## Übung 3: Rosennachfrage, Teil I

1. Erklären Sie im Allgemeinen was Substitute (Substitutionsgüter) sind? Nennen Sie ein Beispiel.

   Als Substitutionsgüter / Substitute bezeichnet man Güter, die dieselben oder ähnliche Bedürfnisse stillen und daher vom Konsumenten als gleichwertiges Ersatzgut angesehen werden. Beispiel: Butter und Margarine (Kunstbutter).

2. Erklären Sie warum der Nelkenpreis einen Einfluss auf die Rosen-Nachfrage hat und deshalb im Regressionsmodell als Regressor aufgenommen werden sollte.

   Wenn der Nelkenpreis relativ zu den Rosen teurer wird, werden die Konsumenten Nelken durch Rosen substituieren, sodass die Rosennachfrage steigen wird, ceteris paribus. Zwischen dem Preis eines Substitutionsgutes (Nelken) und der Rosennachfrage besteht also ein positiver Zusammenhang → positive Kreuzpreiselastizität.

3. Importieren Sie die Daten aus der Excel-Datei „Übung 3_Rosennachfrage.xls".

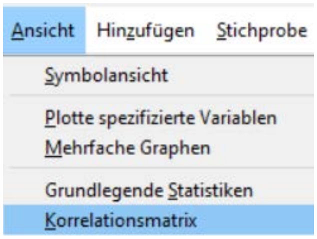
4. Ändern Sie die Namen der Regressoren, um deren Interpretation der Regressionsmodelle zu erleichtern.

| 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
```




   Die Korrelation zwischen Rosenabsatz (Y) und Rosenpreis (PR) beträgt -0.78. Wenn der Rosenpreis steigt, werden weniger Rosen verkauft.

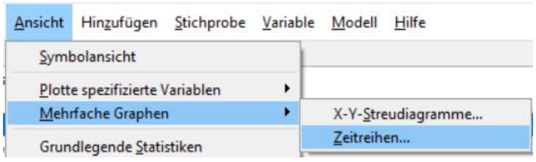
   Die Korrelation zwischen Rosenabsatz (Y) und Nelkenpreis (PN) beträgt -0.022 was darauf hindeutet, dass Nelken kein gutes Substitut für Rosen sind. Wenn der Nelkenpreis steigt, werden die Konsumenten kaum mehr Rosen kaufen.

   Die Korrelation zwischen Rosenpreis (PR) und Nelkenpreis (PN) ist positiv und beträgt 47.2%. Tendenziell steigen die Preise gleichzeitig für beide Blumen.
   Die Inflation könnte diese Tatsache erklären.

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

i. Was stellen Sie fest?

ii. Wie erklären Sie den Anstieg des Rosenpreises?


Die Nachfrage nach Rosen sinkt jährlich bis zum Tiefstpunkt Anfang 1975.

Der Rosenpreis steigt tendenziell über die Zeit.

Gut zu erkennen ist auch der saisonale Effekt: Hohe Preise und niedriger Absatz im Winter, niedrige Preise und hoher Absatz im Sommer.

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

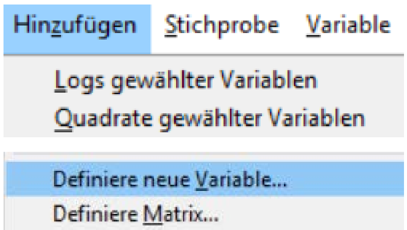
Der relative Preis ist der Preis eines Gutes, ausgedrückt in dem Preis eines anderen Gutes = Verhältnis zweier Preise zueinander. Die relativen Preise beschreiben das Austauschverhältnis von Gütern untereinander.
Wenn der relative Preis PR/PN steigt, werden Rosen relativ zu Nelken teurer.

8. Definieren Sie folgende neue Variable:

RelP= *PR / PN*

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

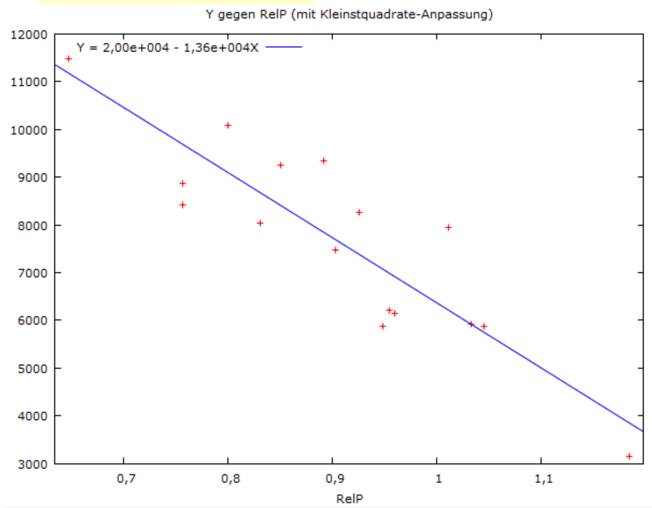
***RelP*** = ***Rel****ativer* ***P****reis*

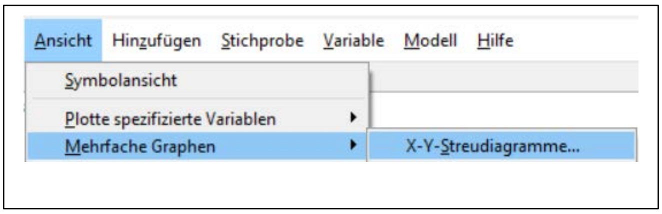



9. Erstellen Sie ein Streudiagramm des Rosenabsatzes gegen den relativen Preis (RelP).

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

Was stellen Sie fest?




Steigt der relative Preis PR/PN d.h. die Rosen werden relativ zu Nelken teurer, sinkt die Nachfrage nach Rosen.

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

1. Modell 1: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + u_t$ $\quad t = 1,\ldots,16$
2. Modell 2: $y_t = \beta_1 + \beta_2 (PR_t / PN_t) + u_t$ $\quad t = 1,\ldots,16$
3. Modell 3: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + \beta_4 EINK_t + u_t$ $\quad t = 1,\ldots,16$
4. Modell 4: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + \beta_4 EINK_t + \beta_5 T + u_t$ $\quad t = 1,\ldots,16$

Es gelte $u_t \sim$ iid $N(0;\sigma^2)$ . iid: **i**ndependent and **i**dentically **d**istributed (unabhängig identisch verteilten Zufallsvariablen)

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

    Positiv für Rosenpreis: Je höher der Rosenpreis, desto kleiner der Rosenabsatz, ceteris paribus (c.p.)

    Positiv für Nelkenpreis: Je höher der Nelkenpreis, desto höher die Rosennachfrage (Substitutionsgut), c.p.

    Positiv für Einkommen: Je höher das Einkommen, desto höher die Nachfrage nach Rosen, c.p.

    Negativ für Zeit: Im Zeitablauf sollte die Nachfrage nach Rosen sinken, wie in der Graphik aus Aufgabe 6 zu erkennen ist.

11. Schätzen Sie für die Regressionsmodelle 1-4.

Modell 1

```
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 2

```
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
```

```
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 3

```
Abhängige Variable: Y

              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  *

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
rho                    -0,230532   Durbin-Watson-Stat        2,333986
```

Modell 4

12. Interpretieren Sie die Regressionskoeffizienten des Regressionsmodells 4 und beurteilen Sie, ob die Parameterschätzungen plausibel sind.

    $b_2$ = –2227.7: Eine Erhöhung des Rosenpreises um $1/Dutzend lässt einen Rückgang des Rosenabsatzes um 2'227.7 x 12 = 26'732 Rosen erwarten, ceteris paribus.

    Das Ergebnis ist sinnvoll → bei steigendem Preis fällt die Nachfrage, ceteris paribus.

    $b_3$ = 1251.14: Eine Erhöhung des Nelkenpreises um $1/Dutzend lässt einen Anstieg des Rosenabsatzes um ca. 1251.14 x 12 = 15'000 Rosen erwarten, ceteris paribus.

    Das Ergebnis ist sinnvoll: Werden die Nelken relativ zu Rosen teurer, werden die Konsumenten substituieren und mehr Rosen kaufen, ceteris paribus.

    $b_4$ = 6.28: Eine Erhöhung des Einkommens um $1 lässt einen Anstieg des Rosenabsatzes um 6 Dutzend Rosen erwarten, ceteris paribus.

    Das Ergebnis ist sinnvoll: Steigt das verfügbares Einkommen, steigt die Nachfrage nach Rosen.

    $b_5$ = –197.4. Mit einem zusätzlichen Quartal reduziert sich der Rosenabsatz im Durchschnitt um ca. 197.4 x 12 = 2369 Rosen, ceteris paribus.

13. Sind die Koeffizienten des Modells 4 statistisch signifikant auf dem 5%-Niveau?
    *Hinweis: Direkt mit gretl-Output beantworten.*

    Nur die Variable PR ist statistisch signifikant auf dem 5%-Signifikanzniveau.

14. Was könnte der Grund dafür sein, dass die erklärenden Variablen Nelkenpreis und Einkommen nicht statistisch signifikant sind?

Gründe für die Nicht-Signifikanz des Nelkenpreises ($b_3$)

a. Die Stichprobe ist relativ klein mit nur 16 Quartalsbeobachtungen, um die statistische Beziehung zwischen Nelkenpreis und Rosennachfrage aufzudecken.
b. Es gibt viele Substitutionsgüter für die Rosen. Nelken sind nur eine Möglichkeit und deshalb ist die Reaktion der Rosennachfrage auf eine Preisänderung der Nelken statistisch kaum nachweisbar.
c. Nelken stellen für bestimmte Konsumenten kein gutes Substitutionsgut für Rosen dar (wer Rosen mag, mag nicht unbedingt auch Nelken).
d. Es besteht zwar ein Zusammenhang, dieser ist aber nicht linear und wird daher durch das Modell schlecht abgebildet.
e. Weil Rosenpreis und Nelkenpreis in der Stichprobe stark korrelieren (vgl. Aufgabe 5) enthalten die Daten zu wenig Information über die Wirkung einer unabhängigen Veränderung des Nelkenpreises (die Variable Rosenpreis kann nicht kontrolliert werden).

Grund für die Nicht-Signifikanz von Einkommen ($b_4$)

Der Anteil der wöchentlichen Ausgaben für Rosenkauf am wöchentlichen verfügbaren Einkommen ist so klein, dass keine statistische Signifikanz zwischen Einkommen und Rosenabsatz vorhanden ist.

15. Berechnen Sie den Standardfehler des Regressionsmodells 4. Wo sehen Sie diese Zahl im gretl-Output?

```
Abhängige Variable: Y

            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  *

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
```

$$s_e^2 = \frac{10'347'220}{16-5} = 940'656$$

$$s_e = \sqrt{s_e^2} = \sqrt{940'656} = 969.87$$

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 |

Modell 4 weist das höchste adjustierte Bestimmtheitsmass auf. Problematisch bei diesem Modell ist, dass nur der Rosenpreis (PR) statistisch signifikant auf dem 5%-Signifikanzniveau ist.
Anhand des Informationskriteriums sollte das Modell 2 vorgezogen werden. Der Vorteil ist, dass der relative Preis statistisch signifikant ist. Für Modell 2 spricht auch, dass es deutlich einfacher ist. Der Wert von adjustiertem $R^2$ ist zudem nur geringfügig kleiner.

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

Es wird geprüft, ob alle Koeffizienten der Regression gemeinsam gleich null sind.
Es wird geprüft, ob das aufgestellte Modell überhaupt sinnvoll ist.

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

Regressionsmodell 4: Modell 4: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + \beta_4 EINK_t + \beta_5 T_t\ \ u_t$

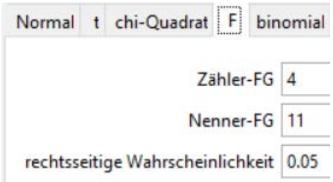
18. Stellen Sie die Nullhypothese und alternative Hypothese auf.

Nullhypothese $H_0$: $\beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

$H_1$: mindestens ein Koeffizient hat einen Einfluss ⇔ $H_1$: $H_0$ ist falsch

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 |




*gretl Hauptfenster: Werkzeuge/Statistische Tabellen/F/*

*rechtsseitige Wahrsch. = 0.05*

Kritischer Wert $F_c(0.95,4,11) = 3.35$

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

$$F = \frac{R^2}{1-R^2}\frac{N-K}{K-1} = \frac{0.834699}{1-0.834699}\cdot\frac{11}{4} = 13.886$$

*Hinweis: Diese Berechnung dient nur dem Umgang mit dieser Formel und ist nicht notwendig, da der F-Wert von gretl automatisch ermittelt wird.*

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

Wenn $F_e > F_c$ → verwerfe $H_0$

$F_e = 13.88 > 3.35$ → verwerfe $H_0$ → mindestens ein Regressionskoeffizient ist von null verschieden.

*Hinweis: Diese Frage dient nur dem Verständnis des Hypothesentests und ist nicht notwendig, da der p-Wert von gretl automatisch ermittelt wird.*

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

Wenn p-Wert < $\alpha$-Signifikanzniveau → verwerfe $H_0$

p-Wert = 0.00028 < 0.05 → verwerfe $H_0$

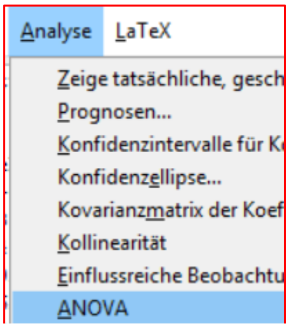
23. Öffnen Sie das Varianzanalyse-Fenster im gretl. Welche Formel wurde benutzt, um den F-Wert zu berechnen?

*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]
```




$$F_e = \frac{\boxed{ESS/(K-1)}}{\boxed{RSS/(N-K)}}$$

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

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: $F = \frac{(RSS_r - RSS)}{RSS}\frac{(N-K)}{L} \approx F_{(L,N-K)}$

L = Anzahl Restriktionen = 2  und $RSS_r$ = 62‘596‘356

N – K = 16- 5 = 11

$$F = \frac{(RSS_r - RSS)}{RSS}\frac{(N-K)}{K-1} = 5.04958 \cdot 2.75 = 13.886$$

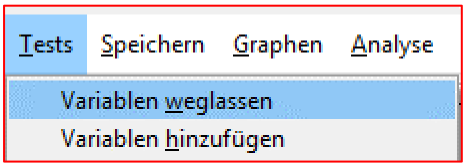
*Hinweis: Diese Berechnung dient nur dem Umgang mit dieser Formel und ist nicht notwendig, da der F-Wert von gretl automatisch ermittelt wird.*

26. Erklären Sie die Intuition hinter dieser Formel

Stehen die Restriktionen der Nullhypothese ($b_2 = b_3 = b_4 = b_5 = 0$) im Einklang mit den beobachteten Daten, dann wird das zusätzliche Residuenquadrat ($RSS_r$ – RSS) klein ausfallen → das Residuenquadrat des restringierten Modells ($RSS_r$) weicht nicht gross vom Residuenquadrat der unrestringierten Modells (RSS) ab! Entsprechend wird F einen kleinen Wert besitzen.
Besteht hingegen ein Widerspruch zwischen den Restriktionen der Nullhypothese und den beobachteten Daten, dann wird $RSS_r$ weitaus grösser sein als RSS und damit wird auch F einen grossen Wert annehmen.

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 / PN und EINK weglassen*

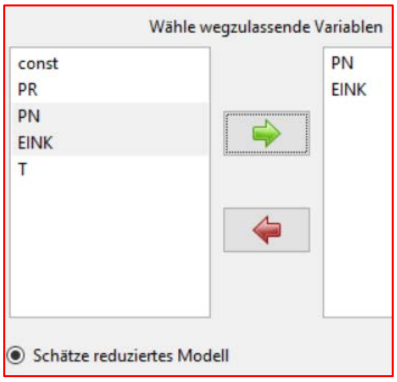


```
 Nullhypothese: Die Regressionskoeffizienten sind Null für die Variablen
   PN, EINK
 Teststatistik: F(2, 11) = 0,621326, p-Wert 0,555066
 Das Weglassen von Variablen verbesserte 3 von 3 Informationskriterien.

Modell 7: KQ, benutze die Beobachtungen 1971:3-1975:2 (T = 16)
Abhängige Variable: Y

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
 ------------------------------------------------------------
 const         14494,1       1559,88         9,292     4,19e-07 ***
 PR           -1509,26        596,726       -2,529     0,0252   **
 T             -254,118        67,4074      -3,770     0,0023   ***

Mittel d. abh. Var.      7645,000   Stdabw. d. abh. Var.     2042,814
Summe d. quad. Res.      11516128   Stdfehler d. Regress.    941,1992
R-Quadrat                0,816026   Korrigiertes R-Quadrat   0,787722
F(2, 13)                 28,83100   P-Wert(F)                0,000017
```




L = Anzahl Restriktionen = 2 Nenner-Freiheitsgrade: N-K = 16- 5 = 11

$F_c$(L,N-K) $F_c$(2,11) = 0.621 < $F_c$ → $H_0$ wird nicht verworfen

p-Wert = 0.55 > 0.05 → $H_0$ wird nicht verworfen

Die Nullhypothese $H_0$: $b_3 = b_4 = 0$ kann nicht verworfen werden.

Konklusion: Beide Variablen sind einzeln nicht statistisch signifikant **und** gemeinsam nicht statistisch signifikant!

*Hinweis: In der Praxis wird dieser Test mittels gretl durchgeführt und alle anderen Berechnungen erübrigen sich!*

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

Modell 4: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + \beta_4 EINK_t + \beta_5 T_t + u_t$

Das Steigen des Rosenpreises um $1 hat die gleiche Wirkung auf den Rosenabsatz, wie das Sinken des Nelkenpreises um $1, ceteris paribus.

Der Koeffizient $\beta_2$ kann somit interpretiert werden als die Veränderung des Rosenabsatzes (in Dutzend) pro zusätzlichen Dollar Preisdifferenz zwischen Rosen und Nelken.

29. Schreiben Sie diese Restriktion in Matrixform.

$$\mathbf{r}'\boldsymbol{\beta} = 0 \Leftrightarrow \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = 0$$

$$\boxed{\begin{array}{l} \beta_2 = -\beta_3 \\ \leftrightarrow \beta_2 + \beta_3 = 0 \end{array}}$$

*Hinweis: Diese Frage ist nicht prüfungsrelevant und illustriert nur die Matrixdarstellung einer Restriktion.*

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

Restriktion: $\beta_2 = -\beta_3 \leftrightarrow \beta_2 + \beta_3 = 0$

Zweiseitiger Test: $H_0$: $\beta_2 + \beta_3 = 0$ $H_1$: $\beta_2 + \beta_3 \neq 0$

Um die Varianz von ($b_2$+$b_3$) zu berechnen, muss die Kovarianzmatrix herangezogen werden.

→ gretl: *Output-Fenster: Analyse / Kovarianz-Matrix der Koeffizienten*

| const | PR | PN | EINK | T | |
|---|---|---|---|---|---|
| 3,58603e+07 | 91554,5 | -2,37314e+06 | -161017 | 132647 | const |
| | 847257 | -788535 | 3347,27 | -73179 | PR |
| | | 1,3387e+06 | -2134,27 | 72245,3 | PN |
| | | | 937,686 | -1334,37 | EINK |
| | | | | 10314,7 | T |

$var(b_2 + b_3) = var(b_2) + var(b_3) + 2cov(b_2, b_3) = 847'257 + 1.3387 \times 10^6 - 2(788'535) = 608'887$

Standardfehler : $se(b_2 + b_3) = (608'887)^{0.5} = 780.3121$

t-Quotient: $t_e = \dfrac{b_2 + b_3}{se(b_2 + b_3)} = \dfrac{-2'227.7 + 1'251.14}{780.3121} = -1.251$

Freiheitsgrade = N- K = 16- 5 = 11 (trotz Restriktion gibt es vier erklärende Variablen + Interzept)

Kritischer Wert: $t_c(0.975, 11) = 2.20$

$t_e = -1.251 < 2.20 \rightarrow H_0$ kann nicht verworfen werden!

Konklusion: Die Daten aus der Stichprobe sind mit der Restriktion $\beta_2 = -\beta_3$ kompatibel.

*Hinweis: Folgender Abschnitt ist nicht prüfungsrelevant und illustriert nur die Berechnung in Matrix-notation. Er kann ohne weiteres übersprungen werden. Diese Berechnung sollte nicht manuell vorgenommen werden und dient nur pädagogischen Zwecken.*

Regressionsmodell in Matrixform: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ wobei $\mathbf{X}$ die Matrix der Beobachtungen darstellt.

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{15} \\ y_{16} \end{pmatrix} = \begin{pmatrix} 1 & 2.26 & 3.49 & 158 & 1 \\ 1 & 2.54 & 2.85 & 173 & 2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 4.24 & 3.58 & 176 & 15 \\ 1 & 3.69 & 3.53 & 188 & 16 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_{15} \\ u_{16} \end{pmatrix}$$

Möglich wäre die Berechnung mittels: $t = \dfrac{\mathbf{r}'\mathbf{b}}{s_e\sqrt{\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}}$

$$\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 38.12 & 0.097 & -2.523 & -0.171 & 0.141 \\ 0.097 & 0.9 & -0.838 & 0.003 & -0.077 \\ -2.522 & -0.838 & 1.423 & -0.002 & 0.076 \\ -0.171 & 0.003 & -0.002 & 0.001 & -0.001 \\ 0.141 & -0.077 & 0.076 & -0.001 & 0.011 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = 0.6473$$

$\mathbf{r}'\mathbf{b} = -976.56$

$$t = \frac{\mathbf{r}'\mathbf{b}}{s_e\sqrt{\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}}} = \frac{-976.56}{969.87\sqrt{0.6473}} = -1.251$$

*Hinweis: Die Matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ ist mit den gretl-Menus nicht ersichtlich.*

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

Regressionsmodell: $y_t = \beta_1 + \beta_2 PR_t + \beta_3 PN_t + \beta_4 EINK_t + \beta_5 T_t + u_t$

Restringiertes Modell: $y = \beta_1 + \beta_2 [PR_t - PN_t] + \beta_4 EINK_t + \beta_5 T_t + u^*$

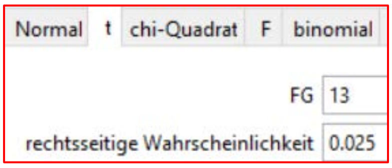
Definieren Sie die neue Variable: Diff = PR – PN

Restringiertes Modell: $y = \beta_1 + \beta_2$ Diff $+ \beta_3 EINK_t + \beta_4 T_t + u^*$

| | Koeffizient | Std.-fehler | t-Quotient | p-Wert | |
|---|---|---|---|---|---|
| const | 7156,72 | 5347,86 | 1,338 | 0,2056 | |
| Diff | -2133,52 | 938,780 | -2,273 | 0,0422 | ** |
| EINK | 8,22845 | 31,2954 | 0,2629 | 0,7971 | |
| T | -198,897 | 103,923 | -1,914 | 0,0798 | * |

| | | | |
|---|---|---|---|
| Mittel d. abh. Var. | 7645,000 | Stdabw. d. abh. Var. | 2042,814 |
| Summe d. quad. Res. | 11820538 | Stdfehler d. Regress. | 992,4942 |
| R-Quadrat | 0,811163 | Korrigiertes R-Quadrat | 0,763953 |
| F(3, 12) | 17,18224 | P-Wert(F) | 0,000122 |

32. Testen Sie die Restriktion mittels t-Tests.



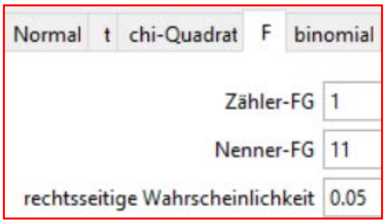

Zweiseitiger Test: $H_0$: Diff $= 0$ $H_1$: Diff $\neq 0$

Freiheitsgrade = N-K = 16 – 3 = 13

t-Wert = -2.273 < $t_c$ = 2.16 → $H_0$ nicht verwerfen

p-Wert ≅ 0.04 < 0.05 → $H_0$ nicht verwerfen

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




Berechnen Sie den F-Wert mittels

$$F = \frac{(RSS_r - RSS)}{RSS}\frac{(N-K)}{L}$$

L = 1 nur eine lineare Restriktion und

Freiheitsgrade = N-K = 16 - 5 = 11

$$F = \frac{(RSS_r - RSS)}{RSS}\frac{(N-K)}{L} = \left(\frac{11'820'538}{10'347'220} - 1\right)\cdot\frac{11}{1} = 1.566$$

$F_e(1,11)$ = 1.566 < $F_c(0.95, 1, 11)$ = 4.84 → $H_0$ nicht ablehnen!

Die Daten aus der Stichprobe sind mit der Restriktion $\beta_2 = -\beta_3$ kompatibel.

*Hinweis: Diese Berechnung dient nur dem Umgang mit dieser Formel und ist in der Praxis nicht erforderlich, da gretl einen eingebauten Test besitzt.*

34. Testen Sie diese Restriktion mittels gretl.

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

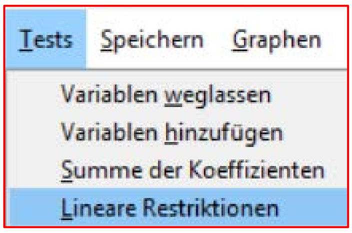
```
Restriktion:
 b[PR] + b[PN] = 0

Teststatistik: F(1, 11) = 1,56627, mit p-Wert = 0,23671

Restringierte Schätzungen:

             Koeffizient   Std.-fehler   t-Quotient   p-Wert
  -----------------------------------------------------------
  const       7156,72       5347,86        1,338      0,2056
  PR         -2133,52        938,780      -2,273      0,0422 **
  PN          2133,52        938,780       2,273      0,0422 **
  EINK           8,22845      31,2954      0,2629     0,7971
  T           -198,897       103,923      -1,914      0,0798 *

  Standardfehler der Regression = 992,494
```




# Restriktionen L = 1 und Nenner-Freiheitsgrade N- K = 16- 5 = 11

$F_e$(L, N-K) = $F_e$(1, 11) = 1.566 < $F_c$(0.095, 1, 11) = 4.84 → $H_0$ nicht ablehnen

p-Wert = 0.236 > $\alpha$ = 5% → $H_0$ nicht ablehnen

Konklusion: Die beobachteten Daten sind mit der Restriktion $\beta_2 = -\beta_3$ im Einklang.

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

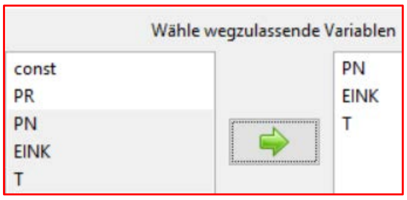
*gretl: Tests / Variablen weglassen →PN, EINK und T auswählen*

```
  Nullhypothese: Die Regressionskoeffizienten sind Null für die Variablen
    PN, EINK, T
  Teststatistik: F(3, 11) = 4,87558, p-Wert 0,0214786
  Das Weglassen von Variablen verbesserte 0 von 3 Informationskriterien.

Modell 25: KQ, benutze die Beobachtungen 1971:3-1975:2 (T = 16)
Abhängige Variable: Y

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
  -----------------------------------------------------------
  const       16899,0       1984,57        8,515     6,57e-07 ***
  PR          -2978,55       629,979      -4,728     0,0003   ***

Mittel d. abh. Var.     7645,000   Stdabw. d. abh. Var.    2042,814
Summe d. quad. Res.     24105953   Stdfehler d. Regress.   1312,194
R-Quadrat               0,614898   Korrigiertes R-Quadrat  0,587391
F(1, 14)                22,35405   P-Wert(F)               0,000324
```




Diese Variablen sind gemeinsam statistisch signifikant.

p-Wert = 2.14% < 5% → $H_0$: $b_3 = b_4 = b_5 = 0$ kann auf 5%-Niveau verworfen werden.

Konklusion: Mindestens eine Variable ist statistisch signifikant.