This useful time: distributions
next time: catalog
& motivate
# eligible voters = 120M

\[ \theta = \text{population proportion favoring Clinton} \]

2 weeks before election

If can achieve random or like-at-random sampling from

\[ p \]

\[ \theta = \frac{\bar{y}}{s} \]

\[ T = \frac{15 + \bar{x}}{s} \]

like at random mean \( \theta \)

mean \( \theta \) \( T \)

IID
\((S_1, \theta) \sim \text{Binomial}(n, \theta)\)

\[E(\theta) = E\left(\frac{S}{n}\right) = \frac{1}{n} E(S) = \frac{n \theta}{n} = \theta\]

So \(\hat{\theta} = \frac{S}{n}\) is an unbiased estimate of \(\theta\).

\[V(\theta) = V\left(\frac{S}{n}\right) = \frac{V(S)}{n^2} = \frac{n \theta (1-\theta)}{n^2} = \frac{\theta (1-\theta)}{n}\]

\[\text{IID} \downarrow \hat{\theta} \downarrow \text{uncertainty} \downarrow \text{about} \ \theta \text{ based on} \ S\]

\[\sup_{\theta} \sqrt{\frac{\theta (1-\theta)}{n}} \Rightarrow o\left(\sqrt{\frac{1}{n}}\right)\]
to cut $SD(\theta)$ in half, need to quadruple the sample size

Q: How big does $n$ have to be to get $SD(\theta) = 0.01$?

A: $\sqrt{\sum_{iid} (\theta) - \frac{\theta(1-\theta)}{n}} = 0.01$

Given 2 candidates & polarization of electorate, $\theta = 0.5$

$f(x) = \frac{0.5}{(1-x^2)^{1/2}}$
\[ \frac{0.5}{\sqrt{\bar{X}}} = 0.01 \rightarrow n = \left( \frac{0.5}{0.01} \right)^2 = 2500 \]

\[ T = 120 \text{ m} \]

\[ n \text{ want: } n \text{ want to use IID case} \]

\[ n \text{ want: } n \text{ want to use IID case} \]

big issue,
however: really hard to achieve like at random

\[
\begin{pmatrix} T \\ c \end{pmatrix} \quad (T, c, \text{ no response})
\]

\[
\text{if } \hat{\theta} \text{ is biased}
\]

\[ \text{RMSE}(\hat{\theta}) = \sqrt{V(\hat{\theta}) + \text{bias}(\hat{\theta})^2} \]

\[ \text{RMSE}(\hat{\theta}) = \sqrt{\text{bias}(\hat{\theta})^2} \quad \text{if no bias} \]

literary digest (4)
1936 < (D) FOR
(R) Alf Landon

In summer of 1936, mailed out 10M letters, only 2.27% responded.

\hat{\theta} - 60% Landon; truth: 59% R

Error: 19 percentage points (!)

How is this possible with n = 2.27M?

\text{MSE} (\hat{\theta}) = \sqrt{\text{Var}(\hat{\theta}) + \text{Bias}^2 (\hat{\theta})}

\text{MSE} (\hat{\theta}) = \sqrt{\frac{(0.6)(0.4)}{2.27M}} = 0.00325

\text{MSE} (\hat{\theta}) = \sqrt{0.00325} = 0.018
George Gallup: correctly estimate LD result to within 1 percentage point before election, correctly estimated FDR's share was about 59%

1. Non-response bias: people who chose to mail postcard back were systematically different in voting behavior than people who chose not to

2.
A sample is representative of the population from which it is taken.

Like TLD or at least

like - TLD

life at - a - place.