CAS Datenanalyse April 2017

Nullserie für Regressions- und Zeitreihenanalyse

Prüfungsdauer: 60 Minuten Total Punkte: 60

Hilfsmittel: Nur Taschenrechner!
Smartphones nicht erlaubt!

Schreiben Sie bitte leserlich!

Aufgabe 1: Terminologie (___/8P)

1. Erklären Sie, was die **Schiefe** ist. Was ist eine rechtsschiefe Verteilung? (___/4P)

____

____

____

____

2. Erklären Sie, was ein **Random Walk** ist. Ist dieser Prozess stationär? Begründen Sie Ihre Antwort. (___/4P)

____

____

____

____

____

Aufgabe 2: Bekanntes Fallbeispiel (___/22P)

Sie wollen den Einfluss der Unternehmensperformance auf das Gehalt der CEOs untersuchen. Sie haben folgende Variablen zur Verfügung:

- SALARY: jährlicher Gehalt in Tausend dollar
- MKTVAL: Börsenkapitalisierung in Mio dollar
- PROFITS: Reingewinn des Unternehmens in Mio dollar
- SALES: Umsatz des Unternehmens in Mio dollar
- CEOTEN: Firmenzugehörigkeit in Jahren im analysierten Unternehmen
- COMTEN: Firmenzugehörigkeit in Jahren
- Profmarg: Gewinnmarge

1. Die **Standardabweichung** der Variable *SALES* beträgt 6'088.7. Interpretieren Sie diese Zahl. (___/2P)

Folgendes Regressionsmodell wurde für Sie geschätzt:

Modell 1: $\ln(salary) = \beta_1 + \beta_2 \ln(sales) + \beta_3 \ln(mktval) + u$

Abhängige Variable: l_SALARY

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	4,62092	0,254408	18,16	4,95e-042	***
l_SALES	0,162128	0,0396703	4,087	6,67e-05	***
l_MKTVAL	0,106708	0,0501240	2,129	0,0347	**

Mittel d. abh. Var.	6,582848	Stdabw. d. abh. Var.	0,606059
Summe d. quad. Res.	45,30965	Stdfehler d. Regress.	0,510294
R-Quadrat	0,299114	Korrigiertes R-Quadrat	0,291057
F(2, 174)	37,12853	P-Wert(F)	3,73e-14
Log-Likelihood	-130,5594	Akaike-Kriterium	267,1188
Schwarz-Kriterium	276,6472	Hannan-Quinn-Kriterium	270,9832

2. Interpretieren Sie die drei geschätzten Regressionskoeffizienten. (___/6P)

b_1:

b_2:

b_3:____

____

Folgendes Regressionsmodell wurde für Sie geschätzt.

Modell 2: $\ln(salary) = \beta_1 + \beta_2 \ln(sales) + \beta_3 \ln(mktval) + \beta_4 profits + u$

Abhängige Variable: l_SALARY

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	4,68692	0,379729	12,34	1,65e-025	***
l_SALES	0,161368	0,0399101			
l_MKTVAL	0,0975286	0,0636886	1,531	0,1275	
PROFITS	3,56601e-05	0,000151960	0,2347	0,8147	

Mittel d. abh. Var.	6,582848	Stdabw. d. abh. Var.	0,606059
Summe d. quad. Res.	45,29524	Stdfehler d. Regress.	0,511686
R-Quadrat	0,299337	Korrigiertes R-Quadrat	0,287186
F(3, 173)	24,63629	P-Wert(F)	2,53e-13
Log-Likelihood	-130,5312	Akaike-Kriterium	269,0625
Schwarz-Kriterium	281,7671	Hannan-Quinn-Kriterium	274,2150

3. Sind die Variablen *profits* und *l_mktval* individuell signifikant auf dem 5%-Signifikanzniveau? Begründen Sie Ihre Antwort mittels **p-Wert**. (___/2P)

____

____

4. Ist die Variable *l_sales* statistisch signifikant auf dem 5%-Signifikanzniveau? Leider wurden der t-Quotient und p-Wert gelöscht. Führen Sie einen t-Test durch und wenden Sie die Faustregel ($t_c = 2$) an. (___/2P)

____

____

5. Interpretieren Sie den geschätzten Koeffizienten von *profits*. (___/2P)

____

____

6. Warum könnte es dennoch Sinn machen, beide Variablen *mktval* und *profits* in die Regression aufzunehmen? (___/6P)

____

____

7. Welches Regressionsmodell würden Sie vorziehen? Begründen Sie Ihre Antwort. (___/4P)

Zusammenstellung der Modelle:

Modell 1: ln(salary) = 4.621 + 0.162 ln(sales) + 0.107 ln(mktval)

Modell 2: ln(salary) = 4.687 + 0.161 ln(sales) + 0.0975 ln(mktval) + 0.0000357 profits

Modell 3: ln(salary) = 4.558 + 0.162 ln(sales) + 0.1018 ln(mktval) + 0.000029 profits + 0.0117ceoten

Modell 4: ln(salary) = 4.441 + 0.164 ln(sales) + 0.0984 ln(mktval) + 0.000039 profits + 0.0452ceoten -0.00121ceoten^2

Modell 5: ln(salary) = 4.621 + 0.158 ln(sales) + 0.112 ln(mktval) – 0.00226 profmarg

Modell 6: ln(salary) = 4.438 + 0.187 ln(sales) + 0.1013 ln(mktval) – 0.0026 profmarg + 0.048ceoten - 0.00114ceoten^2 – 0.008498 comten

Modell 7: ln(salary) = 4.424 + 0.186 ln(sales) + 0.1018 ln(mktval) – 0.0026 profmarg + 0.0477ceoten - 0.00112ceoten^2 – 0.006063 comten – 0.000054 comten^2

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5	Modell 6	Modell 7
# Regressor	3	4	5	6	4	7	8
adj. R^2	0.291	0.2872	0.302	0.324	0.291	0.3522	0.3486
Akaike	267.12	269.06	266.21	261.61	268.01	255.03	256.98

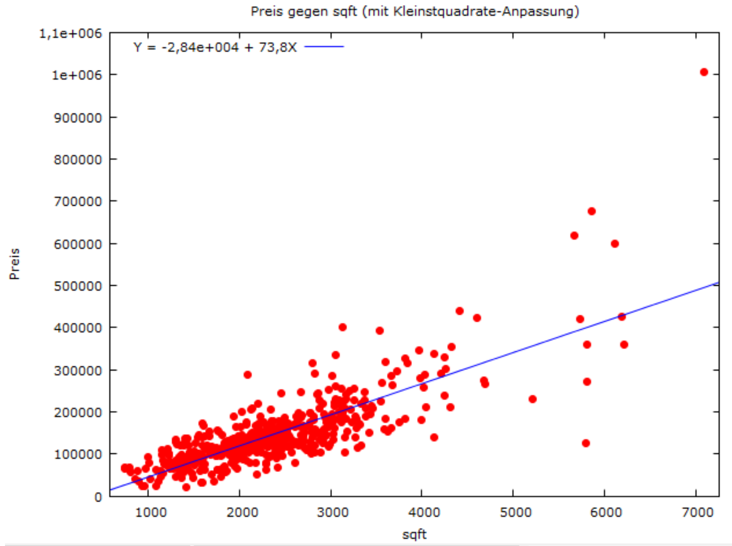
Aufgabe 3: Unbekanntes Fallbeispiel (___/18P)

Sie wollen die Bestimmungsfaktoren für den Hauspreis statistisch analysieren. Sie sammeln Daten über den Verkaufspreis, die Fläche der Häuser in Quadratfuss (square feet) und deren Alter.

Folgende Variablen stehen zur Verfügung:

- Preis: Hauspreis
- sqft: Wohnfläche in Quadratfuss
- age: Alter des Hauses

1. Betrachten Sie folgendes Streudiagramm vom Hauspreis gegen Hausfläche (SQFT) für traditionelle Häuser. Was stellen Sie fest? (___/2P)



____

____

____

Folgendes Modell wurde für Sie geschätzt: Preis = $\beta_1 + \beta_2 SQFT + \beta_3 AGE + u$

```
Modell 4: KQ, benutze die Beobachtungen 1-1080
Abhängige Variable: Preis

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
  -------------------------------------------------------------
  const      -41947,7      6989,64         -6,001      2,67e-09  ***
  sqft           90,9698      2,40310      37,86       4,26e-200 ***
  age          -755,041     140,894        -5,359      1,02e-07  ***

Mittel d. abh. Var.      154863,2   Stdabw. d. abh. Var.    122912,8
Summe d. quad. Res.      6,69e+12   Stdfehler d. Regress.   78814,86
R-Quadrat                0,589592   Korrigiertes R-Quadrat  0,588829
F(2, 1077)               773,6077   P-Wert(F)               5,2e-209
Log-Likelihood          -13707,80   Akaike-Kriterium        27421,59
Schwarz-Kriterium        27436,55   Hannan-Quinn-Kriterium  27427,26
```

2. Interpretieren Sie die Regressionskoeffizienten b_2 und b_3. (___/4P)

b_2: ____

____

b_3: ____

____

3. Erstellen Sie ein 95%-Konfidenzintervall für den Parameter b_2. Der kritische t-Wert beträgt $t_c(0.975, 1077) = 1.96$. Wie viele Beobachtungen liegen zugrunde? (___/4P)

Hinweis: Es reicht, wenn 2 Kommastellen in die Berechnung aufgenommen werden! Konfidenzintervall mit zwei Kommastellen berechnen.

Anzahl Beobachtungen =

4. Interpretieren Sie konkret Ihr 95%-Konfidenzintervall (___/2P)

____

____

5. Testen Sie folgende Vermutung: Wenn ein Haus um 1 Jahr altert, sinkt dessen Preis (P) um weniger als USD 1000.
Wie lautet Ihre Schlussfolgerung mittels t-Test?
Der kritische Wert für das 5%-Signifikanzniveau beträgt $t_c(0.95, 1077) = 1.65$

Stellen Sie die Null- und Alternativhypothese auf: (___/2P)

H_0: H_1:

Berechnen Sie den t-Wert: (___/2P)

$$t_e = \frac{b_3 - c}{se(b_3)} =$$

Schlussfolgerung: (___/2P)

____

____

Aufgabe 4: Zeitreihenanalyse (___/12P)

Das Holt-Verfahren mit α = 0.4 und γ = 0.6 wurde angewandt.

Niveaugleichung: $L_t = \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1})$

Trendgleichung: $b_t = \gamma(L_t - L_{t-1}) + (1-\gamma)b_{t-1}$

$\hat{y}_{19}(18)$: Prognosewert zum Zeitpunkt 18 für Periode 19

Sie bekommen folgende Werte:

$y_{20} - \hat{y}_{20}(19)$	$\hat{y}_{19}(18)$	b_{18}	L_{19}
24	1980	1200	700

1. Welche Gleichung beschreibt am besten den zugrunde liegenden, datenerzeugenden Prozess nach dem Holt Verfahren? (___/2P)

2. Bestimmen Sie den Wert von $\mathbf{L_{18}}$? (___/2P)

3. Bestimmen Sie den Wert von $\mathbf{b_{19}}$ (___/2P)

4. Bestimmen Sie den Wert von $\mathbf{L_{20}}$ (___/2P)

5. Schreiben Sie die Gleichung der einfachen exponentiellen Glättung in Fehlerkorrekturform.

Exponentielle Glättung: $\hat{y}_{t+1}(\mathrm{t}) = \alpha y_t + (1-\alpha)\hat{\mathrm{y}}_{\mathrm{t}}(\mathrm{t}-1)$ (___/2P)

6. Was ist im Allgemeinen die Folge, wenn die einfache exponentielle Glättung für Daten mit einem abnehmenden Trend angewendet wird? (___/2P)