

Übung 3: Rosennachfrage, Teil I

Hintergrundinformation



Die wichtigsten Herkunftsländer für Rosenimporte nach Deutschland nach Importvolumen im Jahr 2014 sind die Niederlande, Kenia, Sambia und Äthiopien. Im Jahr 2014 waren die Niederlande das wichtigste Bezugsland frischer Rosen. Deutschland importierte rund 1,06 Milliarden Stück.

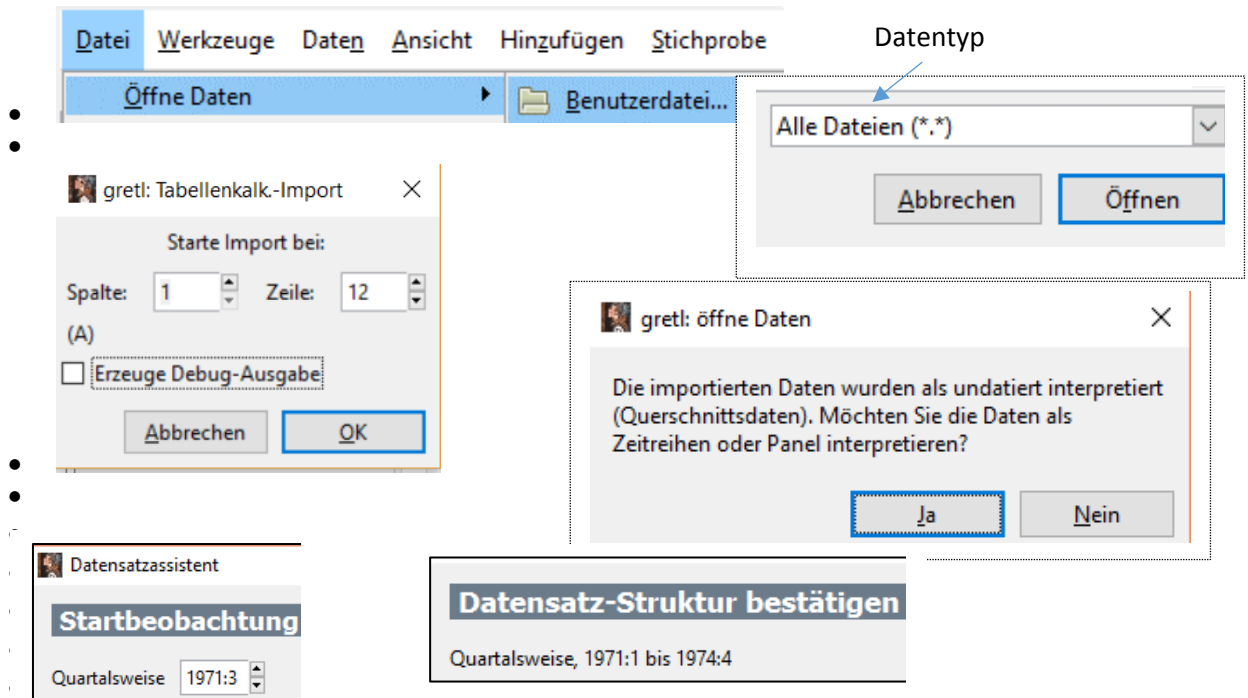
Die grössten Produzenten von Nelken sind Indonesien, Madagaskar und Tansania. Die bekannteste Nelke ist *Dianthus caryophyllus*, die Gartennelke.

Betrachten Sie den Datensatz zum Verkauf von Rosen, abgespeichert unter *Rosen.xls*. Die Daten wurden vierteljährlich vom 3. Quartal 1971 bis zum 2. Quartal 1975 in der amerikanischen Stadt Detroit erhoben. Die einzelnen Variablen bezeichnen:

- y = Absatz von Rosen (in Dutzend)
- x_2 = durchschnittlicher Grosshandelspreis für Rosen (in \$/Dutzend) (wholesale price)
- x_3 = durchschnittlicher Grosshandelspreis für Nelken (in \$/Dutzend)
- x_4 = durchschnittliches Haushaltseinkommen (in \$/Woche)
- x_5 = Zeitkomponente.

1. Erklären Sie im Allgemeinen was **Substitute** (Substitutionsgüter) sind? Nennen Sie ein Beispiel.
2. Erklären Sie warum der **Nelkenpreis** einen Einfluss auf die Rosennachfrage hat und deshalb im Modell als Regressor aufgenommen werden sollte.
3. Importieren Sie die Daten aus der Excel-Datei, Tabellenblatt „Übung 3_Rosennachfrage.xls“. Alternative: Doppelklicken Sie auf das gretl-Workfile „Übung 3_Rosennachfrage.gdt“
 - *Gretl Hinweise: Datei/ öffne Daten / Benutzerdatei*
 - *Datentypen „Alle Dateien“*
 - *Zeitreihenfrequenz: Quartalsweise*

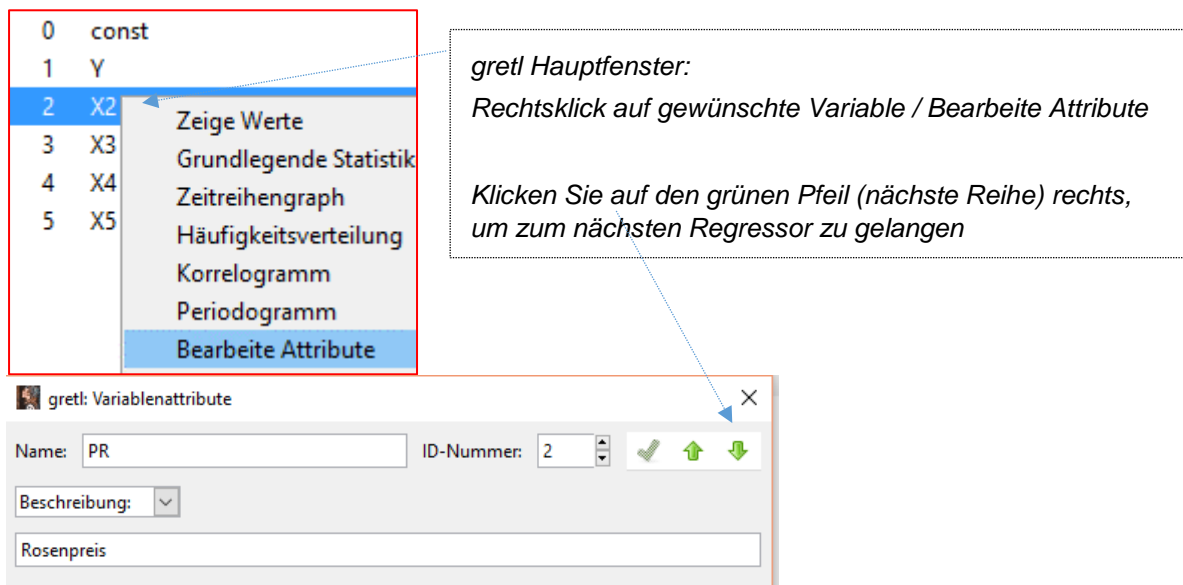
- *Startbeobachtung: 1971.3*



4. Ändern Sie die Namen der Regressoren, um deren Interpretation zu erleichtern.

gretl Hauptfenster: auf Variable recht klicken / Bearbeite Attribute

Klicken Sie auf den grünen Pfeil (nächste Reihe) rechts, um zum nächsten Regressor zu gelangen.



Variable	Benennung	Beschreibung
X2	PR	Rosenpreis
X3	PN	Nelkenpreis
X4	EINK	verfügbares Wocheneinkommen
X5	T	Zeittrend

5. Welche **Korrelationsstruktur** existiert zwischen Rosennachfrage, Rosenpreis und Nelkenpreis? Was stellen Sie fest?

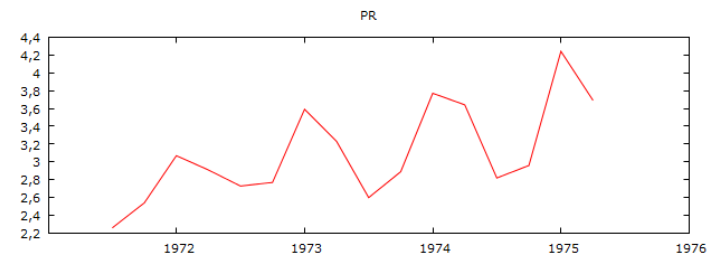
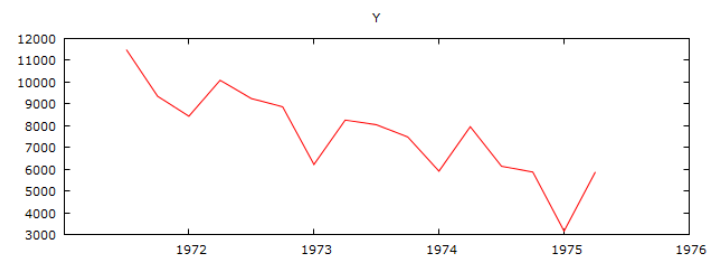
gretl Hauptfenster: Ansicht / Korrelationsmatrix

Y	PR	PN	
1,0000	-0,7842	-0,0227	Y
	1,0000	0,4725	PR
		1,0000	PN

Ansicht	Hinzufügen	Stichprobe
Symbolansicht		
Plotte spezifizierte Variablen		
Mehrfache Graphen		
Grundlegende Statistiken		
Korrelationsmatrix		

6. Betrachten Sie die **Entwicklung** des Rosenabsatzes und des Rosenpreises im Zeitverlauf. *gretl Hinweis: Ansicht / Mehrfache Graphen / Zeitreihen → zu plottende Variablen: Y, PR*

Ansicht	Hinzufügen	Stichprobe	Variable	Modell	Hilfe
Symbolansicht					
Plotte spezifizierte Variablen					
Mehrfache Graphen					
Grundlegende Statistiken					



- Was stellen Sie fest?
- Wie erklären Sie den Anstieg des Rosenpreises?

7. Erklären Sie was ein **relativer Preis** im Allgemeinen ist. Interpretieren Sie **konkret** einen Anstieg des relativen Preises PR/PN.

8. Definieren Sie folgende neue Variable:

$$\text{RelP} = \text{PR} / \text{PN}$$

gretl: Hinzufügen / Definiere neue Variable → RelP = PR / PN

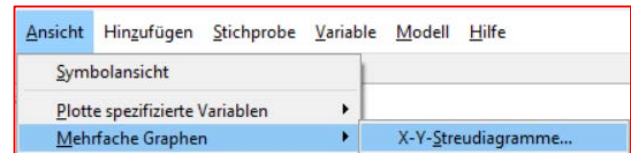
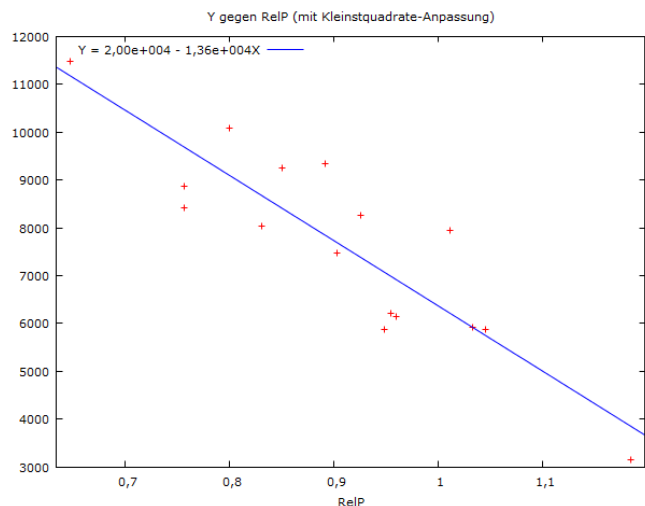
$$\text{RelP} = \text{Relativer Preis}$$

Hinzufügen	Stichprobe	Variable
Logs gewählter Variablen		
Quadrate gewählter Variablen		
Definiere neue Variable...		
Definiere Matrix...		

9. Erstellen Sie ein **Streudiagramm** des Rosenabsatzes gegen den **relativen Preis** (RelP).

gretl: Ansicht / Plotte spezifizierte Variablen / X-Y Streudiagramm / Variablen Y, RelP

Was stellen Sie fest?



Durch Diskussionen mit anderen CAS-Teilnehmern haben Sie folgende Regressionsmodelle gesammelt:

1. Modell 1: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + u_t$ $t = 1, \dots, 16$

2. Modell 2: $y_t = \beta_1 + \beta_2 (PR_t / PN_t) + u_t$ $t = 1, \dots, 16$

3. Modell 3: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + \beta_4 EINK_t + u_t$ $t = 1, \dots, 16$

4. Modell 4: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + \beta_4 EINK_t + \beta_5 T + u_t$ $t = 1, \dots, 16$

Es gelte $u_t \sim \text{iid } N(0; \sigma^2)$. iid: independent and identically distributed (unabhängig identisch verteilten Zufallsvariablen)

10. Welche Vorzeichen für die Regressionskoeffizienten erwarten Sie für das Modell 4?

Rosenpreis:

Nelkenpreis:

Einkommen:

Zeit:

11. Schätzen Sie die Regressionsmodelle 1-4.

Abhängige Variable: Y				
	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	9734,22	2888,06	3,371	0,0050 ***
PR	-3782,20	572,455	-6,607	1,70e-05 ***
PN	2815,25	947,511	2,971	0,0108 **
Mittel d. abh. Var.	7645,000	Stdabw. d. abh. Var.	2042,814	
Summe d. quad. Res.	14356623	Stdfehler d. Regress.	1050,883	
R-Quadrat	0,770648	Korrigiertes R-Quadrat	0,735363	
F(2, 13)	21,84067	P-Wert (F)	0,000070	
Log-Likelihood	-132,3601	Akaike-Kriterium	270,7202	
Schwarz-Kriterium	273,0379	Hannan-Quinn-Kriterium	270,8389	
rho	-0,113813	Durbin-Watson-Stat	2,209999	

Modell 1

Abhängige Variable: Y

	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	20002,8	1759,19	11,37	1,86e-08 ***
RelP	-13638,7	1922,35	-7,095	5,38e-06 ***
Mittel d. abh. Var.	7645,000	Stdabw. d. abh. Var.	2042,814	
Summe d. quad. Res.	13621390	Stdfehler d. Regress.	986,3855	
R-Quadrat	0,782393	Korrigiertes R-Quadrat	0,766850	
F(1, 14)	50,33624	P-Wert (F)	5,38e-06	
Log-Likelihood	-131,9395	Akaike-Kriterium	267,8790	
Schwarz-Kriterium	269,4242	Hannan-Quinn-Kriterium	267,9582	
rho	-0,197084	Durbin-Watson-Stat	2,385343	

Modell 2

	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	10816,0	5988,35	1,806	0,0983 *
PR	-2227,70	920,466	-2,420	0,0340 **
PN	1251,14	1157,02	1,081	0,3027
EINK	6,28299	30,6217	0,2052	0,8412
T	-197,400	101,561	-1,944	0,0780 *

Modell 3

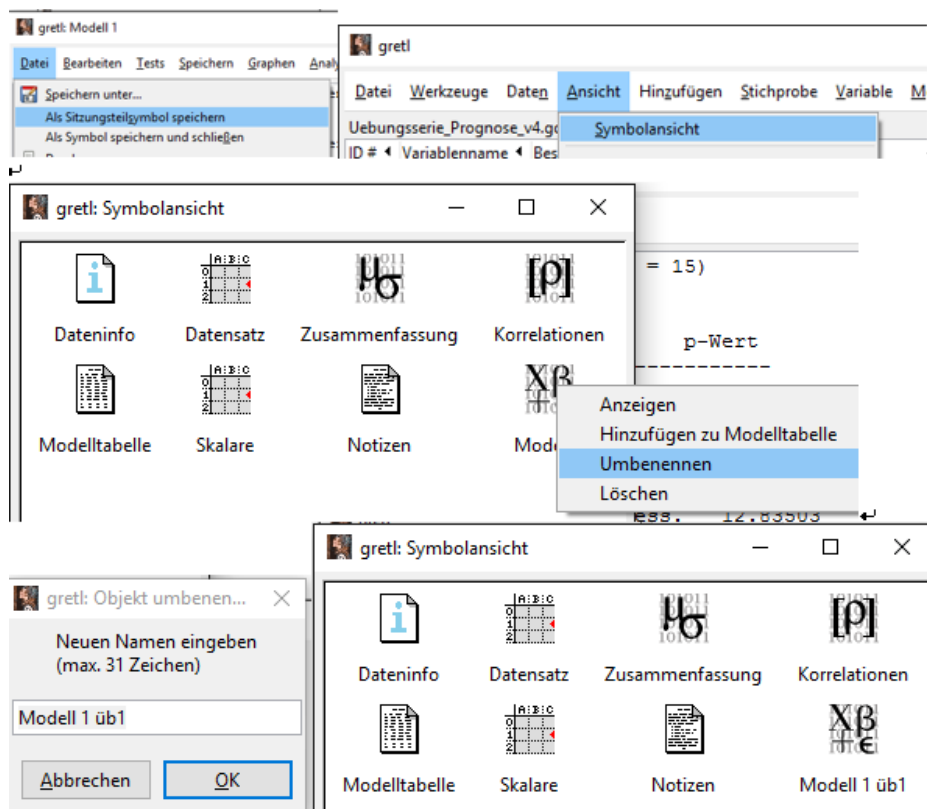
Mittel d. abh. Var.	7645,000	Stdabw. d. abh. Var.	2042,814
Summe d. quad. Res.	10347220	Stdfehler d. Regress.	969,8744
R-Quadrat	0,834699	Korrigiertes R-Quadrat	0,774590
F(4, 11)	13,88635	P-Wert (F)	0,000281
Log-Likelihood	-129,7401	Akaike-Kriterium	269,4803
Schwarz-Kriterium	273,3432	Hannan-Quinn-Kriterium	269,6781

Abhängige Variable: Y

	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	13354,6	6485,42	2,059	0,0619 *
PR	-3628,19	635,628	-5,708	9,79e-05 ***
PN	2633,75	1012,64	2,601	0,0232 **
EINK	-19,2539	30,6946	-0,6273	0,5422
Mittel d. abh. Var.	7645,000	Stdabw. d. abh. Var.	2042,814	
Summe d. quad. Res.	13900824	Stdfehler d. Regress.	1076,291	
R-Quadrat	0,777929	Korrigiertes R-Quadrat	0,722411	
F(3, 12)	14,01227	P-Wert (F)	0,000316	
Log-Likelihood	-132,1020	Akaike-Kriterium	272,2040	
Schwarz-Kriterium	275,2943	Hannan-Quinn-Kriterium	272,3622	
rho	-0,162079	Durbin-Watson-Stat	2,316836	

Modell 4

Hinweis: Speichern Sie ihre Regressionsergebnisse als Sitzungssymbol



12. Interpretieren Sie die Regressionskoeffizienten des Regressionsmodells 4 und beurteilen Sie, ob die Parameterschätzungen plausibel sind.
13. Sind die Koeffizienten des Modells 4 statistisch signifikant auf 5%-Niveau?

Hinweis: Direkt mit gretl-Output beantworten.

14. Was könnte der Grund dafür sein, dass die erklärenden Variablen Nelkenpreis (b_3) und Einkommen (b_4) nicht statistisch signifikant sind?
15. Berechnen Sie den Standardfehler des Regressionsmodells 4. Wo sehen Sie diese Zahl im gretl Output-Fenster?
16. Welches lineare Regressionsmodell würden Sie auswählen. Begründen Sie Ihre Auswahl.

Folgende Tabelle enthält eine Zusammenstellung der zur vergleichenden Kennzahlen

	Modell 1	Modell 2	Modell 3	Modell 4
# Regressoren	K = 3	K = 2	K = 4	K = 5
Regressoren	PR, PN	PR/PN	PR, PN, EINK	PR, PN, EINK, T
\bar{R}^2	0.7353	0.7668	0.7224	0.7745
Akaike	270.72	267.88	272.20	269.48
SIC	273.03	269.42	275.29	273.34

17. Erklären Sie was das Ziel eines F-Tests für eine Mehrfachregression ist.

Sie wollen jetzt das Regressionsmodell 4 mittels F-Test prüfen!

18. Stellen Sie die Nullhypothese und alternative Hypothese auf.

19. Bestimmen Sie den kritischen F-Wert (F_c) auf dem 5%-Signifikanzniveau mittels gretl.

Zähler-Freiheitsgrade	$K-1 = 5 - 1 = 4$
Nenner-Freiheitsgrade	$N - K = 16 - 5 = 11$

Normal	t	chi-Quadrat	F	binomial
Zähler-FG 2				
Nenner-FG 13				
rechtsseitige Wahrscheinlichkeit 0.05				

gretl Hauptfenster: Werkzeuge/Statistische Tabellen/ F/

Kritischer Wert $F_c(0.95, 4, 11) =$

20. Berechnen Sie den F-test mittels Bestimmtheitsmass $F = \frac{R^2}{1 - R^2} \frac{N - k}{L}$

21. Wie lautet die Entscheidungsregel, auf deren Basis Sie Ihre Testentscheidung treffen?

22. Wie lautet die Entscheidungsregel mit dem p-Wert?

23. Öffnen Sie das Varianzanalyse-Fenster im gretl. Welche Formel wurde benutzt, um den F-Wert zu berechnen?

gretl Output-Fenster: Analyse / ANOVA

	Quadratsumme	FG	quad. Mittel
Regression	5,22491e+007	4	1,30623e+007
Residuum	1,03472e+007	11	940656
Total	6,25964e+007	15	4,17309e+006

$R^2 = 5,22491e+007 / 6,25964e+007 = 0,834699$

$F(4, 11) = 1,30623e+007 / 940656 = 13,8864$ [p-Wert 0,0003]

Analyse	LaTeX
Zeige tatsächliche, gesch. Prognosen...	
Konfidenzintervalle für K...	
Konfidenzellipse...	
Kovarianzmatrix der Koef...	
Kollinearität	
Einflussreiche Beobachtu...	
ANOVA	

24. Schätzen Sie das restringierte Modell (Nullhypothesenmodell) um RSS_r zu bestimmen.

Hinweis: Das restringierte Modell stellt das Modell mit den Restriktionen $b_2 = b_3 = b_4 = b_5 = 0$ dar.

	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	7645,00	510,704	14,97	2,00e-010 ***
Mittel d. abh. Var.	7645,000	Stdabw. d. abh. Var.		2042,814
Summe d. quad. Res.	62596356	Stdfehler d. Regress.		2042,814
R-Quadrat	0,000000	Korrigiertes R-Quadrat		0,000000

25. Berechnen Sie den F-Test mittels Formel: $F = \frac{(RSS_r - RSS)}{RSS} \frac{(N - K)}{L} \approx F_{(L, N-K)}$

26. Erklären Sie die Intuition hinter dieser Formel

27. Testen Sie die Nullhypothese $H_0: b_3 = b_4 = 0$ im Modell 4. Benutzen Sie dazu den eingebauten gretl -Test „Weglassen der Variablen“. Was ist Ihre Schlussfolgerung?

gretl Output-Fenster: Test/ Variablen weglassen/

28. Interpretieren Sie konkret folgende Restriktion im Modell 4:
 $\beta_2 = -\beta_3$

Tests	Speichern	Graphen	Analyse
Variablen weglassen			
Variablen hinzufügen			

29. Schreiben Sie diese **Restriktion** in **Matrixform**.

30. Testen Sie anhand des **t-Tests** auf dem 5%-Signifikanzniveau, ob die Restriktion falsch ist.

31. Stellen Sie das **restringierte Modell** auf und schätzen Sie es.

32. Testen Sie die Restriktion anhand des t-Tests

33. Testen Sie anhand des F-Tests auf dem 5%-Signifikanzniveau, ob die Restriktion falsch ist.

Berechnen Sie den F-Wert mittels
$$F = \frac{(RSS_r - RSS)}{RSS} \frac{(N - K)}{L}$$

34. Testen Sie diese Restriktion mittels gretl.

gretl output-Fenster: Test / lineare Restriktionen / $b[2] + b[3] = 0$

35. Testen Sie im **Regressionsmodell 4**, ob die Variablen PN, EINK und T **gemeinsam** statistisch signifikant sind.

gretl: Tests / Variablen weglassen → PN, EINK und T auswählen

