

Quantitative Research Methods: Introduction to correlation and regression

Brendan Halpin, Sociology, University of Limerick

January 2018

Multicollinearity 1

```
. use http://www.stata-press.com/data/r14/bodyfat
(Body Fat)
. corr *
(obs=20)
```

	triceps	thigh	midarm	bodyfat
triceps	1.0000			
thigh	0.9238	1.0000		
midarm	0.4578	0.0847	1.0000	
bodyfat	0.8433	0.8781	0.1424	1.0000

Multicollinearity 2

```
. reg bodyfat tricep
```

Source	SS	df	MS	Number of obs	=	20
Model	352.269824	1	352.269824	F(1, 18)	=	44.30
Residual	143.119689	18	7.95109386	Prob > F	=	0.0000
Total	495.389513	19	26.0731323	R-squared	=	0.7111
				Adj R-squared	=	0.6950
				Root MSE	=	2.8198

bodyfat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
triceps	.8571866	.1287808	6.66	0.000	.5866282 1.127745
_cons	-1.496107	3.319235	-0.45	0.658	-8.46956 5.477347

Multicollinearity 3

```
. reg bodyfat thigh
```

Source	SS	df	MS	Number of obs	=	20
Model	381.965845	1	381.965845	F(1, 18)	=	60.62
Residual	113.423669	18	6.30131492	Prob > F	=	0.0000
Total	495.389513	19	26.0731323	R-squared	=	0.7710
				Adj R-squared	=	0.7583
				Root MSE	=	2.5102

bodyfat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
thigh	.8565467	.1100156	7.79	0.000	.6254124 1.087681
_cons	-23.63449	5.657414	-4.18	0.001	-35.52028 -11.74871

Multicollinearity 4

```
. reg bodyfat midarm
```

Source	SS	df	MS	Number of obs	=	20
Model	10.0516092	1	10.0516092	F(1, 18)	=	0.37
Residual	485.337904	18	26.9632169	Prob > F	=	0.5491
Total	495.389513	19	26.0731323	R-squared	=	0.0203
				Adj R-squared	=	-0.0341
				Root MSE	=	5.1926

bodyfat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
midarm	.1994287	.3266297	0.61	0.549	-.4867949 .8856523
_cons	14.68678	9.095926	1.61	0.124	-4.423052 33.79661

Multicollinearity 5

```
. reg bodyfat triceps thigh midarm
```

Source	SS	df	MS	Number of obs	=	20
Model	396.984607	3	132.328202	F(3, 16)	=	21.52
Residual	98.4049068	16	6.15030667	Prob > F	=	0.0000
Total	495.389513	19	26.0731323	R-squared	=	0.8014
				Adj R-squared	=	0.7641
				Root MSE	=	2.48

bodyfat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
triceps	4.334085	3.015511	1.44	0.170	-2.058512 10.72668
thigh	-2.856842	2.582015	-1.11	0.285	-8.330468 2.616785
midarm	-2.186056	1.595499	-1.37	0.190	-5.568362 1.19625
_cons	117.0844	99.78238	1.17	0.258	-94.44474 328.6136

Multicollinearity 6

```
. estat vif
```

Variable	VIF	1/VIF
triceps	708.84	0.001411
thigh	564.34	0.001772
midarm	104.61	0.009560
Mean VIF	459.26	

Leverage 1

```
. sysuse auto  
(1978 Automobile Data)
```

```
. reg price weight mpg
```

Source	SS	df	MS	Number of obs	=	74
Model	186321280	2	93160639.9	F(2, 71)	=	14.74
Residual	448744116	71	6320339.67	Prob > F	=	0.0000
Total	635065396	73	8699525.97	R-squared	=	0.2934
				Adj R-squared	=	0.2735
				Root MSE	=	2514

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
weight	1.746559	.6413538	2.72	0.008	.467736 3.025382
mpg	-49.51222	86.15604	-0.57	0.567	-221.3025 122.278
_cons	1946.069	3597.05	0.54	0.590	-5226.245 9118.382

Leverage 2

```
. dfbeta
      _dfbeta_1: dfbeta(weight)
      _dfbeta_2: dfbeta(mpg)

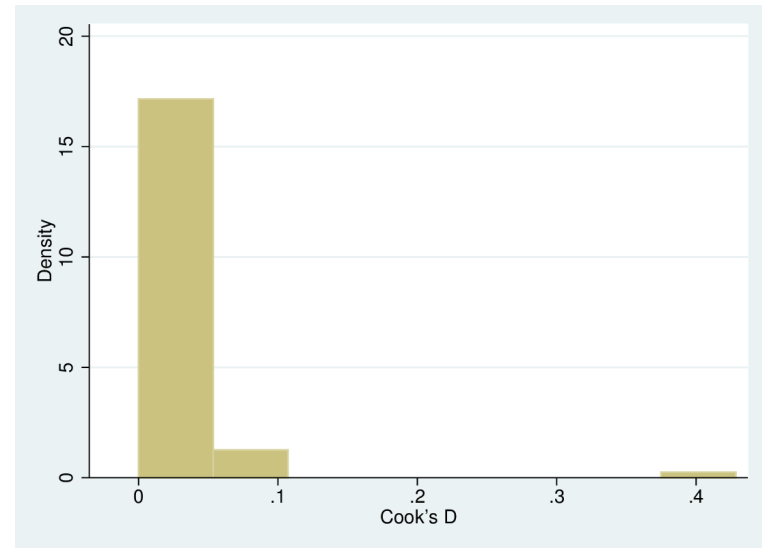
. predict cd, cooks
. su _df* cd
```

Variable	Obs	Mean	Std. Dev.	Min	Max
_dfbeta_1	74	.0032413	.1624324	-.2570012	1.14019
_dfbeta_2	74	.0026526	.1490603	-.4160759	.9075188
cd	74	.0171081	.0526476	2.08e-09	.4285067

```
. list make cd _df* if _dfbeta_1>1
```

	make	cd	_dfbet-1	_dfbet-2
13.	Cad. Seville	.4285067	1.14019	.9075188

Leverage 3



Residuals

```
. sysuse auto
(1978 Automobile Data)
. reg price weight mpg
```

Source	SS	df	MS	Number of obs	=	74
Model	186321280	2	93160639.9	F(2, 71)	=	14.74
Residual	448744116	71	6320339.67	Prob > F	=	0.0000
Total	635065396	73	8699525.97	R-squared	=	0.2934
				Adj R-squared	=	0.2735
				Root MSE	=	2514

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
weight	1.746559	.6413538	2.72	0.008	.467736 3.025382
mpg	-49.51222	86.15604	-0.57	0.567	-221.3025 122.278
_cons	1946.069	3597.05	0.54	0.590	-5226.245 9118.382

```
. predict res, res
. scatter res weight
```

Residuals

