



# Survey analysis week 1

## “the inference wars”

© Peter Lugtig

# Outline for today

- Introduction
- Surveys and samples
- The first inference war:
  - 1948~1960
- The second inference war:
  - 2005~2020
- Election polling
- Class exercise



# But first: Speed-dating!

- Form 2 rows of 12 people
- Ask 1 question (1 min max)
- Get 1 in return (1 min max)
- Row facing the Screen:  
shift one place on signal
- <https://www.youtube.com/watch?v=2dAorgAB0I4>



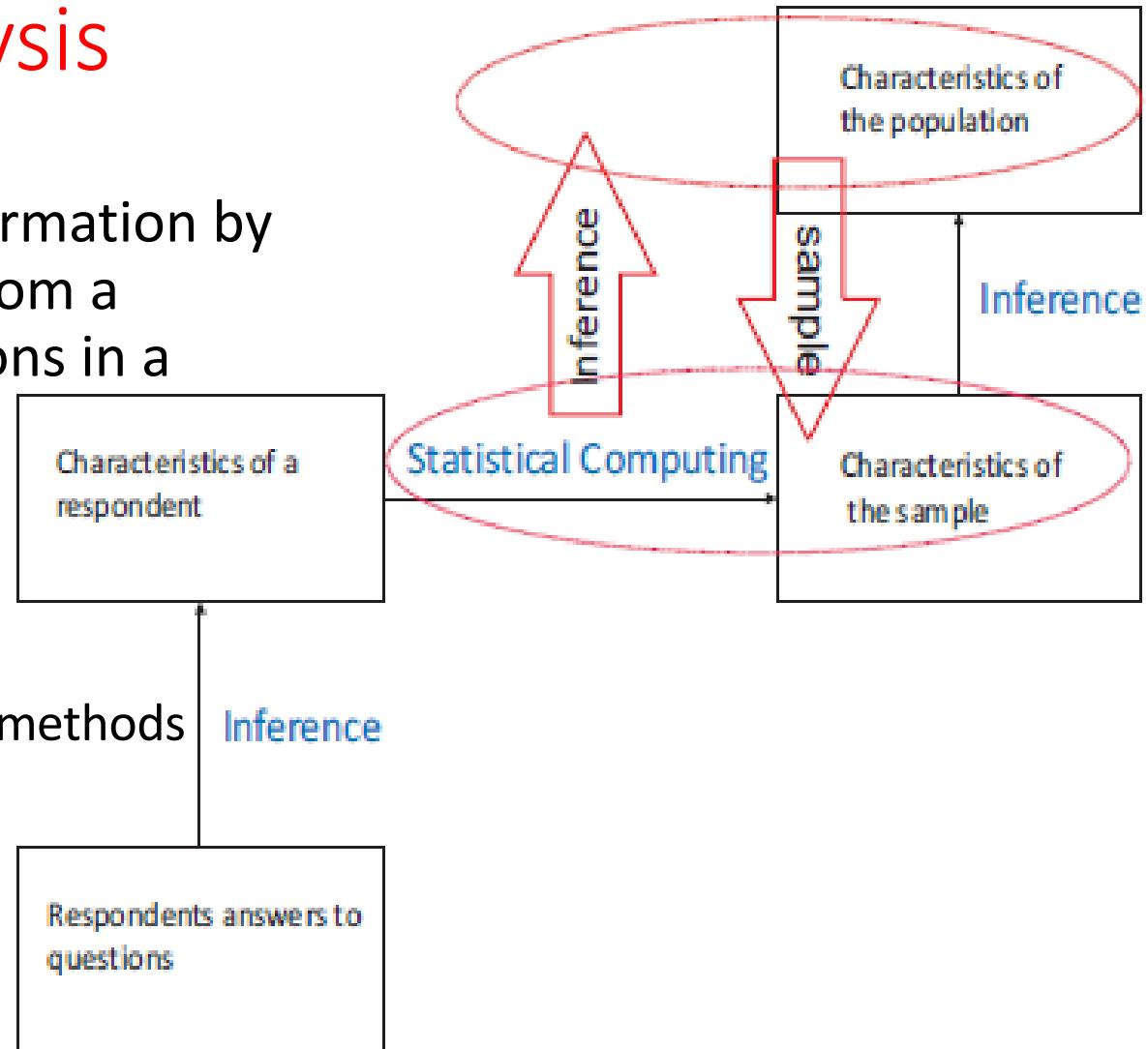
# Survey data analysis

- A method to acquire information by asking people selected from a predefined group questions in a structured manner
  - Dominant method of collecting data in the social and behavioral sciences
  - Learn about the methodology of conducting the survey



# Survey data analysis

- A method to acquire information by asking people selected from a predefined group questions in a structured manner
- Survey **data** analysis
  - Sampling
  - Inference
  - Doing this for all kinds of methods



# Why focus on Sampling?

- Basis of all inferential statistics
  - Standard error in t-test, Anova, regression.
- Sampling errors can be estimated
  - Design based inference
  - (model-based inference later in course)
- Power analysis, efficient design of studies
- Important in understanding other techniques (e.g. Bootstrapping)
- There are not so many people who know about this
  - A specific skill ‘survey statistics’

# Links to other courses

- Multivariate Statistics
- Fundamentals of Statistics
- Computation inference with R
- Survey data analysis:
  - Links/overlap
    - Apply the general linear model under correct inferential design
    - Apply fundamental knowledge about estimators
    - Apply R skills to analyse a real-life dataset and data problem
  - Specific goals: inference, data collection, missing data

# Sample

- NL: steekproef
- DE: stichprobe
- FR: Échantillon
- ES: muestra



+



# The first inference war (1936-1952)



# The 1936 Literary Digest survey

- Correct result since 1916
- Sample: 2.4 million people(!)
- Opt-in sample

Digest poll respondent

Candidate	Yes	No	Total
Roosevelt	42.9 <i>(48.5)</i>	60.9 <i>(66.1)</i>	56.6 <i>(54.8)</i>
Landon	<b>57.1</b> <i>(51.5)</i>	39.1 <i>(33.9)</i>	43.4 <i>(45.2)</i>
Total <sup>a</sup>	23.8 <i>(63.7)</i>	<b>76.2</b> <i>(36.3)</i>	764 <sup>b</sup>

Source: Roper Center for Public Opinion Research 2003b.

Note: This table indicates voting preference *at the time of the Digest poll*; that is, it accounts for those who remembered changing their minds ( $n = 26$ ); those who claimed not to have changed their minds ( $n = 433$ ); those who said that they did not remember ( $n = 13$ ); and those who gave no answer ( $n = 292$ ). The original unweighted results are in parentheses ( $n_{11} = 236$ ;  $n_{12} = 183$ ;  $n_{21} = 251$ ;  $n_{22} = 94$ ). The weighted results are italicized ( $n_{11} = 78$ ;  $n_{12} = 355$ ;  $n_{21} = 104$ ;  $n_{22} = 228$ ). (Weighted cell frequencies do not sum to total due to rounding.) The bolded numbers are known (population) values.

<sup>a</sup>Row percentages.

<sup>b</sup>Total sample size. It is made up of AIPD poll respondents who report having received a *Digest* ballot. Of those, *respondents* are individuals who claim to have returned their straw ballots ( $n = 487$ ), and *nonrespondents* are individuals who said that they did not return theirs ( $n = 246$ ) or did not remember returning it ( $n = 31$ ).

## The Literary Digest

NEW YORK

OCTOBER 21, 1936

### Topics of the day

**LANDON, 1,293,669; ROOSEVELT, 972,897**

Final Returns in The Digest's Poll of Ten Million Voters

Well, the great battle of the ballot is in the Poll of ten million voters, scattered throughout the forty-eight States of the Union. The returns are in, and the figures are in. Below we record the figures arrived at by the hour of going to press.

The returns, apparently as received

from more than one in every five voters polled in the country—they are neither weighted nor adjusted.

Never before in an experience covering

more than a quarter of a century in taking public opinion polls has there been such a variety of criticisms—praise less wary

criticisms from many others—and yet

it has been done, and done so well, and come to an excess time a Poll has been taken.

A telegram from a newspaper in California asks: "Is it true that Mr. Roosevelt purchased 100,000 straw ballots?"

A telegram received only the day before

these lines were written: "Mr. Roosevelt

National Committee purchased Ten

MILLION DOLLARS?" And all types and vari-

ties, including: "Have the Jews purchased

the election?" "Is the Pope involved?"

"Is Mr. Roosevelt in the lead?"

# The 1936 Literary Digest survey

- Sample: 2.4 million people(!)
- Opt-in sample

Candidate	<i>Digest</i> poll respondent		Total
	Yes	No	
Roosevelt	<b>42.9</b>		
Landon	<b>57.1</b>		
Total <sup>a</sup>	<b>23.8</b>		

Source: Roper Center for Public Opinion Research 2003b.

Note: This table indicates voting preference *at the time of the Digest* poll; that is, it accounts for those who remembered changing their minds ( $n = 26$ ); those who claimed not to have changed their minds ( $n = 433$ ); those who said that they did not remember ( $n = 13$ ); and those who gave no answer ( $n = 292$ ). The original unweighted results are in parentheses ( $n_{11} = 236$ ;  $n_{12} = 183$ ;  $n_{21} = 251$ ;  $n_{22} = 94$ ). The weighted results are italicized ( $n_{11} = 78$ ;  $n_{12} = 355$ ;  $n_{21} = 104$ ;  $n_{22} = 228$ ). (Weighted cell frequencies do not sum to total due to rounding.) The bolded numbers are known (population) values.

<sup>a</sup>Row percentages.

<sup>b</sup>Total sample size. It is made up of AIPO poll respondents who report having received a *Digest* ballot. Of those, *respondents* are individuals who claim to have returned their straw ballots ( $n = 487$ ), and *nonrespondents* are individuals who said that they did not return theirs ( $n = 246$ ) or did not remember returning it ( $n = 31$ ).

- The literary Digest

# 1936-1948: Quota Sampling

- Opt-in sample, but:
  - Record characteristics that are important for voting behavior.
  - We want a sample of 1000 respondents
  - Quota:
    - White, male, aged 35-54, blue collar worker      5% of population -> 50 in sample
    - Black, female, aged 18-35, in education      2% of population -> 20 in sample
    - ....
- Choice of variables for quota is important!
- Why does this go wrong in election polling? ----->



# 1948: Truman vs. Dewey

- Gallup, Roper
- Quota sample
- Prediction: +5 for Dewey
- Result: +5 for Truman
  - Problem with quotas
    - Gender, age, race
  - Late undecideds

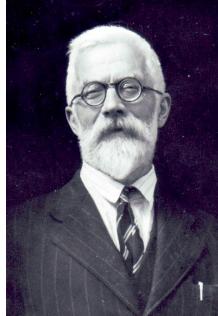


# Developments in statistics 1900-1930

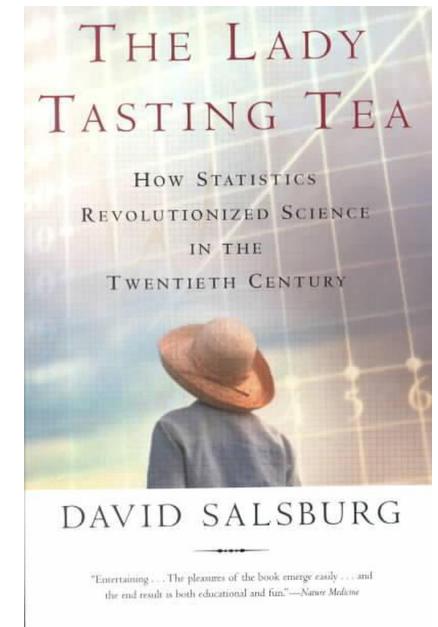
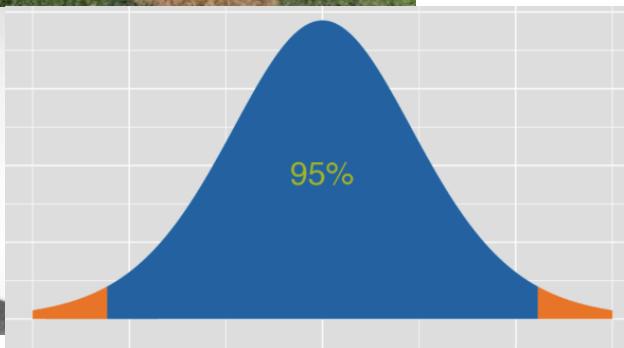
- Gosset  
(Student-T)



- Fisher



- Neyman



# Sampling and the central limit theorem

- Galton board



- Central limit theorem:
  - [https://gallery.shinyapps.io/CLT\\_mean/](https://gallery.shinyapps.io/CLT_mean/)

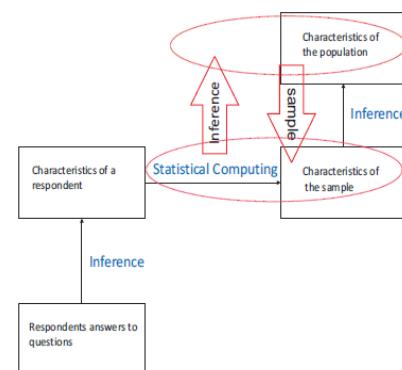
# Sampling and the central limit theorem

- Galton board
- Central limit theorem:
  - [https://gallery.shinyapps.io/CLT\\_mean/](https://gallery.shinyapps.io/CLT_mean/)
  - The distribution of means that are the result from repeatedly sampling from any population distribution will result in a normal distribution
- Law of large numbers: the confidence interval of an estimate from a sample will become smaller the larger the sample size is.

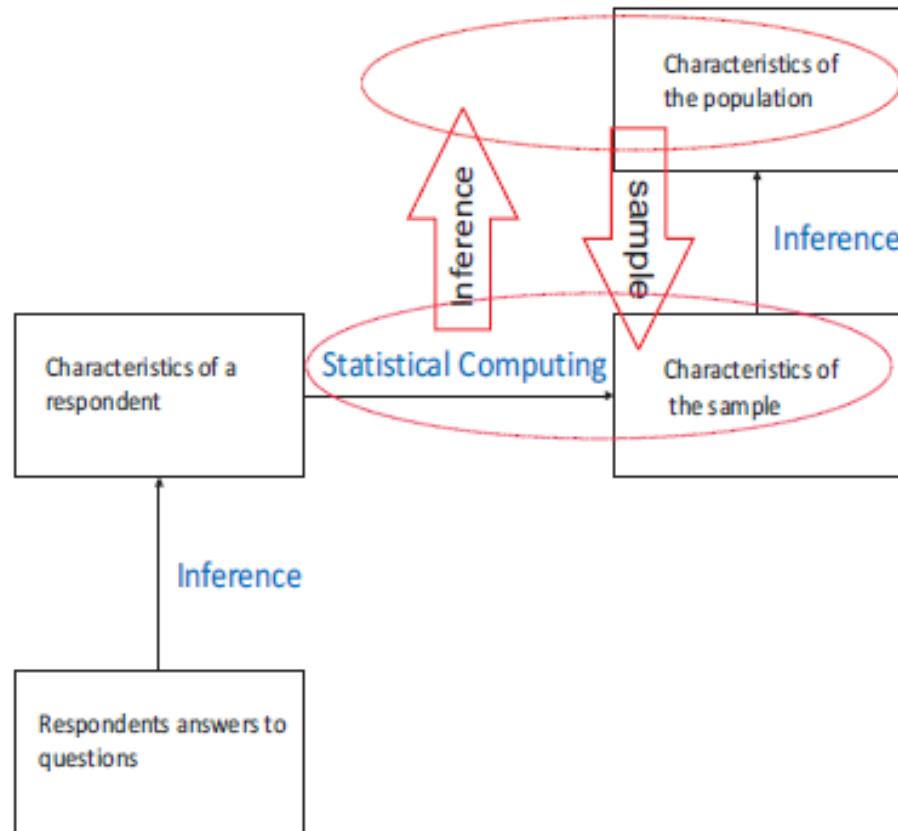
# 1952 onwards: Michigan election studies

Idea: a random sample from the population

- 1. Population
- 2. Frame for population elements
  - Addresses, phone-numbers, etc.
- 3. Sample from frame
- 4. Response from the sample



# Design based inference



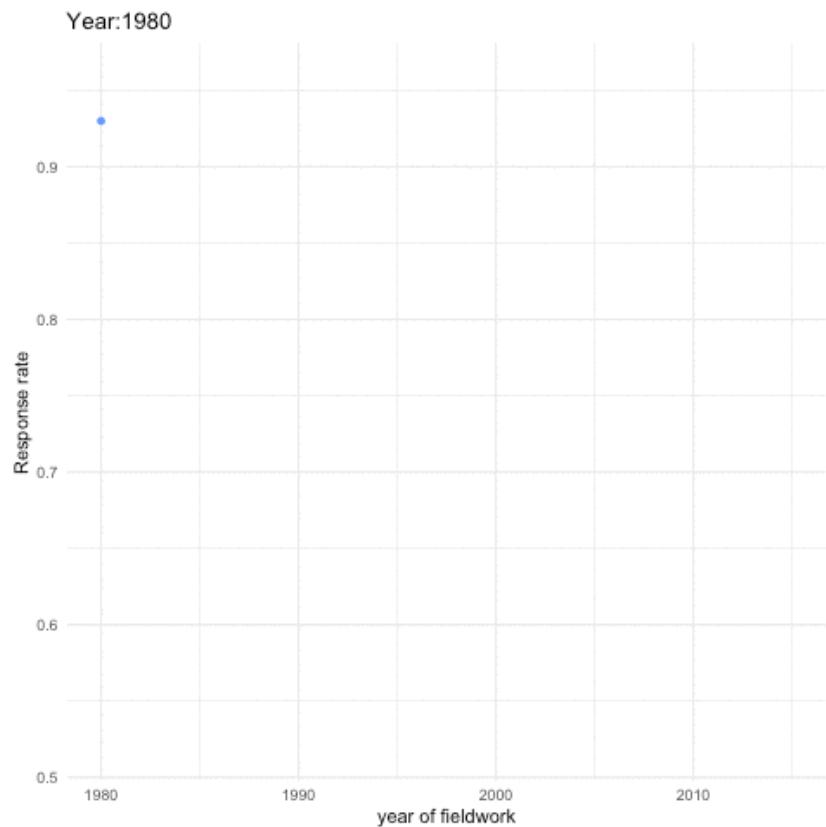
# Elements of design-based inference

- Sampling methods: clustering, stratification
  - Corrections for unequal sampling methods
  - Horvitz-Thompson estimator (1948, 1952) (week 4)
- Corrections for missing data: weighting, imputation
  - Nonresponse (week 44,46,47)
  - Errors on frame (weeks 38,48,50)
- Estimation of uncertainty (errors) (week 39-42)
- Variance estimators (week 39-42)
- 1952-now: development of a framework for inference: **design-based**

# So, what led to the 2nd inference war?

## Problems with surveys

- Lists are getting worse (telephone esp)
- Falling response rates
- Costs!



# So, what led to the 2nd inference war?

- Lists are getting worse
- Falling response rates
- Costs!

- The Internet
- Growing need for information
- Quick and cheap valued more than slow and expensive

Survey X

2 . Depth - The material presented was the right technological depth.

Excellent  
 Very Good  
 Good  
 Fair  
 Ok  
 Bad/No Comments

Submit

# What is the 2nd inference war about?

- Design-based inference:
  - We have worked 60 years to work out sampling theory and survey practice
  - Methods are **unbiased** and **consistent** (week 39)
  - Using opt-in or quota samples is unscientific

Vs.

- **Model-based inference:**
  - Response rates are too low
  - It is just not feasible anymore to do expensive surveys
    - And a lot of other data is just there!
  - You need to model nonresponse anyway
  - Lets model the whole selection process



# Election polling

- Need for fast results
- Can't be too expensive
  - Internet panels: quota samples from pre-recruited panel members
  - Telephone: random digit dialing (USA)
  - Telephone: registers (phone book)
  - IVR vs. in-person calls
- 1000s of polls per election cycle
- Polls are generally accurate (see article by Kennedy et al)
  - But recent high profile misses: US election, British EU-referendum
  - [https://utrecht-university.shinyapps.io/SDA\\_shinyelectionbias/](https://utrecht-university.shinyapps.io/SDA_shinyelectionbias/)

# Inference peace remains as well: design-based dominant



# Class exercise

- The 2016 U.S. elections
  - News stations believed Clinton would easily win
  - Trump won the election (although lost the popular vote)
  - Was there a polling miss?
- [www.fivethirtyeight.com](http://www.fivethirtyeight.com) has a database with about 1,600 polls conducted before the election
  - Sample size, company, date conducted, state conducted, reputation of pollster, whole population vs. likely voters, raw and adjusted % for Trump and Clinton



# Class exercise

- Go to [https://utrecht-university.shinyapps.io/SDA\\_shinyselectionbias/](https://utrecht-university.shinyapps.io/SDA_shinyselectionbias/)

Five groups (5 people each- 20 minutes):

1. Was Trump underestimated? (or was he particularly in the swing states)?
2. Did the quality of the pollster matter? (what is quality?)
3. Was there a difference between sampling likely voters and registered voters?
4. Are larger polls better?
5. Is there a difference between raw and adjusted (modeled) poll estimates?

## Class exercise (1)

Was Trump underestimated? in the swing states? (Michigan, Wisconsin, Pennsylvania)

## Class exercise (2)

Did the quality of the pollster matter?

## Class exercise (3)

Was there a difference between sampling likely voters and registered voters?

## Class exercise (4)

Are larger polls better?

## Class exercise (5)

Is there a difference between raw and modeled polls?

# Next week: Total survey error

- We move into design-based surveys
- Read articles
- Complete take-home exercise and bring to class
  - THE 1- Your adopted survey.