

Week 1: What Is Quantitative Research?

POP88162 Introduction to Quantitative Research Methods

Tom Paskhalis

Department of Political Science, Trinity College Dublin

Overview

- Quantitative research
- Sample vs population
- Research questions
- Research hypotheses
- Designing quantitative study

What Is Quantitative Research?

- Quantitative research uses **data**.
- Data consists of a collection of observed cases.
- Observed cases were selected from a (much) larger pool of cases.
- The observed cases are called the **observations**.
- The pool of cases is called the **population**.
- The collection of observations is called the **sample**.

What Are We Observing?

- One or more properties the cases might have.
- We call these properties the **variables**.
- We care about them because they *vary*.
- The properties vary along a certain dimension.
- In other words, the variables take on different values.

Questions about your household

See www.census.ie

START HERE

H1 When was your house, flat or apartment first built?
Mark — the year in which first built even if the building was subsequently converted, extended or renovated

- 1 ☐ Before 1919
- 2 ☐ 1919 – 1945 inclusive
- 3 ☐ 1946 – 1960 inclusive
- 4 ☐ 1961 – 1970 inclusive
- 5 ☐ 1971 – 1980 inclusive
- 6 ☐ 1981 – 1990 inclusive
- 7 ☐ 1991 – 2000 inclusive
- 8 ☐ 2001 – 2010 inclusive
- 9 ☐ 2011 – 2015 inclusive
- 10 ☐ 2016 or later

H2 Does your household own or rent your accommodation?
Mark — one box only

- 1 ☐ Own with mortgage or loan
- 2 ☐ Own outright
- 3 ☐ Rent
- 4 ☐ Live here rent free

If renting, who is your landlord?
 1 ☐ Private landlord
 2 ☐ Local Authority
 3 ☐ Voluntary/Co-operative housing body

H3 If your accommodation is rented, how much rent does your household pay?
Enter number of the nearest Euro

€ . 0 0

Mark — one box only
 1 ☐ Per week
 2 ☐ Per month
 3 ☐ Per year

H4 How many working smoke alarms are in your accommodation?
Write in number of smoke alarms

H5 How many rooms do you have for use only by your household?

- DO NOT count bathrooms, toilets, kitchenettes, utility rooms, consulting rooms, offices, shops, halls, landings or rooms that can only be used for storage such as cupboards
- Do count all other rooms such as kitchens, living rooms, bedrooms, studies and conservatories you can sit in
- If two rooms have been converted into one, count them as one room

Number of rooms
Of which bedrooms

H6 What is the main type of fuel used by the central heating in your accommodation?
Mark — one box only

- 1 ☐ No central heating
- 2 ☐ Oil
- 3 ☐ Natural gas
- 4 ☐ Electricity
- 5 ☐ Coal (including anthracite)
- 6 ☐ Peat (including turf)
- 7 ☐ Liquid Petroleum Gas (LPG)
- 8 ☐ Wood (including wood pellets)
- 9 ☐ Other

H7 Does your accommodation use any of the following renewable energy sources?
Mark — the boxes that apply

- 1 ☐ No
- 2 ☐ Solar panels for water heating
- 3 ☐ Solar panels for electricity
- 4 ☐ Wind turbine
- 5 ☐ Air source heat pump
- 6 ☐ Ground source heat pump
- 7 ☐ Wood
- 8 ☐ Other

H8 What type of piped water supply does your accommodation have?
Mark — one box only

- 1 ☐ Public supply
- 2 ☐ Public Group Scheme
- 3 ☐ Private Group Scheme
- 4 ☐ Private source (eg well, lake, rainwater tank, etc)
- 5 ☐ No piped water supply

H9 What type of sewerage facility does your accommodation have?
Mark — one box only

- 1 ☐ Public sewer
- 2 ☐ Individual septic tank
- 3 ☐ Individual treatment system other than a septic tank
- 4 ☐ Other sewerage facility
- 5 ☐ No sewerage facility

H10 How many cars or vans are owned or are available for use by one or more members of your household?
Include any company car or van if available for private use
Mark — one box only

- 1 ☐ One
- 2 ☐ Two
- 3 ☐ Three
- 4 ☐ Four or more
- 5 ☐ None

H11 What type of internet connection does your household have?
Mark — the boxes that apply

- 1 ☐ Fixed broadband (eg phone/TV cable, internet cable, etc)
- 2 ☐ Mobile broadband (eg 3G, 4G, dongle, etc)
- 3 ☐ Satellite
- 4 ☐ Other connection
- 5 ☐ No internet connection

If connected, which devices access the internet in your household?

Mark — the boxes that apply

- 1 ☐ Desktop PC
- 2 ☐ Laptop (including notebook, netbook, etc)
- 3 ☐ Tablet
- 4 ☐ Mobile phone
- 5 ☐ Smart TV
- 6 ☐ Video game console
- 7 ☐ Smart domestic appliance
- 8 ☐ Other, write in description

----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

H12 → Go to next page

Page 2

HOUSEHOLD FORM F1102

Household Form

Person 1

5. Please provide information for each person living here. If there is someone living here who pays the rent or owns this residence, start by listing him or her as Person 1. If the owner or the person who pays the rent does not live here, start by listing any adult living here as Person 1.

What is Person 1's name? Print name below.

First Name MI

Last Name(s)

9. What is Person 1's race?
Mark *X* one or more boxes **AND** print origins.

☐ White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc. *X*

☐ Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc. *X*

☐ American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec; Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc. *X*

☐ Chinese ☐ Vietnamese ☐ Native Hawaiian

☐ Filipino ☐ Korean ☐ Samoan

☐ Asian Indian ☐ Japanese ☐ Chamorro

☐ Other Asian – Print, for example, Pakistani, Cambodian, Hmong, etc. *X* ☐ Other Pacific Islander – Print, for example, Tongan, Fijian, Marshallese, etc. *X*

☐ Some other race – Print race or origin. *X*

6. What is Person 1's sex? Mark *X* ONE box.

☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth? For babies less than 1 year old, do not write the age in months. Write 0 as the age.

Age on April 1, 2020 Print numbers in boxes.

years

→ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

☐ No, not of Hispanic, Latino, or Spanish origin

☐ Yes, Mexican, Mexican Am., Chicano

☐ Yes, Puerto Rican

☐ Yes, Cuban

☐ Yes, other Hispanic, Latino, or Spanish origin – Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc. *X*

→ If more people were counted in Question 1 on the front page, continue with Person 2 on the next page.

2

11100021

Example: Longitudinal Election Studies



BO2: Which party did you vote for in the general election? (Modes: CAPI/Online/Paper. Countries: England/Scotland/Wales.)

Value	Label
-999	Not stated
-2	Prefer not to say/Refuse
-1	Don't know
1	Labour Party
2	Conservative Party
3	Liberal Democrats
4	Scottish National Party
5	Plaid Cymru
6	Green Party
7	United Kingdom Independence Party (UKIP)
8	Brexit Party
9	Other
10	An independent candidate
11	Specified name- no party mentioned
12	Spoilt ballot paper
13	None

German Longitudinal Election Study10

Sequence	Variable Name	Question Number	Question Title
1,000	d1	Int01	Gender
2,000	d2a-c	D01.1	Date of Birth: Year, Month, Day
3,000	q1	Vw087	Political Interest
4,000	q2	Vw088	Election Campaign: Interest
5,000	q3	Vw089	Election Campaign: Outcome, Importance
6,000	q4	Vw155	Democracy: Satisfaction
7,000	q5	Vw112	Political Knowledge: First/Second Vote
8,000	q6	Vw001	Vote Participation: Vote Intention, FE 2021
9,000	q7	Vw161	Vote Participation: hypothetical under 18, FE 2021
10,000	q8a-b	Vw002	Vote Intention: FE 2021
11,000	q9a-b	Vw162	Vote Intention: hypothetical under 18, FE 2021
12,000	q10a-b	Vw234	Voting Decision: Absentee Ballot, FE 2021
13,000	q11	Vw008	Vote Intention: Certainty, FE 2021
14,000	q12a-c	Vw624	Attitudes: Elections' Integrity
15,000	q13	Vw030	Economic Situation: current, personal
16,000	q14	Vw029	Economic Situation: retrospective, personal
17,000	q15	Vw014	Economic Situation: Responsibility, personal
18,000	q16	Vw031	Economic Situation: prospective, personal
19,000	q17b-h	Vw351	Political Parties: Scalometer
20,000	q18a-h	Vw352	Leading Politicians: Scalometer
21,000	q19	Vw353	Political Problems: most important
22,000	q20	Vw361	Political Problems: most important, Ability to Solve
23,000	q21	Vw354	Political Problems: second most important
24,000	q22	Vw362	Political Problems: second most important, Ability to Solve
25,000	q23	Vw034	Economic Situation: current, general
26,000	q24	Vw032	Economic Situation: retrospective, general
27,000	q25	Vw033	Economic Situation: Responsibility, general
28,000	q26	Vw035	Economic Situation: prospective, general
29,000	q27a-k	Vw256	Attitudes: Issue Battery
30,000	q28	Vw601	Candidate for Chancellor: Preference
31,000	q29a-b	Vw599	Candidate for Chancellor: Attribute, CDU/CSU
32,000	q30a-b	Vw600	Candidate for Chancellor: Attribute, SPD
33,000	q31a-b	Vw602	Candidate for Chancellor: Attribute, GRÜNE
34,000	q32	Vw597	Right to Vote: Recall, FE 2017

BES Cross-Section 2019

GLS Cross-Section 2021

What Do We Want to Know?

- There is some feature of the population we'd like to know, something involving one or more variables.
- Examples: average annual income, relationship between income and vote cast in last election.
- That feature is called a **parameter** of the population.

How Will We Come to Know it?

- We use the values of the variables from the observations in the sample to calculate a **statistic**.
- We *make inferences* from the statistic to the parameter, from the observed to the unobserved.
- To reiterate: we use data in our sample to construct statistics from which we infer things about the parameters of our population (terminology matters).

Is It a Sample, or Is It A Population?

- Depends on what you are doing with it.
- Are you concerned only with the set of observations itself?
- Or are you using it to make inferences about a larger population?
- If the former, it's a population. If the latter, it's a sample.

Example

- Suppose I calculated the percentage of MSc students in our course from Ireland this year.
- If all I care about is your class, then I am directly observing a parameter of this population.
- But if I use this to estimate the percentage of M.Sc. students from Ireland in a larger set of students, it's a statistic from a sample.
- And which population? I could give different answers, and my sample may be better or worse depending on them.

Should We Use Quantitative Research Methods?

- In this class, yes.
- But in general, we have to ask ourselves two questions first:
 1. What is the problem we want to solve?
 2. Are quantitative research methods the best way to solve this problem? (Depends on the research question)

What Kind of Problems Can We Solve?

- For political scientists, it's usually a problem of *explanation*.
- We want to know why something is happening?
- That means, why is some variable taking on certain values at some times and not others?
- We call this variable the **dependent** (response) variable (Y)

Research Questions

- Formulating a research question should be your first step.
- It should be in principle answerable with data (falsifiable);
- It should be of scientific importance (people care about the answer);
 - The easiest way to check is to see previous academic literature.
- It should be relatively specific (narrow).

Why Did Representative Assemblies Appear in some Places but not Others?

American Political Science Review

Vol. 104, No. 4 November 2010

doi:10.1017/S0003055410000444

When Distance Mattered: Geographic Scale and the Development of European Representative Assemblies

DAVID STASAVAGE *New York University*

*S*cholars investigating European state development have long placed a heavy emphasis on the role played by representative institutions. The presence of an active representative assembly, it is argued, allowed citizens and rulers to contract over raising revenue and accessing credit. It may also have had implications for economic growth. These arguments have in turn been used to draw broad implications about the causal effect of analogous institutions in other places and during other time periods. But if assemblies had such clear efficiency benefits, why did they not become a universal phenomenon in Europe prior to the nineteenth century? I argue that in an era of costly communications and transport, an intensive form of political representative was much easier to sustain in geographically compact polities. This simple fact had important implications for the pattern of European state formation, and it may provide one reason why small states were able to survive despite threats from much larger neighbors. I test several relevant hypotheses using an original data set that provides the first broad view of European representative institutions in the medieval and early modern eras. I combine this with a geographic information system data set of state boundaries and populations in Europe between 1250 and 1750. The results suggest a strong effect of geographic scale on the format of political representation. The broader implication of this result is to provide a reminder that if institutions help solve contracting problems, ultimately, the maintenance of institutions may itself depend on ongoing transactions costs.

Stasavage (2010)

Are Human Rights Practices Getting Better?

American Political Science Review (2019) 113, 3, 868–881

doi:[10.1017/S000305541900025X](https://doi.org/10.1017/S000305541900025X)

© American Political Science Association 2019

Letter

Yes, Human Rights Practices Are Improving Over Time

CHRISTOPHER J. FARISS *University of Michigan*

To document human rights, monitoring organizations establish a standard of accountability, or a baseline set of expectations that states ought to meet in order to be considered respectful of human rights. If the standard of accountability has meaningfully changed, then the categorized variables from human rights documents will mask real improvements. Cingranelli and Filippov question whether the standard of accountability is changing and whether data on mass killings are part of the same underlying conceptual process of repression as other abuses. These claims are used to justify alternative models, showing no improvement in human rights. However, by focusing on the coding process, the authors misunderstand that the standard of accountability is about how monitoring organizations produce documents in the first place and not how academics use published documents to create data. Simulations and latent variables that model time in a substantively meaningful way validate the conclusion that human rights are improving.

Fariss (2019)

Does Democracy Matter for Normatively Desirable Outcomes?



Annual Review of Political Science

Does Democracy Matter?

John Gerring,¹ Carl Henrik Knutsen,² and Jonas Berge²

¹Department of Government, University of Texas at Austin, Austin, Texas, USA;
email: jgerring@austin.utexas.edu

²Department of Political Science, University of Oslo, Oslo, Norway;
email: c.h.knutsen@stv.uio.no, jonasvb96@gmail.com

Gerring et al. (2022)

Does the Internet Facilitate Selective Exposure to Politically Congenial Content?



AMERICAN JOURNAL
of POLITICAL SCIENCE

(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets

Andrew M. Guess

Princeton University

Abstract: *Does the internet facilitate selective exposure to politically congenial content? To answer this question, I introduce and validate large-N behavioral data on Americans' online media consumption in both 2015 and 2016. I then construct a simple measure of media diet slant and use machine classification to identify individual articles related to news about politics. I find that most people across the political spectrum have relatively moderate media diets, about a quarter of which consist of mainstream news websites and portals. Quantifying the similarity of Democrats' and Republicans' media diets, I find nearly 65% overlap in the two groups' distributions in 2015 and roughly 50% in 2016. An exception to this picture is a small group of partisans who drive a disproportionate amount of traffic to ideologically slanted websites. If online "echo chambers" exist, they are a reality for relatively few people who may nonetheless exert disproportionate influence and visibility.*

Guess (2021)

How Do We Solve This Problem?

- We try to explain the values assumed by the dependent variable using another variable
- We call that the **independent** (explanatory) variable (X)
- We hypothesize that the value of the dependent variable depends on the value of the independent variable, which doesn't depend on anything else (at least in theory).

Hypothesised relationship I

- At its core, hypothesis takes the following form: $X \rightarrow Y$
- Or “ Y depends on X ”
- Or “ Y is associated with X ”
- Simply put, variables X and Y take on different values according to some relationship between them.

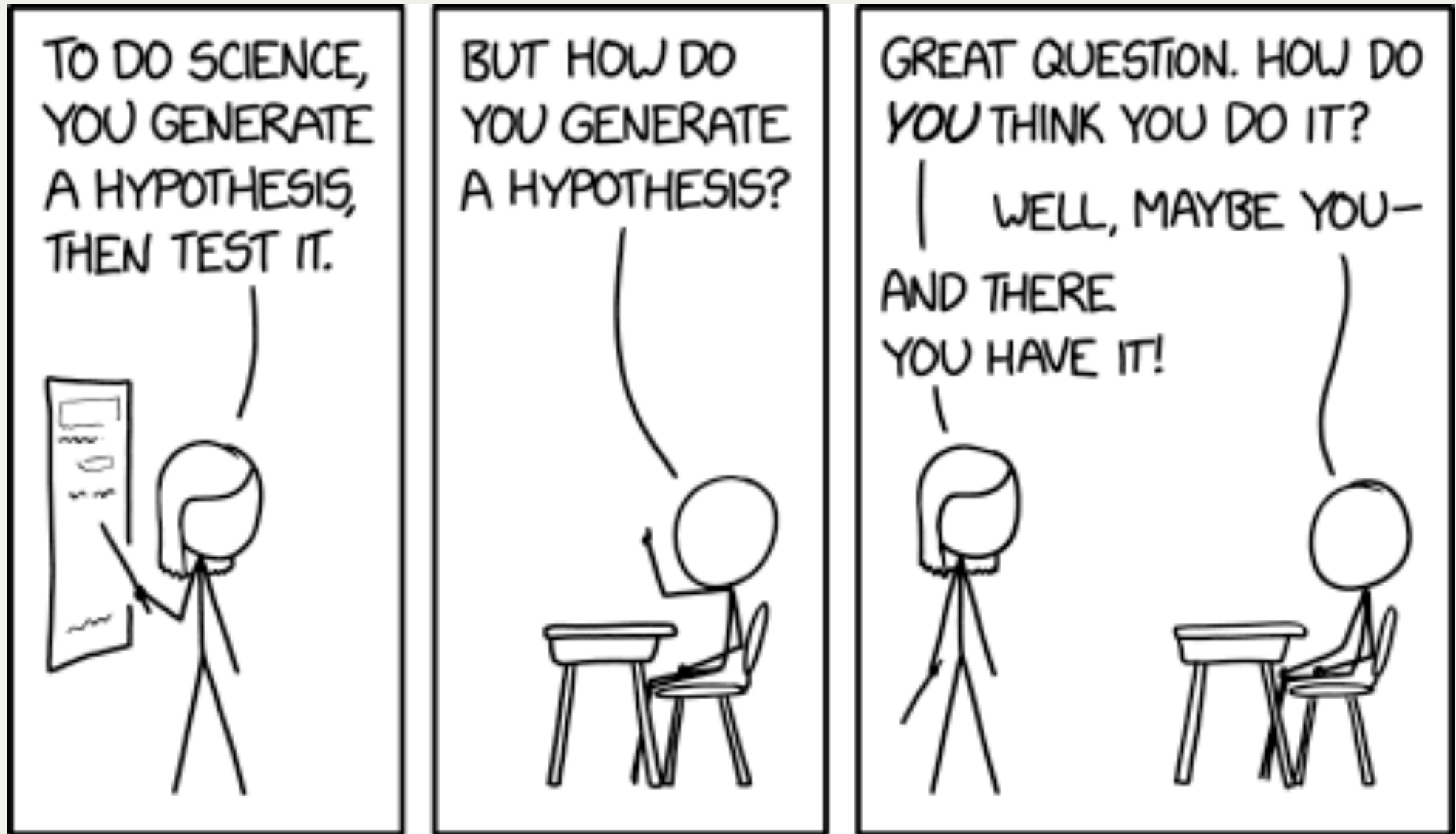
Hypothesised relationship II

- The world is not that simple and deterministic.
- Thus, a hypothesis is more likely to look like this: $X \rightarrow Y : Z$
- Or “ Y depends on X in the presence of Z ”
- Or “ Y is associated with X conditional on Z ”
- Re-writing as an equation gives us: $Y = X + Z + \epsilon$
- ϵ means that the relationship is not perfect (one-to-one)!
- There is always some error involved.

Example Hypotheses

- Democracies are less likely to go to war with each other than non-democracies.
- Democracies have higher economic output than non-democracies.
- More left-leaning and economically deprived voters are more likely to support higher taxes.
- The share of female legislators increased after the introduction of gender quotas.
- Human rights practices improve over time.

Designing Quantitative Study



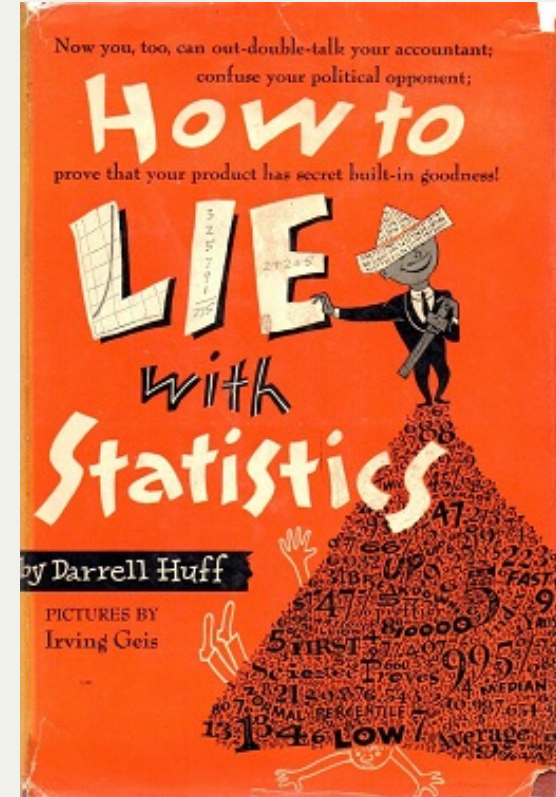
Steps in Designing Quantitative Study

- Identify your problem (formulate a research question);
- Specify your dependent variable (Y);
- Explain why it is a significant problem (i.e., why should anybody care);
- Explain how much we already know about the problem (the literature review);
- Formulate one or more hypotheses;
- Design a model to test your hypotheses - to explain why or how your dependent variable varies the way it does;
- Identify a dataset suitable to testing your model/hypotheses;
- Measure your variables;
- Perform statistical tests on your data.

Trust, But Verify



Wikipedia



Wikipedia

- Numbers can be wrong (poor data source, lots of bias)
- Be critical and scrutinize them.

Next

- Workshop:
 - Tuesday 13:00-14:00 in M20 Museum Building
 - R Overview
- Next week:
 - Descriptive Statistics