NTNU
Det skapende universitet

Institutt for sosiologi og statsvitenskap

Eksamensoppgave i SOS3003 Anvendt statistisk dataanalyse i samfunnsvitenskap

Faglig kontakt under eksamen: Johan Fredrik Rye

Telefon: 992 73 088

Eksamensdato: 10.06.2016

Eksamenstid: 6 timer

Hjelpemiddelkode/Tillatte hjelpemidler: Alle kalkulatorer

Sensurdato: 01.07.2016

Målform/språk/language: Bokmål/Nynorsk

Antall sider: 1 side bokmål / 1 side Nynorsk

Antall sider vedlegg: 11 (s. 3-14)

BOKMÅL
Besvar følgende tre oppgaver (Hver besvarelse teller en tredjedel av den samlede karakteren)

Oppgave 1
Vedlegg 1 gjengir en loggfil fra statistikkprogrammet Stata, og viser en analyse av data fra det norske utvalget av European Social Survey fra 2014. Beskriv oppbyggingen av modellene og estimatene fra modell 1 og modell 2 i vedlegg 1 (s. 4-10).

Oppgave 2
Beskriv de viktigste forutsetningene for OLS-regresjon, og drøft i hvilken grad disse forutsetningene er oppfylt ut fra testene i vedlegg 1 (s. 11-14).

Oppgave 3
Vi foretrekker ofte å bruke OLS-regresjon når vi skal analysere en kontinuerlig avhengig variabel i et sannsynlighetsutvalg. Hva er problemene med å bruke OLS-regresjon når vi skal analysere data med en dikotom (todelt) avhengig variabel, strukturerte data med klyngestruktur og dermed avhengighet mellom observasjonene, og når vi har tidsseriedata som for eksempel månedlige registreringer av andelen arbeidsløse i forhold til den totale arbeidsstyrken?

NYNORSK
Svar på alle tre oppgåvene (Kvart svar tel ein tredjedel av den samla karakteren)

Vedlegg 1 viser ei loggfil frå statistikkprogrammet Stata, og presenterer ein analyse av data frå det norske utvalet av European Social Survey frå 2014. Beskriv oppbygginga av modellane og estimata frå modell 1 og modell 2 i vedlegg 1 (s. 4-10).

Oppgåve 2
Beskriv dei viktigaste føresetnadane for OLS-regresjon, og drøft i kva grad disse føresetnadane er oppfylte ut frå testane i vedlegg 1 (s. 11-14).

Oppgåve 3
Vi føretrekker ofte å bruke OLS-regresjon når vi skal analysere ein kontinuerlig avhengig variabel i eit sannsyneutval. Kva er problema med å bruke OLS-regresjon når vi skal analysere data med ein dikotom (todelt) avhengig variabel, strukturerte data med klyngestruktur og dermed avhengighet mellom observasjonane, og når vi har tidsseriedata som for eksempel månadlege registreringer av andelen arbeidslause i forhold til den totale arbeidsstyrken?

Vedlegg 1

```
. * TASK 1
. * Dependent variable
. tab1 trstprl trstlgl trstplc trstplt trstprt

-> tabulation of trstprl
```

Trust in country's parliament	Freq.	Percent	Cum.
0. No trust at all	19	1.33	1.33
1. 1	16	1.12	2.45
2. 2	29	2.03	4.48
3. 3	51	3.57	8.05
4. 4	79	5.53	13.58
5. 5	149	10.43	24.00
6. 6	184	12.88	36.88
7. 7	321	22.46	59.34
8. 8	338	23.65	83.00
9. 9	142	9.94	92.93
10. Complete trust	101	7.07	100.00
Total	1,429	100.00	

```
-> tabulation of trstlgl
```

Trust in the legal system	Freq.	Percent	Cum.
0. No trust at all	13	0.91	0.91
1. 1	8	0.56	1.47
2. 2	25	1.75	3.22
3. 3	47	3.29	6.50
4. 4	42	2.94	9.44
5. 5	141	9.86	19.30
6. 6	136	9.51	28.81
7. 7	258	18.04	46.85
8. 8	385	26.92	73.78
9. 9	245	17.13	90.91
10. Complete trust	130	9.09	100.00
Total	1,430	100.00	

-> tabulation of trstplc

Trust in the police	Freq.	Percent	Cum.
0. No trust at all	17	1.19	1.19
1. 1	7	0.49	1.67
2. 2	17	1.19	2.86
3. 3	31	2.16	5.02
4. 4	40	2.79	7.82
5. 5	107	7.47	15.28
6. 6	132	9.21	24.49
7. 7	276	19.26	43.75
8. 8	395	27.56	71.32
9. 9	267	18.63	89.95
10. Complete trust	144	10.05	100.00
Total	1,433	100.00	

-> tabulation of trstplt

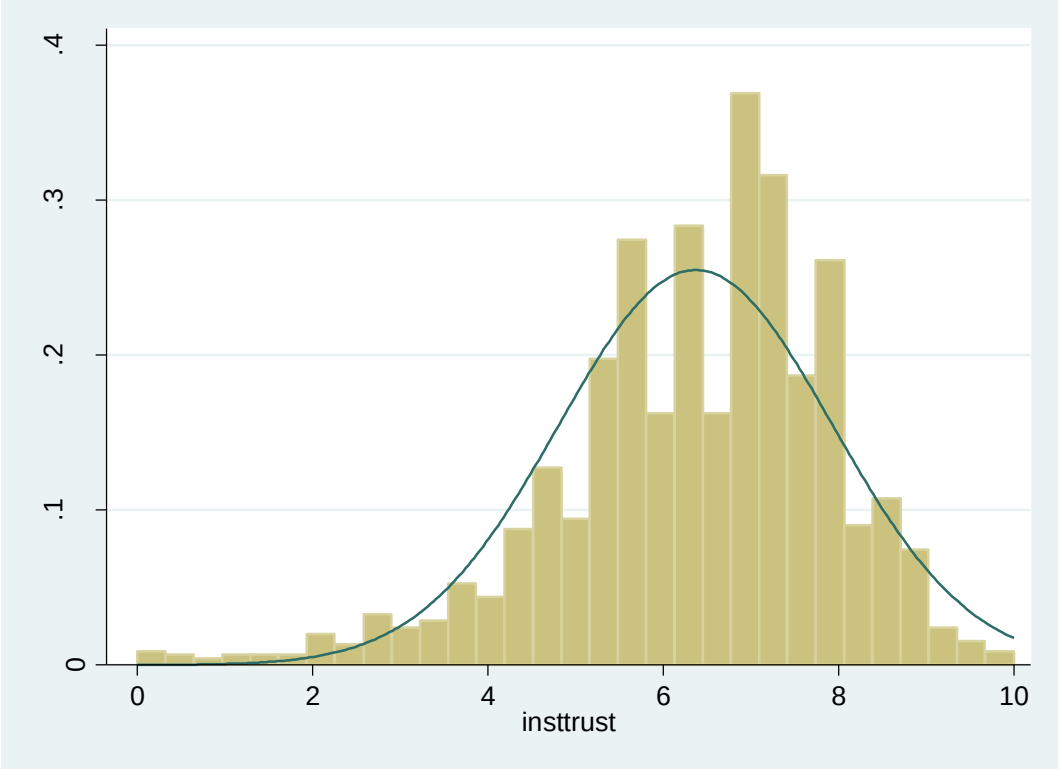
Trust in politicians	Freq.	Percent	Cum.
0. No trust at all	31	2.17	2.17
1. 1	29	2.03	4.20
2. 2	79	5.53	9.73
3. 3	114	7.98	17.70
4. 4	175	12.25	29.95
5. 5	310	21.69	51.64
6. 6	289	20.22	71.87
7. 7	256	17.91	89.78
8. 8	115	8.05	97.83
9. 9	20	1.40	99.23
10. Complete trust	11	0.77	100.00
Total	1,429	100.00	

-> tabulation of trstprt

Trust in political parties	Freq.	Percent	Cum.
0. No trust at all	26	1.83	1.83
1. 1	19	1.34	3.17
2. 2	72	5.07	8.23
3. 3	116	8.16	16.40
4. 4	179	12.60	28.99
5. 5	339	23.86	52.85
6. 6	272	19.14	71.99
7. 7	231	16.26	88.25
8. 8	125	8.80	97.04
9. 9	29	2.04	99.09
10. Complete trust	13	0.91	100.00
Total	1,421	100.00	

```
. alpha trstprl trstlgl trstplc trstplt trstprt

Test scale = mean(unstandardized items)

Average interitem covariance:     2.090183
Number of items in the scale:            5
Scale reliability coefficient:      0.8484

. generate insttrust=(trstprl+trstlgl+trstplc+trstplt+trstprt)/5
(25 missing values generated)

. summarize insttrust
```

Variable	Obs	Mean	Std. Dev.	Min	Max
insttrust	1411	6.371368	1.564346	0	10



```
. * Continous independent variables
. summarize agea eduyrs
```

Variable	Obs	Mean	Std. Dev.	Min	Max
agea	1436	46.76671	18.68344	15	104
eduyrs	1434	13.85216	3.717394	0	30

```
. * Categorical independent variables
. tab1 female regions polintr

-> tabulation of female
```

RECODE of gndr (Gender)	Freq.	Percent	Cum.
Male	764	53.20	53.20
Female	672	46.80	100.00
Total	1,436	100.00	

```
-> tabulation of regions
```

Region	Freq.	Percent	Cum.
Eastern Norway	738	51.39	51.39
Agder and Rogaland	180	12.53	63.93
Western Norway	248	17.27	81.20
Trøndelag	140	9.75	90.95
Northern Norway	130	9.05	100.00
Total	1,436	100.00	

```
-> tabulation of polintr
```

How interested in politics	Freq.	Percent	Cum.
1. Very interested	142	9.89	9.89
2. Quite interested	570	39.69	49.58
3. Hardly interested	595	41.43	91.02
4. Not at all interested	129	8.98	100.00
Total	1,436	100.00	

```
. * Model 1
. regress insttrust female agea eduyrs i.regions i.polintr
```

Source	SS	df	MS		
Model	260.574021	10	26.0574021	Number of obs	= 1410
Residual	3189.81121	1399	2.2800652	F(10, 1399)	= 11.43
				Prob > F	= 0.0000
Total	3450.38523	1409	2.44881848	R-squared	= 0.0755
				Adj R-squared	= 0.0689
				Root MSE	= 1.51

insttrust	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
female	.0217259	.0817042	0.27	0.790	-.13855	.1820018
agea	-.0087248	.0022769	-3.83	0.000	-.0131914	-.0042583
eduyrs	.0469327	.011629	4.04	0.000	.0241206	.0697448
regions						
Agder and Rogaland	-.1766634	.1268619	-1.39	0.164	-.4255235	.0721967
Western Norway	-.0885928	.112733	-0.79	0.432	-.3097368	.1325512
Trøndelag	.1572054	.1401817	1.12	0.262	-.1177836	.4321944
Northern Norway	-.4279995	.1459855	-2.93	0.003	-.7143735	-.1416254
polintr						
2. Quite interested	-.3296141	.1451707	-2.27	0.023	-.6143898	-.0448384
3. Hardly interested	-.5882118	.1499914	-3.92	0.000	-.882444	-.2939795
4. Not at all interested	-1.259279	.1992371	-6.32	0.000	-1.650115	-.8684434
_cons	6.663423	.2739974	24.32	0.000	6.125932	7.200913

```
. testparm i.polintr

 ( 1)  2.polintr = 0
 ( 2)  3.polintr = 0
 ( 3)  4.polintr = 0

       F(  3,  1399) =   15.48
            Prob > F =    0.0000

. testparm i.regions

 ( 1)  2.regions = 0
 ( 2)  3.regions = 0
 ( 3)  4.regions = 0
 ( 4)  5.regions = 0

       F(  4,  1399) =    3.15
            Prob > F =    0.0138
```

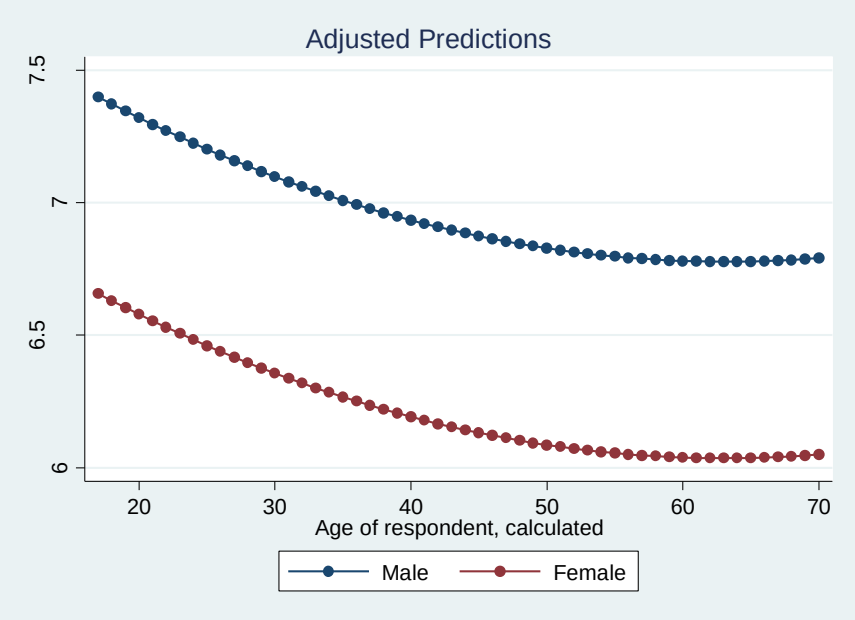
```
. * Model 2
. regress insttrust female agea c.agea#c.agea eduyrs i.regions i.polintr i.polintr#c.female
```

Source	SS	df	MS		
Model	302.513525	14	21.6081089	Number of obs	= 1410
Residual	3147.87171	1395	2.25653886	F(14, 1395)	= 9.58
				Prob > F	= 0.0000
Total	3450.38523	1409	2.44881848	R-squared	= 0.0877
				Adj R-squared	= 0.0785
				Root MSE	= 1.5022

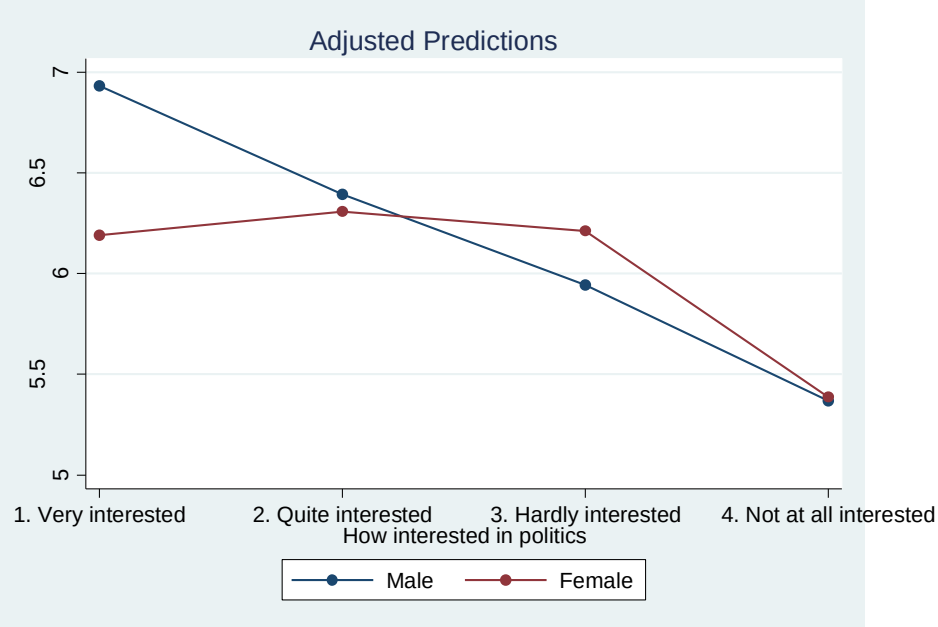
insttrust	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
female	-.7422764	.2708637	-2.74	0.006	-1.27362	-.2109323
agea	-.0368826	.0112304	-3.28	0.001	-.0589129	-.0148522
c.agea#c.agea	.0002921	.0001155	2.53	0.012	.0000655	.0005187
eduyrs	.0560367	.0121948	4.60	0.000	.0321145	.0799589
regions						
Agder and Rogaland	-.1656987	.1263012	-1.31	0.190	-.4134594	.082062
Western Norway	-.0888885	.112281	-0.79	0.429	-.3091462	.1313692
Trøndelag	.1803555	.1398191	1.29	0.197	-.0939228	.4546337
Northern Norway	-.4323831	.1452908	-2.98	0.003	-.717395	-.1473711
polintr						
2. Quite interested	-.5406308	.1805272	-2.99	0.003	-.8947648	-.1864967
3. Hardly interested	-.9912959	.1871873	-5.30	0.000	-1.358495	-.624097
4. Not at all interested	-1.565529	.2589123	-6.05	0.000	-2.073429	-1.05763
polintr#c.female						
2. Quite interested	.6571547	.2991533	2.20	0.028	.0703158	1.243994
3. Hardly interested	1.012231	.2981446	3.40	0.001	.4273709	1.597091
4. Not at all interested	.7611896	.3862881	1.97	0.049	.0034213	1.518958
_cons	7.381691	.3366289	21.93	0.000	6.721338	8.042045

```
. testparm agea c.agea#c.agea

 ( 1)  agea = 0
 ( 2)  c.agea#c.agea = 0

       F(  2,  1395) =   11.18
            Prob > F =    0.0000

. testparm i.polintr#c.female

 ( 1)  2.polintr#c.female = 0
 ( 2)  3.polintr#c.female = 0
 ( 3)  4.polintr#c.female = 0

       F(  3,  1395) =    4.21
            Prob > F =    0.0057
```

```
. * Conditional effect plot from Model 2
. quietly: margins, at(eduyrs=(10) regions=(1) polintr=(1) agea=(17/70) female=(0 1))

. marginsplot, noci

  Variables that uniquely identify margins: agea female
```




```
. do "C:\Users\arible\AppData\Local\Temp\STD0e000000.tmp"

. quietly: margins, at(eduyrs=(10) agea=(40) regions=(1) polintr=(1 2 3 4) female=(0 1))

. marginsplot, noci

  Variables that uniquely identify margins: polintr female
```



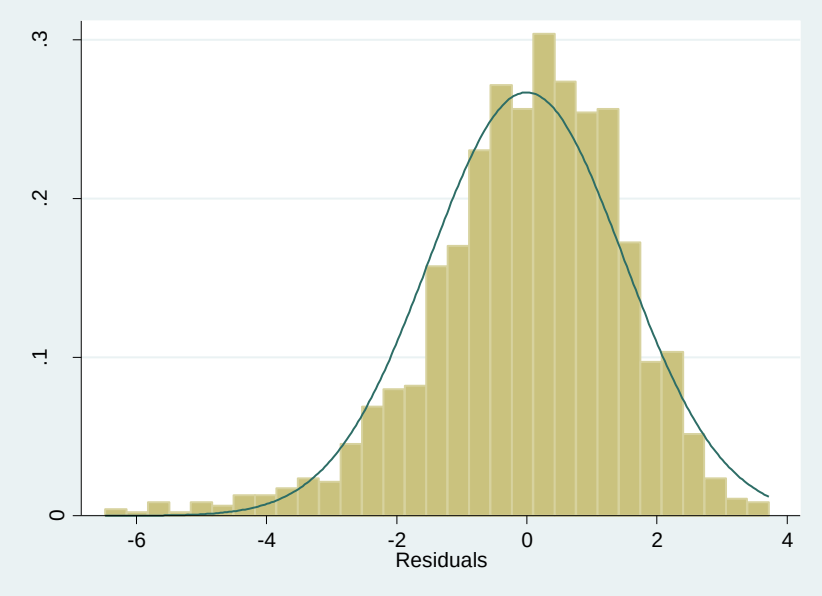
```
. * TASK 2
. * Link test for model specification
. linktest
```

Source	SS	df	MS		
Model	306.600519	2	153.30026	Number of obs	= 1410
Residual	3143.78471	1407	2.23438857	F(2, 1407)	= 68.61
				Prob > F	= 0.0000
				R-squared	= 0.0889
				Adj R-squared	= 0.0876
Total	3450.38523	1409	2.44881848	Root MSE	= 1.4948

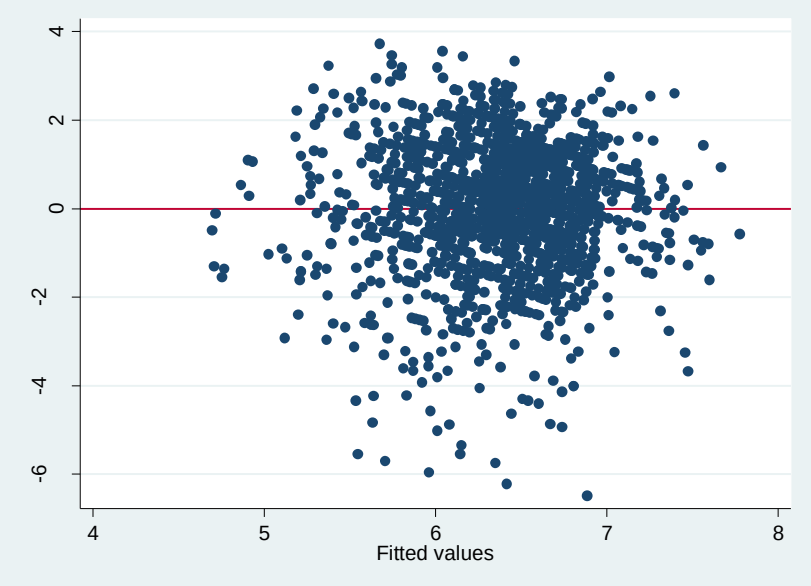
insttrust	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_hat	3.066591	1.530443	2.00	0.045	.0643943	6.068787
_hatsq	-.1645522	.1216692	-1.35	0.176	-.4032248	.0741204
_cons	-6.451812	4.801933	-1.34	0.179	-15.87153	2.967907

```
. * Ramsey's regression specification error test
. ovtest

Ramsey RESET test using powers of the fitted values of insttrust
       Ho:  model has no omitted variables
                 F(3, 1392) =      1.44
                   Prob > F =      0.2304

. * Breusch-Pagan (1979) and Cook-Weisberg (1983) test for heteroskedasticity
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of insttrust

         chi2(1)      =    32.26
         Prob > chi2  =   0.0000
```

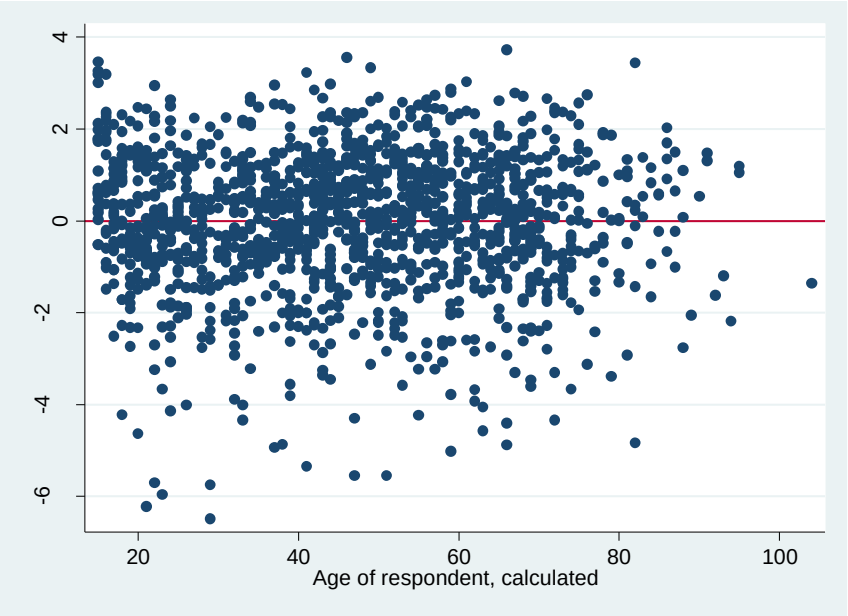
```
. * Tests of residual from Model 2
. predict residual, residual
(26 missing values generated)

. histogram residual, normal
(bin=31, start=-6.4846292, width=.32928309)
```



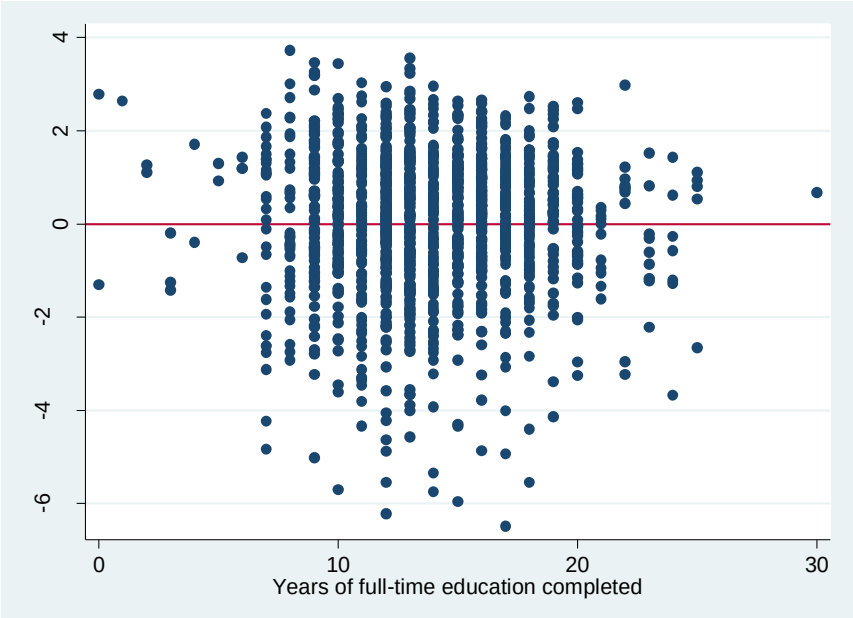

```
. * Residual-versus-fitted plot from Model 2
. rvfplot, yline(0)
```



```
. * Residual-versus-predictor plots from Model 2
. rvpplot agea, yline(0)
```




```
. rvpplot eduyrs, yline(0)
```



```
. * Test of collinearity in Model 2
. vif
```

Variable	VIF	1/VIF
female	11.41	0.087626
agea	27.18	0.036798
c.agea#		
c.agea	27.38	0.036521
eduyrs	1.28	0.781119
regions		
2	1.09	0.913926
3	1.11	0.898778
4	1.09	0.921232
5	1.08	0.925029
polintr		
2	4.89	0.204610
3	5.33	0.187781
4	3.26	0.306610
polintr#		
c.female		
2	8.21	0.121833
3	9.33	0.107228
4	3.74	0.267505
Mean VIF	7.60	

```
.
end of do-file
```