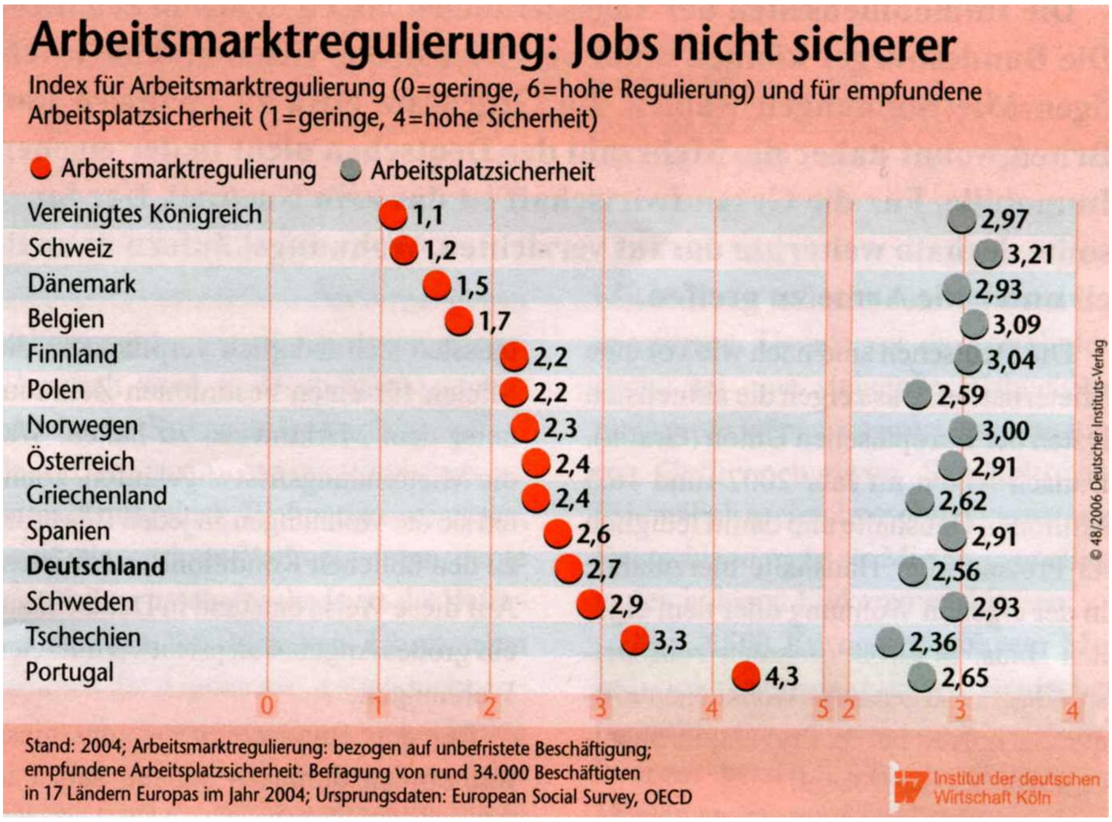
Prof. Dr. Peter v. d. Lippe

Aufgabe zur einfachen linearen Regression mit E-Views

Zusammenhang zwischen Arbeitsmarktregulierung und Arbeitsplatzsicherheit

1) Daten

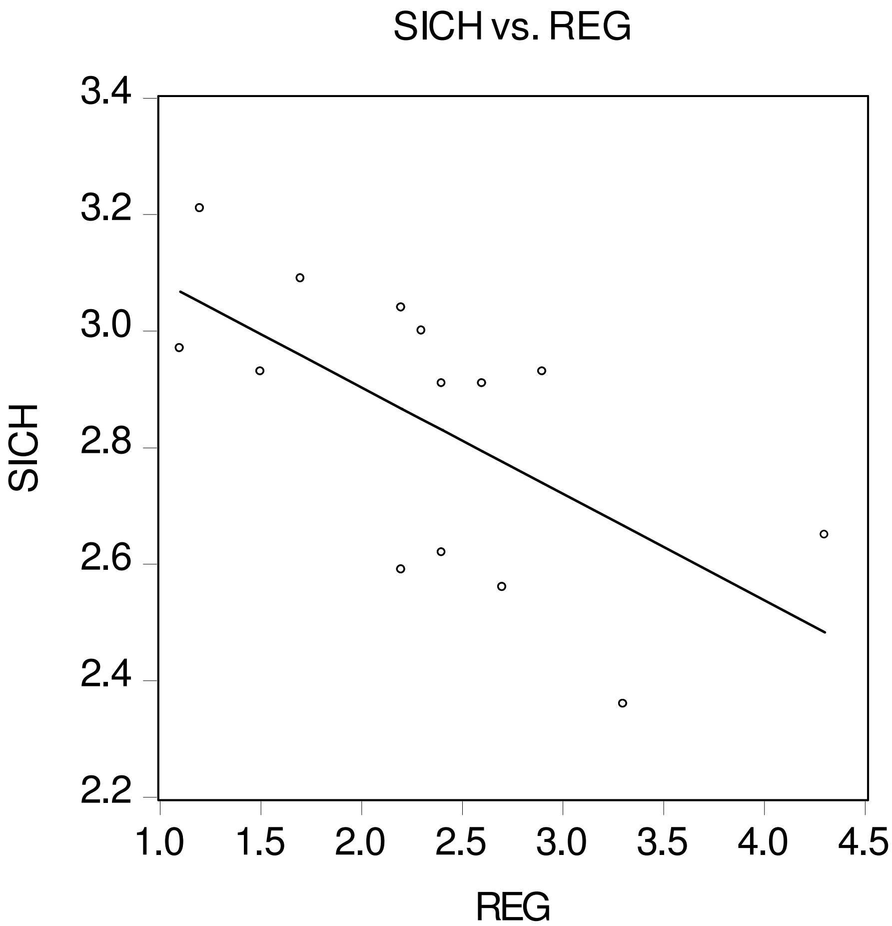
a) als Bild aus iwd-48/2006



b) als E-Views "group" (Gruppe von Variablen, hier zwei Variablen REG und SICH)

obs	REG	SICH
1	1.100000	2.970000
2	1.200000	3.210000
3	1.500000	2.930000
4	1.700000	3.090000
5	2.200000	3.040000
6	2.200000	2.590000
7	2.300000	3.000000
8	2.400000	2.910000
9	2.400000	2.620000
10	2.600000	2.910000
11	2.700000	2.560000
12	2.900000	2.930000
13	3.300000	2.360000
14	4.300000	2.650000

2) Streuungsdiagramm (E-Views output) (weitere graphische Optionen möglich)



3) Ergebnisse der Regressionsanalyse (der Befehl heißt: "make equation")
in Word übertragen und formatiert [z.B. mit farblicher Hervorhebung]

Dependent Variable: SICH				
Method: Least Squares				
Date: 11/08/07 Time: 16:58				
Sample: 1 14				
Included observations: 14				
Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
REG	**- 0.182486**	**0.063930**	**- 2.854484**	**0.0145**
C	**3.268253**	**0.158559**	**20.61225**	**0.0000**
R-squared	**0.404410**	**Mean dependent var**		**2.840714**
Adjusted R-squared	**0.354777**	**S.D. dependent var**		**0.242375**
S.E. of regression	**0.194690**	**Akaike info criterion**		**-0.303258**
Sum squared resid	**0.454848**	**Schwarz criterion**		**-0.211964**
Log likelihood	**4.122804**	**F-statistic**		**8.148076**
Durbin-Watson stat	**3.264928**	**Prob(F-statistic)**		**0.014502**

4) Weitere Auswertungen

Die **Regressionsgerade** $\hat{y}_t = \hat{\alpha} + \hat{\beta} x_t$ (oder $\hat{y} = a + bx$) lautet demnach

SICH = 3,268253 – 0,182486 (REG)

Der **Korrelationskoeffizient** beträgt - 0,63593 (= $-\sqrt{0.404410}$) aus $\hat{\beta} < 0$ und dem Streuungsdiagramm [scatter plot] ist ersichtlich, dass die Korrelation negativ ist.

Aus $t_{\hat{\beta}} = \hat{\beta} / \hat{\sigma}_{\hat{\beta}}$ folgt - 2,854484 = - 0,182486/0,063930[1] (entsprechender Zusammenhang für $t_{\hat{\alpha}}$). Die Werte der Spalte **t-Statistic** lassen sich also aus den anderen beiden Spalten errechnen.[2]

Aus entsprechenden Tabellen erhält man die folgenden Werte für t bei der **t-Verteilung** mit T - K - 1 = T - 2 = 14 - 2 = 12 Freiheitsgraden (K = Anzahl der Regressoren)

einseitig		zweiseitig	
0,975	2,18	0,95	1,78
0,995	3,05	0,99	2,68

Weil die angegebenen t-Werte alle (bei α und β) erheblich größer sind als diese Tabellenwerte, sind die Schätzwerte $\hat{\alpha}$ und $\hat{\beta}$ **hochsignifikant** (H_0 jeweils α = 0 und β = 0).[3] Damit lässt sich auch ein zweiseitiges **Konfidenzintervall** berechnen für β (entsprechend für α und entsprechend auch bei einseitigen Konfidenzintervallen).

Grenzen des Konfidenzintervalls bei 95%

unten – 0,182486 – 1,78*0,063930 = - 0,294361

oben – 0,182486 + 1,78*0,063930 = -0,0706085.

Die Grenzen des Konfidenzintervalls bei 99% sind entsprechend zu berechnen mit 2,68 statt 1,78. Ergebnis: unten – 0.3538184 oben – 0,1111536. Da der Wert 0 *nicht* enthalten ist, ist – wie schon beim Vergleich von $t_{\hat{\beta}}$ mit dem Tabellenwert der t-Verteilung ersichtlich – β signifikant verschieden von Null (H_0: β = 0).

5) Konfidenzintervall für die Varianz σ^2 (der Störgröße)[4]

Der **S.E.** (standard error) **of Regression**, d.h. die Größe 0,194690 und die **Sum** (of) **squared resid.** (= $\sum \hat{u}_t^2 = S_{\hat{u}\hat{u}}$) = 0,454848 hängen bei T = 14 wie folgt zusammen

$$\hat{\sigma}_u = \sqrt{\frac{1}{T-2}\sum \hat{u}_t^2} = \sqrt{\frac{0{,}454848}{12}} = 0{,}194690$$ (dies ist der erwartungstreue Schätzer für σ^2).

Man kann diesen Wert benutzen um ein Konfidenzintervall für σ^2, die Varianz der Störgröße zu berechnen (jetzt mit der χ^2 **Verteilung** mit T-2 Freiheitsgraden arbeiten[5]. Ergebnis: **Untergrenze** $S_{\hat{u}\hat{u}}/G_o$ = und 0.4548/23,3 = 0,01952 und

Obergrenze $S_{\hat{u}\hat{u}}/G_u$ = 0.4548/4,4 = 0,10337,

[1] (Rundungsfehler!) mit dem oben angegebenen Bruch - 0,182486/0,063930 erhält man -2,854466 statt – 2,584484 (aber die Werte 0,182486 und 0,063930 sind ja ihrerseits auch wieder gerundet !!).
[2] Wenn H_0 lautet β = 0. Die H_0 kann aber auch lauten β = - 0,2 (nur als ein Beispiel).
[3] Erklärung der Spalte "prob." (= prob. value) in der Vorlesung.
[4] Zur Intervallschätzung von σ^2 vgl. Teil C des downloads Nr. 2 (Übersicht über Schätz- und Testtheorie ...)
[5] zu den Größen G_u und G_o siehe nächste Seite.

G_u und G_o sind der in der χ^2 Tabelle mit T-K-1 = T-2 = 12 Freiheitsgraden angegebenen untere und obere Wert (Signifikanzschranke) für das 95% Konfidenzintervall, also $G_u = \chi^2_{0.025}$ = 4,40 und $G_o = \chi^2_{0.975}$ = 23,30.

6) ANOVA (Analysis of Variance [F Test]) und weitere Zusammenhänge zwischen den von E-Views ausgegebenen Rechenergebnissen

Aus **S.D.** (standard deviation) **of dependent variable** $\sqrt{\frac{1}{T-1}S_{yy}}$ = 0,242375 folgt S_{yy} = 0.76369. Man hat damit alle Bestandteile der **ANOVA-Tabelle** (Varianzzerlegung), die wie folgt aussieht

Variation (sum of squares)	d.f. (degrees of freedom)	Varianz (mean squares)
explained $S_{\hat{y}\hat{y}} = 0{,}308842$*	K = 1	$S_{\hat{y}\hat{y}}/K = 0{,}308842$
residual $S_{\hat{u}\hat{u}} = 0{,}454848$	T - K - 1 = T - 2 = 12	$S_{\hat{u}\hat{u}}/(T-K-1)$ = 0,454848/12 = 0,037904
total $S_{yy} = 0{,}76369$	T – 1 = K + (T-K-1)	0,76369/13 = 0,05874 = $(0{,}24237)^2$= (S.D.depend.var.)^2 (**nicht** die Summe von 0,3088 und 0,0379)

* als Differenz bestimmt: 0,76369 - 0,454848

Die F-Statistik $F = \frac{\text{explained mean square}}{\text{residual mean square}} = \frac{S_{\hat{y}\hat{y}}/K}{S_{\hat{u}\hat{u}}/(T-K-1)}$

ist F verteilt mit K=1 und T-K-1 = 12 Freiheitsgraden (Symbol $F_{K,T-K-1} = F_{1,12}$). Sie beträgt in diesem Fall F = 0,308842/0,037904 = 8,1480055 (E-Views [siehe oben unter Nr. 3] gibt an 8,148076). Die entsprechenden F-Werte der Tabelle sind bei 5% Signifikanzniveau 4,75 < 8,14 und bei 1% 9,33 > 8,14 (prob.value ist 1,45%).

Man beachte, dass bei einfacher Regression gilt $\sqrt{F} = t_{\hat{\beta}}$ (im Beispiel $\sqrt{8{,}148076} = 2{,}854$) und dass die F-Verteilung mit *einem* Freiheitsgrad im Zähler, also $F_{1,v}$ identisch ist mit der t-Verteilung t_v (t-Verteilung mit v Freiheitsgraden). Der Test der Hypothese H_0: β = 0 und der Hypothese H_0: $\rho^2 = 0$ (Unkorreliertheit) sind also identisch.[6]

Zwischen R^2 und dem korrigierten (adjusted) R^2 (mit $\bar{R}^2$ bezeichnet) besteht folgender Zusammenhang

$$R^2 = r^2 = \frac{S_{\hat{y}\hat{y}}}{S_{yy}} = 1 - \frac{S_{\hat{u}\hat{u}}}{S_{yy}}$$ = 0,3088/0,76369 = 0,4044075 (vgl. E-Views Angabe oben 0,404410)

$$\bar{R}^2 = R^2 - \frac{(1-R^2)K}{T-K-1} = 1 - \frac{S_{\hat{u}\hat{u}}/(T-K-1)}{S_{yy}/(T-1)}$$ (man verifiziere das für das Zahlenbeispiel).

7) Weitere Bemerkung

Der Durbin-Watson Koeffizient (oder DW statistic) in Höhe von d = 3.264928 weist auf eine negative Autokorrelation der Residuen (Störgrößen u_t, u_{t-1},…) von etwa r = 1 - d/2 = - 0,6325 hin (mehr dazu später beim Thema Autokorrelation).

[6] in der Stichprobe spricht man von r und r^2. In der Grundgesamtheit werden griechische Symbole verwendet also ρ (Rho) bzw. ρ^2.