
Z-Axis (Sales, $)
b1 = is the
intersection of the
hen x and y are zero
z = b1
ex: sales = b1
The change in z with no change in y: z = b1 +b2(x)
ex., as price pushes sales down slope is negative.
with no advertisement (y=0), thus
sales = b1 - b2(price),
b2= sales drop/increase in price
b2(price)
green dot = observed sales (z)
black dot = estimated sales (z^)
red line = residuals (error) = z - z^
e = z - ((b1+b2(x)+b3(y))
= observed sales - estimated sales
= sales - (b1-b2(price) + b3(advertisement))
Multiple regression estimate of z = geometric plane
Plane: z^ = b1 + b2(x) + b3(y)
Estimated sales = b1 - b2(price) + b3(advertisement)
The plane above is the change in
change in z = + b3(y)
advertisement has a positive influence
sales with no change in price, thus
added sales = + b3(advertisement)
b3 = $ sales increase/advertisement $
X-Axis (Price, $)


1
2
3
Z-Axis (Sales, $)
z = b1
ex: sales = b1
green dot = observed sales (z)
black dot = estimated sales (z^)
red line = residuals (error) = z - z^
e = z - ((b1+b2(x)+b3(y))
= observed sales - estimated sales
= sales - (b1-b2(price)+b3(advertisement))
The change in z with no change in y: z = b1 +b2(x)
ex., as price pushes sales down slope is negative.
with no advertisement (y=0), thus
sales = b1 - b2(price),
b2= sales drop/increase in price
b2(price)
Multiple regression estimate of z = geometric plane
Plane: z^ = b1 + b2(x) + b3(y)
Estimated sales = b1 - b2(price) + b3(advertisement)
X-Axis (Price, $)

1

1

Why do we need to move from bivariate to multivariate regression?

How do we do so?

How do we **interpret** our results?

1

1. A credible **causal mechanism**

2. Ruling out **endogeneity**

3. **Covariation**

4. Controlling for **confounding variables** that may make current association **spurious**?

Source: Kelstedt & Whitten (2018: 55)

1

Taking this step to multiple regression finally enables us to potentially pass the **fourth hurdle** for the first* time.

* not counting experimental methods

1

Experimental research designs control for other theoretically relevant factors through **random assignment** into treatment and control groups.

This is the **gold standard**.

Observational research designs control for other factors by **adding them** to the regression model.

This involves (1) **theoretically identifying** relevant factors (e.g., **culture**) and (2) **finding observable/measurable indicators** of them.

1

$$Y_i = \alpha + \beta X_i + u_i$$

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.
u = population error term/residual

1

$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_1$$

Where:

Y is the outcome you are trying to explain.
X is the main explanatory variable.
Z is an additional explanatory/control variable
α (alpha) is the value of Y when X=0 & Z=0.
β_1 (beta) is the estimated effect of X on Y holding constant the effects of Z.
β_2 (beta) is the estimated effect of Z on Y holding constant the effects of X.
u = population error term/residual

1

$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_1$$

Where:

Y is the outcome you are trying to explain.
X is the main explanatory variable.
Z is an additional explanatory/control variable.
α **(alpha) is the value of Y when X=0 & Z=0.**
β_1 **(beta) is the estimated effect of X on Y holding constant the effects of Z.**
β_2 **(beta) is the estimated effect of Z on Y holding constant the effects of X.**
u = population error term/residual

1

Bivariate——$Y_i = \alpha + \beta X_i + u_i$
Multivariate—$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_i$

Notice the **subscripts** for the variables and the slope coefficients (βs).

The subscript **"i"** tells us that the equation is for each observation of our variables from *i* to *n*.

Variables by themselves (e.g., **Y, X, Z**) actually represent a **vector** (i.e.values of each variable).

1

Bivariate——-—$Y_i = \alpha + \beta X_i + u_i$
Multivariate—$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_1$

Adding in $\beta_2 Z_i$ to the equation changes the estimation of $\beta_1 X_i$, sometimes by a little, sometimes a lot.

Four possible changes

1. β_1 was statistically significant but is **no longer statistically significant**
2. β_1 was not statistically significant but is **now statistically significant**
3. β_1 value is larger than before (i.e., potentially **more substantively significant**)
4. β_1 value is smaller than before (i.e., potentially **less substantively significant**)

1

Two-variable regression slope: $\beta = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1} (X_i - \bar{X})^2}$

Multivariate regression slope: $\widehat{\beta_1} = \frac{\sum_{i=1}^{n} (X_i - \widehat{X_i})(Y_i - \widehat{Y_i^*})}{\sum_{i=1} (X_i - \widehat{X_i})^2}$

where $\widehat{Y}_i^* = \widehat{\alpha}^* + \widehat{\beta}^* Z_i$

And $\widehat{X_i}$ is the predicted value of X based on Z.

1

Perfect multicollinearity is when there is an linear relationship between two independent variables.

Some **examples** of perfect multicollinearity include **spatial** (e.g., EU country/non-EU country) or **temporal** (Cold War/Post-Cold War) dummy variables.

If there is **perfect multicollinearity**, one of the variables will be **automatically dropped** by your software.

If there is **high collinearity** between independent variables, this will **distort your parameter estimates**.

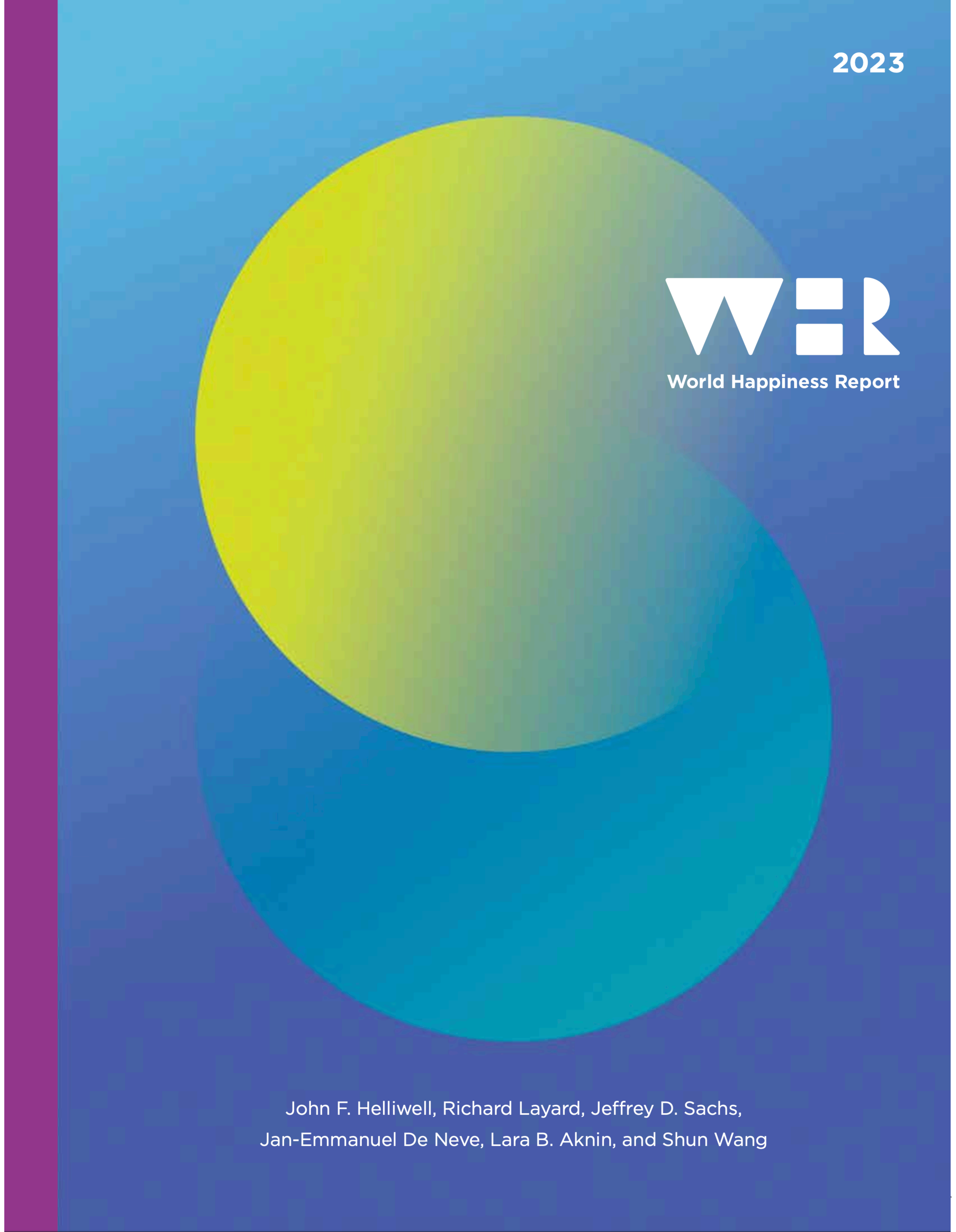
There are **tests for multicollinearity** (e.g., variance inflation factors), but initially I would suggest creating a **correlation table** of your independent variables.


2

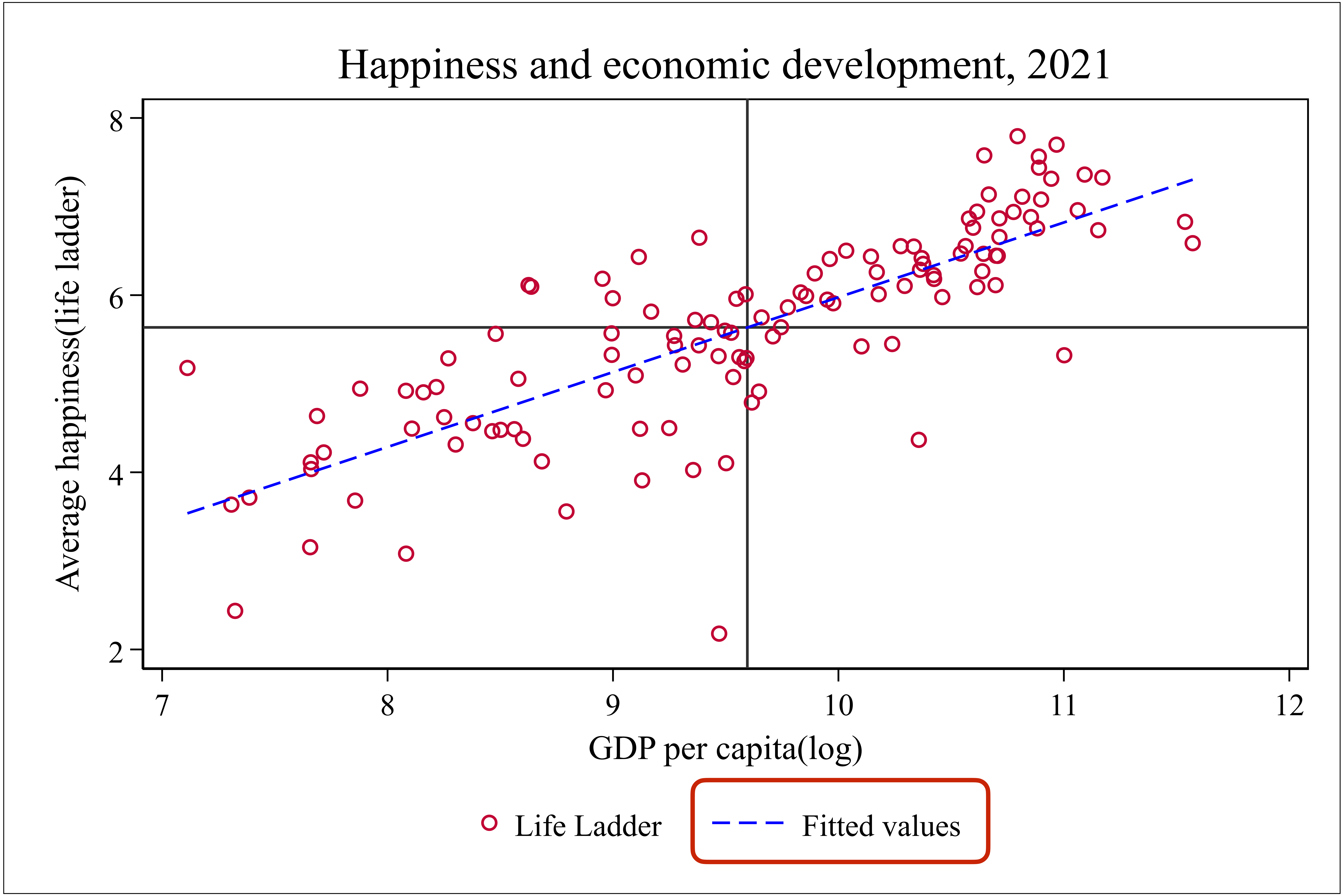
2

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

2


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

2


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

Y=Happiness; **X**=GDP; **Z**=Freedom

Bivariate: $Y_i = \alpha + \beta X_i$

= -2.47 + 0.85**X**

$\widehat{Y_{Australia}}$= -2.47 + 0.85(10.82) = 7.27 (actual value is 7.11)

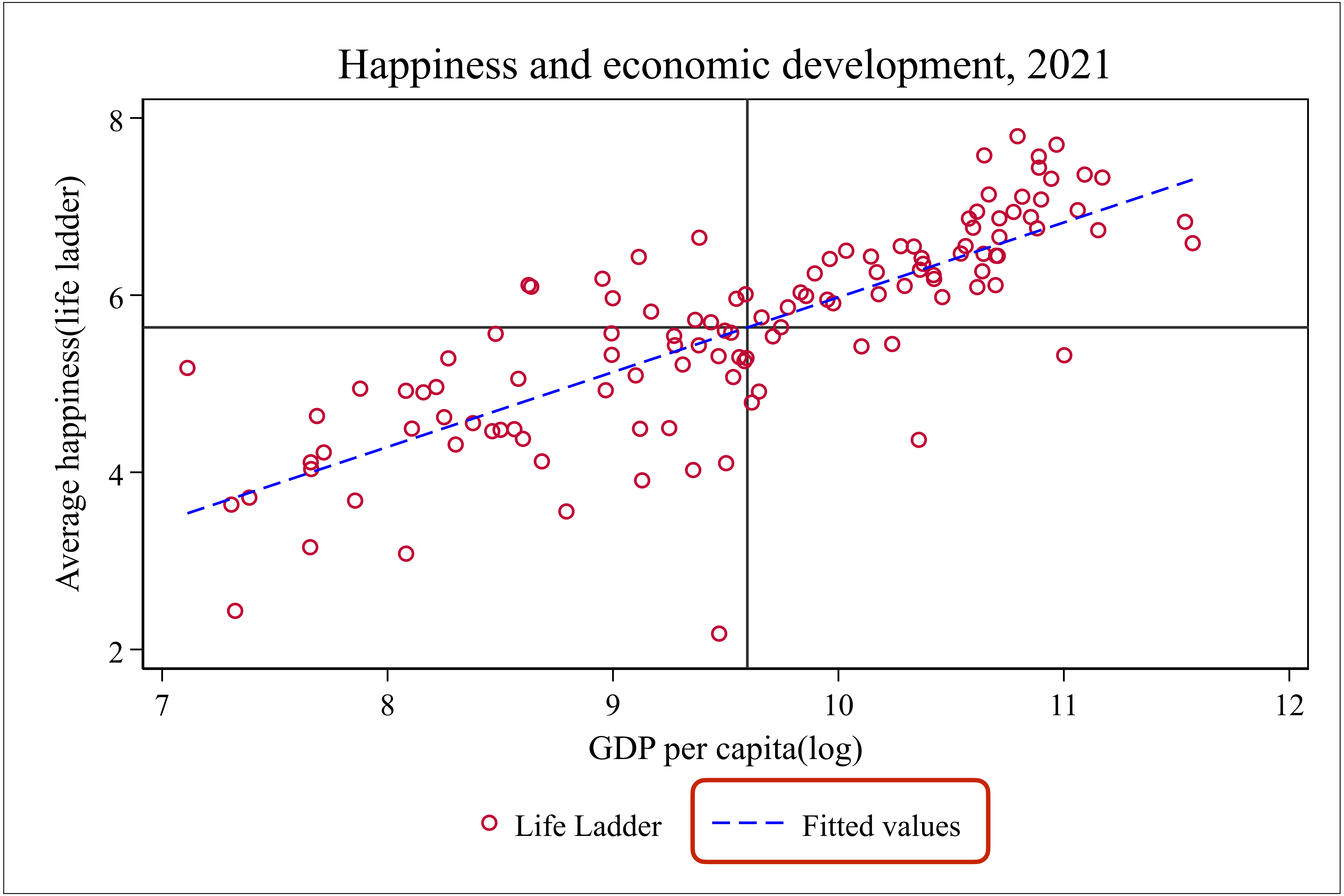
Multivariate: $Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + u_1$

$\widehat{Y_i}$ = -4.19 + 0.72X + 3.74Z

$\widehat{Y_{Australia}}$= -4.19 + 0.72(10.82) + 3.74(0.91) = 7.38 (actual value is 7.11)

All intercepts and slope coefficients are statistically significant at the 0.001 level.

2



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

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.914511457
R Square	0.836331204
Adjusted R Square	0.823861201
Standard Error	0.488555351
Observations	114

ANOVA

	df	SS	MS	F	Significance F
Regression	8	128.064648	16.008081	67.0674395	8.62998E-38
Residual	105	25.0620647	0.23868633		
Total	113	153.126713			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-2.839184217	0.90619544	-3.1330816	0.00224225	-4.636002371	-1.0423661	-4.6360024	-1.0423661
gdp	0.425234417	0.11208314	3.79391955	0.00024821	0.202994254	0.64747458	0.20299425	0.64747458
socialsupport	3.061116942	0.74453716	4.11143607	7.825E-05	1.584837285	4.5373966	1.58483729	4.5373966
life_expectancy	0.000490025	0.01918002	0.02554871	0.97966579	-0.037540405	0.03852045	-0.0375404	0.03852045
freedom	1.459257323	0.61422166	2.37578291	0.01932505	0.241369236	2.67714541	0.24136924	2.67714541
generosity	-0.048945192	0.33205007	-0.147403	0.88309657	-0.707339136	0.60944875	-0.7073391	0.60944875
corruption	-0.722962505	0.30732942	-2.3524025	0.02051661	-1.332339977	-0.113585	-1.33234	-0.113585
positiveaffect	1.850464041	0.61997154	2.98475642	0.00353085	0.621174996	3.07975309	0.621175	3.07975309
negativeaffect	0.23832172	0.78087249	0.30519928	0.76081856	-1.310004172	1.78664761	-1.3100042	1.78664761

2

Table 2.1: Regressions to Explain Average Happiness across Countries (Pooled OLS)

	Dependent Variable			
Independent Variable	Cantril Ladder (0-10)	Positive Affect (0-1)	Negative Affect (0-1)	Cantril Ladder (0-10)
Log GDP per capita	0.359	-.015	-.001	0.392
	(0.067)***	(0.009)	(0.007)	(0.065)***
Social support (0-1)	2.526	0.318	-.337	1.865
	(0.356)***	(0.056)***	(0.046)***	(0.35)***
Healthy life expectancy at birth	0.027	-.0005	0.003	0.028
	(0.01)***	(0.001)	(0.001)***	(0.01)***
Freedom to make life choices (0-1)	1.331	0.371	-.090	0.505
	(0.297)***	(0.041)***	(0.039)**	(0.278)*
Generosity	0.537	0.088	0.027	0.33
	(0.256)**	(0.032)***	(0.027)	(0.245)
Perceptions of corruption (0-1)	-.716	-.009	0.094	-.712
	(0.262)***	(0.027)	(0.022)***	(0.249)***
Positive affect (0-1)				2.285
				(0.331)***
Negative affect (0-1)				0.185
				(0.388)
Year fixed effects	Included	Included	Included	Included
Number of countries	156	156	156	156
Number of observations	1,964	1,959	1,963	1,958
Adjusted R-squared	0.757	0.439	0.334	0.782

Notes: This is a pooled OLS regression for a tattered panel explaining annual national average Cantril ladder responses from all available surveys from 2005 through 2022. See Technical Box 2 for detailed information about each of the predictors. Coefficients are reported with robust standard errors clustered by country (in parentheses). ***, **, and * indicate significance at the 1, 5, and 10 percent levels respectively.

Source: World Happiness Report 2023: 38

2

Table 2.1: Regressions to Explain Average Happiness across Countries (Pooled OLS)

	Dependent Variable		
Independent Variable	Cantril Ladder (0-10)	[redacted]	Cantril Ladder (0-10)
Log GDP per capita	0.359		0.392
	(0.067)***		(0.065)***
Social support (0-1)	2.526		1.865
	(0.356)***		(0.35)***
Healthy life expectancy at birth	0.027		0.028
	(0.01)***		(0.01)***
Freedom to make life choices (0-1)	1.331		0.505
	(0.297)***		(0.278)*
Generosity	0.537		0.33
	(0.256)**		(0.245)
Perceptions of corruption (0-1)	-.716		-.712
	(0.262)***		(0.249)***
Positive affect (0-1)			2.285
			(0.331)***
Negative affect (0-1)			0.185
			(0.388)
Year fixed effects	Included		Included
Number of countries	156		156
Number of observations	1,964		1,958
Adjusted R-squared	0.757		0.782

Notes: This is a pooled OLS regression for a tattered panel explaining annual national average Cantril ladder responses from all available surveys from 2005 through 2022. See Technical Box 2 for detailed information about each of the predictors. Coefficients are reported with robust standard errors clustered by country (in parentheses). ***, **, and * indicate significance at the 1, 5, and 10 percent levels respectively.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.915
R Square	0.836
Adjusted R Square	0.824
Standard Error	0.489
Observations	114

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	128.065	16.008	67.067	0.000
Residual	105	25.062	0.239		
Total	113	153.127			

	Coefficients	*S.E.*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-2.839	0.906	-3.133	0.002	-4.636	-1.042
gdp	0.425	0.112	3.794	0.000	0.203	0.647
socialsupport	3.061	0.745	4.111	0.000	1.585	4.537
life_expectancy	0.000	0.019	0.026	0.980	-0.038	0.039
freedom	1.459	0.614	2.376	0.019	0.241	2.677
generosity	-0.049	0.332	-0.147	0.883	-0.707	0.609
corruption	-0.723	0.307	-2.352	0.021	-1.332	-0.114
positiveaffect	1.850	0.620	2.985	0.004	0.621	3.080
negativeaffect	0.238	0.781	0.305	0.761	-1.310	1.787

Source: World Happiness Report 2023: 38

3

3

Theoretically, you want to:

(1) **build on the best existing research** and show you can replicate/approximate it,

(2) demonstrate whether your results support or fail to support your **alternate hypothesis(es)**, and

(3) demonstrate whether or not your results are **robust** to alternate theoretical and practical specifications.

3

You want to include **enough information** to allow readers to:

(1) **understand** what you did,

(2) **reach their own conclusions** as to whether your results are statistically and substantively significant,

(3) **replicate** your research if they are interested.

3


POLITICAL ECONOMY RESEARCH INSTITUTE
PERI
University of Massachusetts Amherst

Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Thomas Herndon, Michael Ash and Robert Pollin

April 2013

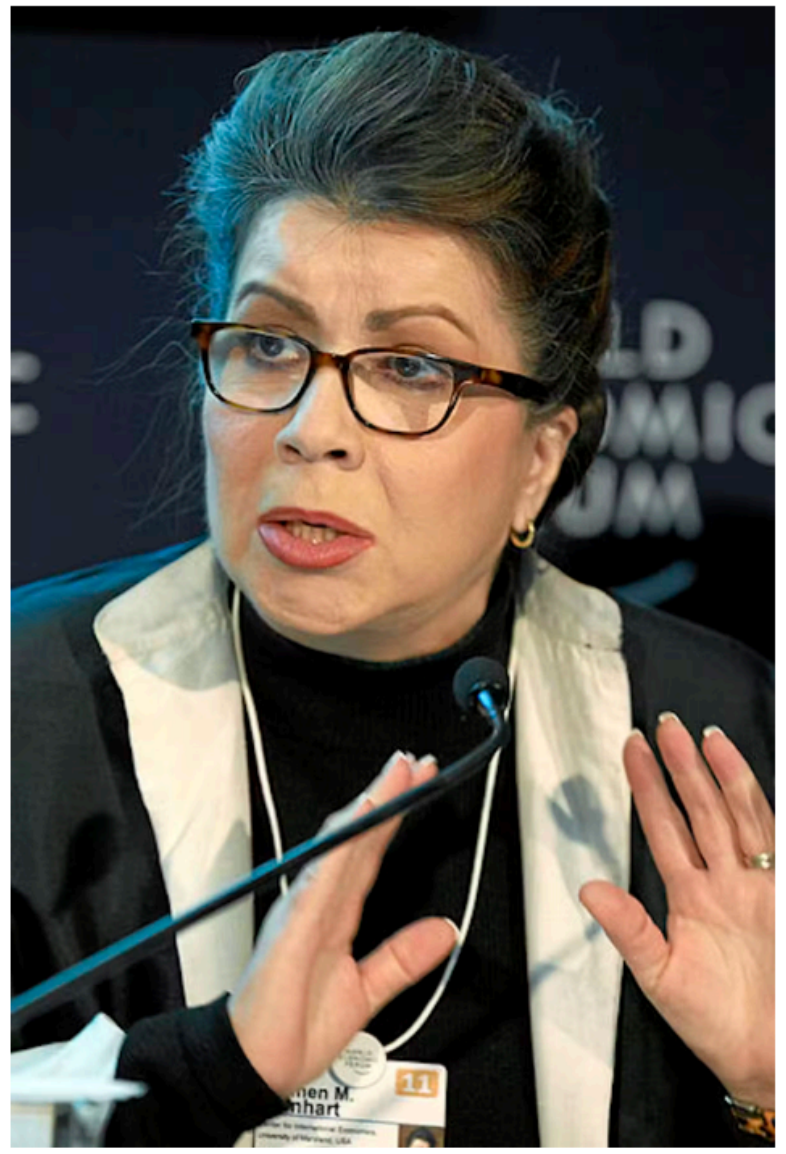
WORKINGPAPER SERIES
Number 322

POLITICAL ECONOMY RESEARCH INSTITUTE

Gordon Hall
418 North Pleasant Street
Amherst, MA 01002

Phone: 413.545.6355
Fax: 413.577.0261
peri@econs.umass.edu
www.peri.umass.edu

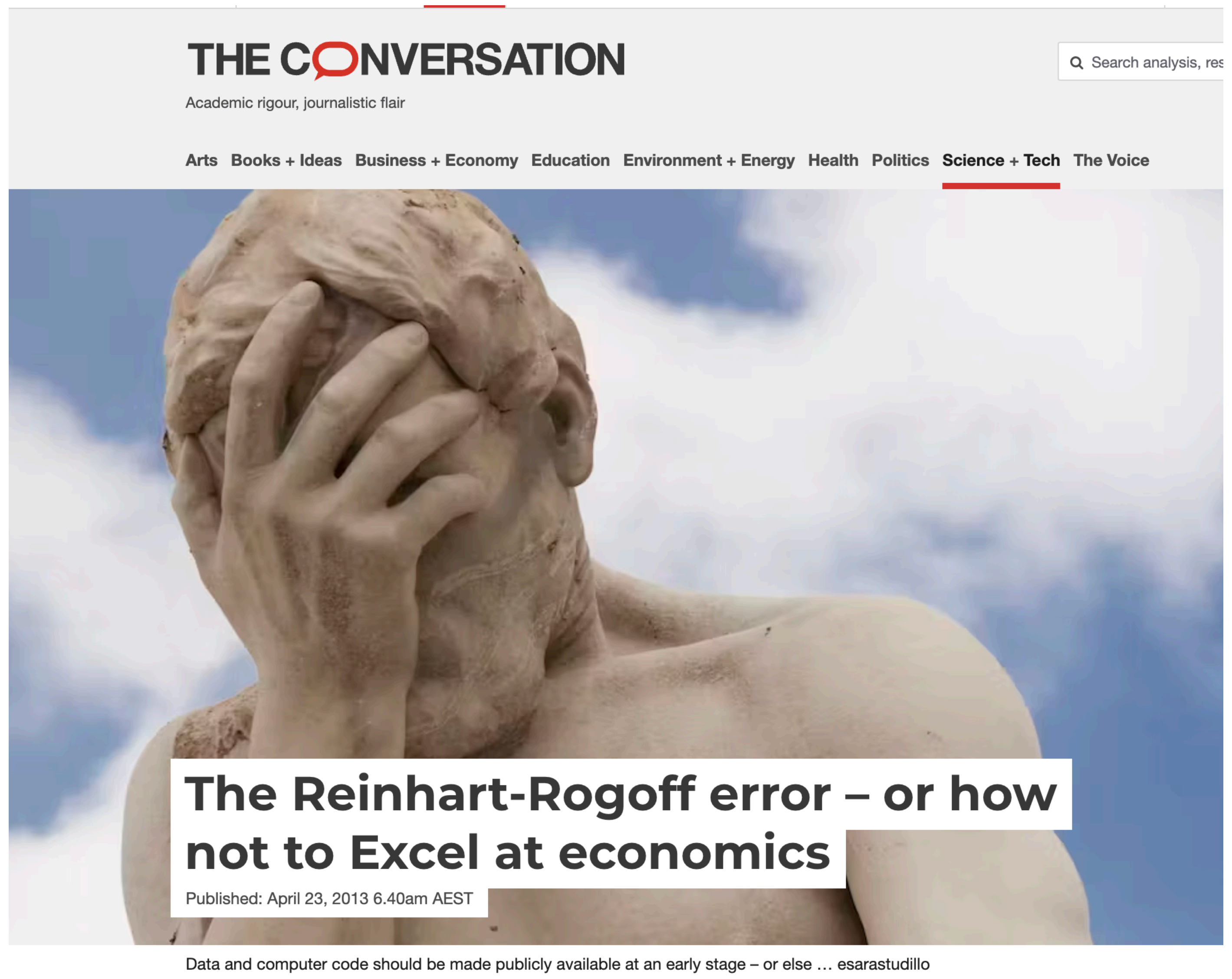
UMASS


Carmen Reinhart. Wikimedia Commons

During their analysis, Herndon, Ash and Pollin obtained the actual spreadsheet that Reinhart and Rogoff used for their calculations; and after analysing this data, they identified three errors.

The most serious was that, in their Excel spreadsheet, Reinhart and Rogoff had not selected the entire row when averaging growth figures: they omitted data from Australia, Austria, Belgium, Canada and Denmark.

In other words, they had accidentally only included 15 of the 20 countries under analysis in their key calculation.



Source: https://theconversation.com/the-reinhart-rogoff-error-or-how-not-to-excel-at-economics-13646

3

Katalin Karikó was born in 1955 in Szolnok, Hungary. She received her PhD from Szeged's University in 1982 and performed postdoctoral research at the Hungarian Academy of Sciences in Szeged until 1985. She then conducted postdoctoral research at Temple University, Philadelphia, and the University of Health Science, Bethesda. In 1989, she was appointed Assistant Professor at the University of Pennsylvania, where she remained until 2013. After that, she became vice president and later senior vice president at BioNTech RNA Pharmaceuticals. Since 2021, she has been a Professor at Szeged University and an Adjunct Professor at Perelman School of Medicine at the University of Pennsylvania.

Demoted from U.Penn in 1995 when unable to secure grants. "Not of faculty quality".

Seminal paper desk rejected from *Nature* in 2005.

3

Research is an often **messy**, time-intensive, stressful, and confusing process.

There are usually **multiple ways to define** your dependent variable (e.g., continuous, dichotomous, change, logged).

Often people will study **multiple variations** of their dependent variable and only report one's results.

3

What is/are your main **independent** variables?

What are **other factors** (i.e., control variables) that theoretically affect your outcome?

What are the most theoretically grounded way of **measuring** these factors (e.g., absolute value, % GDP, % population, logged)?

3

Usually, there is a **standard/influential model** in your research area.

Report your replication of those results.

Then compare the results with **your best model**.

Add additional models to incorporate other hypotheses, control variables, or methodological concerns.

3

Make sure your **sample** includes what you think it includes.

Think about how it represents/fails to represent the **population** you are theorising about.

Explore the data for potential **outliers** or cases with **missing data**.

Think about important **cross-temporal or cross-spatial differences**.

2

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.915
R Square	0.836
Adjusted R Square	0.824
Standard Error	0.489
Observations	114

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	128.065	16.008	67.067	0.000
Residual	105	25.062	0.239		
Total	113	153.127			

	Coefficients	*S.E.*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-2.839	0.906	-3.133	0.002	-4.636	-1.042
gdp	0.425	0.112	3.794	0.000	0.203	0.647
socialsupport	3.061	0.745	4.111	0.000	1.585	4.537
life_expectancy	0.000	0.019	0.026	0.980	-0.038	0.039
freedom	1.459	0.614	2.376	0.019	0.241	2.677
generosity	-0.049	0.332	-0.147	0.883	-0.707	0.609
corruption	-0.723	0.307	-2.352	0.021	-1.332	-0.114
positiveaffect	1.850	0.620	2.985	0.004	0.621	3.080
negativeaffect	0.238	0.781	0.305	0.761	-1.310	1.787

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.915
R Square	0.836
Adjusted R Square	0.824
Standard Error	0.489
Observations	114

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	128.065	16.008	67.067	0.000
Residual	105	25.062	0.239		
Total	113	153.127			

	Coefficients	*S.E.*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-2.839	0.906	-3.133	0.002	-4.636	-1.042
gdp	0.425	0.112	3.794	0.000	0.203	0.647
socialsupport	3.061	0.745	4.111	0.000	1.585	4.537
life_expectancy	0.000	0.019	0.026	0.980	-0.038	0.039
freedom	1.459	0.614	2.376	0.019	0.241	2.677
generosity	-0.049	0.332	-0.147	0.883	-0.707	0.609
corruption	-0.723	0.307	-2.352	0.021	-1.332	-0.114
positiveaffect	1.850	0.620	2.985	0.004	0.621	3.080
negativeaffect	0.238	0.781	0.305	0.761	-1.310	1.787

Table 2.1: Regressions to Explain Average Happiness across Countries (Pooled OLS)

	Dependent Variable		
Independent Variable	Cantril Ladder (0-10)	[redacted]	Cantril Ladder (0-10)
Log GDP per capita	0.359	[redacted]	0.392
	(0.067)***	[redacted]	(0.065)***
Social support (0-1)	2.526	[redacted]	1.865
	(0.356)***	[redacted]	(0.35)***
Healthy life expectancy at birth	0.027	[redacted]	0.028
	(0.01)***	[redacted]	(0.01)***
Freedom to make life choices (0-1)	1.331	[redacted]	0.505
	(0.297)***	[redacted]	(0.278)*
Generosity	0.537	[redacted]	0.33
	(0.256)**	[redacted]	(0.245)
Perceptions of corruption (0-1)	-.716	[redacted]	-.712
	(0.262)***	[redacted]	(0.249)***
Positive affect (0-1)		[redacted]	2.285
		[redacted]	(0.331)***
Negative affect (0-1)		[redacted]	0.185
		[redacted]	(0.388)
Year fixed effects	Included	[redacted]	Included
Number of countries	156	[redacted]	156
Number of observations	1,964	[redacted]	1,958
Adjusted R-squared	0.757	[redacted]	0.782

Notes: This is a pooled OLS regression for a tattered panel explaining annual national average Cantril ladder responses from all available surveys from 2005 through 2022. See Technical Box 2 for detailed information about each of the predictors. Coefficients are reported with robust standard errors clustered by country (in parentheses). ***, **, and * indicate significance at the 1, 5, and 10 percent levels respectively.

Source: World Happiness Report 2023: 38

3

Tell your readers in words **what you want them to take away** from your table.

Often focus is on both **statistical** and **substantive** significance.

Connect results back to your **theory** and **hypotheses**.

3

Why do we need to move from bivariate to multivariate regression?

How do we do so?

How do we **interpret** our results?

4