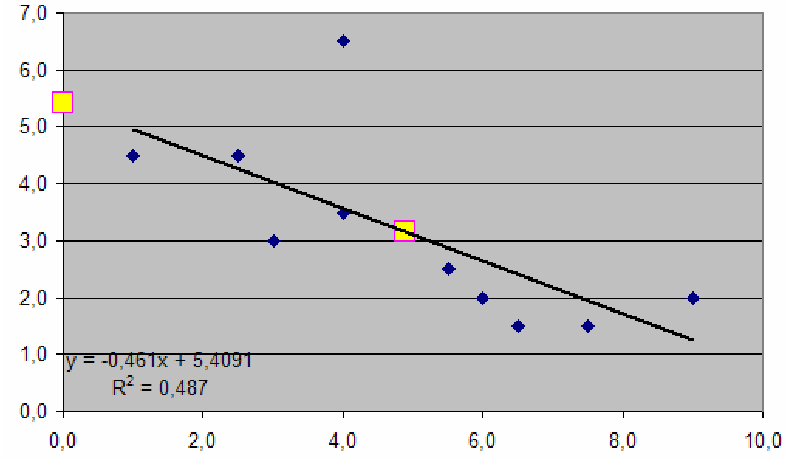
AUFGABE 1 10 PUNKTE

Druck x(i) N/cm²	Konzentration y(i) g/m³	$x_i * y_i$	x_i^2	y_i^2	$x_i^2 \cdot y_i$	$(y_i - \bar{y})^2$
1,0	4,5	4,50	1,00	20,25	4,50	1,82
2,5	4,5	11,25	6,25	20,25	28,13	1,82
3,0	3,0	9,00	9,00	9,00	27,00	0,02
4,0	3,5	14,00	16,00	12,25	56,00	0,12
4,0	6,5	26,00	16,00	42,25	104,00	11,22
5,5	2,5	13,75	30,25	6,25	75,63	0,42
6,0	2,0	12,00	36,00	4,00	72,00	1,32
6,5	1,5	9,75	42,25	2,25	63,38	2,72
7,5	1,5	11,25	56,25	2,25	84,38	2,72
9,0	2,0	18,00	81,00	4,00	162,00	1,32
49,0	31,5	129,50	294,00	122,75	677,00	23,53
4,9	3,15	= Mittelwerte		Diese Spalten sind überflüssig!		

$$m = \frac{10 \cdot 129,5 - 49 \cdot 31,5}{10 \cdot 294 - 49^2} = \frac{-248,5}{539} = -0,461. \qquad b = 31,5 - (-0,461) \cdot 4,9 = 5,409$$

$$r = \frac{-248,5}{\sqrt{(10 \cdot 294 - 49^2) \cdot (10 \cdot 122,75 - 31,5^2)}} = \frac{-248,5}{\sqrt{539 \cdot 235,25}} = \frac{-248,5}{356,09} = -0,698$$

$$F_{empirisch} = \frac{0,487}{1 - 0,487} \cdot 8 = 7,595$$



b)

b) m = -0,461. b = 5,409. $\hat{y}(x)$ = -0,461 x + 5,409

c) r = -0,698 Zwischen Druck und Konzentration besteht eine negative 69,8-prozentige Korrelation.

r^2 = 0,487 Zu 48,7% ist die Änderung bei der Konzentration auf die Veränderung des Druckes zurückzuführen. Oder umgekehrt. 51,3% der Ursachen bleiben unerklärt.

d) $xF_{empir} = 7,595 > xF_{crit} = 5,318$ der Zusammenhang bei einem Sicherheitsgrad von 95% statistisch gesichert, weil $xF_{empir} > xF_{crit}$

$xF_{empir} = 7,595 < xF_{crit} = 11,259$ der Zusammenhang bei einem Sicherheitsgrad von 99% nicht statistisch gesichert, weil $xF_{empir} < xF_{crit}$.

e) Schätzwert $\hat{y}(5) = -0,461 \cdot 5 + 5,409 = 3,104$ [g/m³]

AUFGABE 2 10 PUNKTE

Man möchte untersuchen, ob und wie in einem bestimmten Bezirk die drei Faktoren:
zeitliche Entwicklung t [Jahre],
Baustellenstrecken u [km],
Verkehrdichte v [Autos pro Stunde auf zufällig ausgewählten Strecken]
die Anzahl der Verkehrsunfälle y beeinflussen.

Das Ergebnis einer Datenanalyse liegt vor:

Zeitraum	Bau-strecken	Verkehrs-dichte	Unfälle
t	u	v	y
Jahre	km	Stück/h	Stück/Jahr
1	23	766	75
2	32	774	88
3	11	791	40
4	31	780	81
5	8	775	30
6	18	786	68
7	27	779	75
8	16	790	70
9	29	787	83
10	26	780	77
11	14	776	69

Korrelationsmatrix

	Periode	Baustr.	VerkDichte	Unfälle
Periode	1			
Baustr.	-0,09366101	1		
VerkDichte	0,34090500	-0,20491432	1	
Unfälle	0,13146524	0,88499372	-0,16529166	1

Regressions-Statistik	
Multipler Korrelatio	0,912646395
Bestimmtheitsmaß	0,832923442
Adjustiertes Bestim	0,761319202
Standardfehler	8,73969078
Beobachtungen	11

ANOVA

	Freiheitsgrade (d	dratsummen (	Quadratsumn	Prüfgröße (F)	F krit
Regression	3	2665,50645	888,502151	11,63232023	0,0041394
Residue	7	534,675365	76,3821949		
Gesamt	10	3200,18182			

	Koeffizienten	Standardfehle	t-Statistik	P-Wert	Untere 95%	Obere 95%
Schnittpunkt	135,298626	306,622356	0,44125493	0,672336225	-589,747515	860,344767
X Variable 1	1,276158436	0,88669057	1,43923763	0,19325901	-0,8205301	3,37284697
X Variable 2	1,911418796	0,33747238	5,66392654	0,000763353	1,11342398	2,70941361
X Variable 3	-0,14744813	0,39327953	-0,37491942	0,718821366	-1,07740577	0,78250951

AUSGABE: RESIDUENPLOT

Beobachtung	Schätzung für	Residuen	e²	(e(t)-e(t-1)²
1	67,5921481	7,4078519	54,8762698	
2	84,89149065	3,10850935	9,66283037	18,48434638
3	43,52123613	-3,52123613	12,3991039	43,95352514
4	84,64769994	-3,64769994	13,3057148	0,015993095
5	42,69846672	-12,6984667	161,251057	81,91637928
6	61,46688367	6,53311633	42,681609	369,8537865
7	80,97794819	-5,97794819	35,7358646	156,5267355
8	59,60657043	10,3934296	108,023378	268,02201
9	86,17351761	-3,17351761	10,071214	184,0620559
10	82,74755658	-5,74755658	33,0344066	6,625676605
11	61,67648198	7,32351802	53,6339162	170,852991
		Summen:	534,675365	1300,313499

noch Aufgabe 2

a) Formulieren Sie die Gleichung der Regressionsfunktion. [1 P.]

b) Mit Hilfe der Regressionsfunktion versucht man die Anzahl der Verkehrsunfälle für ein bestimmtes Szenario zu schätzen. Bestimmen Sie den Prognosewert für $t = 12$; $u = 25$ [km]; $v = 770$ [Autos/Std.]
Geben Sie die Einheiten an, runden Sie sinnvoll. [1,5 P.]

c) Ist das Regressionsmodell statistisch gesichert ? Prüfen Sie mit Hilfe des FISHER-Prüfmaßes bei einem Sicherheitsgrad von 95 %.
Kann man die Nullhypothese auch bei einem Sicherheitsgrad von 99% verwerfen? [2 P.]

d) Prüfen Sie mit der Student-t-Prüfgröße, welche Einflussvariablen einen signifikanten Beitrag für diesen untersuchten Zusammenhang liefern, bei $\alpha = 0{,}02$. [2 P.]

e) Prüfen Sie auf Interkorrelation zwischen den Einflussfaktoren.
Interpretieren Sie Ihr Ergebnis. [1,5 P.]

f) Prüfen Sie auf Autokorrelation der Ordnung 1 mit Hilfe des DURBIN-WATSON-Maßes. Interpretieren Sie das Ergebnis. [2 P.]

Lösung:

a) $y = 135{,}299 + 1{,}276\,t + 1{,}911\,u - 0{,}147\,v$

b) *Prognose* $\hat{y} = 84{,}86 \approx 85$ *Unfälle*

c) *Das Regressionsmodell ist statistisch gesichert, weil*
$xF_{(crit.\,|\,\alpha=0{,}05\,|\,v=7\,|\,p=3)} = 4{,}347 < xF_{empir.}$. $v = 11-3-1 = 7$. $p = 3$. $F_{empir.} = 11{,}63$
Ja, die Null-Hypothese kann auch bei einem Sicherheitsgrad von 99% verworfen werden, weil $8{,}451 < 11{,}63$ *d.h.* $xF_{empir.} > xF_{crit}$

d) $t_{crit\,|\,\alpha=0{,}02\,|\,v=7} = 2{,}517$. *Nur die "Baustellenstrecken" liefern einen signifikanten Beitrag für diesen untersuchten Zusammenhang, weil* $t_{empir.} = 5{,}66 > t_{crit.}$

e) *Interkorrelation: Die Korrelationsmatrix zeigt, dass keine Interkorrelation zwischen den Einflussgrößen besteht, da alle Interkorrelations-Koeffizienten (betragsmäßig) wesentlich kleiner als 0,5 sind.*
Der Korrelationskoeffizient $r_{Baustellen,\,Unfälle} = 0{,}885$ *zwischen u und y bestätigt das Ergebnis aus dem o.g. t-Prüfmaß.*

f) *DW1-Maß:* $DW1 = \dfrac{1300{,}31}{534{,}68} = 2{,}43$ *liegt im Intervall* $[1{,}30\ ;\ 2{,}70]$
Es liegt keine Autokorrelation der Ordnung 1 vor.

AUFGABE 3 10 PUNKTE

A'-Regression

x(i)	y(i)	$\sqrt{e^x}$	$\left(\sqrt{e^x}\right)^2$	$y_i \cdot \sqrt{e^x}$	$\hat{y}$	$(\hat{y}_i - \bar{y})^2$	$(y_i - \bar{y})^2$
1	3	1,649	2,718	4,946	2,372	3,199	1,346
5	7	12,182	148,413	85,277	7,040	8,295	8,066
4	5	7,389	54,598	36,945	4,916	0,571	0,706
3	4	4,482	20,086	17,927	3,627	0,284	0,026
2	1,8	2,718	7,389	4,893	2,846	1,728	5,570
15	20,8	28,420	233,204	149,989		14,076	15,712

a) $A=\sum_{i=1}^{n}(y_i-a\sqrt{e^x}-b)^2$

b) $\begin{cases} \dfrac{\partial A}{\partial a}=\sum 2\cdot\left(y_i-a\sqrt{e^x})-b\right)\cdot\left(-\sqrt{e^x}\right) \\ \dfrac{\partial A}{\partial b}=\sum 2\cdot\left(y_i-a\sqrt{e^x}-b\right)\cdot(-1) \end{cases}$

c) $\begin{cases} a\sum\left(\sqrt{e^x}\right)^2+b\sum\sqrt{e^x}=\sum y_i\sqrt{e^x} \\ a\sum\sqrt{e^x}+nb=\sum y_i \end{cases} \Rightarrow \begin{cases} 233{,}204a+28{,}420b=149{,}989 \\ 28{,}420a+5b=20{,}8 \end{cases}$

d) $a=\dfrac{n\sum y_i\cdot\sqrt{e^x}-\sum y_i\cdot\sum\sqrt{e^x}}{n\sum\left(\sqrt{e^x}\right)^2-\left(\sum\sqrt{e^x}\right)^2}=\dfrac{5\cdot 149{,}989-20{,}8\cdot 28{,}42}{5\cdot 233{,}204-28{,}42^2}=0{,}443$

$b=\dfrac{1}{n}\sum y_i-\dfrac{a}{n}\sum\sqrt{e^x}=20{,}8/5-0{,}443/5\cdot 28{,}42=1{,}641$

Mit Hilfe der Normalgleichungen:

233,204	28,420	149,989	
28,420	5	20,8	
1	0,122	0,643	
0	1,536	2,521	28,4202422
1	0	0,44319571	0,122
0	1	1,64085409	

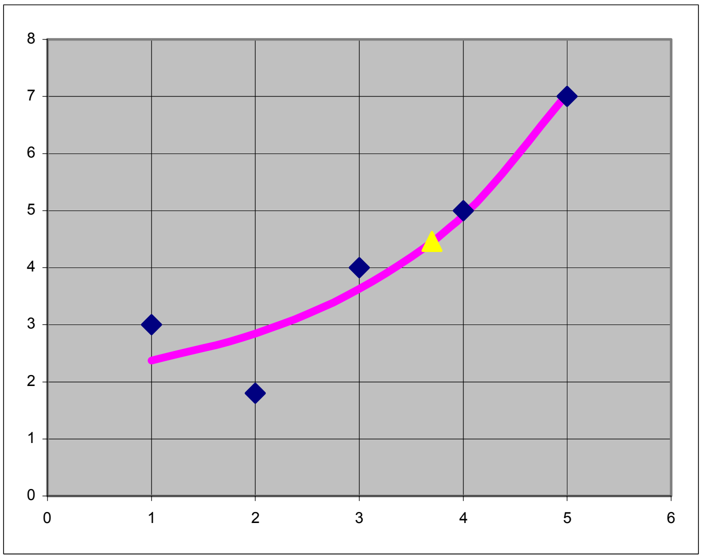
d) $\hat{y}=0{,}443\cdot\sqrt{e^x}+1{,}641$

e) $\hat{y}(3{,}7)$ =	4,459
a =	0,44319571
b =	1,64085409
$\bar{y}$ =	4,16
f) r^2 =	0,896

g)
89,6% des Anstiegs der Variable Y ist auf die Steigerung der Variablen X zurückzuführen, 10,4% des Anstiegs sind durch die Regression nicht erklärt.

$$r^2=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{14{,}076}{15{,}712}=0{,}896$$

Wertetabelle siehe oben



Punkt (3,7 | 4,459) eingetragen

AUFGABE 4 10 PUNKTE

In einer Arztpraxis wurde die Häufigkeit einer bestimmten Ultraschall-Untersuchung über 13 Quartale aufgezeichnet.

a) Bestimmen Sie die Funktionsgleichung der Trendgeraden. [2 P.]
b) Bestimmen Sie die saisonbereinigten Trendwerte. [2 P.]
c) Bestimmen Sie die trendbereinigten Saisonwerte. [2 P.]
d) Bestimmen Sie den Prognosewert für das 14.Quartal. [1 P.]
e) Bestimmen Sie die irregulären Komponenten. [2 P.]
f) Die irreguläre Komponente im 9.Quartal ist –3, interpretieren Sie dies. [1 P.]

Rechnen Sie auf zwei Nachkommastellen genau; Ergebnisse mit sinnvollen Einheiten.

Zeitraum	Anzahl der Untersuchungen				Trendwerte	Residuen s_i		irreguläre Komponente
x [Quartale]	y [Stück]	x(i) * y(i)	x(i) ²	y(i) ²	$\hat{y} = mx + b$	$y_i - \hat{y}$	Quartal	$y_i - \hat{y} - \bar{s}_j$
1	48	48	1	2304	47,429	0,57	I	3,03
2	43	86	4	1849	48,434	-5,43	II	0,02
3	53	159	9	2809	49,440	3,56	III	0,36
4	55	220	16	3025	50,445	4,55	IV	-0,98
5	49	245	25	2401	51,451	-2,45	I	0,01
6	47	282	36	2209	52,456	-5,46	II	0,00
7	58	406	49	3364	53,462	4,54	III	1,33
8	62	496	64	3844	54,467	7,53	IV	2,00
9	50	450	81	2500	55,473	-5,47	I	-3,01
10	51	510	100	2601	56,478	-5,48	II	-0,02
11	59	649	121	3481	57,484	1,52	III	-1,69
12	63	756	144	3969	58,489	4,51	IV	-1,02
13	57	741	169	3249	59,495	-2,49	I	-0,03
91	695,00	5048	819	37605			II	
							III	

Steigung m =	1,005	
y-Abschnitt b =	46,423	
Funktionsgleichung	$\hat{y}$ = 1,01 x + 46,42	

d) Mittlere Quartalswerte $\bar{s}_j$

I. Quartale	-2,462
II. Quartale	-5,46
III. Quartale	3,21
IV. Quartale	5,53

e) Prognose für das Quartal i = 14: $\hat{p}_{14} = \hat{y}(x_{14}) + \bar{s}_{II} =$ 55,04

f) Der für das 9. Quartal berechnete Wert ŷ(9) weicht um 3 Untersuchungen nach unten ab. Diese Abweichung kann nicht durch die Zeitreihenuntersuchung erklärt werden.