Gov 51: Confidence Intervals

Matthew Blackwell

Harvard University

Confidence intervals

- Awesome: sample proportion is correct on average.
- Awesomer: get an range of plausible values.
- **Confidence interval**: way to construct an interval that will contain the true value in some fixed proportion of repeated samples.

CLT

$$\overline{Y} - p = \text{chance error}$$

- How can we figure out a range of plausible chance errors?
    - Find a range of plausible chance errors and add them to $\overline{Y}$
- Central limit theorem:

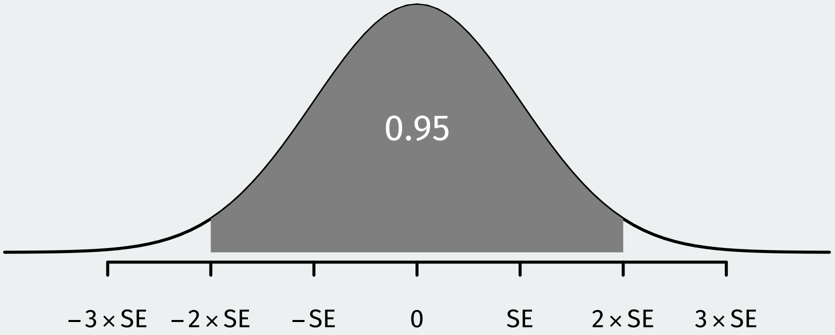
$$\overline{Y} \overset{\text{approx}}{\sim} N\left(\mathbb{E}(Y_i), \frac{\mathbb{V}(Y_i)}{n}\right)$$

- In this case:

$$\overline{Y} \overset{\text{approx}}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

- Chance error: $\overline{Y} - p$ is approximately normal with mean 0 and SE equal to $\sqrt{p(1-p)/n}$

Chance errors



- We know 95% of chance errors will be within $\approx 2 \times SE$
  - (actually it's $1.96 \times SE$)
- $\rightsquigarrow$ range of plausible chance errors is $\pm 1.96 \times SE$

Confidence interval

- First, choose a **confidence level**.
    - What percent of chance errors do you want to count as "plausible"?
    - Convention is 95%.
- $100 \times (1 - \alpha)\%$ confidence interval:

$$CI = \overline{Y} \pm z_{\alpha/2} \times SE$$

    - In polling, $\pm z_{\alpha/2} \times SE$ is called the **margin of error**
- $z_{\alpha/2}$ is the $N(0,1)$ z-score that would put $\alpha/2$ of the probability density above it.
    - $\mathbb{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = \alpha$
    - 90% CI $\rightsquigarrow \alpha = 0.1 \rightsquigarrow z_{\alpha/2} = 1.64$
    - 95% CI $\rightsquigarrow \alpha = 0.05 \rightsquigarrow z_{\alpha/2} = 1.96$
    - 99% CI $\rightsquigarrow \alpha = 0.01 \rightsquigarrow z_{\alpha/2} = 2.58$

Standard normal z-scores in R

- `qnorm(x, lower.tail = FALSE)` will find the value of z so that $\mathbb{P}(Z < z)$ is equal to `x`, where Z is $N(0, 1)$:

```
qnorm(0.05, lower.tail = FALSE)
```

```
## [1] 1.64
```

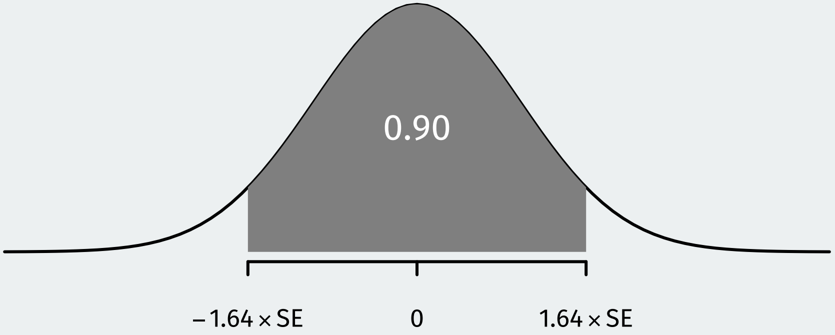
```
qnorm(0.025, lower.tail = FALSE)
```

```
## [1] 1.96
```

```
qnorm(0.005, lower.tail = FALSE)
```

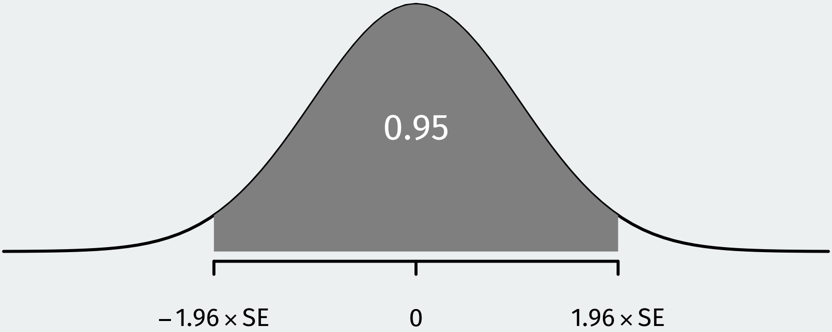
```
## [1] 2.58
```

Z-values



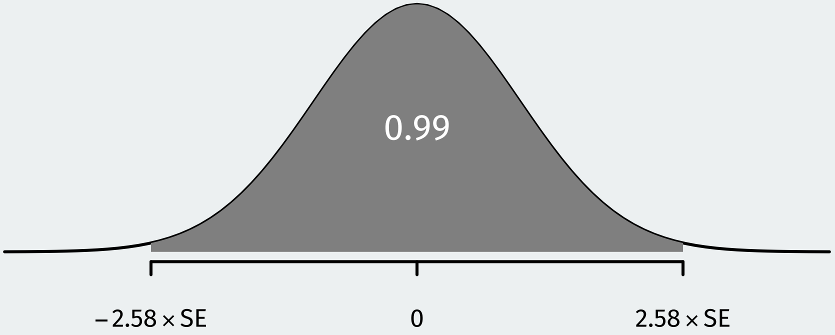
$$CI_{90} = \overline{Y} \pm 1.64 \times SE$$

Z-values



$$CI_{95} = \overline{Y} \pm 1.96 \times SE$$

Z-values



$$CI_{99} = \overline{Y} \pm 2.58 \times SE$$

CIs for the Gallup poll

- Gallup poll: $\overline{Y} = 0.37$ with an SE of 0.012.
- 90% CI:

$$[0.37 - 1.64 \times 0.012,\ 0.37 + 1.64 \times 0.012] = [0.350, 0.389]$$

- 95% CI:

$$[0.37 - 1.96 \times 0.012,\ 0.37 + 1.96 \times 0.012] = [0.346, 0.394]$$

- 99% CI:

$$[0.37 - 2.58 \times 0.012,\ 0.37 + 2.58 \times 0.012] = [0.339, 0.401]$$

- More confidence $\rightsquigarrow$ wider intervals

Interpretation and simulation

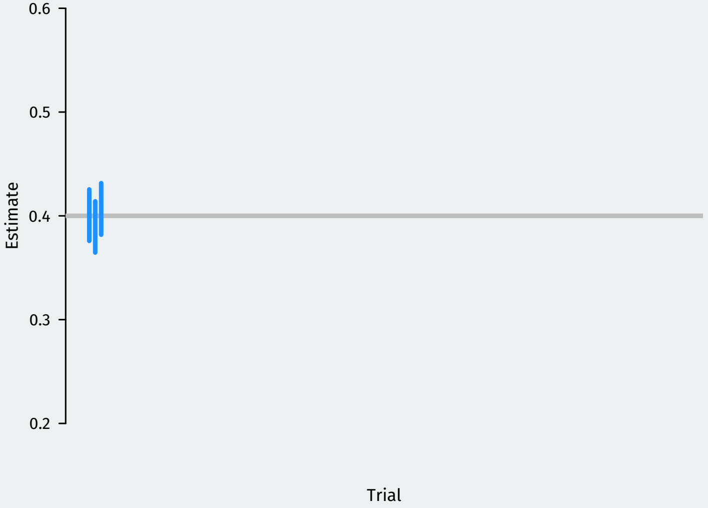
- Be careful about interpretation:
  - A 95% confidence interval will contain the true value in 95% of repeated samples.
  - For a particular calculated confidence interval, truth is either in it or not.
- A simulation can help our understanding:
  - Draw samples of size 1500 assuming population approval for Trump of $p = 0.4$.
  - Calculate 95% confidence intervals in each sample.
  - See how many overlap with the true population approval.

Plotting the CIs

Plotting the CIs

Plotting the CIs

Plotting the CIs



Plotting the CIs