

Übung 2: Die Vorhersage von US-Präsidentschaftswahlen

Es wird versucht, anhand eines Regressionsmodells, den Ausgang von Präsidentschaftswahlen in den USA vorherzusagen.

Im Zentrum des Regressionsmodells steht die zu erklärende Variable *VOTE*. Sie ist in Prozentpunkten gemessen und gibt für jede Präsidentschaftswahl den Anteil der Wählerstimmen an, der jeweils auf die amtierende Partei (*incumbent party*) entfällt - also auf die Partei, die zum Zeitpunkt der Wahl (noch) regiert. Betrachtet werden dabei nur die beiden grossen Parteien, Demokraten und Republikaner.

Variablen

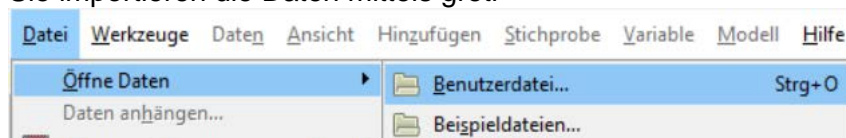
- **V (Vote)**: Stimmenanteil für die amtierende Partei
- **VD (Vote Democrat)**: Stimmenanteil für die Demokraten
- **G (Growth)**: Jährliche Wachstumsrate des realen pro-Kopf BIP während der ersten drei Quartale des jeweiligen Wahljahres.
BIP: Bruttoinlandsprodukt → gibt den Gesamtwert aller Güter (Waren und Dienstleistungen) an, die innerhalb eines Jahres in einem Land hergestellt wurden.
- **P (Price)**: Jährliche Inflationsrate (BIP-Deflator) während der ersten 15 Quartale der Regierungsperiode ausser für die Wahljahre 1920, 1944 und 1948.
- **GOODNEWS** = Anzahl der Quartale in den ersten 15 Quartalen der Regierungsperiode in denen die jährliche Wachstumsrate des realen pro-Kopf BIP höher ist als 3.2% p.a. ausser für die Wahljahre 1920, 1944 und 1948.

Ergebnisse der letzten sieben Präsidentschaftswahlen:

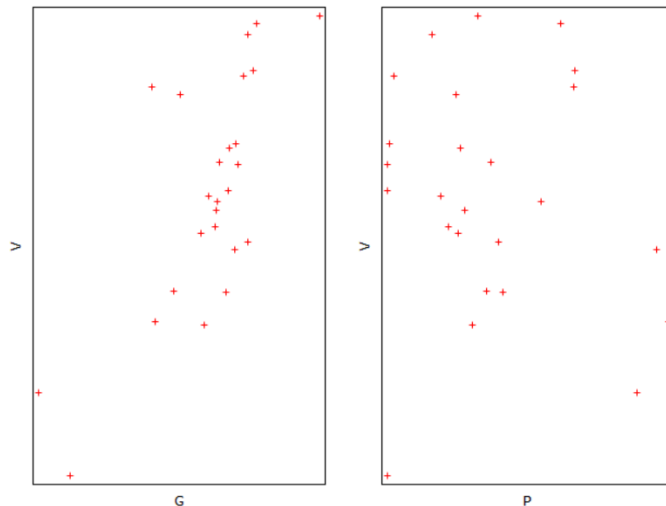
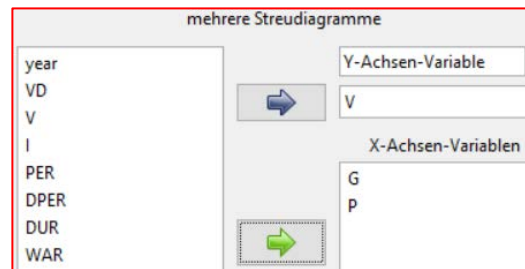
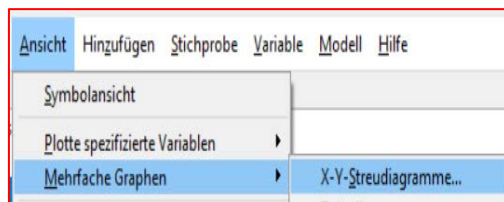
Wahl	Winner		Demokraten	Republikaner
2016	Donald Trump	Hillary Clinton	43.1%	56.9%
2012	Barack Obama	Mitt Romney	51.06%	
2008	Barack Obama	John McCain	52.93%	
2004	George W. Bush	John Kerry	50.73%	
2000	George W. Bush	Al Gore		47.87%
1996	Bill Clinton	Bob Dole	49.23%	
1992	Bill Clinton	George H.W. Bush	43.01%	

1. Laden Sie die Daten aus Moodle herunter. Sie haben zwei Möglichkeiten, die Daten in gretl einzuspeisen:

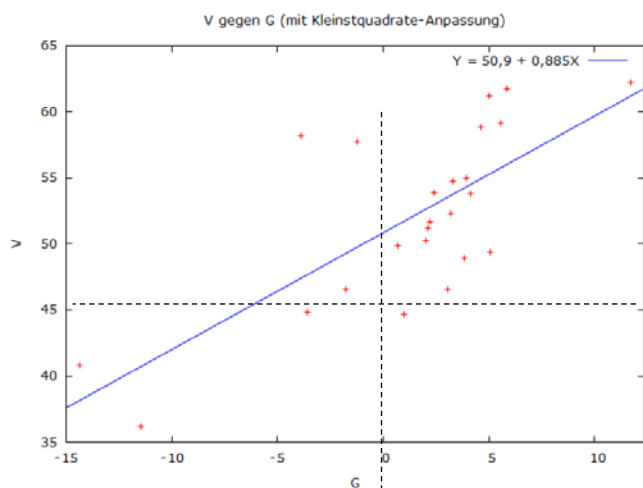
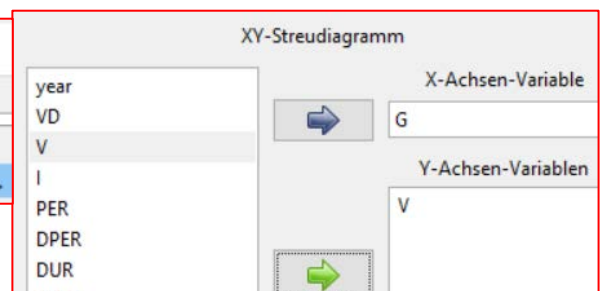
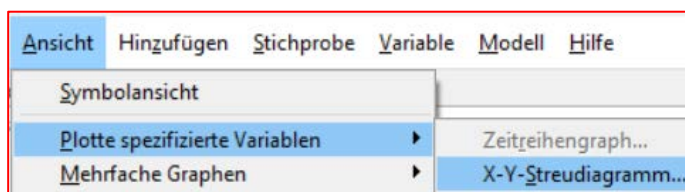
- Doppelklicken Sie auf das gretl-Workfile „Übung 2_Präsidentenwahl.gdt“
- Sie importieren die Daten mittels gretl



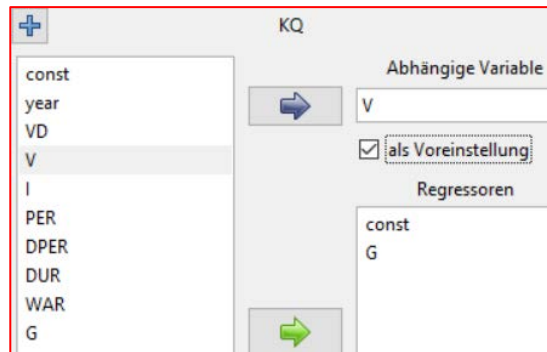
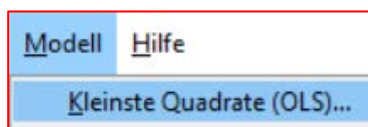
2. Erstellen Sie ein Streudiagramm der Variable V (Vote) gegen G (Growth) und P (Inflation). Ist ein Zusammenhang zwischen den Variablen ersichtlich? Geben Sie eine ökonomische Begründung dafür?



3. Zeigen Sie die Reihen G (Growth) und V (VOTE) als *Scatter-Plot* mit Regressionslinie an. Wurde die jeweilige Regierungspartei bei negativen Wachstumsraten häufiger abgewählt als im Amt bestätigt?



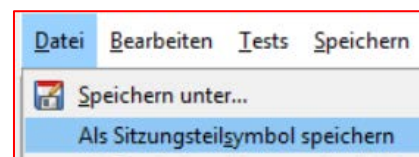
4. Die Stichprobenperiode liegt zwischen den Wahljahren 1916 und 2012. Aus welchen Gründen wurden die Wahljahre vor 1916 nicht berücksichtigt?
5. Schätzen Sie das Modell 1: $V = \beta_1 + \beta_2 G + u$



Modell 1: KQ, benutze die Beobachtungen 1-25
Abhängige Variable: V

	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	50,8555	0,961783	52,88	1,59e-025 ***
G	0,884594	0,176370	5,016	4,49e-05 ***
Mittel d. abh. Var.	52,08388	Stdabw. d. abh. Var.	6,587316	
Summe d. quad. Res.	497,4019	Stdfehler d. Regress.	4,650395	
R-Quadrat	0,522384	Korrigiertes R-Quadrat	0,501618	

Hinweis: Speichern Sie dieses Regressionsmodell

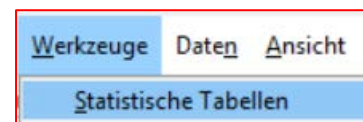


6. Interpretieren Sie die geschätzten Regressionsparameter.

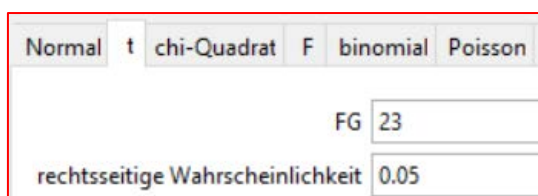
7. Ist der geschätzte Steigungsparameter signifikant auf dem 5%-Signifikanzniveau?

8. Geben Sie dazu die Null- und Alternativhypothese.

9. Ermitteln Sie den kritischen Wert der t-Statistik mittels gretl.



Kritischer Wert: $t_c(0.095, 23) = 1.7138$



t(23)
rechtsseitige Wahrscheinlichkeit = 0,05
komplementäre Wahrscheinlichkeit = 0,95
zweiseitige Wahrscheinlichkeit = 0,1

Kritischer Wert = 1,71387

10. Berechnen Sie manuell den **t-Wert** anhand des Standardfehlers.

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

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

13. Interpretieren Sie den p-Wert für Steigungsparameter.

14. Bestimmen Sie das 95%- Konfidenzintervall für den Parameter β_2 mittels gretl.

		$t(23, 0,025) = 2,069$			
Analyse	LaTeX	VARIABLE	KOEFFIZIENT	95% KONFIDENZ-INTERVALL	
Zeige tatsächliche, geschätzte, Residuen		const	50,8555	48,8659	52,8451
Prognosen...		G	0,884594	0,519744	1,24944
Konfidenzintervalle für Koeffizienten					

15. Bestimmen Sie den kritischen Wert t_c für die Berechnung eines 95%-Konfidenzintervalls

Konfidenzniveau = $1 - \alpha = 95\%$
$$t_c(1-\alpha/2, df) = t_c(0.975, 23) = 2.069 \text{ (aufgerundet)}$$

16. Bestimmen Sie manuell ein 95%- Konfidenzintervall für den Parameter β_2 . Vergleichen Sie es mit dem gretl Intervall (Frage 14). Interpretieren Sie konkret das 95%-Konfidenzintervall.

17. Interpretieren Sie konkret den R^2 -Wert

18. Schätzen Sie das Modell 2: $V = \beta_1 + \beta_2 P + u$

3. Schätzen Sie das Modell 2: $V = \beta_1 + \beta_2 P + u$

Modell 2: KQ, benutze die Beobachtungen 1-25
Abhängige Variable: V

	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	53,2450	2,14039	24,88	3,96e-018 ***
P	-0,408531	0,589489	-0,6930	0,4952

Mittel d. abh. Var.	52,08388	Stdabw. d. abh. Var.	6,587316
Summe d. quad. Res.	1020,123	Stdfehler d. Regress.	6,659819
R-Quadrat	0,020455	Korrigiertes R-Quadrat	-0,022134

Hinweis: Speichern Sie dieses Regressionsmodell

Modell Hilfe

Kleinste Quadrate (OLS)...

Datei Bearbeiten Tests Speichern

Speichern unter...

Als Sitzungsteilsymbol speichern

19. Interpretieren Sie den geschätzten Regressionskoeffizienten b_2 .

20. Ist der geschätzte b_2 -Koeffizient auf dem 5%-Signifikanzniveau signifikant?

21. Stellen Sie die Null- und Alternativhypothese für b_2 auf.

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

23. Beurteilen Sie die **Anpassungsgüte** dieses Modells?

24. Testen Sie folgende Nullhypothese: „Wenn die Inflationsrate null ist, beträgt der erwartete Stimmenanteil der amtierenden Partei mindestens 50%“.

$$E(V | P = 0) =$$

25. Bestimmen Sie ein 95%- Konfidenzintervall für den erwarteten Stimmenanteil der amtierenden Partei (VOTE) wenn $P = 2$ (2% Inflation). Interpretieren Sie konkret Ihr 95%-Konfidenzintervall.

Erwarteter Stimmenanteil wenn Inflationsrate $P = 2$

$$E(V | P = 2) = b_1 + 2b_2 =$$

$$\text{var}(b_1 + 2b_2) = \text{var}(b_1) + 4\text{var}(b_2) + 4\text{cov}(b_1, b_2)$$

$$t_c(0.975, 23) =$$

95%-Konfidenzintervall:

Analyse	LaTeX
Zeige tatsächliche, geschätzte, Res	
Prognosen...	
Konfidenzintervalle für Koeffiziente	
Konfidenzellipse...	
Kovarianzmatrix der Koeffizienten	

Kovarianzmatrix

const	P
4,58125	-0,987658
0,347498	P

Ein Kollege von Ihnen schlägt das Modell 3 vor: $VD = \beta_1 + \beta_2 G + u$

26. Schätzen Sie dieses Modell.

Modell 1: KQ, benutze die Beobachtungen 1-25				
Abhängige Variable: VD				
	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	49,0343	1,41427	34,67	2,32e-021 ***
G	0,287214	0,259346	1,107	0,2795
Mittel d. abh. Var.	49,43312	Stdabw. d. abh. Var.	6,870431	
Summe d. quad. Res.	1075,517	Stdfehler d. Regress.	6,838245	
R-Quadrat	0,050625	Korrigiertes R-Quadrat	0,009348	

27. Ist der geschätzte b_2 -Koeffizient signifikant auf dem 5%-Signifikanzniveau?

28. Warum ist dieses Regressionsmodell nicht geeignet?

29. Schätzen Sie das Modell 4: $V = \beta_1 + \beta_2 G + \beta_3 P + u$

Modell 4: KQ, benutze die Beobachtungen 1-25				
Abhängige Variable: V				
	Koeffizient	Std.-fehler	t-Quotient	p-Wert
const	50,6291	1,62087	31,24	1,02e-019 ***
G	0,892169	0,185294	4,815	8,26e-05 ***
P	0,0759775	0,432453	0,1757	0,8621
Mittel d. abh. Var.	52,08388	Stdabw. d. abh. Var.	6,587316	
Summe d. quad. Res.	496,7050	Stdfehler d. Regress.	4,751579	
R-Quadrat	0,523053	Korrigiertes R-Quadrat	0,479694	

30. Sind die geschätzten Regressionskoeffizienten b_2 , b_3 signifikant auf dem 10%-Signifikanzniveau?

31. Stellen Sie dazu die Null- und Alternativhypothese auf. Wie lautet Ihre Konklusion?

32. Was ist mit dem Vorzeichen von P (Inflationsrate) passiert?

33. Nehmen Sie an, dass die Inflationsrate 2% beträgt. Was ist die Vorhersage, wenn die Wachstumsrate i) -2% ii) 0% iii) 2% beträgt.

$VOTE(-2\%, 2\%) =$

$VOTE(0\%, 2\%) =$

$VOTE(2\%, 2\%) =$

34. Schätzen Sie das Modell 5: $V = \beta_1 + \beta_2 G + \beta_3 P + \beta_4 \text{GOODNEWS} + u$

Modell 5: KQ, benutze die Beobachtungen 1-25
Abhängige Variable: V

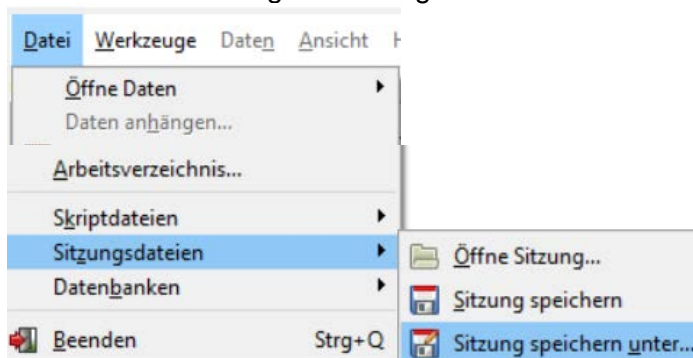
	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	48,2467	1,89574	25,45	2,89e-017	***
G	0,746028	0,186440	4,001	0,0006	***
P	-0,202761	0,424743	-0,4774	0,6380	
GOODNEWS	0,703645	0,338275	2,080	0,0500	**
Mittel d. abh. Var.	52,08388	Stdabw. d. abh. Var.	6,587316		
Summe d. quad. Res.	411,8483	Stdfehler d. Regress.	4,428524		
R-Quadrat	0,604534	Korrigiertes R-Quadrat	0,548039		

Hinweis: Speichern Sie Ihre Sitzung

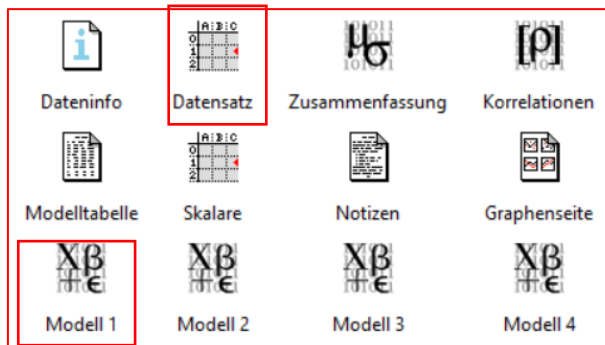
35. Sind die geschätzten Regressionskoeffizienten signifikant auf dem 5%-Signifikanzniveau?

36. Interpretieren Sie den geschätzten Regressionskoeffizienten b_4 ?

Schliessen Sie Ihre gretl-Sitzung und öffnen Sie sie anschliessend erneut.



Das Sitzungs-Fenster, Symbolansicht genannt, enthält eine Reihe von anklickbaren Symbolen. Wenn sie das Symbol „Modell 1“ anklicken, öffnet sich das Output-Fenster.



Wenn Sie auf das Symbol Datensatz anklicken, öffnet sich die entsprechende Datentabelle.

37. Schätzen Sie das Modell 6: $V = \beta_1 + \beta_2 G + \beta_3 \text{GOODNEWS} + u$

	coefficient	std. error	t-ratio	p-value	
const	47.8750	1.69784	28.20	9.18e-019	***
G	0.774813	0.173296	4.471	0.0002	***
GOODNEWS	0.652698	0.315315	2.070	0.0504	*
Mean dependent var	52.08388	S.D. dependent var	6.587316		
Sum squared resid	416.3175	S.E. of regression	4.350118		
R-squared	0.600243	Adjusted R-squared	0.563901		

38. Sind die geschätzten Regressionskoeffizienten signifikant auf dem 5%-Signifikanzniveau?

39. Welches Regressionsmodell würden Sie anhand des adjustierten Bestimmtheitsmasses auswählen?

Folgende Tabelle enthält eine Zusammenstellung der zur vergleichenden Kennzahlen.

	Modell 1	Modell 2	Modell 4	Modell 5	Modell 6
Variable	G	P	G, P	G, P, GoodN	G, GoodN
Adj. R^2	50.16	-2.2	47.96	54.8	56.39

40. Testen Sie folgende alternative Hypothese für Modell 6: "Die amtierende Partei erlangt die Stimmenmehrheit, wenn die Wachstumsrate 2% und die Anzahl Quartale mit einer Wachstumsrate höher als 3.2% 2 beträgt". Nehmen Sie ein Signifikanzniveau von 5% an.

Kovarianzmatrix:

const	G	GOODNEWS	
2,88265	0,0385666	-0,454013	const
	0,0300316	-0,0167226	G
		0,0994237	GOODNEWS

Analyse	LaTeX
Zeige tatsächliche, geschätzte, Res	
Prognosen...	
Konfidenzintervalle für Koeffiziente	
Konfidenzellipse...	
Kovarianzmatrix der Koeffizienten	

41. Versetzen Sie sich nun zurück ins Jahr 2016 kurz vor der Präsidentschaftswahl vom 8. November 2016. Zu diesem Zeitpunkt haben Sie noch keine Information über den Wahlausgang. Berechnen Sie die Vorhersage des Wählerstimmenanteils bei folgenden Werten mittels Modellen 5 und 6: $G = 0.97\%$ $P = 1.42\%$ und Goodnews = 2 im Jahr 2016.

42. Berechnen Sie den Prognosefehler für beide Modelle 5 und 6. Die Wahlergebnisse (popular vote) waren 48.02% für Clinton und 46.05% für Trump.

Prognosefehler für Modell 5: $f_5 = V_{16} - \hat{V}_{16} =$

Prognosefehler für Modell 6: $f_6 = V_{16} - \hat{V}_{16} =$

43. Welches Modell hat die beste Prognose des Wählerstimmenanteils geliefert?

44. Warum wurde Trump als Präsident gewählt, obwohl Hillary Clinton mehr Stimmen (popular vote) bekommen hat?

Versetzen Sie sich nun zurück ins Jahr 2012 kurz vor der Präsidentschaftswahl vom 6. November 2012. Zu diesem Zeitpunkt haben Sie noch keine Information über den Wahlausgang. Der demokratische Amtsinhaber Barack Obama trat gegen den Republikaner Mitt Romney an.

45. Reduzieren Sie die Stichprobe für die Regression auf die Zeitperiode 1916-2008.

Stichprobe Variable
Bereich wählen...
Gesamtbereich wie

Wähle Stichprobenbereich

Start: 1 Ende: 24

Beobachtungen: 24

Die Beobachtung 24 entspricht den Ergebnissen des Wahljahres 2012

Hinweis: Nach der Regression vergessen Sie nicht, den Gesamtbereich (1-25 Beobachtungen) wieder herzustellen, sonst wird gretl für andere Regressionen den Stichprobenbereich 1-24 Beobachtungen wieder verwenden.

46. Schätzen Sie das Modell für die Zeitperiode 1916-2008 (Beobachtungen 1 bis 24).
Modell 7: $V = \beta_1 + \beta_2 G + \beta_3 \text{GOODNEWS} + u$

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	47,5216	1,83133	25,95	1,94e-017	***
G	0,765626	0,176732	4,332	0,0003	***
GOODNEWS	0,706547	0,333740	2,117	0,0464	**
Mittel d. abh. Var.	52,08696	Stdabw. d. abh. Var.	6,728976		
Summe d. quad. Res.	409,9065	Stdfehler d. Regress.	4,418072		
R-Quadrat	0,606396	Korrigiertes R-Quadrat	0,568910		

47. Sind die geschätzten Regressionskoeffizienten signifikant auf dem 5%-Niveau?
48. Berechnen Sie die Vorhersage des Wählerstimmenanteils der amtierenden Partei mit den folgenden Werten: $G = 1.42$, $P = 1.47$ und $\text{Goodnews} = 1$.
49. Berechnen Sie unter Berücksichtigung des tatsächlichen Wahlergebnisses den Vorhersagefehler (f) für 2012.
Wahlergebnis: 51.1% für Obama gegen 47.2% für Mitt Romney