

QS101: Introduction to Quantitative Methods in Social Science

Week 15: Measures of Association – Correlation

Dr. Florian Reiche

Teaching Fellow in Quantitative Methods

Course Director BA Politics and Sociology

Deputy Director of Student Experience and Progression

February 6, 2015

Assessment 2

Formatting Tables and Graphs

Missing Values

Correlation

Assessment 2

Assessment: Task

- ▶ Which socio-demographic factors are you looking at?
- ▶ Which variables are you choosing for this?

Formatting Tables and Graphs

Missing Values

Reasons for Missing Data

- ▶ Interviewee refused the answer
- ▶ Interviewer forgot to ask
- ▶ Equipment failed
- ▶ ...

Coding of Missing Data

- ▶ When you code data, you might just give it a random number, such as -9
- ▶ Stata recognises missing data only in the form of a dot, for example `.` or `.a`
- ▶ The letters allow you to code different reasons for why the data are missing
- ▶ Stata command is **mvdecode**, see Acock book, chapter 3

Example

rgdpch	grgdpch	land	pop	death	birth	fuelx	agri	yrtsch	conc	oilexp
3210.802	-3.938805	1246700	5885.455	.	52.1404491	0
3008.608	-6.297315	1246700	5943.466484	0
2943.087	-2.177778	1246700	6163.714	24.981	52.366484	0
2757.634	-6.301311	1246700	6287.1835	0
2730.219	-9.941382	1246700	6452.54635	0
2516.246	-7.837218	1246700	6743.08	24.7938	52.672635	0
2384.475	-5.236808	1246700	6879.574	0
2453.734	2.904586	1246700	7018.526	24.669	52.877874	0
2474.558	.8486423	1246700	7240.535874	0
2739.029	10.6876	1246700	7442.448874	0
2633.564	-3.85043	1246700	7584.709	24.6666	53.0168	.	13.4	.	.874	0
2830.181	7.465817	1246700	7748.632	.	.	.	14.2	.	.874	0
2971.034	4.976814	1246700	7878.396	24.665	53.11	.	12.8	.	.887	0
3158.241	6.301059	1246700	8023.39	.	.	.	15.8	.	.9	0
3015.413	-4.5224	1246700	8153.844	.	.	.	19.1	.	.9055	0
2962.883	-1.742024	1246700	8296.606	24.1208	52.7212	.	17.9	.	.911	0
3181.842	7.390052	1246700	8496.81
3049.962	-4.14477	1246700	8741.838	23.758	52.462
2045.445	-32.93538	1246700	8966.258
2142.351	4.737625	1246700	9172.28
2473.534	15.45888	1246700	9419.124	23.0506	50.3782
2566.742	3.768212	1246700	9650.607
2996.92	16.75968	1246700	9855.579	22.579	48.989
2522.227	-15.83937	1246700	10039.17
3058.969	21.28048	1246700	10215.44
2901.421	-5.15034	1246700	10377.27	22.2856	48.7574
2570.318	-11.41177	1246700	10538.39
2992.343	16.41919	1246700	10768.51	22.09	48.603

Implications

- ▶ Stata uses listwise deletion for missing values
- ▶ This means, that as soon as one value within an observation is missing, Stata drops that entire observation
- ▶ This can potentially decimate the number of observations for analysis drastically
- ▶ This in turn has implications for inference

What to do?

- ▶ Be vigilant!
- ▶ There are statistical methods to deal with this, such as multiple imputation (too advanced for now)

Correlation

Stata Commands

- ▶ Stata has two commands:
 - ▶ **correlate**
 - ▶ **pwcorr**

correlate

- ▶ Uses listwise deletion if any value of an observation is missing.
- ▶ Disadvantage: very few options
- ▶ Advantage: You are always operating with the same sub-sample

- ▶ Much more powerful command than **correlate**
- ▶ Advantage: You can apply different options
- ▶ Disadvantage: You need to pay more attention

How to do it

- ▶ It's simple!
- ▶ **correlate** *var1 var2 var3*
- ▶ **pwcorr** *var1 var2 var3*, **listwise sig star(5)**

Example

- ▶ `pwcorr a_fimngrs_dv a_agegr10_dv a_sex a_jbhrs, listwise sig star(5)`

```
. pwcorr a_fimngrs_dv a_agegr10_dv a_sex a_jbhrs, listwise sig star(5)
```

	a_fimng~	a_ageg~v	a_sex	a_jbhrs
a_fimngrs_dv	1.0000			
a_agegr10_dv	-0.0028 0.5945	1.0000		
a_sex	-0.1795* 0.0000	-0.0067 0.2073	1.0000	
a_jbhrs	0.4205* 0.0000	-0.3085* 0.0000	-0.1095* 0.0000	1.0000

Queries

- ▶ What does the coefficient on **a_fimngrs_dv** and **a_sex** tell us?

Queries

- ▶ What does the coefficient on **a_fimngrs_dv** and **a_sex** tell us?
- ▶ Is it significant?

Queries

- ▶ What does the coefficient on **a_fimngrs_dv** and **a_sex** tell us?
- ▶ Is it significant?
- ▶ What does that mean?

Queries

- ▶ What does the coefficient on **a_fimngrs_dv** and **a_sex** tell us?
- ▶ Is it significant?
- ▶ What does that mean?
- ▶ What does the coefficient on **a_fimngrs_dv** and **a_jbhrs** tell us?

Queries

- ▶ What does the coefficient on **a_fimngrs_dv** and **a_sex** tell us?
- ▶ Is it significant?
- ▶ What does that mean?
- ▶ What does the coefficient on **a_fimngrs_dv** and **a_jbhrs** tell us?
- ▶ Is it significant?

Queries

- ▶ What does the coefficient on **a_fimngrs_dv** and **a_sex** tell us?
- ▶ Is it significant?
- ▶ What does that mean?
- ▶ What does the coefficient on **a_fimngrs_dv** and **a_jbhrs** tell us?
- ▶ Is it significant?
- ▶ What does that mean?

REPLICATE!

Task

