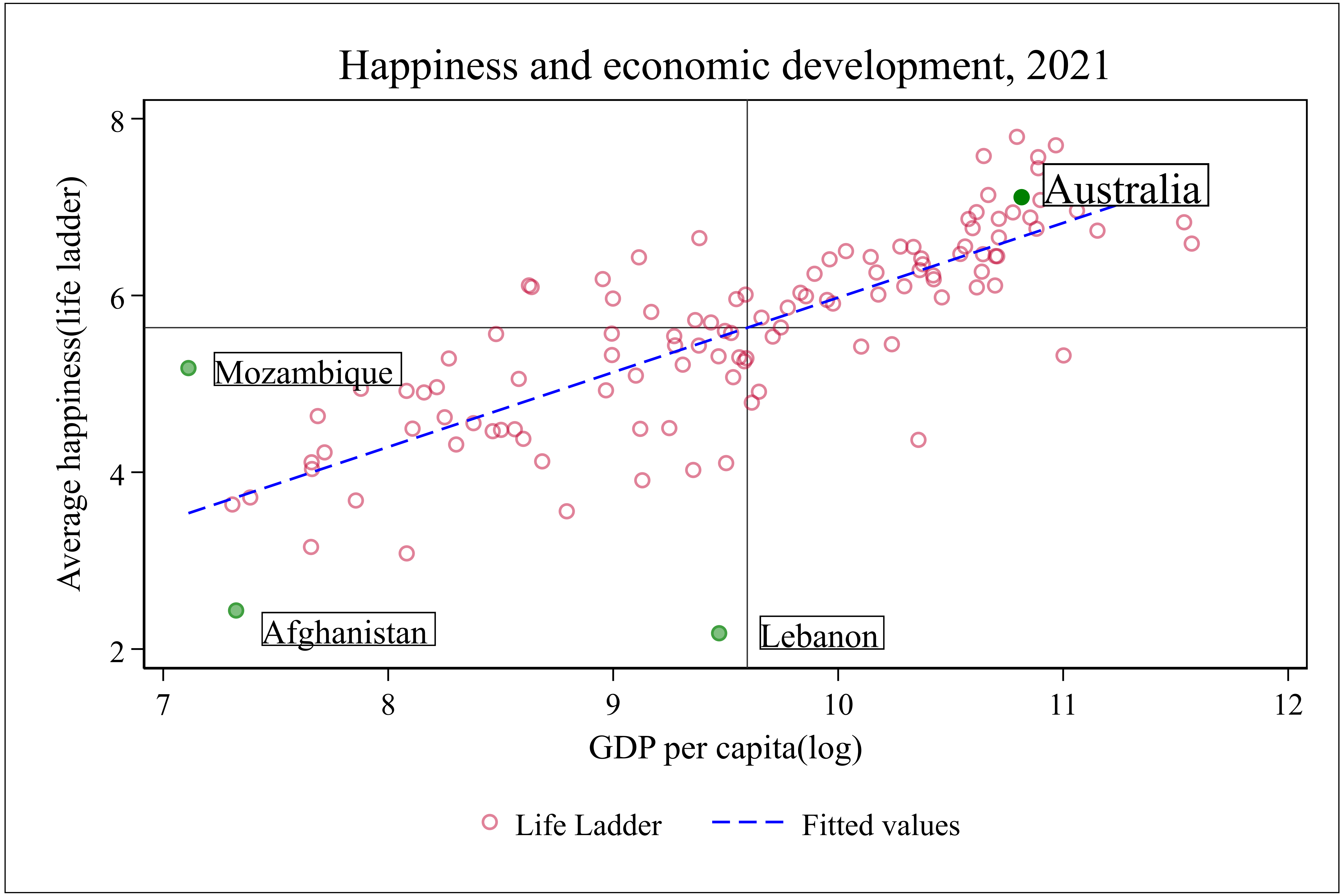

Happiness and economic development, 2021
Average happiness(life ladder)
GDP per capita(log)
8
6
4
2
7
8
9
10
11
12
Australia
Mozambique
Afghanistan
Lebanon
Life Ladder
Fitted values

1

2

3

1

1

Why and how do we run a **bivariate regression**?

Why and how do you **interpret regression results** (both yours and others)?

1

What if we are interested not just if there is a statistically significant difference in a sample **(goodness of fit**) or pairs of samples (**difference of means test**) or whether two variables are **correlated**?

Rather we want a more complex understanding of the **directionality** and **significance** in the relationship between an X and Y?

Or perhaps we want to **predict** our outcome as we vary values of our independent variable?

1

Latin for “other things equal.”

Also short for “all other things being equal.”

Regression helps us control for other factors to better isolate the effect of the variable we care about.

1

Political Behavior
https://doi.org/10.1007/s11109-021-09720-y

ORIGINAL PAPER

Check for updates

Correlates of Voter Turnout

Richard W. Frank[1] · Ferran Martínez i Coma[2]

Accepted: 11 May 2021

Abstract
Despite decades of research, there is no consensus as to the core correlates of national-level voter turnout. We argue that this is, in part, due to the lack of comprehensive, systematic empirical analysis. This paper conducts such an analysis. We identify 44 articles on turnout from 1986 to 2017. These articles include over 127 potential predictors of voter turnout, and we collect data on seventy of these variables. Using extreme bounds analysis, we run over 15 million regressions to determine which of these 70 variables are robustly associated with voter turnout in 579 elections in 80 democracies from 1945 to 2014. Overall, 22 variables are robustly associated with voter turnout, including compulsory voting, concurrent elections, competitive elections, inflation, previous turnout, and economic globalization.

Keywords Elections · Turnout · Extreme bounds analysis · Meta-analysis

Introduction

A common challenge in the study of comparative politics is balancing theoretical and empirical comprehensiveness with substantive importance. Consider voter turnout. If we ask what the most statistically significant and substantively important predictors of national-level voter turnout in democratic elections are, even after more than 50 years of comparative voter turnout research, there are few certainties beyond the fact that compulsory voting increases turnout. For example, several studies including Radcliff and Davis (2000) find larger district magnitudes increase turnout while others like Tavits (2008) find either no significant relationship or even a negative one (Fumagalli & Narciso, 2012).

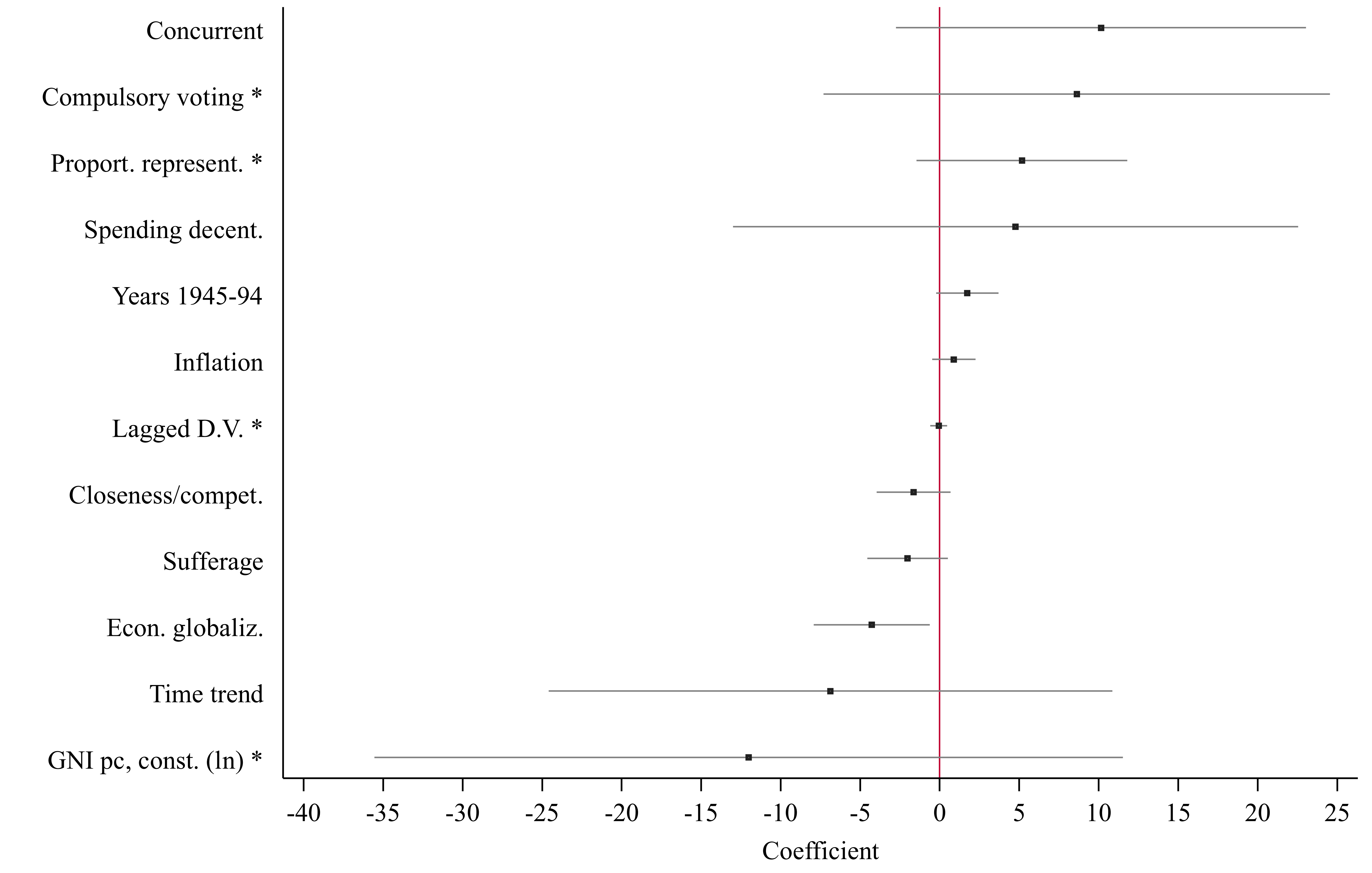
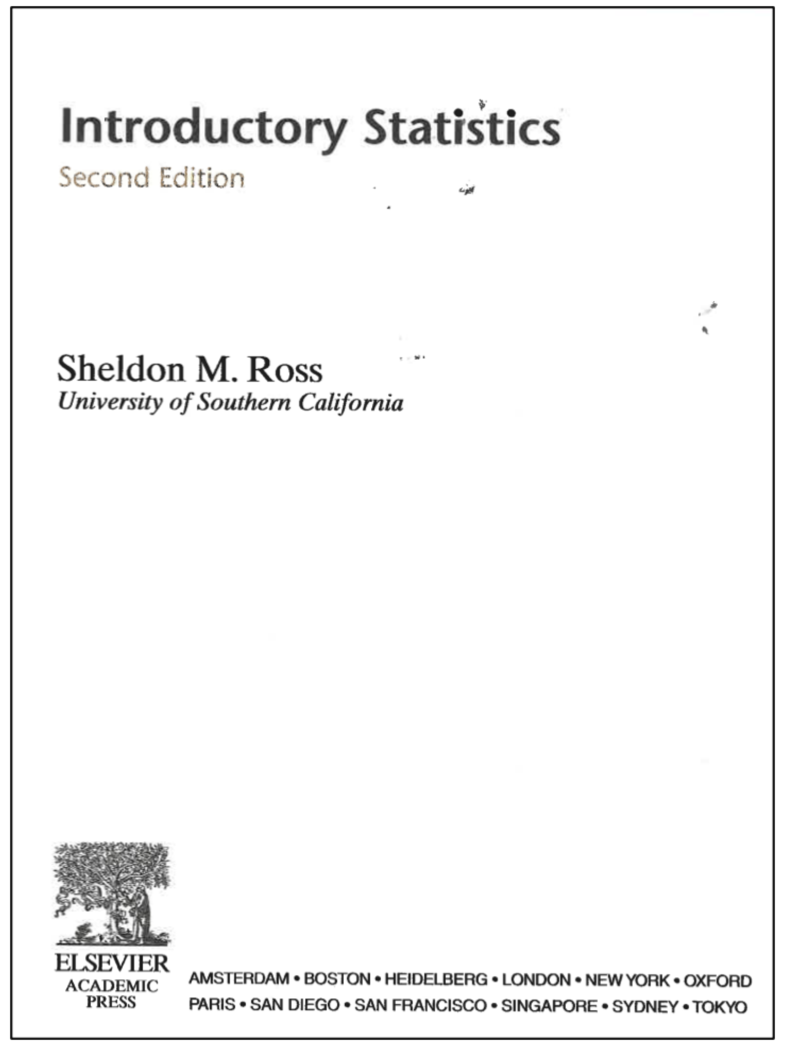


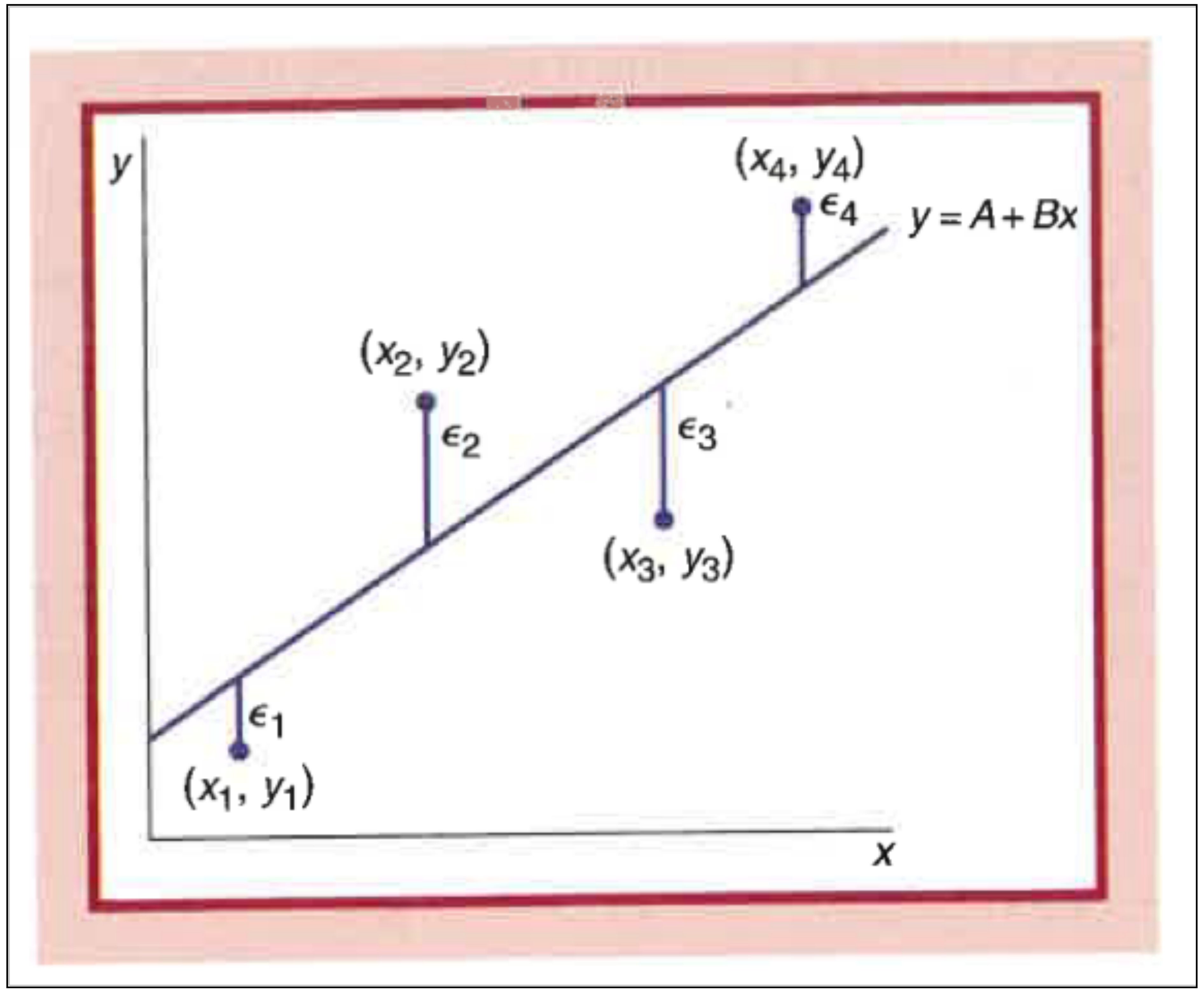
Fig. 2 Robust predictors of voter turnout, mean coefficients and extreme bounds. *Note* Results from fixed effects models reported in Table 5. Table A7 includes complete results. *Identifies core variables in all fixed effects models. The only core variable found to be (Sala-i-Martin) robust is proportional representation. Population is a core variable but not included in figure due to the disproportionate size of its extreme bounds

1


Introductory Statistics
Second Edition

Sheldon M. Ross
University of Southern California

ELSEVIER ACADEMIC PRESS
AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO



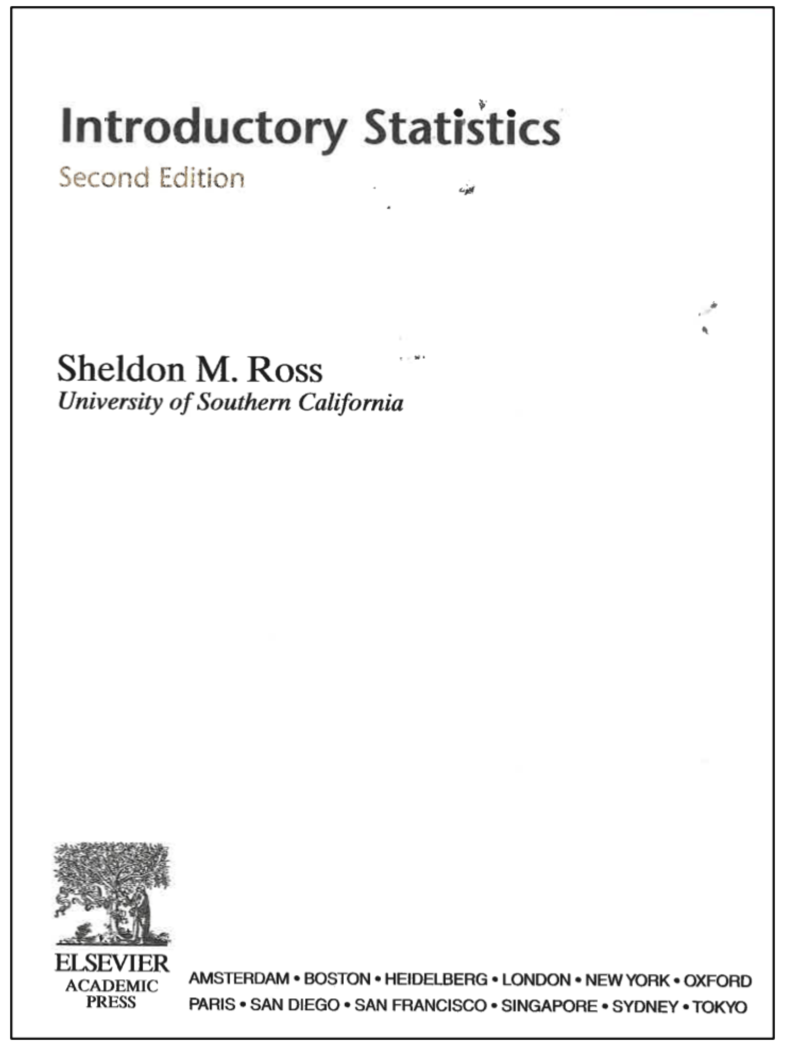
1

$$Y = \alpha + \beta X + \epsilon + \varepsilon$$

Where:

Y is the outcome you are trying to explain.
X is the main explanatory variable.
α (alpha) is the value of Y when X=0.
β (beta) is the estimated relationship between X and Y.
ϵ is the systematic error.
ε is the random error.

1


Introductory Statistics
Second Edition

Sheldon M. Ross
University of Southern California

ELSEVIER ACADEMIC PRESS
AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

It can be shown that the least-squares estimators of α and β, which we call $\hat{\alpha}$ and $\hat{\beta}$, are given by

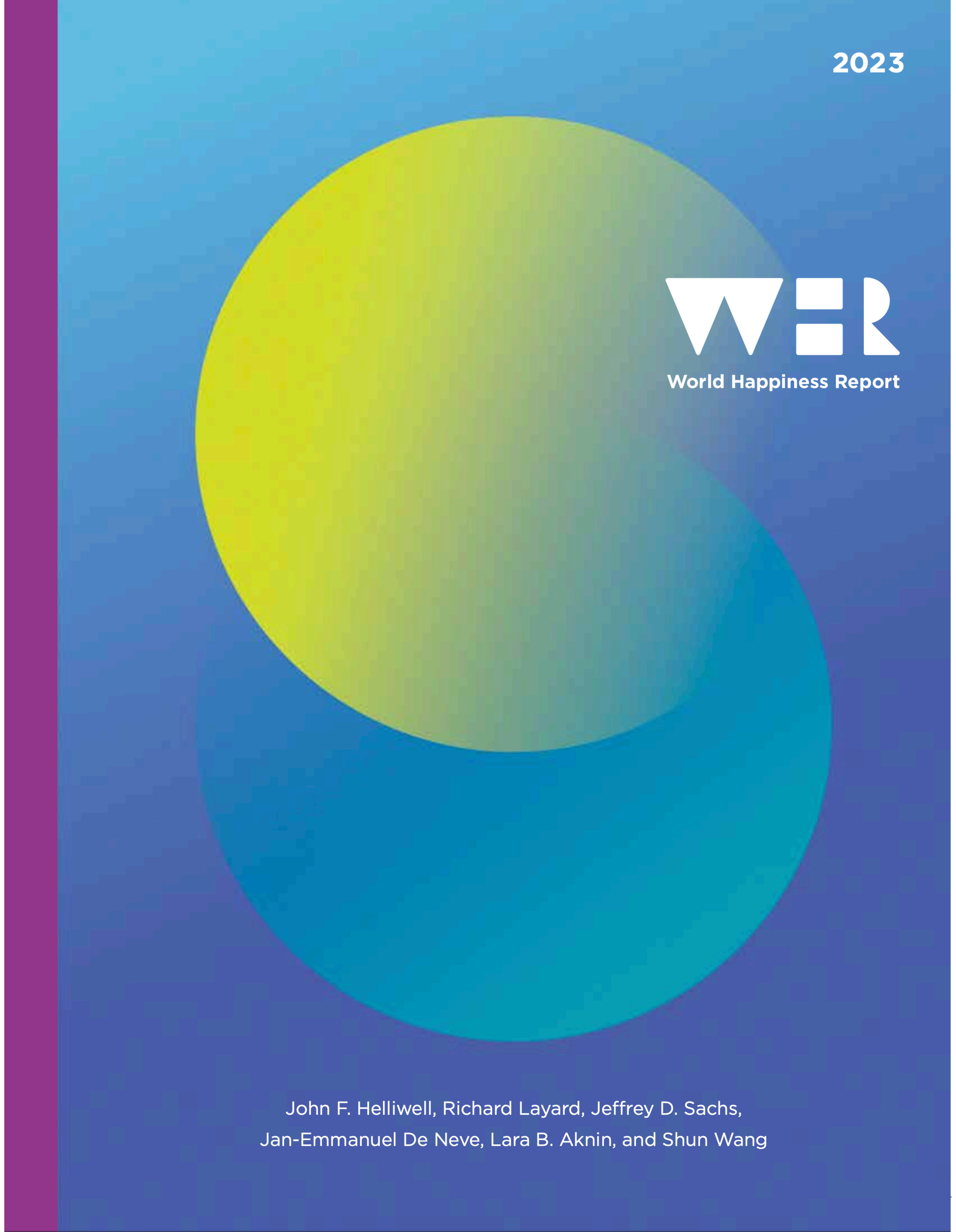
$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\,\overline{x}$$

where

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{and} \quad \overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

1


2023
World Happiness Report
John F. Helliwell, Richard Layard, Jeffrey D. Sachs,
Jan-Emmanuel De Neve, Lara B. Aknin, and Shun Wang

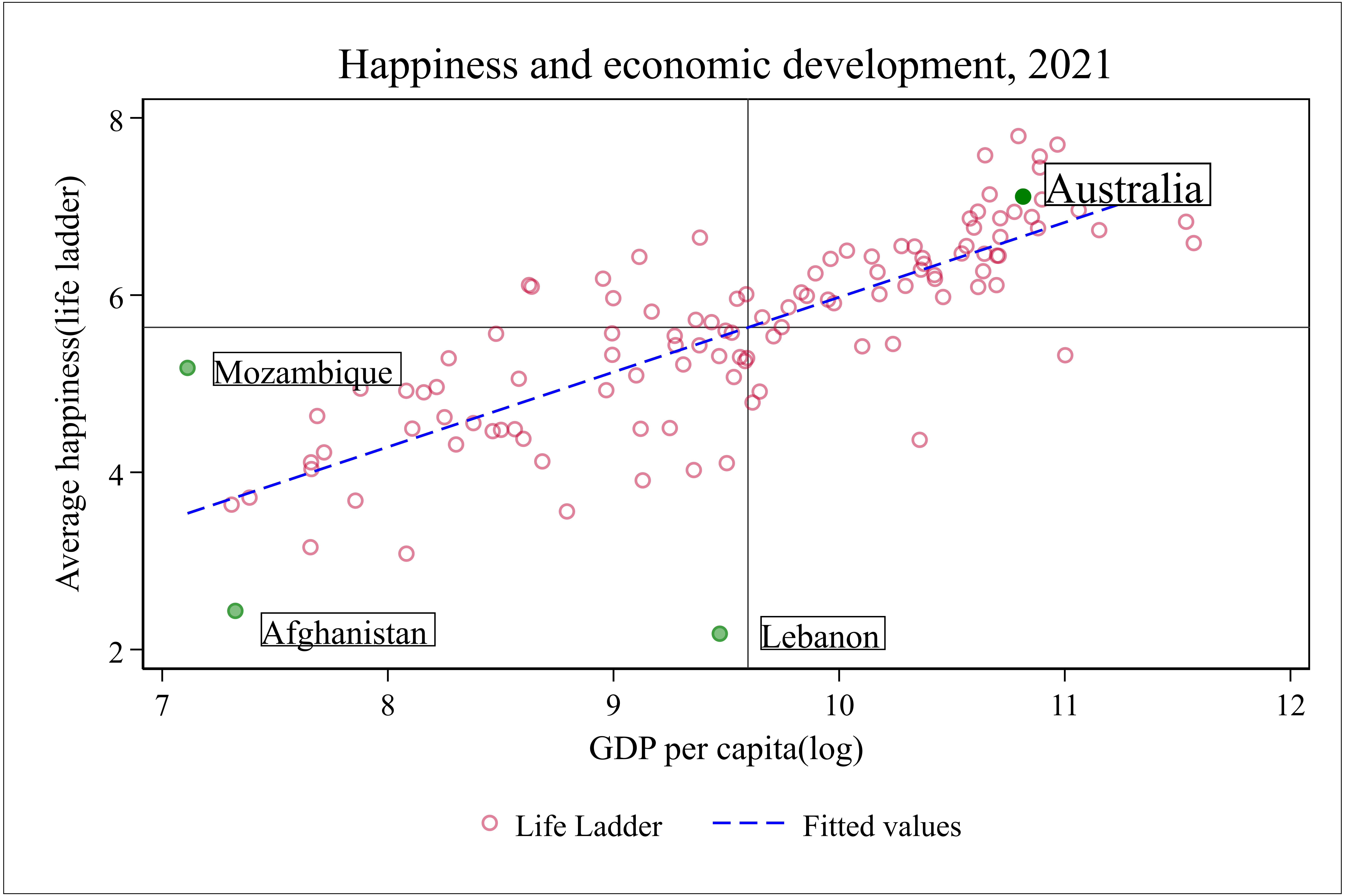
1

Happiness and economic development, 2021
Average happiness(life ladder)
8
6
4
2
7
8
9
10
11
12
GDP per capita(log)

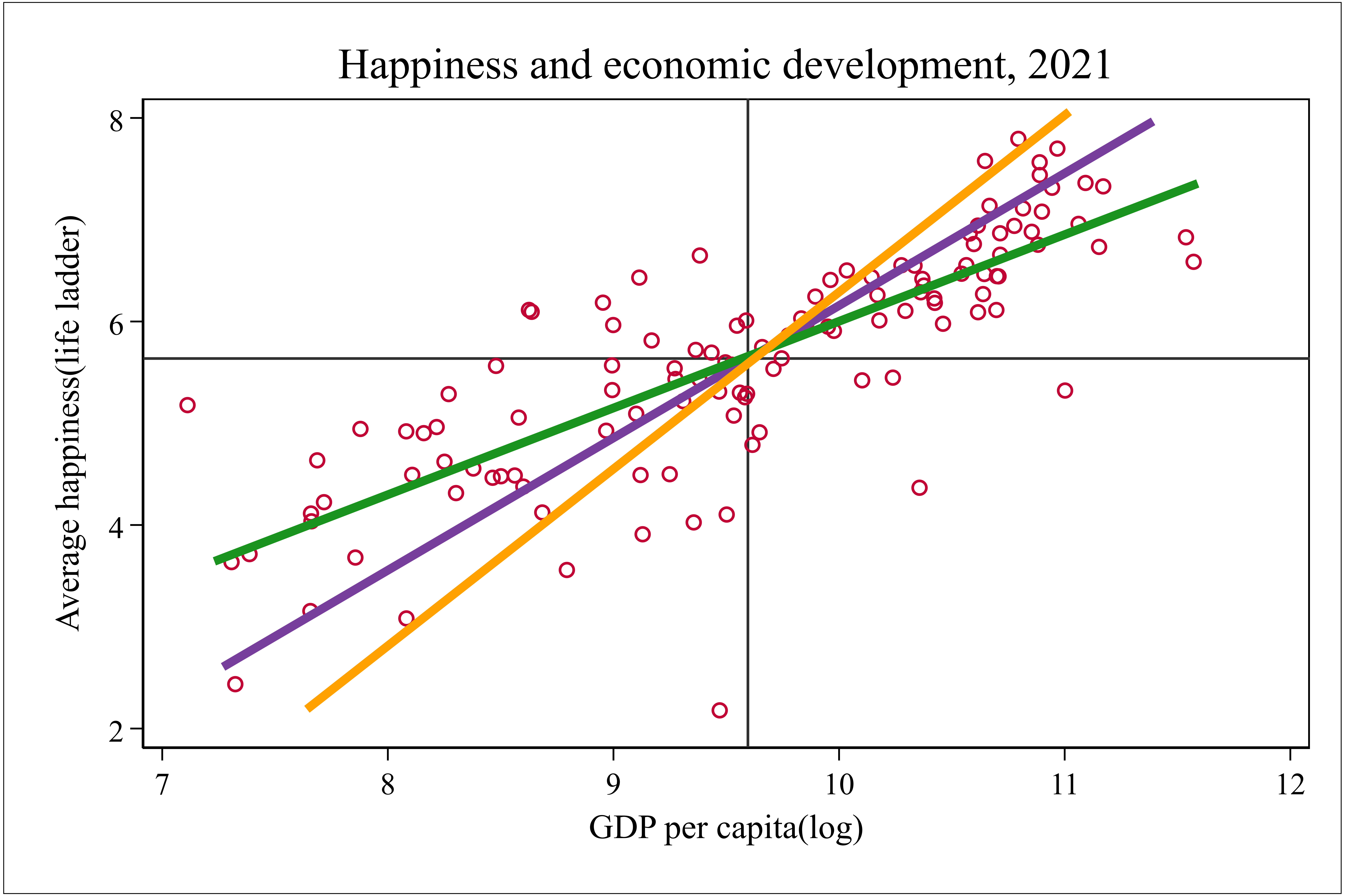
1

Happiness and economic development, 2021
Average happiness(life ladder)
GDP per capita(log)
8
6
4
2
7
8
9
10
11
12

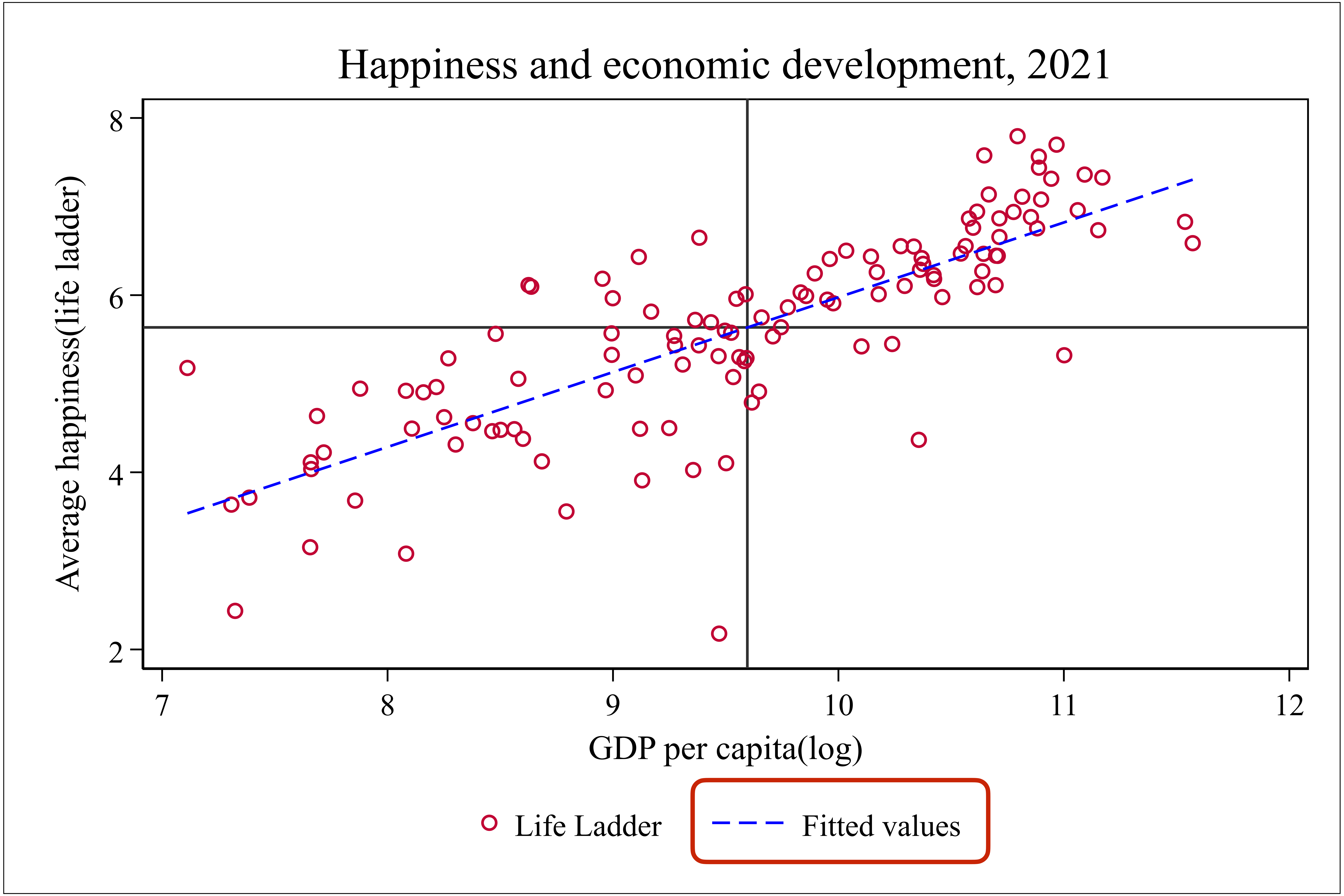
1



Happiness and economic development, 2021
Average happiness(life ladder)
8
6
4
2
Australia
Mozambique
Afghanistan
Lebanon
7
8
9
10
11
12
GDP per capita(log)
Life Ladder
Fitted values


Happiness and economic development, 2021
Average happiness(life ladder)
8
6
4
2
7
8
9
10
11
12
GDP per capita(log)

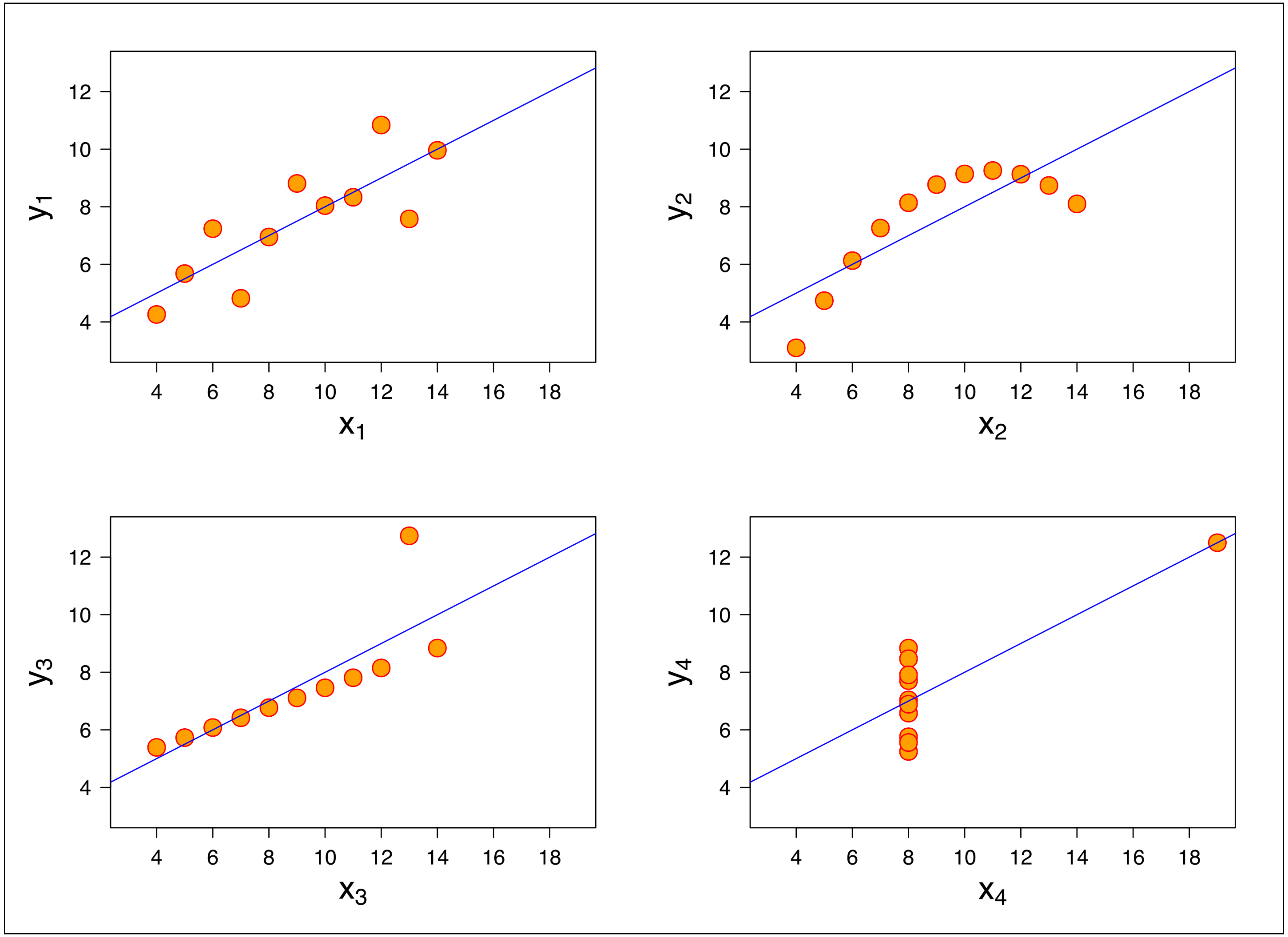
1



Happiness and economic development, 2021
Average happiness(life ladder)
8
6
4
2
7
8
9
10
11
12
GDP per capita(log)
Life Ladder
Fitted values

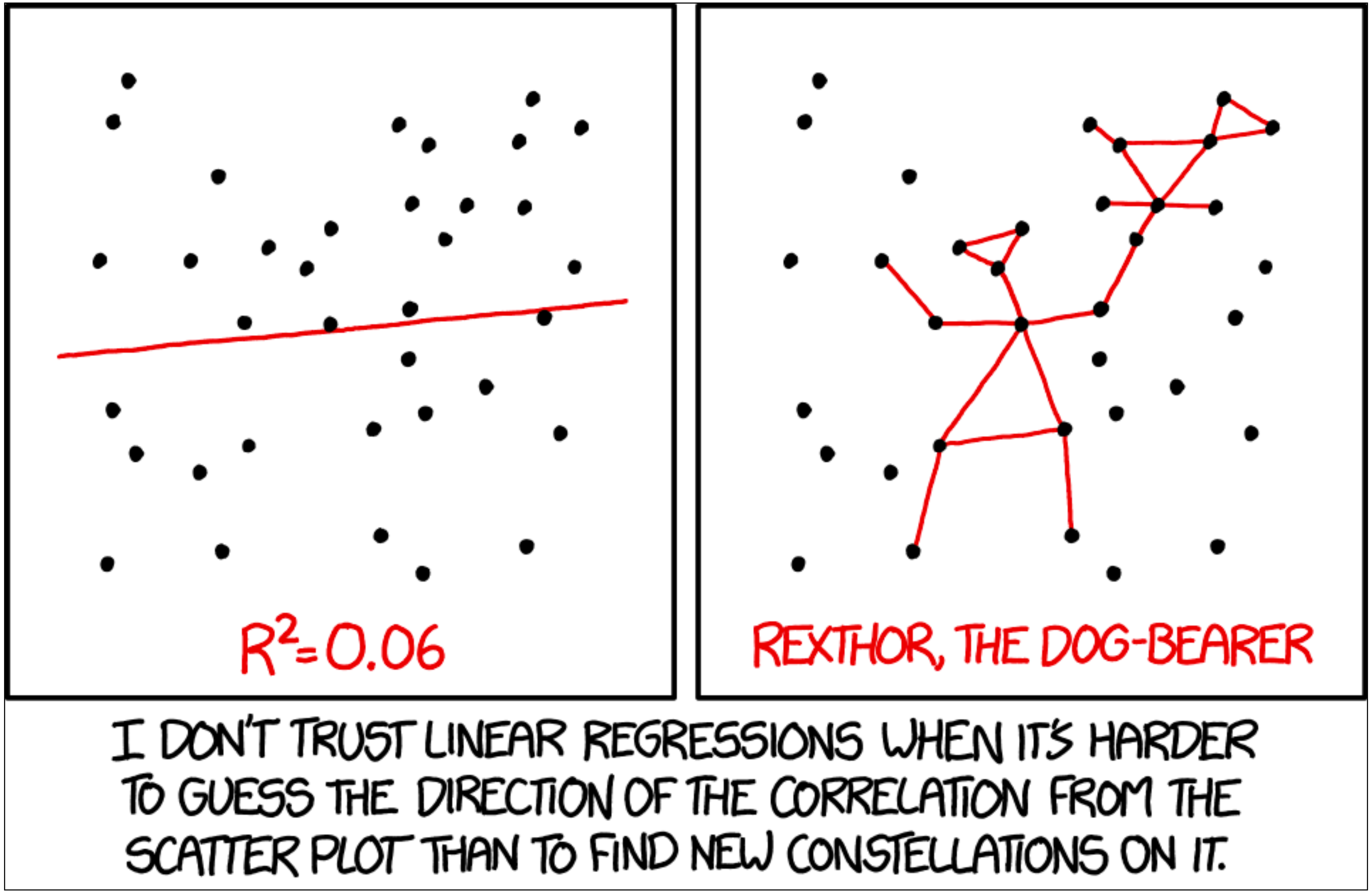
1

Almost identical descriptive statistics but very different underlying value distributions.



Source: https://en.wikipedia.org/wiki/Anscombe%27s_quartet#/media/File:Anscombe's_quartet_3.svg

1



Source: https://xkcd.com/1725/

1

Here are my regression results for happiness regressed on GDP: β **= 0.845; se= 0.060**.

My theory's main empirical hypotheses are:

H0 (null hypothesis): β = 0

H1 (alternative hypothesis): $\beta \neq$ 0

To test these hypotheses we do a **t-test**, in this case we set β_{null} = 0.

$$t = \frac{\beta - \beta null}{se(\hat{\beta})}$$

t **= (0.845-0)/0.06 = 14.083.**

With ~118 degrees of freedom, with a two-tailed test at the 0.05 level the threshold t statistic is 1.984. The estimated ***p-value*** *is 0.000*. I therefore **reject the null hypothesis** in favour of the alternate hypothesis.

1

We can estimate confidence intervals using the following equations:

$$\widehat{\beta} +/- [t * se(\widehat{\beta})]$$

$$\widehat{\alpha} +/- [t * se(\widehat{\alpha})]$$

So my **slope's confidence interval** is [0.726, 0.963].

My **intercept's confidence interval** is [-3.627, -1.324].

1

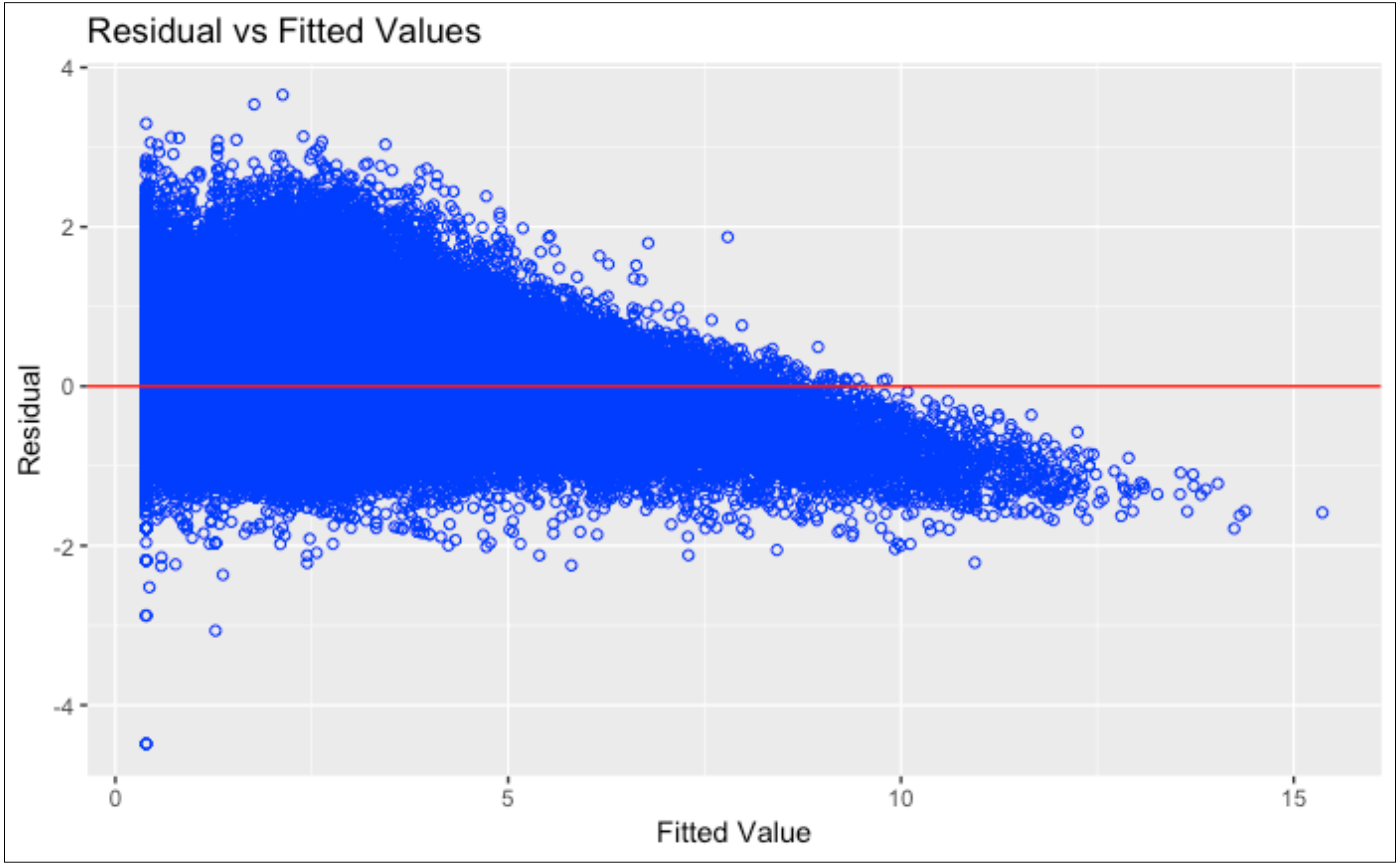
1. The **population stochastic component** is distributed normally with a mean of 0 and a variance of σ^2 where $\sigma^2 = \dfrac{\sum_i^n \widehat{u^2}}{n-2}$.

This allows us to use the t-table to make probabilistic inferences about population regression given sample regression.

It also assumes that the expected **population errors** are **not biased** one way or another.

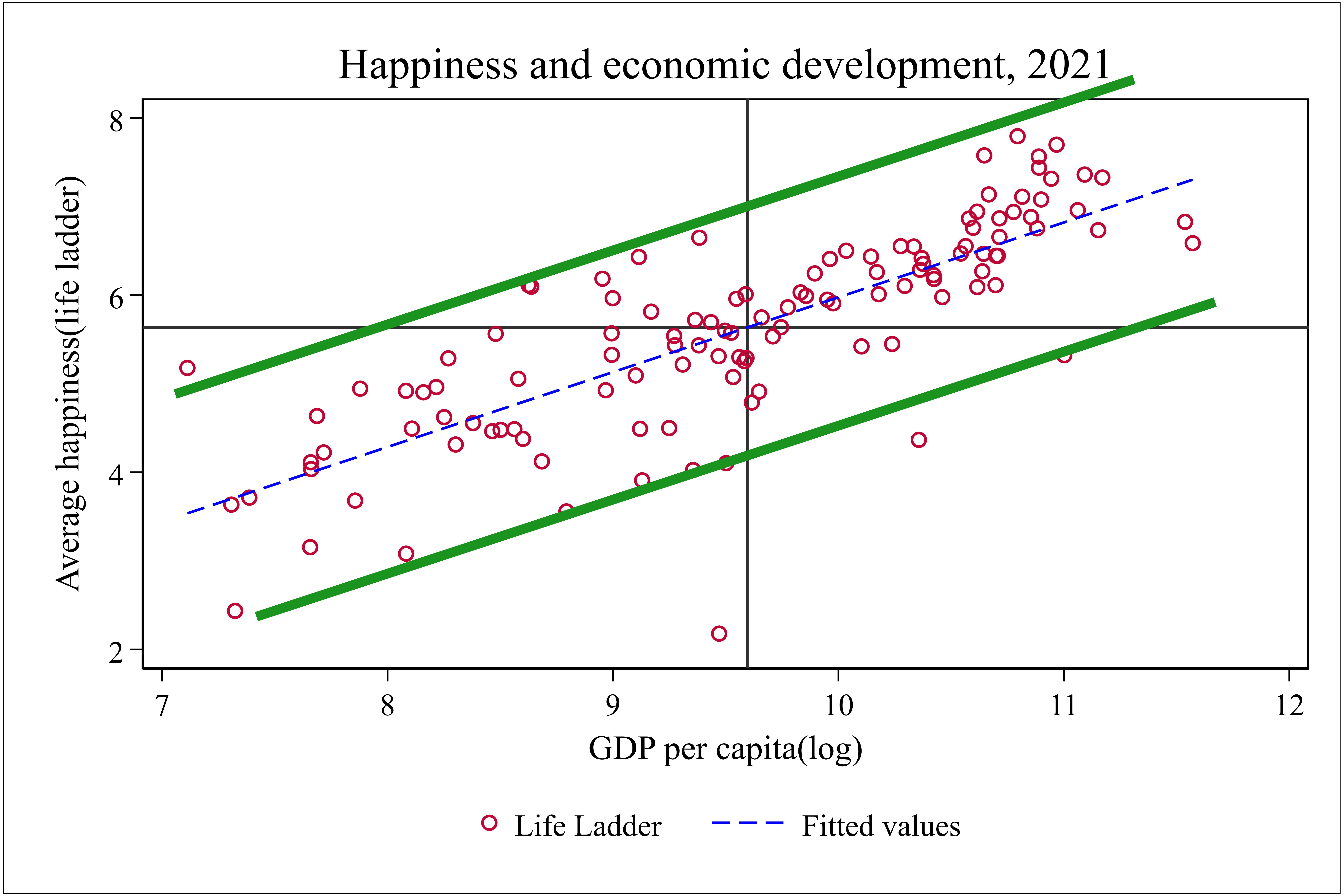
The **variance** is assumed to be **the same across values** of our Y and X.

1



Source: https://stats.stackexchange.com/questions/389127/does-it-make-sense-to-use-residuals-as-an-independent-variable

1

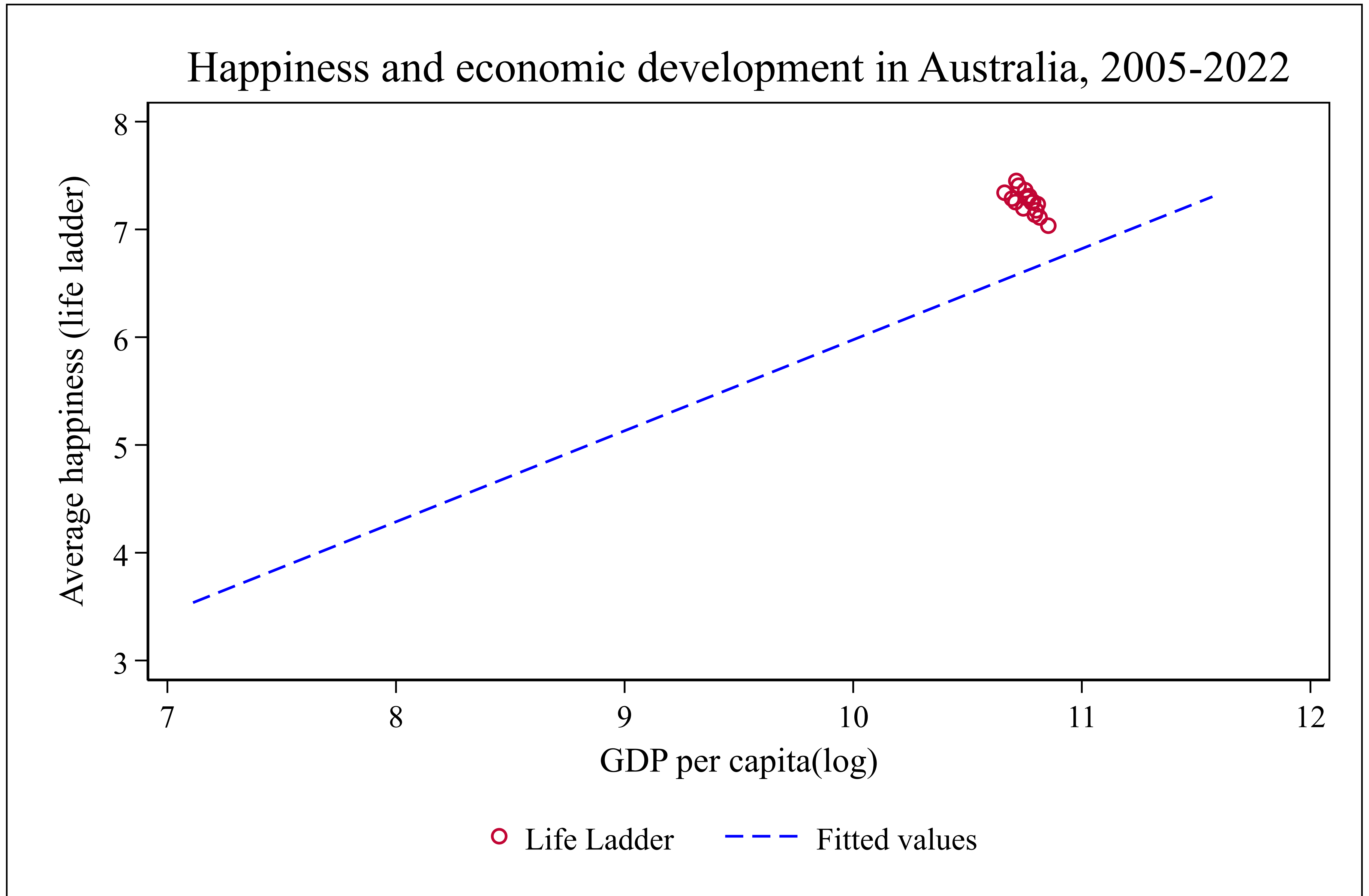


Happiness and economic development, 2021
Average happiness(life ladder)
2
4
6
8
7
8
9
10
11
12
GDP per capita(log)
Life Ladder
Fitted values

1

2. There is no **autocorrelation** in the population random error terms.

This is important to think about when you have multiple observations of the same units (e.g., people or countries) often through time-series data.



1

3. Independent variable **(X) values measured without error**. This allows us to assume that all deviations from the expected values are due to the population stochastic component (u_i) rather than measurement error.

1

4. **No causal variables left out** and no non-causal variables included in our model.

5. The relationship (β) between Y and X **stay the same across all values of X**.

6. Our **independent variable must vary**.

7. There must be **more cases than parameters** (e.g., α and β).

1

Ordinary least squares regression is about **fitting a line that minimises the (squared) distance between sample values and the line**.

A basic regression provides us with two important estimates:

(1) the **slope** of the line summarising the relationship between X and Y

(2) the **intercept** (expected value of Y when X=0).

Multiple regression enables us to control for other factors that might understate or overstate our X-Y relationship if we do not include them.

1

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	98.0363889	98.0363889	198.417241	4.89759E-27
Residual	118	58.3028664	0.49409209		
Total	119	156.339255			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

2

2

Now we have some regression results, **what do we do with them?**

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	98.0363889	98.0363889	198.417241	4.89759E-27
Residual	118	58.3028664	0.49409209		
Total	119	156.339255			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	98.0363889	98.0363889	198.417241	4.89759E-27
Residual	118	58.3028664	0.49409209		
Total	119	156.339255			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

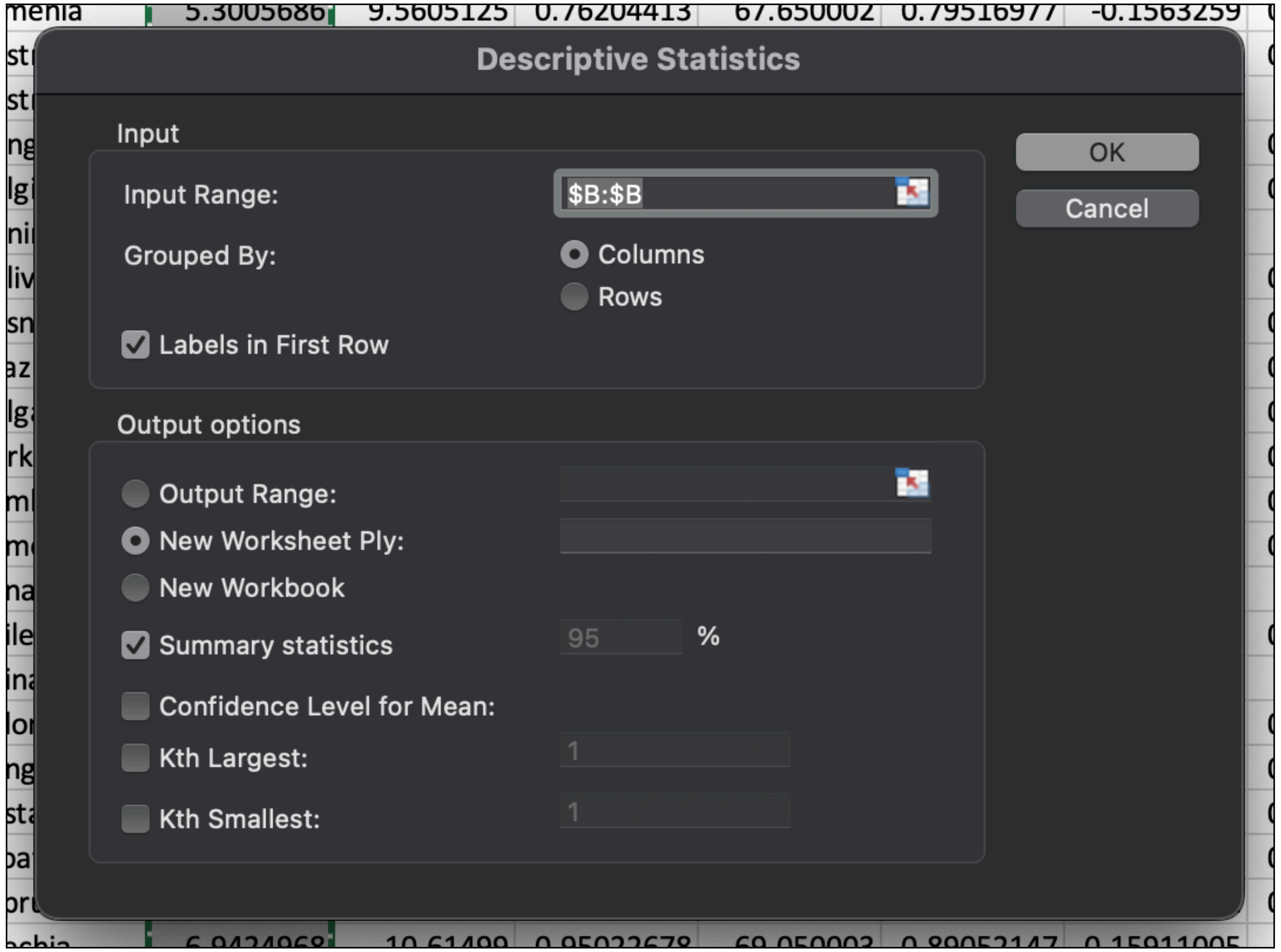
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	98.0363889	98.0363889	198.417241	4.89759E-27
Residual	118	58.3028664	0.49409209		
Total	119	156.339255			

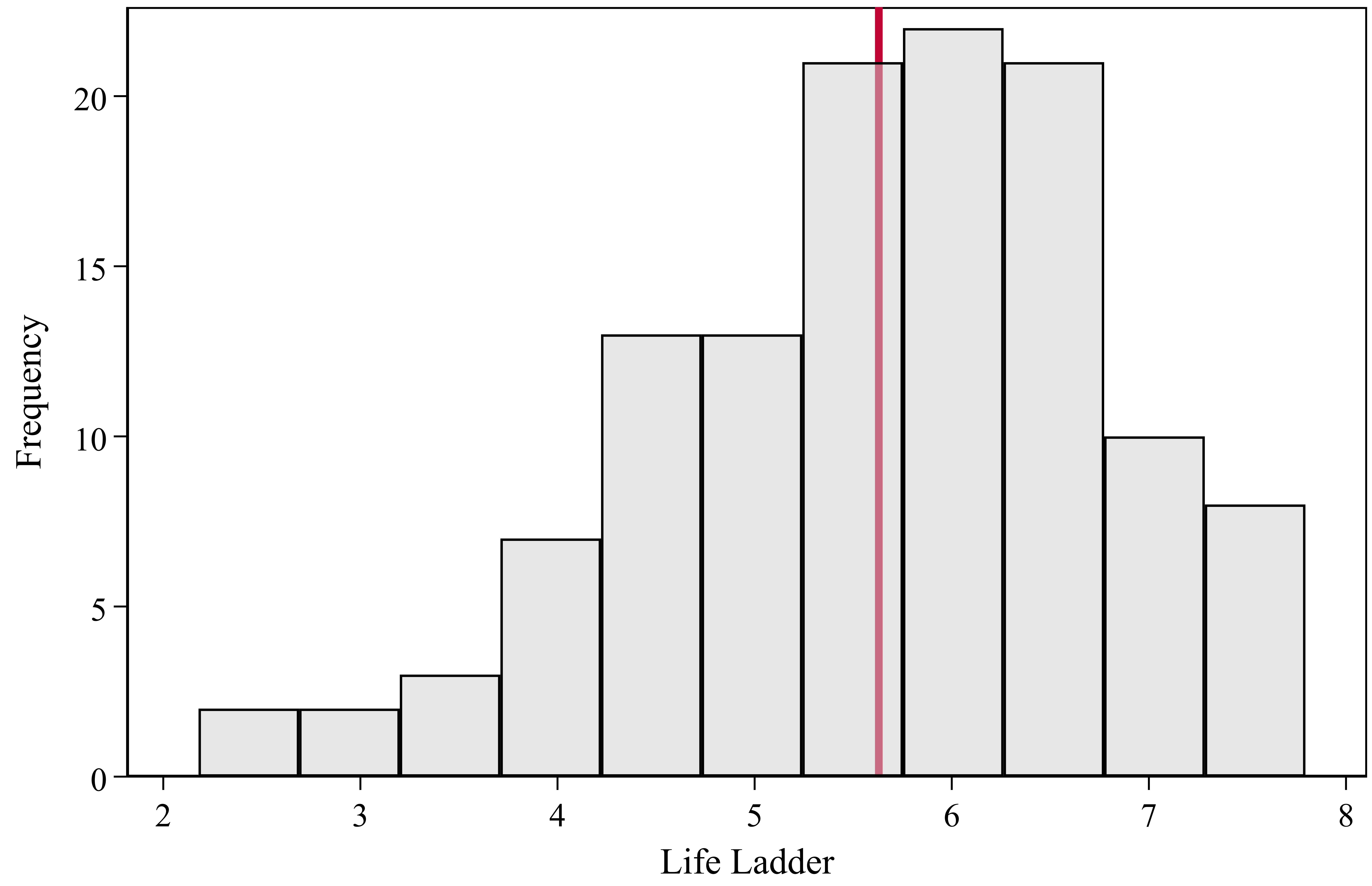
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

My outcome is a country's average **happiness**.

2



	A	B
1		*happiness*
2		
3	Mean	5.635510239
4	Standard Error	0.104633298
5	Median	5.7812784
6	Mode	#N/A
7	Standard Deviation	1.146200355
8	Sample Variance	1.313775254
9	Kurtosis	0.037870895
10	Skewness	-0.522553251
11	Range	5.6155684
12	Minimum	2.1788094
13	Maximum	7.7943778
14	Sum	676.2612287
15	Count	120
16		
17		



Note: Red line represents mean value

2

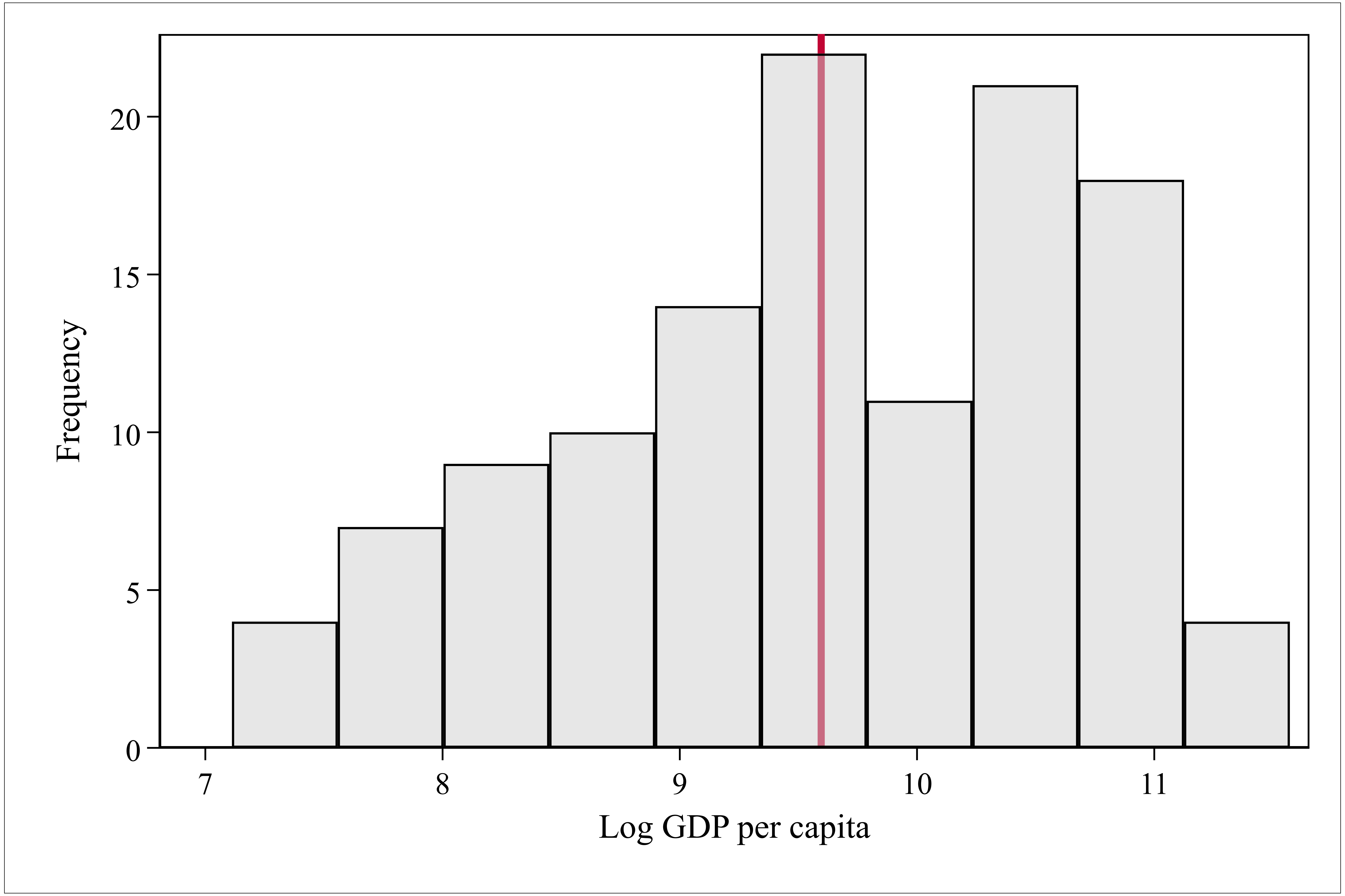
Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

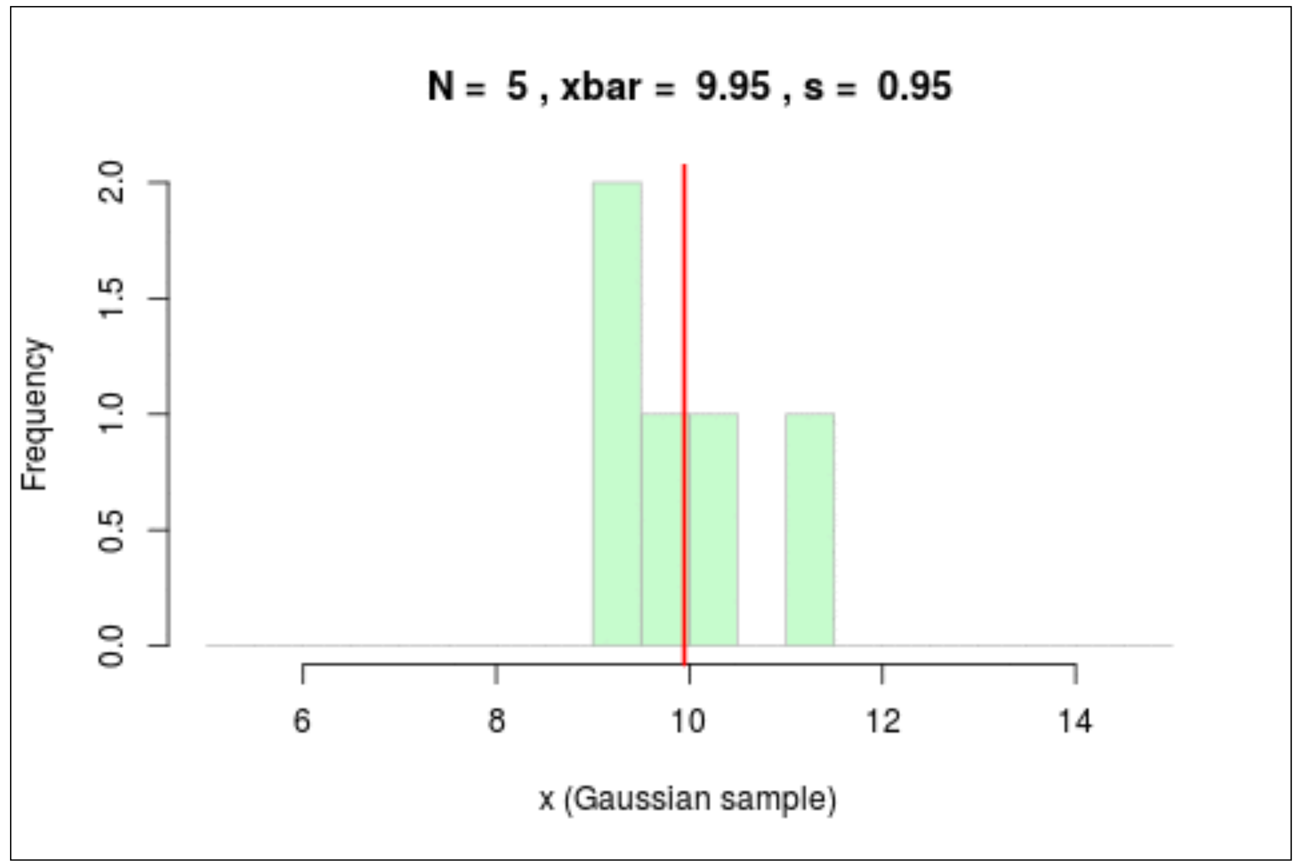
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	98.0363889	98.0363889	198.417241	4.89759E-27
Residual	118	58.3028664	0.49409209		
Total	119	156.339255			

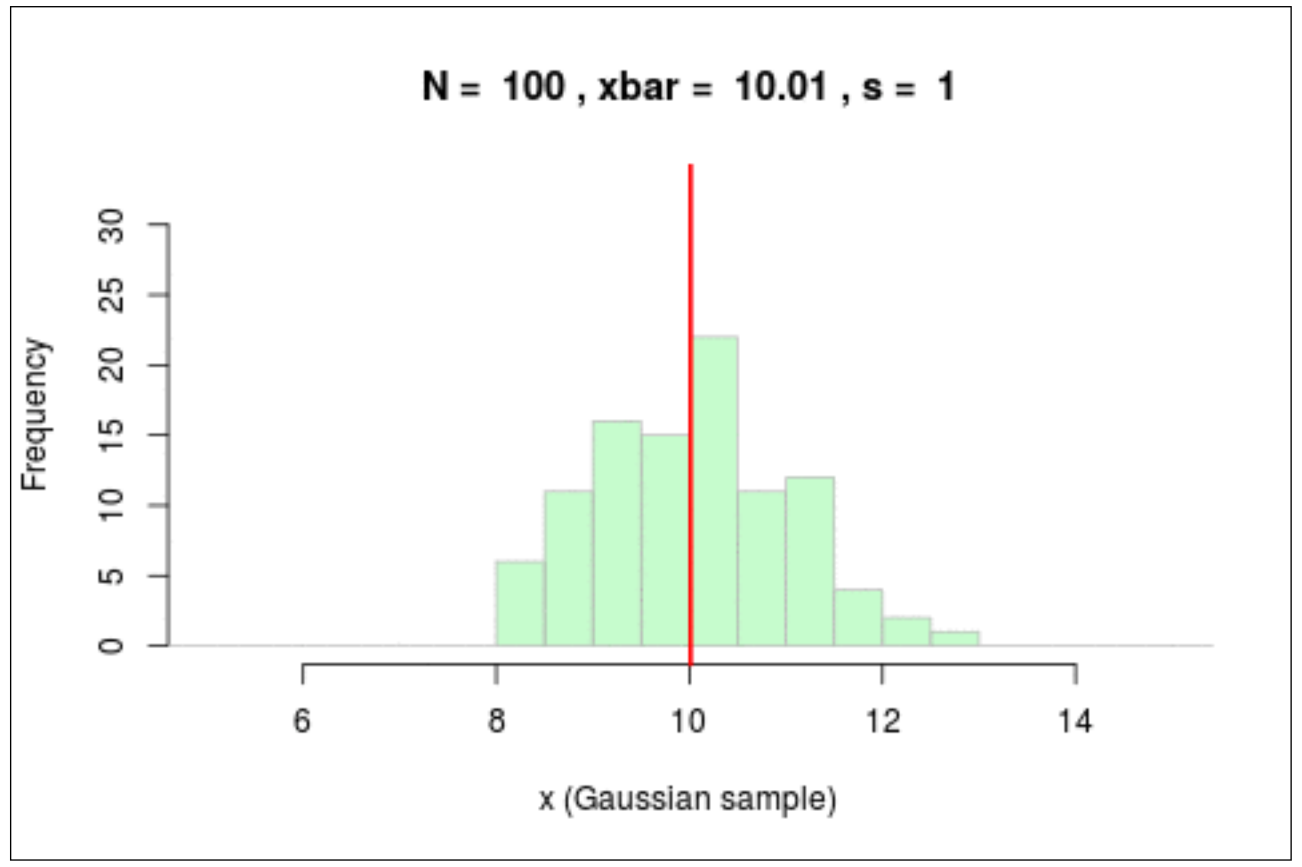
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

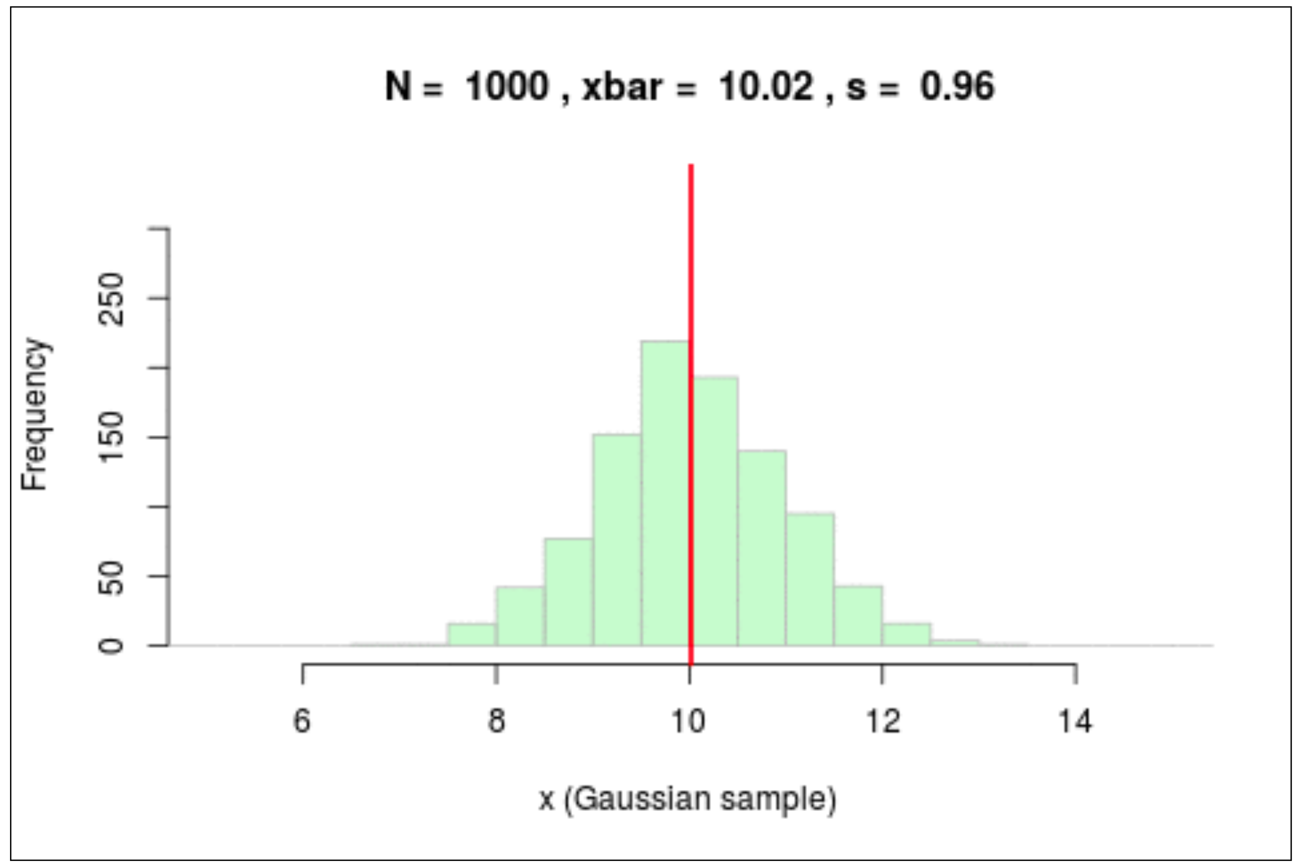
2

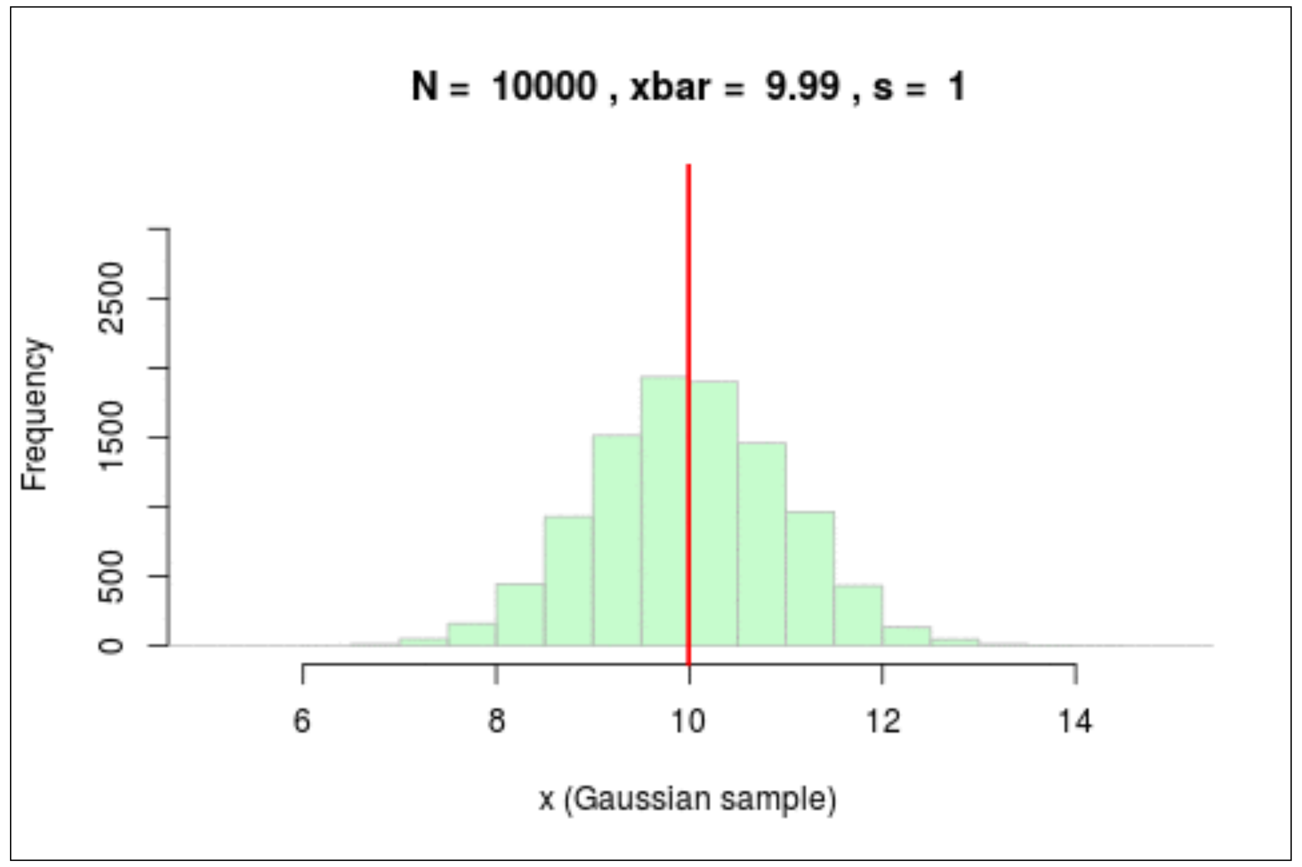


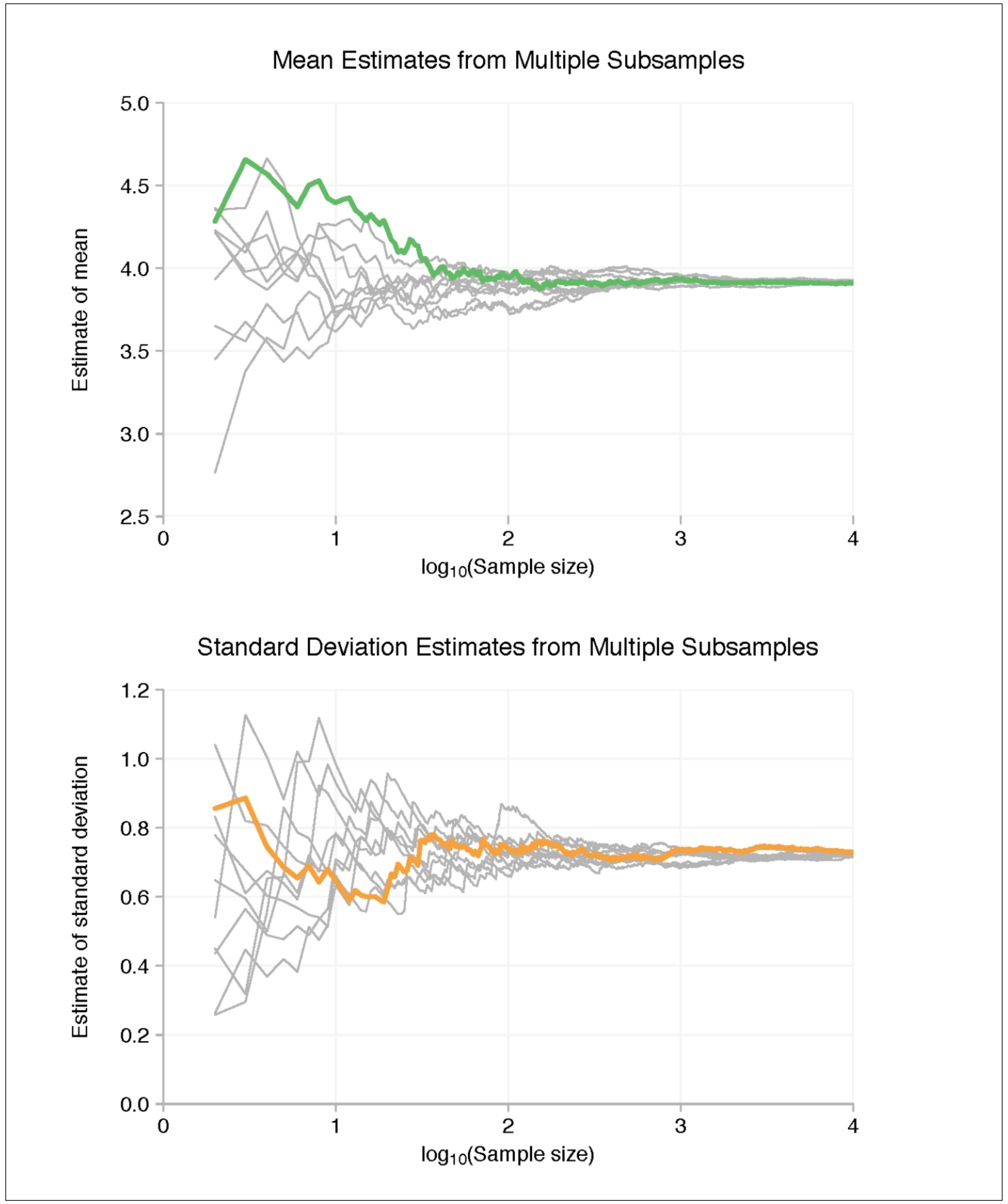
2











Sources: https://aiexplorations.in/2015/09/05/animated-mean-and-sample-size/; https://www.infragistics.com/community/blogs/b/tim_brock/posts/visual-explorations-of-sample-size

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

	df	*SS*	*MS*	
Regression	1	98.03[illegible]3889	98.0363889	19
Residual	118	58.[illegible]028664	0.49409209	
Total	119	1[illegible]6.339255		

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.470[illegible]756	0.57901[illegible]6	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

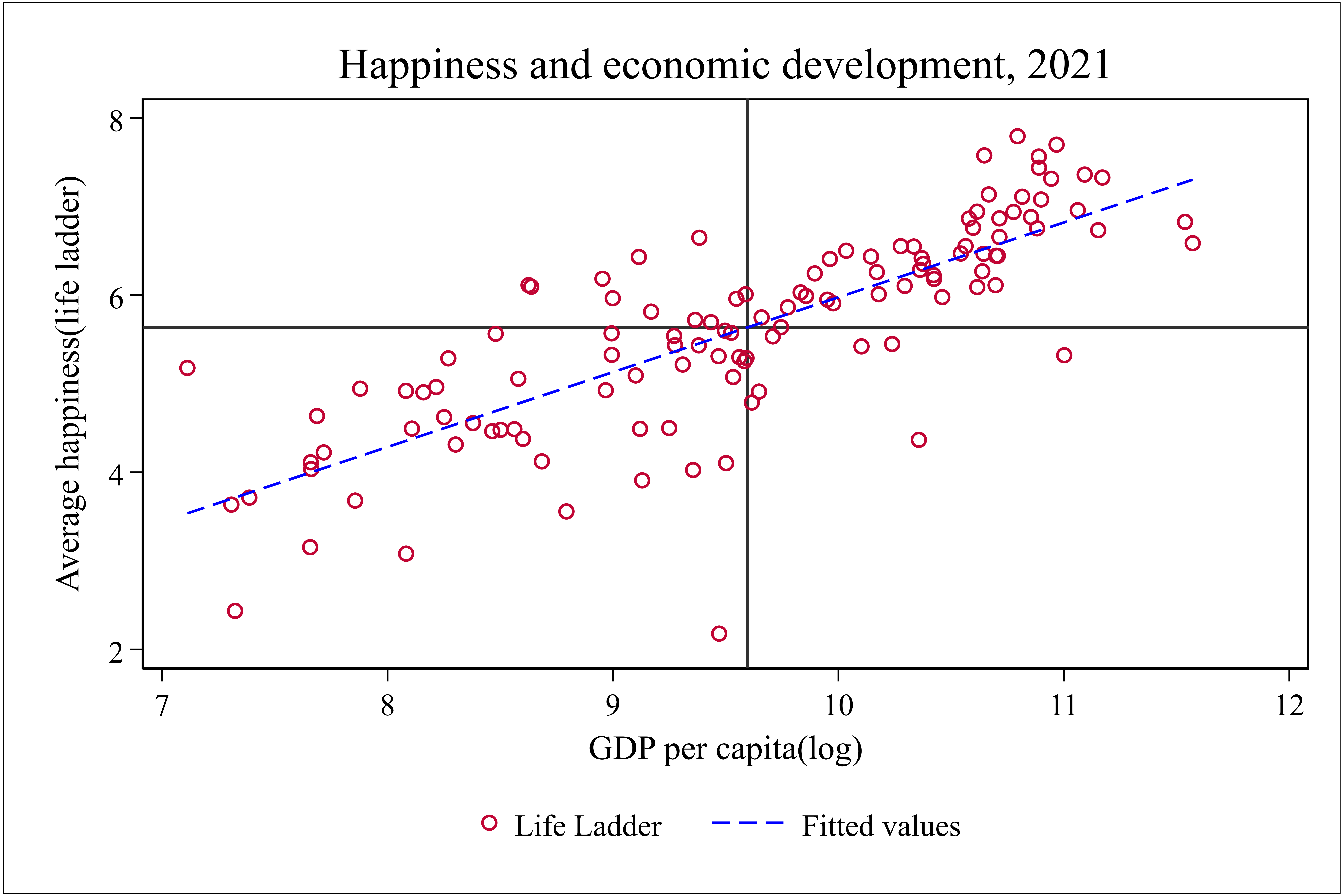
The slope of the regression line (β) is also called the estimated coefficient.

The β and its standard error (α) lets you a hypothesis test using the same t-statistic approach as last week to see if we can conclude that it is statistically significant.

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2},$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}.$$

1



Happiness and economic development, 2021
Average happiness(life ladder)
8
6
4
2
7
8
9
10
11
12
GDP per capita(log)
Life Ladder
Fitted values

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	98.0363889	98.0363889	198.417241	4.89759E-27
Residual	118	58.3028664	0.49409209		
Total	119	156.339255			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

The **Multiple R** in this case is the correlation between our DV and IV.

If we had more than one IV, this would be the multiple correlation between the DV and the IVs.

ANOVA

	df	SS	MS
Regression	1	98.0363889	98.0363889
Residual	118	58.3028664	0.49409209
Total	119	156.339255	

	A	B	C
1		*happiness*	*gdp*
2	happiness	1	
3	gdp	0.79188047	1
4			

	Coefficients	Standard Error	t Stat					95.0%
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

The **R-square** or R^2 is the **coefficient of determination**. In other words the proportion of the DV variation accounted for by the model.

ANOVA

	df	SS
Regression	1	98.0363[illegible]
Residual	118	58.3028[illegible]
Total	119	156.339[illegible]

	Coefficients	Standard Error	...	Upper 95.0%
Intercept	-2.4702756	0.57901[illegible]	[illegible]	[illegible]6709
gdp	0.84466562	0.05996402	[illegible]	[illegible]4119

$$R^2 = \frac{MSS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})}{\sum_{i=1}^{n}(Y_1 - \bar{Y})^2}$$

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

	df	SS	MS
Regression	1	98.0363889	98.03638
Residual	118	58.3028664	0.494092
Total	119	156.339255	

	Coefficients	Standard Error	t Stat
Intercept	-2.4702756	0.57901366	-4
gdp	0.84466562	0.05996462	14

The **Adjusted R Square** is the R^2 adjusted for the number of predictors (e.g., IVs) in the model. It will always be less than the R^2.

$$\overline{R^2} = 1 - \frac{(n-1)}{(n-k)}(1 - R^2)$$

Where n is the number of observations and k is the number of independent variables.

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

ANOVA

	df	SS	MS	F	Significance F
Regression	1	98.0363889	98.036388		
Residual	118	58.3028664	0.494092		
Total	119	156.339255			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-2.4702756	0.57901366	-4.2663511	4.0311E-05	-3.616880299	-1.3236709	-3.6168803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4.8976E-27	0.725919338	0.9634119	0.72591934	0.9634119

The regression's **standard error** is average distance that the observed values fall from the regression line.

The better the regression fit the smaller this value will be.

$$Regression\ standard\ error = \frac{\sum_{i=1}^{n}\left(Y_1 - \hat{Y}_i\right)}{n}$$

2

Regression Statistics	
Multiple R	0.79188047
R Square	0.62707468
Adjusted R Square	0.6239143
Standard Error	0.70291684
Observations	120

The regression's **F statistic** is a measure of the regression's overall significance measured using analysis of variance (ANOVA).

ANOVA

	df	SS	MS	F	Significance F
Regression	1	98.0363889	98.0363889	198.417241	4.89759E-27
Residual	118	58.3028664	0.49409209		
Total	119	156.339255			

With the F statistic, you can do a statistical significance test using the F-distribution for 1 and n-2 degrees of freedom

	Coefficients	tandard Erroi	t Stat	I	5.0%	Upper 95.0%
Intercept	-2.4702756	0.57901366	-4.2663511	4	8803	-1.3236709
gdp	0.84466562	0.05996462	14.0860655	4	1934	0.9634119

For a two-variable regression:

$$F = \frac{(\widehat{\beta_1} \sum y_i x_{1i})}{\sum \widehat{u_i^2}/(n-2)}$$

Equation source: Gujarati (2003: 140)

3

3

Significance

Sign

Size

Table A.1: Observer Effects on Ballot Stuffing

	Ballot stuffing	Confidence Intervals
Observer Present (OP)	**-0.037** (0.025) [-1.51]	(-.09, .01)
Medium Saturation	**0.022** (0.024) [0.92]	(-.03, .07)
High Saturation	**0.010** (0.016) [0.63]	(-.02, .04)
Competition	**0.019** (0.018) [1.03]	(-.02, .06)
Urban	**-0.007** (0.017) [-0.41]	(-.04, .03)
Constant/Intercept	**0.052**** (0.021) [2.55]	**(.01, .09)**
Observations	2,004	
R-squared	0.011	
F(5,59)	1.43,	*p*-value=.223

Note: Robust standard errors in parentheses. t-statistics in square brackets.
*** p<0.01, ** p<0.05, * p<0.1.
Note: Although we provide all statistics here, generally *either* t-statistics *or* standard errors are provided. Confidence intervals are often not provided.

Table source: Long (2016)

3

Long, Abby. 2016. *10 Things to Know About Reading a Regression Table*. Evidence in Governance and Politics (EGAP). Available from https://egap.org/resource/10-things-to-know-about-reading-a-regression-table/ (accessed 1 July 2022).

Abstract

This guide[1] gives basic information to help you understand how to interpret the results of ordinary least squares (OLS) regression in social science research. The guide focuses on regression but also discusses general concepts such as confidence intervals.

The table below that will be used throughout this methods guide is adapted from a study done by EGAP members Miriam Golden, Eric Kramon and their colleagues (J. Asunka et al., "Protecting the Polls: The Effect of Observers on Election Fraud"). The authors performed a field experiment in Ghana in 2012 to test the effectiveness of domestic election observers on combating two common electoral fraud problems: ballot stuffing and overvoting. Ballot stuffing occurs when more ballots are found in a ballot box than are known to have been distributed to voters. Overvoting occurs when more votes are cast at a polling station than the number of voters registered. This table reports a multiple regression (this is a concept that will be further explained below) from their experiment that explores the effects of domestic election observers on ballot stuffing. The sample consists of 2,004 polling stations.

Table A.1: Observer Effects on Ballot Stuffing

	Ballot stuffing	Confidence Intervals
Observer Present (OP)	**-0.037** (0.025) [-1.51]	(-.09, .01)
Medium Saturation	**0.022** (0.024)	(-.03, .07)

Table A.1: Observer Effects on Ballot Stuffing

	Ballot stuffing	Confidence Intervals
Observer Present (OP)	**-0.037** (0.025) [-1.51]	(-.09, .01)
Medium Saturation	**0.022** (0.024) [0.92]	(-.03, .07)
High Saturation	**0.010** (0.016) [0.63]	(-.02, .04)
Competition	**0.019** (0.018) [1.03]	(-.02, .06)
Urban	**-0.007** (0.017) [-0.41]	(-.04, .03)
Constant/Intercept	**0.052**** (0.021) [2.55]	**(.01, .09)**
Observations	2,004	
R-squared	0.011	
F(5,59)	1.43,	*p*-value=.223

Note: Robust standard errors in parentheses. t-statistics in square brackets.
*** p<0.01, ** p<0.05, * p<0.1.
Note: Although we provide all statistics here, generally *either* t-statistics *or* standard errors are provided. Confidence intervals are often not provided.

3

International Journal of Research & Method in Education
Vol. 31, No. 2, July 2008, 155–167

Anxiety in undergraduate research methods courses: its nature and implications

Elena C. Papanastasiou[a]* and Michalinos Zembylas[b]

[a]*University of Nicosia and European Psychometric Services, Nicosia, Cyprus;* [b]*Open University of Cyprus, Nicosia, Cyprus*

(*Received 1 February 2007; final version received 15 February 2008*)

The study reported in this article examines the nature of anxiety that undergraduate students experience in a research methods course and explores some of the factors that influence their anxiety levels. Two questionnaires measuring the attitudes towards research and the anxiety level were administered to 472 students enrolled in a research methods course at the University of Cyprus between the fall of 2002 and the spring of 2005. The results showed that students' self-perceptions seemed to influence the level of anxiety in such courses, while the grades that students were expecting to earn did not predict students' anxiety. Another important finding was that students who considered research to be important for their profession had higher levels of anxiety. Finally, the implications of this study are discussed and teaching interventions are suggested to assist students deal with their anxiety.

Keywords: research methods; attitudes; research anxiety; Attitudes Toward Research scale (ATR)

3

International Journal of Research & Method in Education
Vol. 31, No. 2, July 2008, 155–167

Routledge
Taylor & Francis Group

Anxiety in undergraduate research methods courses: its nature and implications

Elena C. Papanastasiou[a]* and Michalinos Zembylas[b]

[a]*University of Nicosia and European Psychometric Services, Nicosia, Cyprus;* [b]*Open University of Cyprus, Nicosia, Cyprus*

(*Received 1 February 2007; final version received 15 February 2008*)

The study reported in this article examines the nature of anxiety that undergraduate students experience in a research methods course and explores some of the factors that influence their anxiety levels. Two questionnaires measuring the attitudes towards research and the anxiety level were administered to 472 students enrolled in a research methods course at the University of Cyprus between the fall of 2002 and the spring of 2005. The results showed that students' self-perceptions seemed to influence the level of anxiety in such courses, while the grades that students were expecting to earn did not predict students' anxiety. Another important finding was that students who considered research to be important for their profession had higher levels of anxiety. Finally, the implications of this study are discussed and teaching interventions are suggested to assist students deal with their anxiety.

Keywords: research methods; attitudes; research anxiety; Attitudes Toward Research scale (ATR)

The literature suggests that statistics anxiety negatively affects course performance.

(Zeidner 1991; Onwueg- buzie and Seaman 1995; Zanakis and Valenza 1997)

3

Finally, there are several implications from previous research on statistics anxiety in relation to teaching and learning strategies that can alleviate anxiety. For example, Gal and Ginsburg (1994) emphasized that in order to make statistics less threatening and more effective, attention should be focused on students' beliefs and attitudes. Other researchers report specific strategies that help reduce students' anxiety levels; these strategies include: encouraging students, using humour, teaching gimmicks, helping students to understand the course objectives, administering open book exams, using performance assessments, using effective teaching style, provide extensive feedback, addressing ways to relieve anxiety, applying statistics to real world examples and assigning students to work in groups (Onwuegbuzie and Wilson 2003).

Source: Papnastasiou and Zembylas (2008: 158)

3

Research methods

The research questions that are examined in this study are the following:

(1) What are the levels of anxiety experienced by undergraduate students enrolled in a research methods class?
(2) What is the relationship between research methods anxiety and other anxiety types and attitudes towards research?
(3) What variables can explain and predict the anxiety levels of these students?
(4) How does the students' anxiety affect their achievement in the course?

Source: Papnastasiou and Zembylas (2008: 160)

3

"AMAS-C is a 49-item self-report measure designed to assess chronic, manifest anxiety in the college student population. The students had to respond to the AMAS-C on a nominal true/false scale. The construct validity of the scale that was obtained through a factor analysis revealed four subscales: worry anxiety, physiological anxiety, test anxiety and social anxiety."

Papnastasiou and Zembylas (2008: 160)

Note the scale is reversed in some models so that higher values suggest lower anxiety levels.

3

32 questions measured on a seven-point Likert Scale ranging from **1 (strongly disagree)** to **7 (strongly agree)**. These questions are combined into **five sub scales**:

Usefulness of research to students' profession

Research anxiety

Positive attitudes to research

Relevance of research to students' personal lives

Research difficulty

Papnastasiou and Zembylas (2008: 160)

3

472 students enrolled in an undergraduate methods course for education students at the University of Cyprus from **2002 to 2005**.

What **population** do you think this sample is part of?

3

Table 3. Predicting anxiety from the other ATR scales.

	Unstandardized coefficients		Standardized coefficients		
Subscales	*β*	Std. error	*β*	*t*	Sig.
Constant	0.033	0.374		0.087	0.931
Usefulness for the profession	–0.333	0.089	–0.255	–3.755	0.000
Positive attitudes	0.570	0.073	0.472	7.805	0.000
Relevance to life	0.118	0.081	0.086	1.465	0.144
Research difficulty	0.477	0.054	0.419	8.852	0.000

$F = 62.258$, $p = 0.000$.

$R^2 = .452$

Separate analyses found **gender differences in anxiety** but not in difficulty.

Source: Papanastasiou and Zembylas (2008: 162)

3

Table 6. Predicting the research methods course grade.

	Unstandardized coefficients		Standardized coefficients		
	β	Std. error	*β*	*t*	Sig.
Constant	6.840	0.451		15.183	0.000
Usefulness for the profession	0.387	0.109	0.318	3.546	0.000
Anxiety	–0.177	0.070	–0.190	–2.520	0.012
Positive attitudes	0.075	0.096	0.066	0.774	0.439
Relevance to life	–0.128	0.097	–0.101	–1.319	0.188
Research difficulty	0.066	0.073	0.062	0.901	0.368

$F = 6.56$, $p = 0.000$

$R^2 = 0.102$.

Source: Papanastasiou and Zembylas (2008: 163)

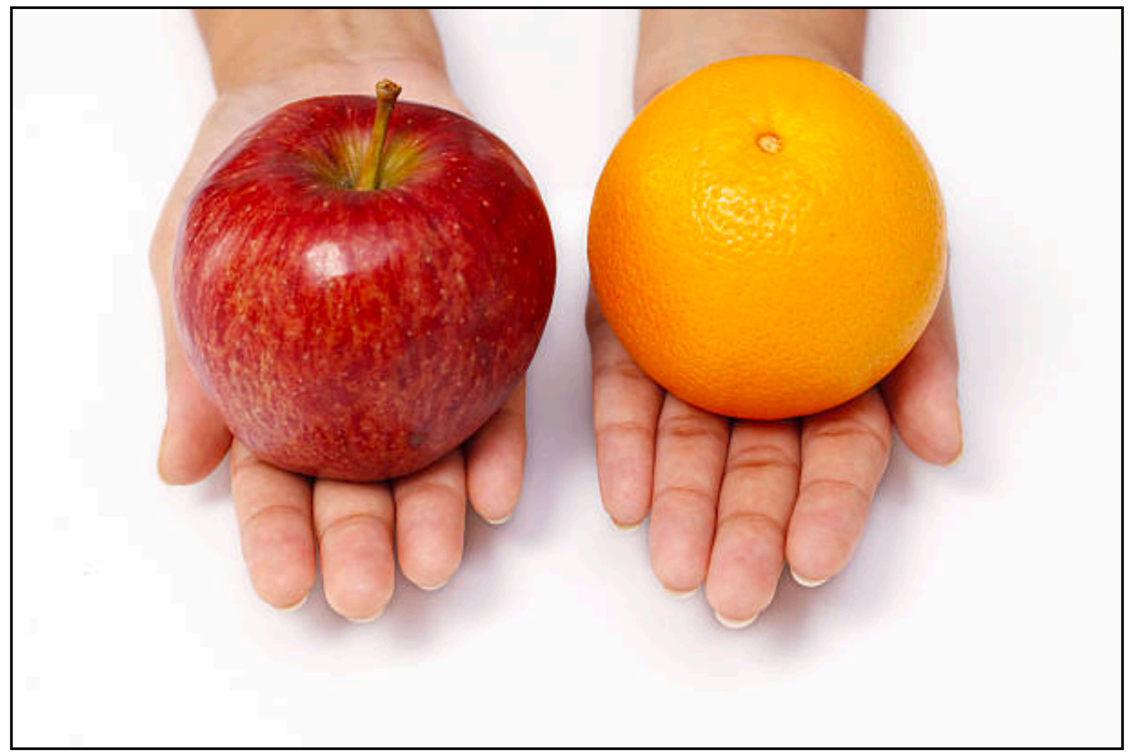
3

Most of the time our independent variables use different measurement units.

This makes direct comparison of regression coefficients difficult.

Standardising the coefficients puts the coefficients on the same scale, which aids comparability.

This comes at a cost of easily understanding one unit change in the independent variable.



Source: https://ccnb.com.au/comparing-apples-with-oranges/

3

$$\beta_1^* = \hat{\beta}_1 \left(\frac{sd_{x_1}}{sd_y} \right)$$

Where:

$\hat{\beta}_1$ equals the unstandardised coefficient for variable x_1.

sd_{x_1} equals the sample standard deviation for variable x_1.

sd_y equals the sample standard deviation for the outcome variable y.

Side note: You can also standardise the variable instead of the coefficient.

Source: Gujarati, Damodar N. 2003. *Basic Econometrics*. 4th edition. New York: McGraw Hill: 175.

4

Why and how do we run a regression?

Why and how do you interpret regression results?

4