

Institutt for sosiologi og statsvitenskap

**Eksamensoppgave i SOS3003 Anvendt statistisk
dataanalyse i samfunnsvitenskap
Examination paper for SOS3003 Applied Social Statistics**

Faglig kontakt under eksamen: Arild Blekesaune

Telefon: 911 89 768

Eksamensdato: 29.05.2015

Eksamenstid: 6 timer

Hjelpemiddelkode/Tillatte hjelpemidler: Alle kalkulatorer

Sensurdato: 18.06.2015

Målform/språk/language: Bokmål/Nynorsk/English

Antall sider: 1 side Bokmål – 1 side Nynorsk – 1 page English

Antall sider vedlegg / Number of pages Appendix: 11 (pp. 5-15)

BOKMÅL

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

Oppgave 1

Beskriv de viktigste forutsetningene for OLS-regresjon, og presenter noen metoder som kan brukes for å teste om modellforutsetningene er tilfredsstillt.

Oppgave 2

Beskriv de viktigste forutsetningene for logistisk regresjon, og presenter noen metoder som kan brukes for å teste om modellforutsetningene er tilfredsstillt.

Oppgave 3

Vedlegg 1 gjengir en loggfil fra statistikkprogrammet Stata, og viser en analyse av data fra det norske utvalget av European Social Survey fra 2012. Beskriv hva denne loggfilen viser.

NYNORSK

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

Oppgåve 1

Grei ut om dei viktigaste føresetnadene for OLS-regresjon, og vis nokre metodar som kan brukast for å teste om desse modellføresetnadene er tilfredsstilt.

Oppgåve 2

Grei ut om dei viktigaste føresetnadene for logistisk regresjon, og vis nokre metodar som kan brukast for å teste om desse modellføresetnadene er tilfredsstilt.

Oppgåve 3

Vedlegg 1 viser ei loggfil frå statistikkprogrammet Stata, og presenterer en analyse av data frå det norske utvalet frå European Social Survey frå 2012. Grei ut om kva denne loggfila viser.

ENGLISH

Answer the following three questions (Each assignment counts a third of the total grade)

Question 1

Describe the most important prerequisites for OLS regression and show some methods that can be used to test whether the model assumptions are satisfied.

Question 2

Describe the most important prerequisites for logistic regression, and show some methods that can be used to test whether the model assumptions are satisfied.

Question 3

Appendix 1 reproduces a log file from the statistical program Stata, and shows an analysis of data from the Norwegian sample from European Social Survey 2012. Describe what this log file shows.

Vedlegg 1 / Appendix 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	24	1.51	1.51
1. 1	17	1.07	2.58
2. 2	33	2.07	4.65
3. 3	91	5.72	10.36
4. 4	126	7.91	18.28
5. 5	240	15.08	33.35
6. 6	237	14.89	48.24
7. 7	327	20.54	68.78
8. 8	315	19.79	88.57
9. 9	105	6.60	95.16
10. Complete trust	77	4.84	100.00
Total	1,592	100.00	

-> tabulation of trstlgl

Trust in the legal system	Freq.	Percent	Cum.
0. No trust at all	14	0.88	0.88
1. 1	8	0.50	1.38
2. 2	20	1.26	2.64
3. 3	42	2.64	5.28
4. 4	68	4.27	9.55
5. 5	132	8.29	17.84
6. 6	151	9.48	27.32
7. 7	304	19.10	46.42
8. 8	447	28.08	74.50
9. 9	279	17.53	92.02
10. Complete trust	127	7.98	100.00
Total	1,592	100.00	

-> tabulation of trstplc

Trust in the police	Freq.	Percent	Cum.
0. No trust at all	8	0.50	0.50
1. 1	11	0.69	1.19
2. 2	20	1.26	2.45
3. 3	38	2.39	4.84
4. 4	72	4.52	9.36
5. 5	127	7.98	17.34
6. 6	188	11.81	29.15
7. 7	346	21.73	50.88
8. 8	416	26.13	77.01
9. 9	237	14.89	91.90
10. Complete trust	129	8.10	100.00
Total	1,592	100.00	

-> tabulation of trstplt

Trust in politicians	Freq.	Percent	Cum.
0. No trust at all	37	2.32	2.32
1. 1	30	1.88	4.21
2. 2	86	5.40	9.61
3. 3	164	10.30	19.91
4. 4	234	14.70	34.61
5. 5	345	21.67	56.28
6. 6	299	18.78	75.06
7. 7	248	15.58	90.64
8. 8	114	7.16	97.80
9. 9	22	1.38	99.18
10. Complete trust	13	0.82	100.00
Total	1,592	100.00	

-> tabulation of trstprt

Trust in political parties	Freq.	Percent	Cum.
0. No trust at all	28	1.76	1.76
1. 1	23	1.44	3.20
2. 2	89	5.59	8.79
3. 3	147	9.23	18.03
4. 4	224	14.07	32.10
5. 5	401	25.19	57.29
6. 6	296	18.59	75.88
7. 7	249	15.64	91.52
8. 8	92	5.78	97.30
9. 9	28	1.76	99.06
10. Complete trust	15	0.94	100.00
Total	1,592	100.00	

```
. alpha trstprl trstlgl trstplc trstplt trstprt
```

```
Test scale = mean(unstandardized items)
```

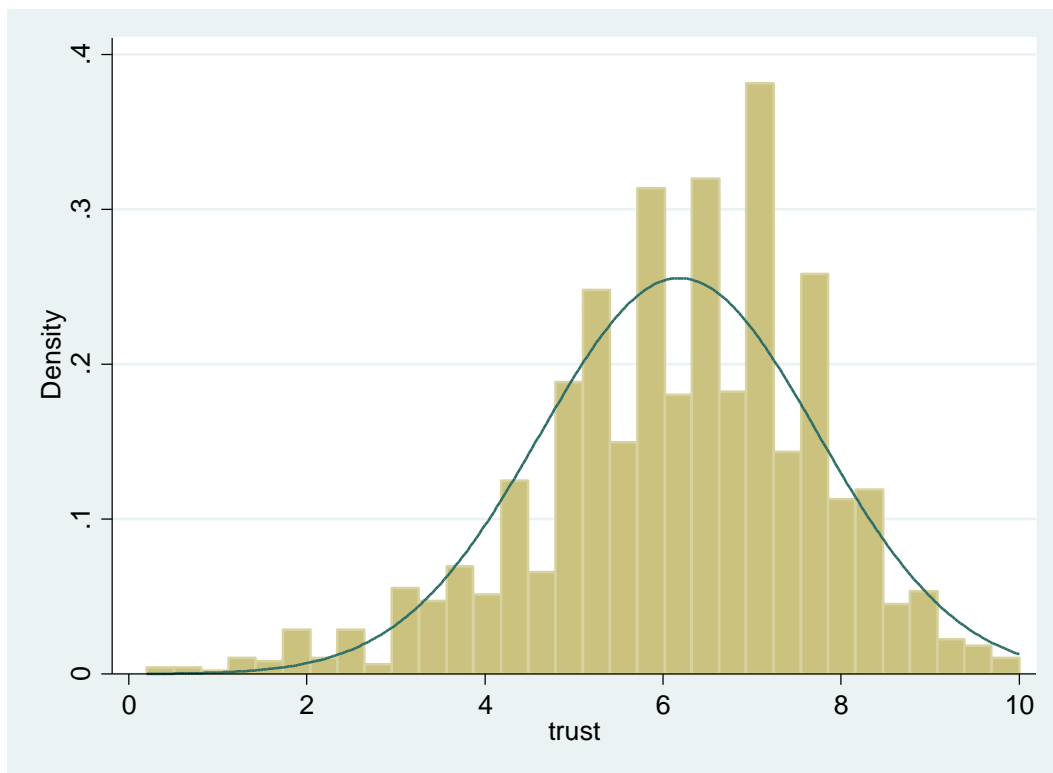
```
Average interitem covariance:    2.085772  
Number of items in the scale:    5  
Scale reliability coefficient:    0.8570
```

```
. generate trust=(trstprl+trstlgl+trstplc+trstplt+trstprt)/5
```

```
. summarize trust
```

Variable	Obs	Mean	Std. Dev.	Min	Max
trust	1592	6.18304	1.560106	.2	10

```
. histogram trust, normal  
(bin=32, start=.2, width=.30625)
```



. * Continous independent variables
 . summarize agea eduyrs

Variable	Obs	Mean	Std. Dev.	Min	Max
agea	1592	46.13191	18.07315	15	95
eduyrs	1592	13.09925	4.376694	0	28

. * Categorical independent variables
 . tab1 female polintr prtclano

-> tabulation of female

RECODE of gndr (Gender)	Freq.	Percent	Cum.
0. Male	843	52.95	52.95
1. Female	749	47.05	100.00
Total	1,592	100.00	

-> tabulation of polintr

How interested in politics	Freq.	Percent	Cum.
1. Very interested	173	10.87	10.87
2. Quite interested	645	40.52	51.38
3. Hardly interested	674	42.34	93.72
4. Not at all interested	100	6.28	100.00
Total	1,592	100.00	

-> tabulation of prtclano

Which party feel closer to, Norway	Freq.	Percent	Cum.
1. The Party Red (RØDT)	10	0.63	0.63
2. Socialist Left Party (SV)	73	4.59	5.21
3. Labour Party (A)	347	21.80	27.01
4. Liberal party (V)	42	2.64	29.65
5. Christian Democratic Party (KRF)	65	4.08	33.73
6. Centre Party (SP)	60	3.77	37.50
7. Conservative Party (H)	308	19.35	56.85
8. Progress Party (FRP)	106	6.66	63.51
10. Other	11	0.69	64.20
66. Not applicable	549	34.48	98.68
77. Refusal	16	1.01	99.69
88. Don't know	5	0.31	100.00
Total	1,592	100.00	


```
. * Model 1
. regress trust female agea eduyrs i.polintr i.prtclano
```

Source	SS	df	MS	Number of obs =	1592
Model	538.256271	17	31.6621336	F(17, 1574) =	14.95
Residual	3334.12582	1574	2.1182502	Prob > F =	0.0000
				R-squared =	0.1390
				Adj R-squared =	0.1297
Total	3872.38209	1591	2.43392966	Root MSE =	1.4554

	trust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]		
	female	-.0489335	.075156	-0.65	0.515	-.1963498	.0984829	
	agea	-.0080424	.0021259	-3.78	0.000	-.0122122	-.0038726	
	eduyrs	.0239309	.008788	2.72	0.007	.0066936	.0411683	
	polintr							
	2. Quite interested	.1377641	.1277617	1.08	0.281	-.112837	.3883652	
	3. Hardly interested	-.0566408	.1307119	-0.43	0.665	-.3130286	.1997469	
	4. Not at all interested	-.818044	.1951033	-4.19	0.000	-1.200734	-.4353543	
	prtclano							
	2. Socialist Left Party (SV)	.647024	.4976767	1.30	0.194	-.3291551	1.623203	
	3. Labour Party (A)	.63161	.4744432	1.33	0.183	-.2989972	1.562217	
	4. Liberal party (V)	-.1259836	.5183921	-0.24	0.808	-1.142795	.8908283	
	5. Christian Democratic Party ..)	.2269619	.5029184	0.45	0.652	-.7594986	1.213422	
	6. Centre Party (SP)	.1466904	.5052269	0.29	0.772	-.8442982	1.137679	
	7. Conservative Party (H)	-.1703456	.4746572	-0.36	0.720	-1.101373	.7606814	
	8. Progress Party (FRP)	-1.300249	.4878329	-2.67	0.008	-2.25712	-.3433784	
	10. Other	-1.004744	.6387639	-1.57	0.116	-2.257661	.248174	
	66. Not applicable	-.1131736	.4733536	-0.24	0.811	-1.041644	.8152964	
	77. Refusal	-.3372925	.5967933	-0.57	0.572	-1.507886	.8333011	
	88. Don't know	-.0447874	.8021448	-0.06	0.955	-1.618172	1.528597	
	_cons	6.273369	.4914885	12.76	0.000	5.309327	7.23741	

```
. * Model 2
. regress trust female agea eduyrs c.agea#c.agea ib4.polintr ib8.prtclano
```

Source	SS	df	MS	Number of obs =	1592
Model	553.562545	18	30.7534747	F(18, 1573) =	14.58
Residual	3318.81955	1573	2.10986621	Prob > F =	0.0000
				R-squared =	0.1430
				Adj R-squared =	0.1331
Total	3872.38209	1591	2.43392966	Root MSE =	1.4525

	trust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	female	-.052982	.0750222	-0.71	0.480	-.200136	.094172
	agea	-.0359689	.0105832	-3.40	0.001	-.0567275	-.0152103
	eduyrs	.0308586	.0091399	3.38	0.001	.0129309	.0487863
	c.agea#c.agea	.0002955	.0001097	2.69	0.007	.0000803	.0005106
	polintr						
	1. Very interested	.8330259	.1947962	4.28	0.000	.4509383	1.215113
	2. Quite interested	.9926875	.1661194	5.98	0.000	.6668488	1.318526
	3. Hardly interested	.7863958	.15981	4.92	0.000	.4729328	1.099859
	prtclano						
	1. The Party Red (RØDT)	1.293494	.4868729	2.66	0.008	.3385056	2.248482
	2. Socialist Left Party (SV)	1.910465	.2257157	8.46	0.000	1.46773	2.3532
	3. Labour Party (A)	1.913336	.1629961	11.74	0.000	1.593624	2.233049
	4. Liberal party (V)	1.154765	.2664533	4.33	0.000	.6321242	1.677406
	5. Christian Democratic Party ..)	1.484425	.232463	6.39	0.000	1.028455	1.940395
	6. Centre Party (SP)	1.439534	.2355013	6.11	0.000	.9776047	1.901464
	7. Conservative Party (H)	1.108788	.1649962	6.72	0.000	.7851525	1.432424
	10. Other	.2862737	.4627792	0.62	0.536	-.6214554	1.194003
	66. Not applicable	1.157838	.1564571	7.40	0.000	.8509518	1.464725
	77. Refusal	.9215103	.3937253	2.34	0.019	.1492287	1.693792
	88. Don't know	1.228738	.6673705	1.84	0.066	-.0802914	2.537767
	_cons	4.625189	.2940898	15.73	0.000	4.04834	5.202039

```
. * Model 3
. regress trust female agea eduyrs c.agea#c.agea ib4.polintr ib8.prtclano c.agea#c.eduyrs
```

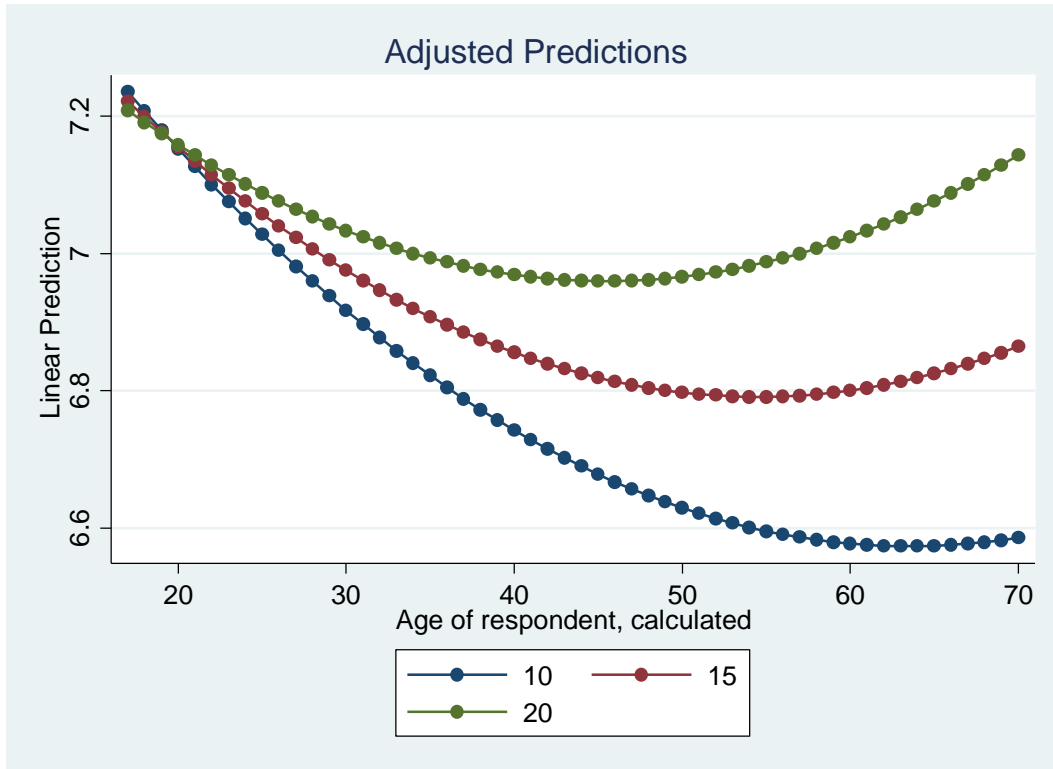
Source	SS	df	MS				
Model	564.229722	19	29.6963011	Number of obs =	1592		
Residual	3308.15237	1572	2.10442263	F(19, 1572) =	14.11		
				Prob > F =	0.0000		
				R-squared =	0.1457		
				Adj R-squared =	0.1354		
Total	3872.38209	1591	2.43392966	Root MSE =	1.4507		

	trust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]		
	female	-.0551471	.0749315	-0.74	0.462	-.2021233	.0918291	
	agea	-.0498927	.0122459	-4.07	0.000	-.0739127	-.0258727	
	eduyrs	-.0214594	.0249662	-0.86	0.390	-.0704301	.0275112	
	c.agea#c.agea	.000306	.0001097	2.79	0.005	.0000909	.0005211	
	polintr							
	1. Very interested	.8479042	.194657	4.36	0.000	.4660896	1.229719	
	2. Quite interested	1.011298	.1661107	6.09	0.000	.6854758	1.33712	
	3. Hardly interested	.8070201	.1598664	5.05	0.000	.4934463	1.120594	
	prtclano							
	1. The Party Red (RØDT)	1.260921	.4864596	2.59	0.010	.3067429	2.215099	
	2. Socialist Left Party (SV)	1.890478	.2255991	8.38	0.000	1.447971	2.332985	
	3. Labour Party (A)	1.921428	.1628254	11.80	0.000	1.60205	2.240806	
	4. Liberal party (V)	1.163888	.2661401	4.37	0.000	.641861	1.685915	
	5. Christian Democratic Party ..)	1.497692	.2322377	6.45	0.000	1.042164	1.95322	
	6. Centre Party (SP)	1.425333	.2352819	6.06	0.000	.9638335	1.886832	
	7. Conservative Party (H)	1.100413	.1648252	6.68	0.000	.7771123	1.423713	
	10. Other	.3096746	.4622987	0.67	0.503	-.5971124	1.216462	
	66. Not applicable	1.159504	.1562569	7.42	0.000	.8530103	1.465998	
	77. Refusal	.9094363	.3932536	2.31	0.021	.1380795	1.680793	
	88. Don't know	1.186244	.6667762	1.78	0.075	-.1216199	2.494109	
	c.agea#c.eduyrs	.0011023	.0004896	2.25	0.024	.000142	.0020627	
	_cons	5.252916	.4049717	12.97	0.000	4.458574	6.047258	

```
. * Conditional effect plot from Model 3
. quietly: margins, at(female=(0) polint=(1) prtclano=(3) age=(17/70) eduys=(10 15 20))

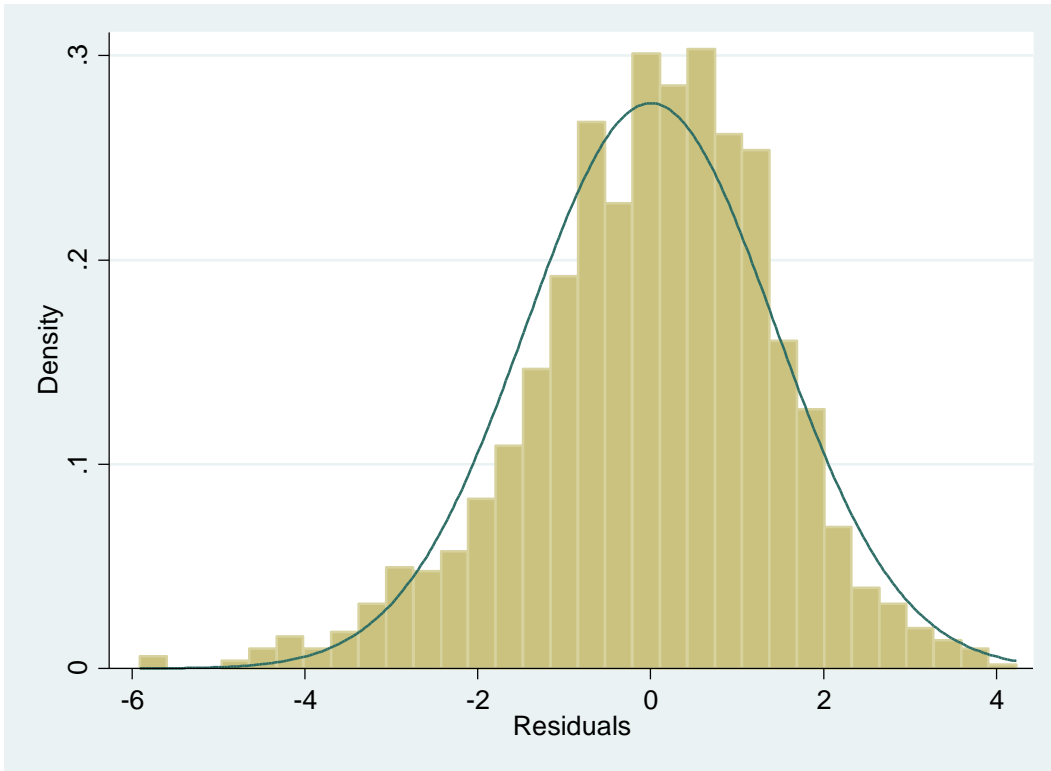
. marginsplot, noci

Variables that uniquely identify margins: agea eduys
```

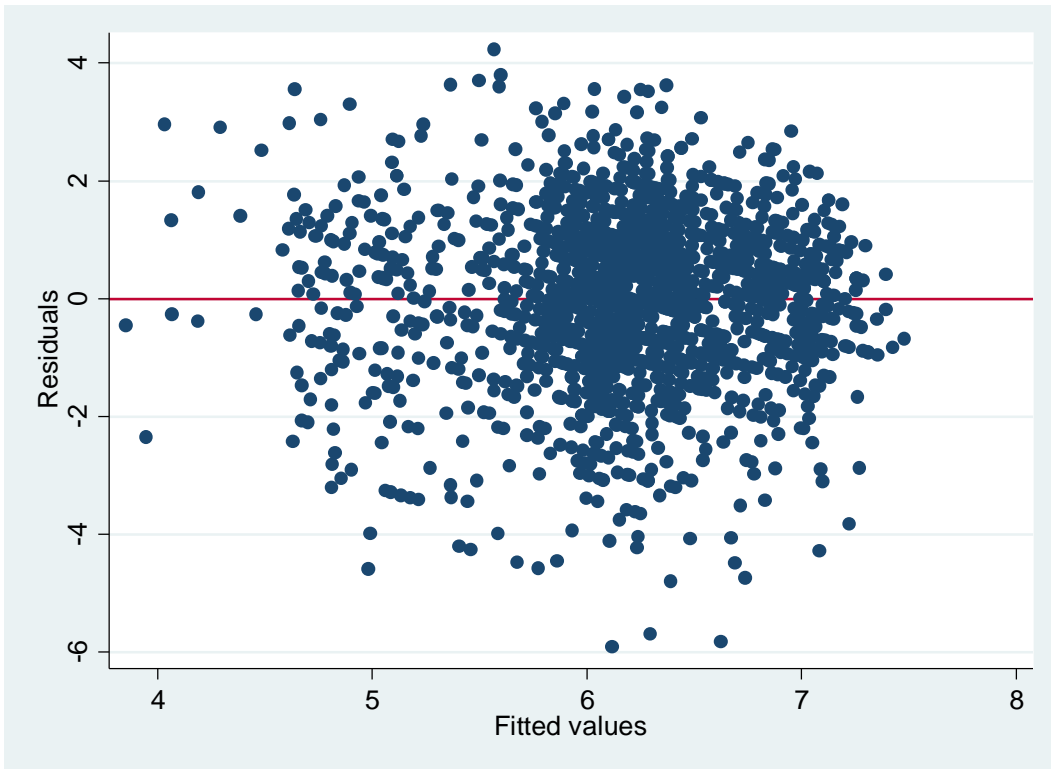


```
. * Tests of residual from Model 3
. predict residual, residual

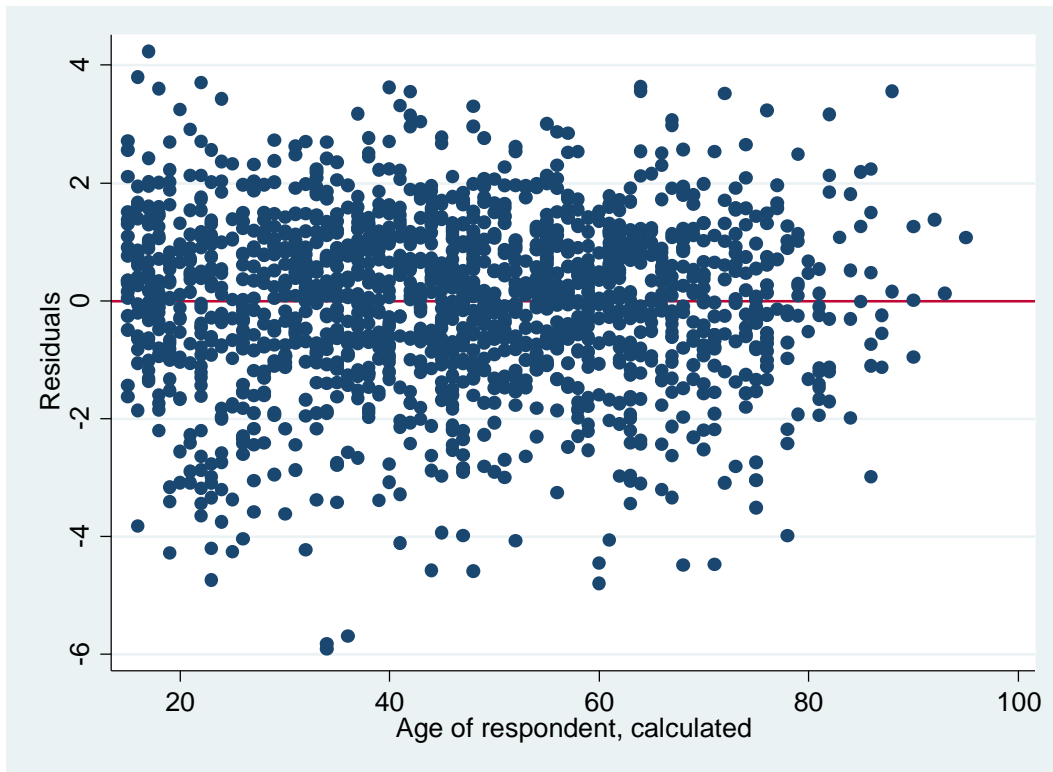
. histogram residual, normal
(bin=32, start=-5.9171095, width=.31708685)
```



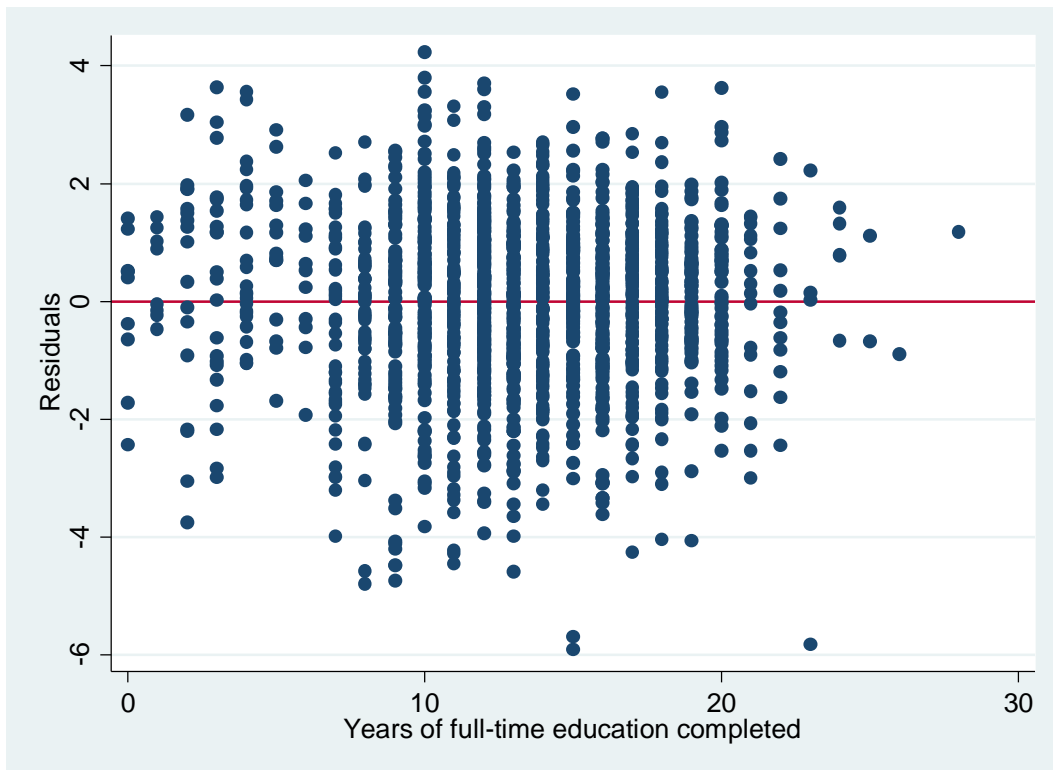
```
. * Residual-versus-fitted plot from Model 3
. rvfplot, yline(0)
```



```
. * Residual-versus-predictor plots from Model 3  
. rvpplot agea, yline(0)
```



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



```

. * Cook's D from Model 3
. predict cook, cooksD

. sort cook

. summarize cook, d

```

Cook's D

Percentiles		Smallest		
1%	6.61e-08	4.07e-11		
5%	1.17e-06	6.03e-10		
10%	5.85e-06	1.69e-09	Obs	1592
25%	.0000376	2.60e-09	Sum of Wgt.	1592
50%	.0001689		Mean	.0006096
		Largest	Std. Dev.	.0016106
75%	.0005466	.0132838		
90%	.0014485	.0158598	Variance	2.59e-06
95%	.002673	.0223062	Skewness	11.30139
99%	.0058992	.0375138	Kurtosis	208.0185

```

. * Test of collinearity in Model 3
. vif

```

Variable	VIF	1/VIF
female	1.06	0.945013
agea	37.03	0.027003
eduyrs	9.03	0.110781
c.agea#		
c.agea	27.76	0.036027
polintr		
1	2.78	0.360170
2	5.03	0.198779
3	4.72	0.211865
prtclano		
1	1.12	0.894903
2	1.68	0.593637
3	3.42	0.292505
4	1.38	0.726565
5	1.60	0.625833
6	1.52	0.658399
7	3.21	0.311827
10	1.11	0.901376
66	4.17	0.239630
77	1.16	0.859121
88	1.05	0.949662
c.agea#		
c.eduyrs	15.44	0.064771
Mean VIF	6.54	

```

.
end of do-file

```