

Institutt for sosiologi og statsvitenskap

Eksamensoppgave i SOS3003 Anvendt statistisk dataanalyse i samfunnsvitenskap

Faglig kontakt under eksamen: Arild Blekesaune

Telefon: 911 89 768

Eksamensdato: 10.12.2015

Eksamenstid: 6 timer

Hjelpemiddelkode/Tillatte hjelpemidler: Alle kalkulatorer

Sensurdato: 08.01.2016

Målform/språk/language: Bokmål

Antall sider: 1 side Bokmål

Antall sider vedlegg: 12 (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. 3-8).

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. 9-12).

Oppgave 3

Beskriv estimatene fra modell 3 i vedlegg 1 (s. 13), og drøft i hvilken grad modell 3 tilfredsstillere forutsetningene for logistisk regresjon.

Vedlegg 1

```
. * TASK 1
. * Dependent variable
. tab1 ppltrst pplfair pplhlp
```

-> tabulation of ppltrst

Most people can be trusted or you can't be too careful	Freq.	Percent	Cum.
0. You can't be too careful	6	0.42	0.42
1. 1	9	0.63	1.05
2. 2	15	1.05	2.09
3. 3	59	4.11	6.20
4. 4	65	4.53	10.73
5. 5	235	16.38	27.11
6. 6	157	10.94	38.05
7. 7	380	26.48	64.53
8. 8	376	26.20	90.73
9. 9	96	6.69	97.42
10. Most people can be trusted	37	2.58	100.00
Total	1,435	100.00	

-> tabulation of pplfair

Most people try to take advantage of you, or try to be fair	Freq.	Percent	Cum.
0. Most people try to take advantage of	6	0.42	0.42
1. 1	2	0.14	0.56
2. 2	13	0.91	1.46
3. 3	35	2.44	3.91
4. 4	59	4.11	8.02
5. 5	166	11.58	19.60
6. 6	171	11.92	31.52
7. 7	409	28.52	60.04
8. 8	390	27.20	87.24
9. 9	133	9.27	96.51
10. Most people try to be fair	50	3.49	100.00
Total	1,434	100.00	

-> tabulation of pplhlp

Most of the time people helpful or mostly looking out for themselves	Freq.	Percent	Cum.
0. People mostly look out for themselve	7	0.49	0.49
1. 1	6	0.42	0.91
2. 2	22	1.54	2.44
3. 3	111	7.75	10.19
4. 4	128	8.93	19.12
5. 5	264	18.42	37.54
6. 6	245	17.10	54.64
7. 7	320	22.33	76.97
8. 8	231	16.12	93.09
9. 9	76	5.30	98.39
10. People mostly try to be helpful	23	1.61	100.00
Total	1,433	100.00	

```
. alpha ppltrst pplfair pplhlp

Test scale = mean(unstandardized items)

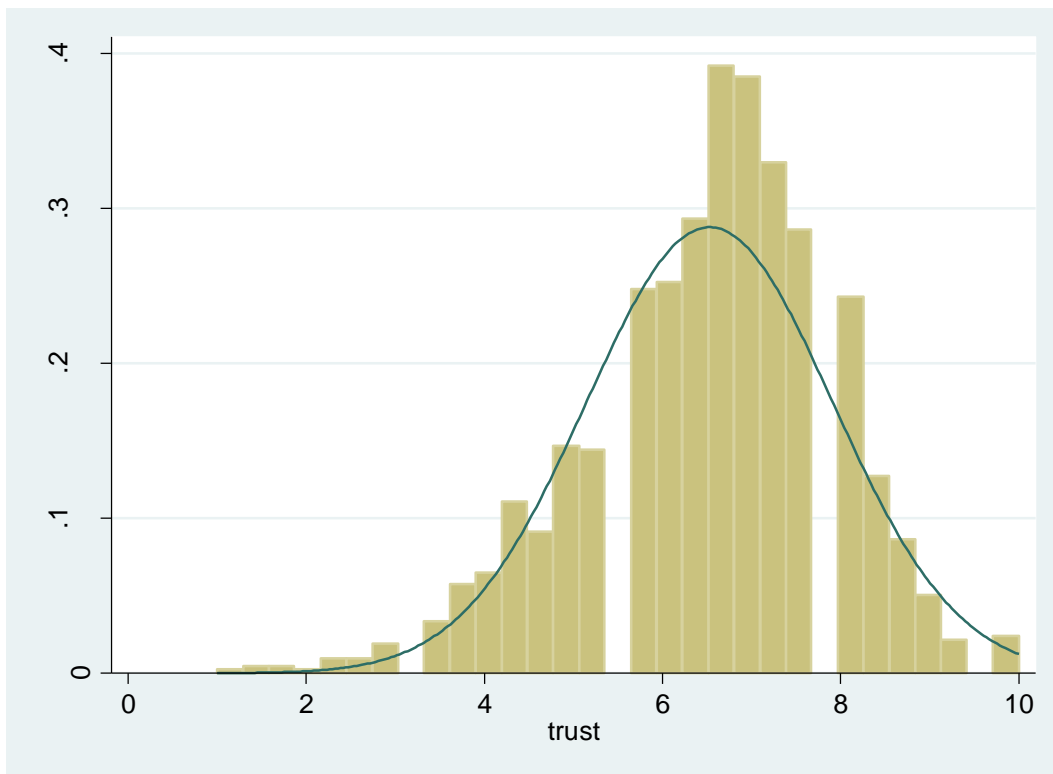
Average interitem covariance:    1.328456
Number of items in the scale:    3
Scale reliability coefficient:    0.6931

. generate trust=(ppltrst+pplfair+pplhlp)/3
(5 missing values generated)

. summarize trust
```

Variable	Obs	Mean	Std. Dev.	Min	Max
trust	1431	6.528302	1.385053	1	10

```
. histogram trust, normal
(bin=31, start=1, width=.29032258)
```



. * Continous independent variables
 . summarize agea eduysr

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

. * Categorical independent variables
 . tab1 female prtclbno

-> 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 prtclbno

Which party feel closer to, Norway	Freq.	Percent	Cum.
1. The Party Red (RØDT)	9	0.63	0.63
2. Socialist Left Party (SV)	46	3.20	3.83
3. Labour Party (A)	371	25.84	29.67
4. Liberal party (V)	46	3.20	32.87
5. Christian Democratic Party (KRF)	47	3.27	36.14
6. Centre Party (SP)	65	4.53	40.67
7. Conservative Party (H)	228	15.88	56.55
8. Progress Party (FRP)	90	6.27	62.81
9. Coastal Party (KYST)	3	0.21	63.02
10. Green Party (MDG)	23	1.60	64.62
11. Other	8	0.56	65.18
66. Not applicable	476	33.15	98.33
77. Refusal	18	1.25	99.58
88. Don't know	6	0.42	100.00
Total	1,436	100.00	

```
. * Model 1
. regress trust female agea eduyrs i.prtclbno
```

Source	SS	df	MS	Number of obs =	1429
Model	278.462872	16	17.4039295	F(16, 1412) =	9.98
Residual	2462.76917	1412	1.7441708	Prob > F =	0.0000
				R-squared =	0.1016
				Adj R-squared =	0.0914
Total	2741.23204	1428	1.91963028	Root MSE =	1.3207

	trust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
	female	.1193801	.0708808	1.68	0.092	-.0196629	.258423
	agea	.0105	.0019285	5.44	0.000	.006717	.014283
	eduyrs	.0379086	.0097748	3.88	0.000	.0187339	.0570833
	prtclbno						
	2. Socialist Left Party (SV)	-.4892913	.481851	-1.02	0.310	-1.434512	.4559295
	3. Labour Party (A)	-.9492705	.4456242	-2.13	0.033	-1.823427	-.0751138
	4. Liberal party (V)	-1.104005	.4820196	-2.29	0.022	-2.049557	-.1584536
	5. Christian Democratic Party (KRF)	-.4127026	.4807733	-0.86	0.391	-1.355809	.5304043
	6. Centre Party (SP)	-.944277	.4698458	-2.01	0.045	-1.865948	-.0226062
	7. Conservative Party (H)	-1.137263	.4490375	-2.53	0.011	-2.018116	-.256411
	8. Progress Party (FRP)	-2.047237	.462436	-4.43	0.000	-2.954372	-1.140101
	9. Coastal Party (KYST)	-3.13376	.8809798	-3.56	0.000	-4.86193	-1.40559
	10. Green Party (MDG)	-1.148011	.5197627	-2.21	0.027	-2.167601	-.1284205
	11. Other	-1.025179	.6663622	-1.54	0.124	-2.332345	.2819873
	66. Not applicable	-1.193812	.4445835	-2.69	0.007	-2.065927	-.3216971
	77. Refusal	-1.100897	.5395327	-2.04	0.041	-2.159269	-.0425257
	88. Don't know	-2.72959	.6960709	-3.92	0.000	-4.095034	-1.364146
	_cons	6.569124	.4737384	13.87	0.000	5.639817	7.498431

```
. testparm i.prtclbno
```

- (1) 2.prtclbno = 0
- (2) 3.prtclbno = 0
- (3) 4.prtclbno = 0
- (4) 5.prtclbno = 0
- (5) 6.prtclbno = 0
- (6) 7.prtclbno = 0
- (7) 8.prtclbno = 0
- (8) 9.prtclbno = 0
- (9) 10.prtclbno = 0
- (10) 11.prtclbno = 0
- (11) 66.prtclbno = 0
- (12) 77.prtclbno = 0
- (13) 88.prtclbno = 0

```
F( 13, 1412) = 7.49
Prob > F = 0.0000
```

```
. * Model 2
. regress trust female agea eduysr ib8.prtclbno c.agea#c.female
```

Source	SS	df	MS	Number of obs =	1429
Model	285.262661	17	16.7801565	F(17, 1411) =	9.64
Residual	2455.96938	1411	1.74058779	Prob > F =	0.0000
				R-squared =	0.1041
				Adj R-squared =	0.0933
Total	2741.23204	1428	1.91963028	Root MSE =	1.3193

	trust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	female	-.2294699	.1901714	-1.21	0.228	-.6025191 .1435793
	agea	.0069753	.0026252	2.66	0.008	.0018257 .012125
	eduysr	.0391117	.0097837	4.00	0.000	.0199195 .0583039
	prtclbno					
	1. The Party Red (RØDT)	2.019413	.4621752	4.37	0.000	1.112789 2.926038
	2. Socialist Left Party (SV)	1.553868	.2421934	6.42	0.000	1.07877 2.028966
	3. Labour Party (A)	1.086268	.1567092	6.93	0.000	.7788603 1.393676
	4. Liberal party (V)	.940708	.2431125	3.87	0.000	.4638072 1.417609
	5. Christian Democratic Party (KRF)	1.592032	.240443	6.62	0.000	1.120368 2.063697
	6. Centre Party (SP)	1.105116	.2155743	5.13	0.000	.6822355 1.527997
	7. Conservative Party (H)	.903429	.1658437	5.45	0.000	.5781022 1.228756
	9. Coastal Party (KYST)	-1.080134	.7749966	-1.39	0.164	-2.600403 .4401358
	10. Green Party (MDG)	.87582	.3106164	2.82	0.005	.2665003 1.48514
	11. Other	1.017347	.5198622	1.96	0.051	-.0024387 2.037133
	66. Not applicable	.8398829	.1541121	5.45	0.000	.5375694 1.142196
	77. Refusal	.9202773	.3417265	2.69	0.007	.2499307 1.590624
	88. Don't know	-.6720859	.5567679	-1.21	0.228	-1.764268 .420096
	c.agea#c.female	.0074795	.0037842	1.98	0.048	.0000563 .0149027
	_cons	4.678607	.2323685	20.13	0.000	4.222782 5.134432

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

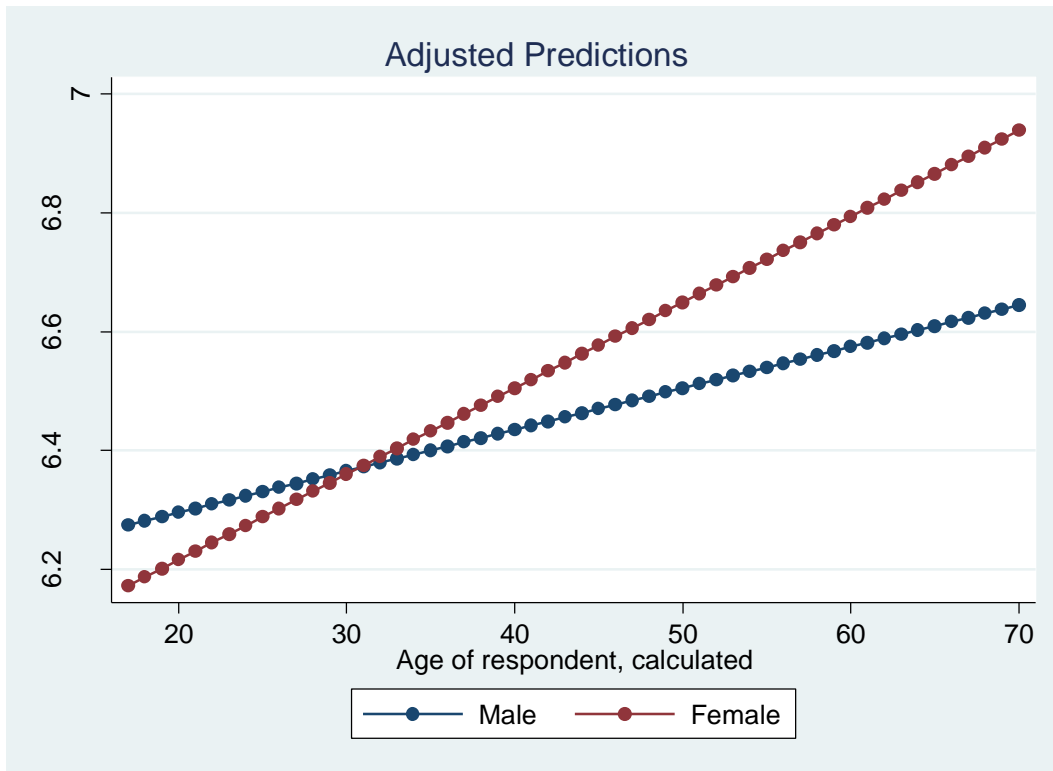
```
( 1) c.agea#c.female = 0
```

```
F( 1, 1411) = 3.91
Prob > F = 0.0483
```

```
. * Conditional effect plot from Model 2
. quietly: margins, at(eduysr=(10) prtclbno=(3)agea=(17/70) female=(0 1))

. marginsplot, noci

Variables that uniquely identify margins: agea female
```




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

Source	SS	df	MS	Number of obs =	1429
Model	286.957912	2	143.478956	F(2, 1426) =	83.37
Residual	2454.27413	1426	1.72108985	Prob > F =	0.0000
Total	2741.23204	1428	1.91963028	R-squared =	0.1047
				Adj R-squared =	0.1034
				Root MSE =	1.3119

trust	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
_hat	-.1106351	1.121757	-0.10	0.921	-2.311106	2.089835
_hatsq	.0866961	.087354	0.99	0.321	-.0846602	.2580523
_cons	3.538383	3.601283	0.98	0.326	-3.525998	10.60276

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

```
Ramsey RESET test using powers of the fitted values of trust
Ho: model has no omitted variables
F(3, 1408) = 4.60
Prob > F = 0.0033
```

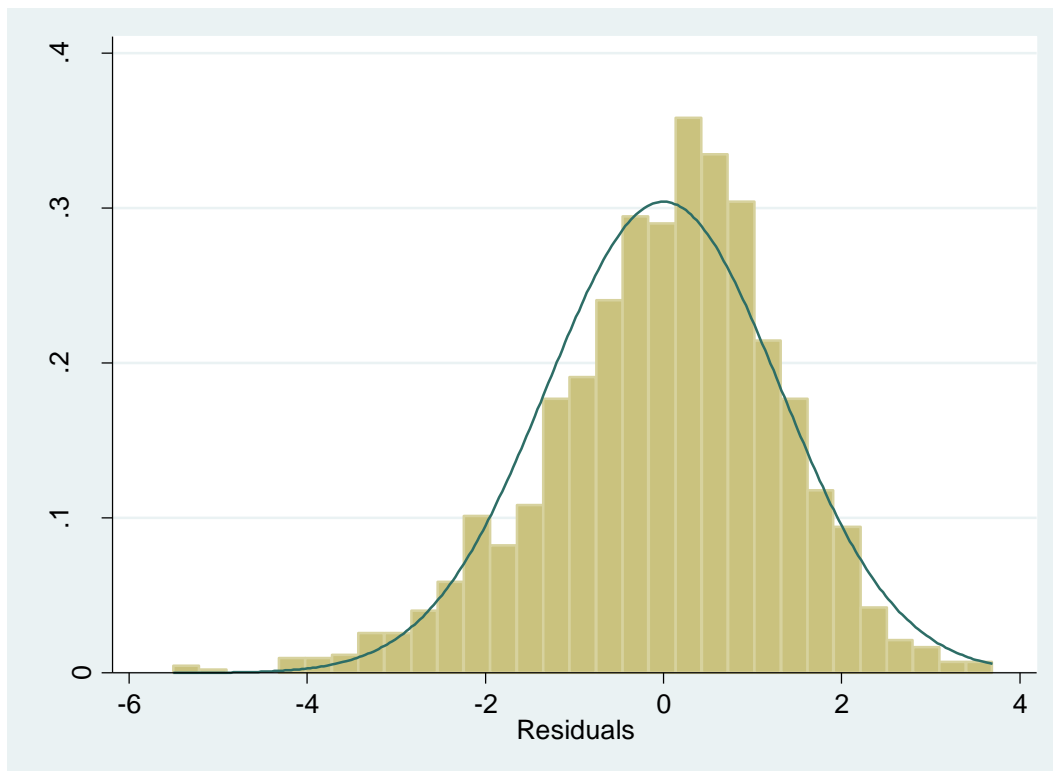
```
. * Breusch-Pagan (1979) and Cook-Weisberg (1983) test for heteroskedasticity
. estat hettest
```

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

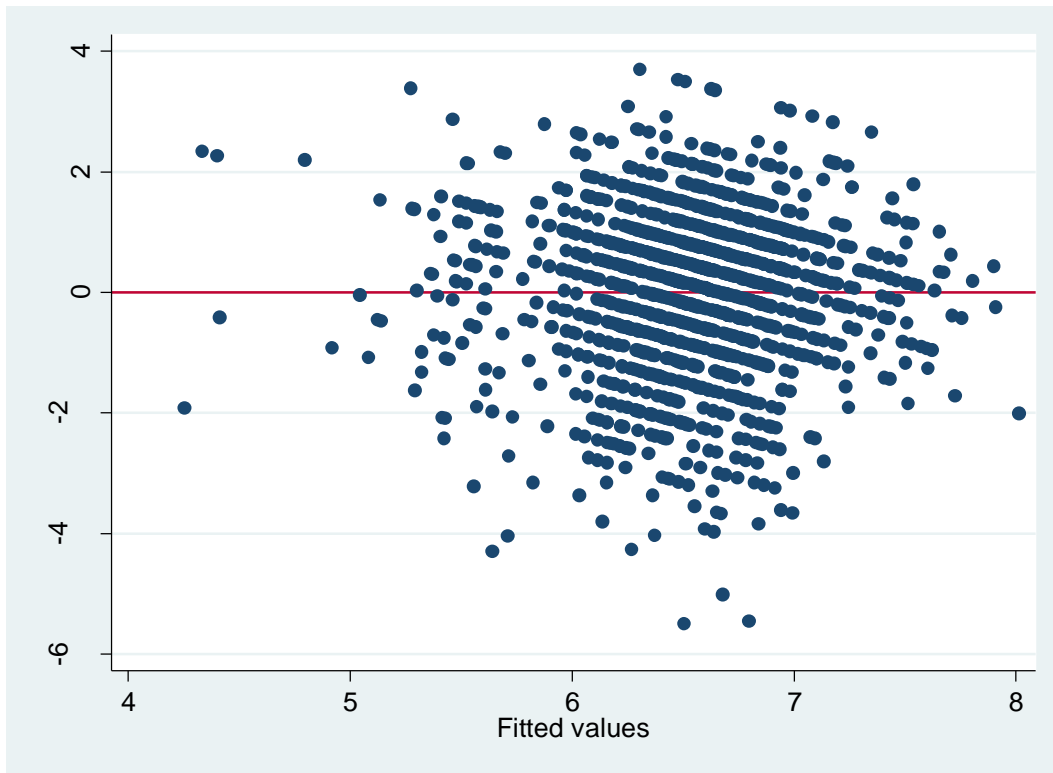
chi2(1) = 13.49
Prob > chi2 = 0.0002
```

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

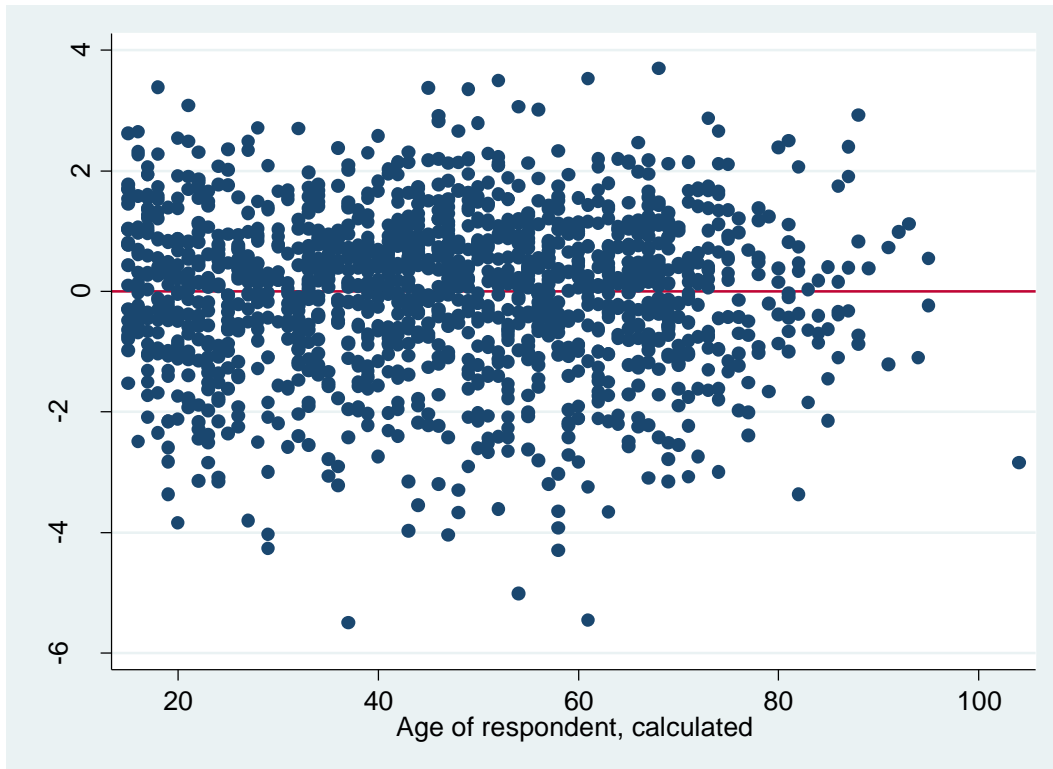
. histogram residual, normal
(bin=31, start=-5.505023, width=.29675212)
```



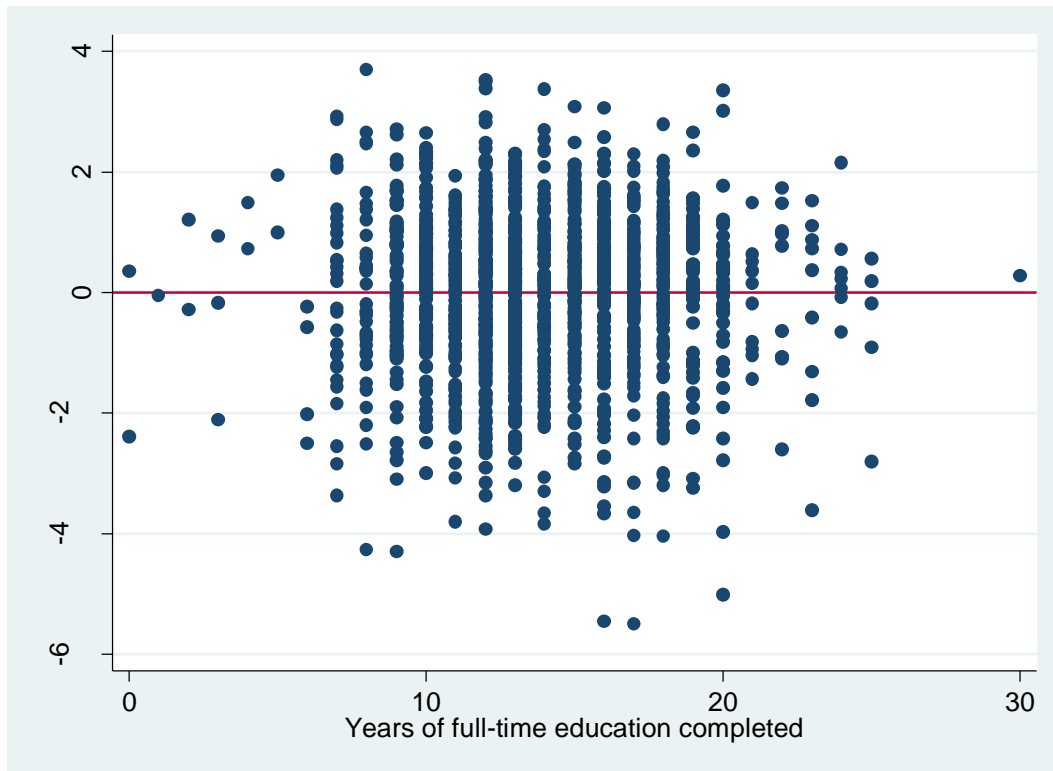
```
. * Residual-versus-fitted plot from Model 2  
. rvfplot, yline(0)
```



```
. * 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	7.39	0.135269
agea	1.97	0.508485
eduyrs	1.08	0.922924
prtclbno		
1	1.10	0.911138
2	1.50	0.666536
3	3.85	0.259404
4	1.51	0.661506
5	1.51	0.662366
6	1.66	0.603680
7	3.03	0.330257
9	1.03	0.968028
10	1.25	0.797198
11	1.08	0.924600
66	4.32	0.231352
77	1.19	0.838632
88	1.06	0.939775
c.agea#		
c.female	8.47	0.118016
Mean VIF	2.53	

```

. * TASK 3
. * Model 3: Logistic Regression Model
. recode trust 1/6.5=0 6.5/10=1, generate(trustdummy)
(1431 differences between trust and trustdummy)

. logit trustdummy female agea eduyrs ib8.prtclbno c.agea#c.female

```

```

Iteration 0: log likelihood = -978.23668
Iteration 1: log likelihood = -928.47577
Iteration 2: log likelihood = -928.17205
Iteration 3: log likelihood = -928.17113
Iteration 4: log likelihood = -928.17112

```

```

Logistic regression                Number of obs   =       1429
                                LR chi2(17)       =       100.13
                                Prob > chi2        =       0.0000
Log likelihood = -928.17112       Pseudo R2      =       0.0512

```

	trustdummy	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
	female	-.4854933	.2990496	-1.62	0.104	-1.07162 .1006331
	agea	.0088992	.0041121	2.16	0.030	.0008397 .0169587
	eduyrs	.0691304	.0158373	4.37	0.000	.0380899 .1001709
	prtclbno					
	1. The Party Red (RØDT)	2.821525	1.089749	2.59	0.010	.6856571 4.957393
	2. Socialist Left Party (SV)	2.067339	.4438224	4.66	0.000	1.197463 2.937215
	3. Labour Party (A)	1.232253	.25619	4.81	0.000	.7301298 1.734376
	4. Liberal party (V)	1.087933	.3894902	2.79	0.005	.324546 1.85132
	5. Christian Democratic Party (KRF)	1.778978	.4197314	4.24	0.000	.9563194 2.601636
	6. Centre Party (SP)	1.367893	.3485299	3.92	0.000	.684787 2.050999
	7. Conservative Party (H)	.9011462	.2686728	3.35	0.001	.3745572 1.427735
	9. Coastal Party (KYST)	.3327816	1.249737	0.27	0.790	-2.116658 2.782221
	10. Green Party (MDG)	1.064774	.4942968	2.15	0.031	.0959703 2.033578
	11. Other	1.414589	.8810458	1.61	0.108	-.3122291 3.141407
	66. Not applicable	.9321164	.2517935	3.70	0.000	.4386102 1.425623
	77. Refusal	1.096302	.5409448	2.03	0.043	.0360695 2.156534
	88. Don't know	.0215257	.9145613	0.02	0.981	-1.770981 1.814033
	c.agea#c.female	.0113072	.0060118	1.88	0.060	-.0004758 .0230902
	_cons	-2.161383	.379185	-5.70	0.000	-2.904572 -1.418194

```
. * Hosmer-Lemeshows goodnes-of-fit Statistics
. estat gof, table group(10)
```

Logistic model for trustdummy, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

Group	Prob	Obs_1	Exp_1	Obs_0	Exp_0	Total
1	0.3936	49	46.2	94	96.8	143
2	0.4645	68	62.3	75	80.7	143
3	0.5068	58	69.5	85	73.5	143
4	0.5413	69	74.9	74	68.1	143
5	0.5707	80	79.6	63	63.4	143
6	0.5985	83	83.6	60	59.4	143
7	0.6316	95	88.0	48	55.0	143
8	0.6676	85	92.9	58	50.1	143
9	0.7215	106	98.7	37	44.3	143
10	0.9235	115	112.1	27	29.9	142

```
number of observations = 1429
number of groups = 10
Hosmer-Lemeshow chi2(8) = 11.34
Prob > chi2 = 0.1830
```

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