Tutorial 05, Michaelmas Term

Research Methods for Political Science (PO3600)

Stefan Müller

14 November 2017

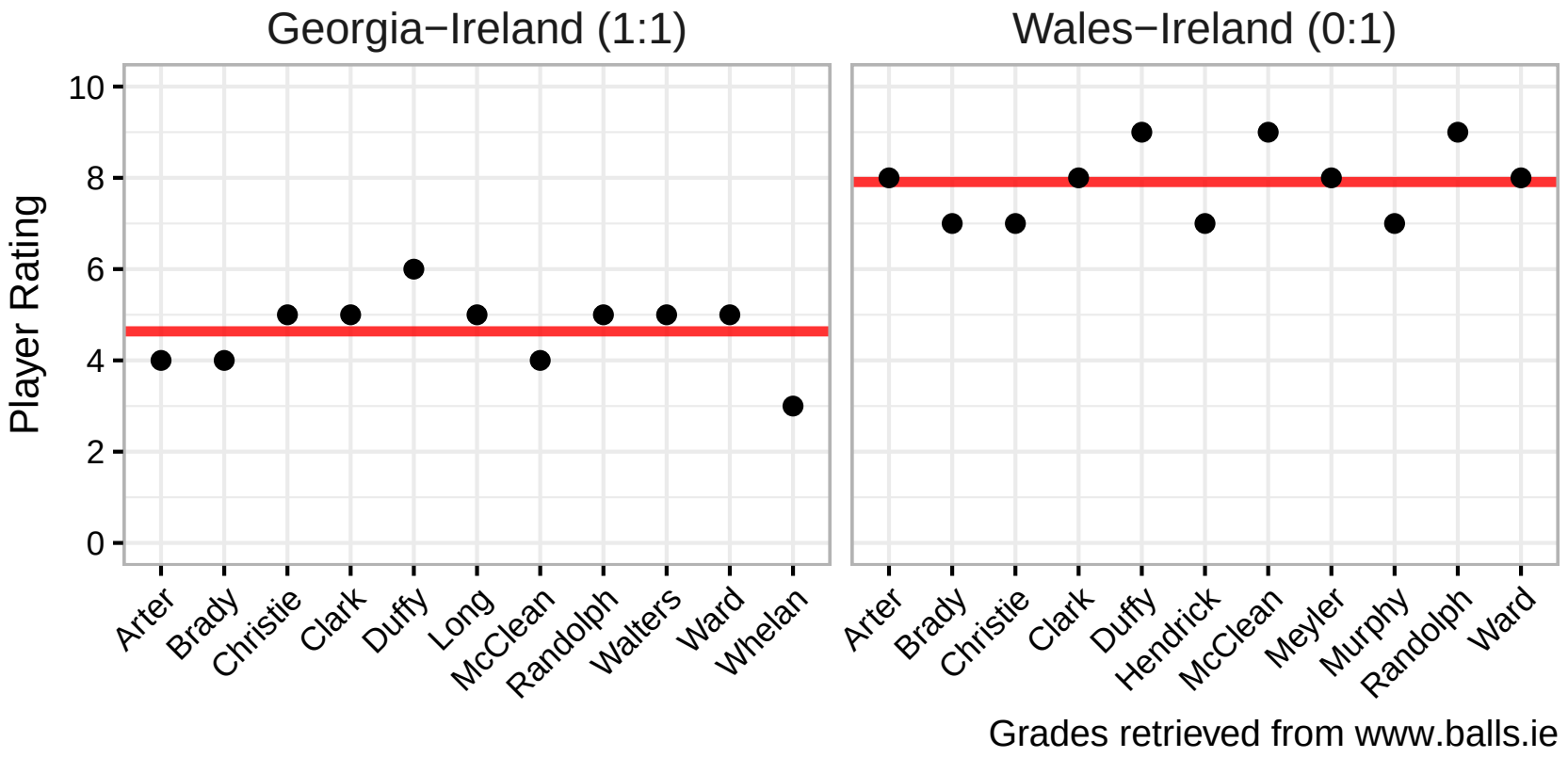
Trinity College Dublin
http://muellerstefan.net/research-methods

Standard deviation and standard error

Mean, Standard Deviation and Standard Error

Application: Mean and standard deviation (#COYBIG)

Application: Mean and Standard Deviation (#COYBIG)



Grades retrieved from www.balls.ie

Match	$\bar{x}$	σ
Georgia–Ireland	4.63	0.81
Wales–Ireland	7.91	0.83

Confidence Intervals

Confidence intervals

For a given statistic calculated from a sample, the confidence interval is a range of values around that statistic that are believed to contain, with a certain probability, the true value of that statistic (population value).

A 95% confidence interval will contain the population mean 19 out of 20 times.

Calculate confidence intervals

Sample mean ($\bar{x}$): 170(cm)

Sample standard deviation (σ): 10

Sample size (n): 30

Standard error of the mean ($sd(\bar{X})$)

Task: Estimate the 95 % confidence intervals

$$\text{Standard error} = sd(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{30}} = 1.82$$

$$CI = 170 \pm 1.96 * 1.83$$

$$CI_{low} = 170 - 1.96 * 1.83 = 166.41$$

$$CI_{high} = 170 + 1.96 * 1.83 = 173.58$$

T-Test

One-sample t-test (I)

IQ in general population is 100. We take a random sample of 30 high school students and find $\bar{x} = 110$, $\sigma = 10$.

H_0: The observed average IQ equals the general population's IQ.

H_1: The observed average IQ differs from the general population IQ.

We ask: if the null hypothesis were true, how likely would we be to collect the data we have?

Selecting sampling distribution and critical region: Population standard deviation (σ) is unknown: t-distribution with $n - 1$ degrees of freedom ($n = 30, df = 29$). $\alpha = 0.05$

One-sample t-test (II)

Calculate test-statistic: $t = \frac{observed\ \ value - expected\ \ value\ \ under\ \ H_0}{standard\ \ error}$

$$t = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$t = \frac{110 - 100}{\frac{10}{\sqrt{30}}}$$

$$t = 5.47$$

Problems in Homework Assignments

Homework Assignments: Problems

Formatting:

- low quality graphs
- screen-shots of SPSS output
- equation mode not used

Homework Assignments: Problems

Formatting:

- low quality graphs
- screen-shots of SPSS output
- equation mode not used
- table formatting

Example: Equations

$$\sqrt{\frac{30}{9-1}} = 1.94$$

root(30/(9-1)) = 1.936492

Bad Example: Tables

Six-fold classification of regimes Cheibub 2000

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Parliamentary democracy	56	29.3	29.6	29.6
	Mixed democracy	21	11.0	11.1	40.7
	Presidential democracy	37	19.4	19.6	60.3
	Civilian dictatorship	38	19.9	20.1	80.4
	Military dictatorship	24	12.6	12.7	93.1
	Monarchic dictatorship	13	6.8	6.9	100.0
	Total	189	99.0	100.0	
Missing	System	2	1.0		
Total		191	100.0		

Good Example: Tables

Manifesto	Total Sentences in Manifesto	Mean Expert Evaluations: Natural Sequence	Mean Expert Evaluations: Random Sequence	Total Expert Evaluations	Mean Crowd Evaluations	Total Crowd Evaluations
Con 1987	1,015	6.0	2.4	7,920	44	36,594
LD 1987	878	6.0	2.3	6,795	22	24,842
Lab 1987	455	6.0	2.3	3,500	20	11,087
Con 1992	1,731	5.0	2.4	11,715	6	28,949
LD 1992	884	5.0	2.4	6,013	6	20,880
Lab 1992	661	5.0	2.3	4,449	6	23,328
Con 1997	1,171	6.0	2.3	9,107	20	11,136
LD 1997	873	6.0	2.4	6,847	20	5,627
Lab 1997	1,052	6.0	2.3	8,201	20	4,247
Con 2001	748	5.0	2.3	5,029	5	3,796
LD 2001	1,178	5.0	2.4	7,996	5	5,987
Lab 2001	1,752	5.0	2.4	11,861	5	8,856
Con 2005	414	5.0	2.3	2,793	5	2,128
LD 2005	821	4.1	2.3	4,841	5	4,173
Lab 2005	1,186	4.0	2.4	6,881	5	6,021
Con 2010	1,240	4.0	2.3	7,142	5	6,269
LD 2010	855	4.0	2.4	4,934	5	4,344
Lab 2010	1,349	4.0	2.3	7,768	5	6,843
Total	18,263	91,400	32,392	123,792		215,107

Source: Benoit et al. (2016, p. 284)

Homework Assignments: Problems

Stating a Research Question:

- **Don't**: Is there a correlation between X and Y.
- **Don't**: What are the factors that resulted in Y.
- **Don't**: The impact of X on A, B, C, D

Conceptualisation of variables

- Be precise.
- Define what you want to measure.
- Don't mix up independent and dependent variable

Confidence intervals

Example: 1500 respondents in survey asked how much money they spend on electronic devices per year. $\bar{x} = 432$; $\sigma = 120$

Sample average = population mean + random error

Calculate standard error of the mean (standard deviation of sample average)

$\frac{s}{\sqrt{n}} = \frac{120}{\sqrt{1500}} = 3.09$

The Normal Approximation tells us, for example, that for 95% of all large samples, the sample average will be within two standard errors of the true population average.

$432 \pm 1.96 * 3.09 = [425.94, 438.06]$

Interpreting Confidence Intervals

WRONG: There is a 95% probability that the population mean lies between 425 and 438.

Why: The population mean is assumed to be fixed, thus there is no randomness. The probability that the confidence intervals cover the true mean is 0 or 1.

Correct interpretation of 95 % confidence intervals: 95% of the time, when we calculate a confidence interval in this way, the true mean will be between the two values. 5% of the time, it will not. Because the true mean (population mean) is an unknown value, we don't know if we are in the 5% or the 95%.