Inference

Week 7

27 February, 2016

Prof. Andrew Eggers

What we're trying to understand today

Dependent variable: Nobel Prizes awarded per capita (in log scale)

	(1)	(2)	(3)
Intercept	-1.629* (0.509)	-3.166* (0.511)	-2.982* (0.527)
Chocolate consumption per capita (log scale)	2.092* (0.298)	1.026* (0.326)	0.709 (0.415)
GDP/capita (thousands of USD)		0.105* (0.024)	0.106* (0.024)
NW Europe			0.549 (0.452)
R^2	0.70	0.85	0.86
N	34	34	34

Standard errors in parentheses. * Indicates p<0.05

- What do the stars mean on regression tables? Numbers in parentheses?
- What is the "margin of error" of a poll?
- What statistical findings are reliable? Which might be just a fluke?

What we're trying to understand today

270 EFFECTIVE GOVERNMENT AND POLICY-MAKING

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34
Internal conflict risk (1990–2004)	0.346**	2.097	32
Weighted domestic conflict index (1981–2009)	−105.0*	1.611	30
Weighted domestic conflict index (1990–2009)	−119.7**	2.177	33
Deaths from domestic terrorism (1985–2010)	−2.357**	1.728	33

* Statistically significant at the 10 percent level (one-tailed test)
** Statistically significant at the 5 percent level (one-tailed test)
*** Statistically significant at the 1 percent level (one-tailed test)

Source: Based on data in Kaufmann, Kraay, and Mastruzzi 2010; PRS Group 2004; Banks, 2010; and GTD Team 2010

- What do the stars mean on regression tables? "t-values"?
- What is the "margin of error" of a poll?
- What statistical findings are reliable? Which might be just a fluke?

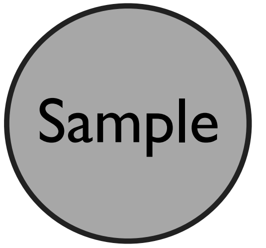
General points to remember

General points to remember

Generally, we have data from a **sample**

General points to remember

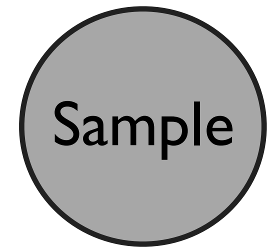
Generally, we have data from a **sample**



General points to remember

Generally, we have data from a **sample**

but we want to say something about a (larger) **population**.

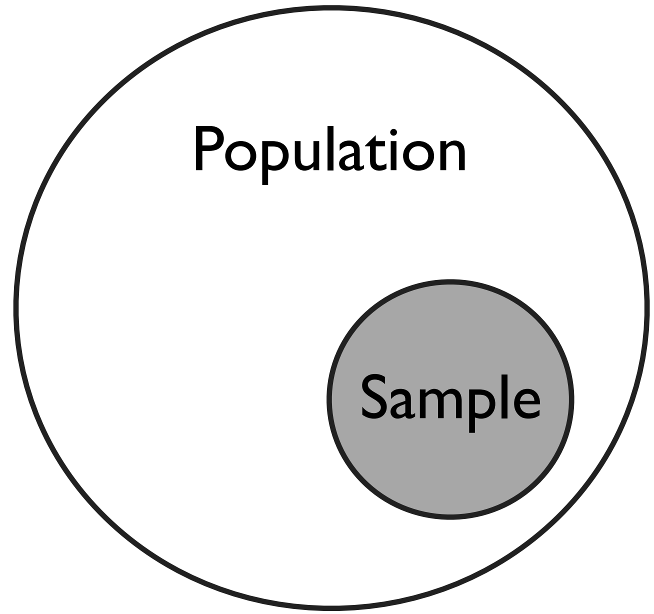


General points to remember

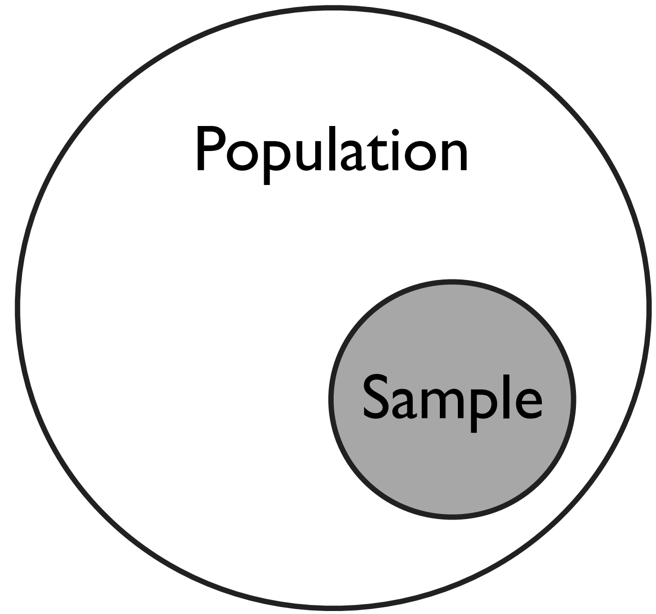
Generally, we have data from a **sample**

but we want to say something about a (larger) **population**.

This is **statistical inference**.



General points to remember



Generally, we have data from a **sample**

but we want to say something about a (larger) **population**.

This is **statistical inference**.

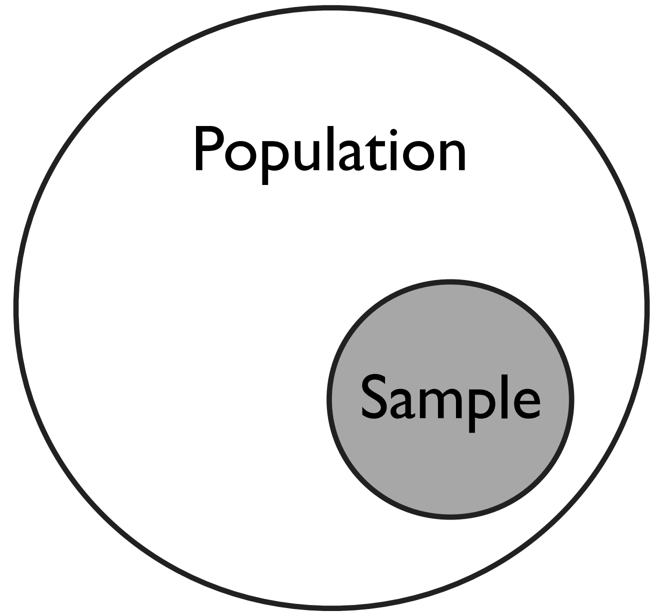
In hypothesis testing, we use data from a **sample** to assess conjectures about the **population**.

General points to remember

Generally, we have data from a **sample**

but we want to say something about a (larger) **population**.

This is **statistical inference**.



In hypothesis testing, we use data from a **sample** to assess conjectures about the **population**.

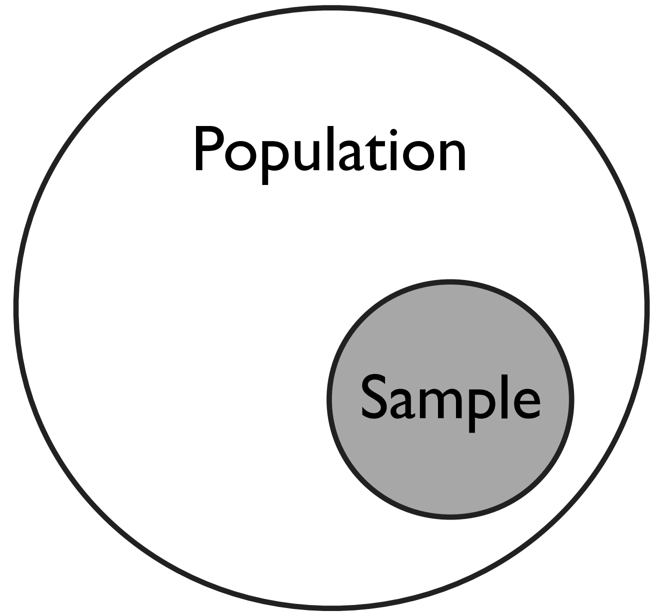
Because the sample is not the population:

General points to remember

Generally, we have data from a **sample**

but we want to say something about a (larger) **population**.

This is **statistical inference**.



In hypothesis testing, we use data from a **sample** to assess conjectures about the **population**.

Because the sample is not the population:

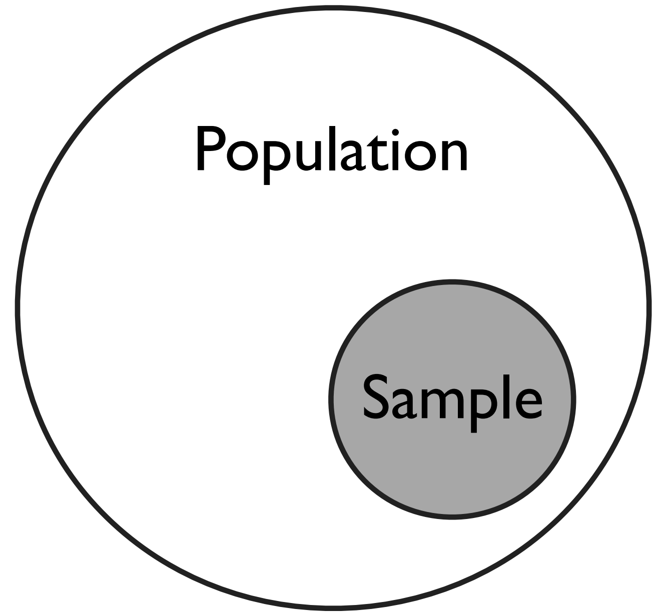
- polls have a **margin of error**

General points to remember

Generally, we have data from a **sample**

but we want to say something about a (larger) **population**.

This is **statistical inference**.



In hypothesis testing, we use data from a **sample** to assess conjectures about the **population**.

Because the sample is not the population:

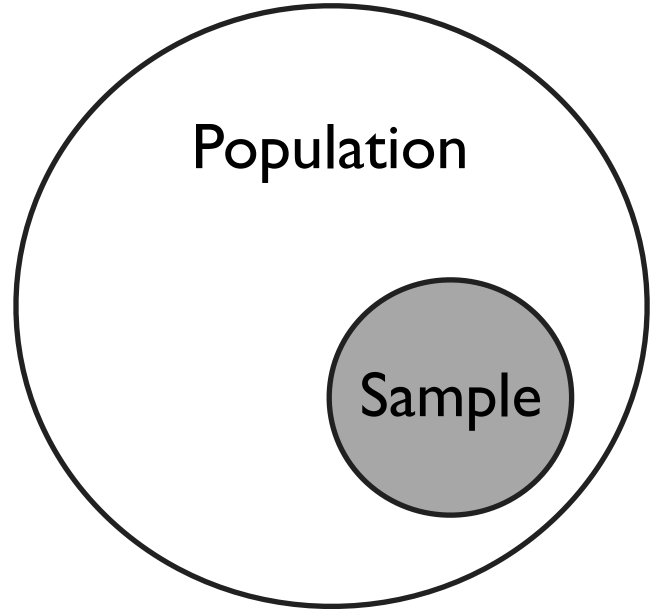
- polls have a **margin of error**
- regression coefficients have **standard errors**

General points to remember

Generally, we have data from a **sample**

but we want to say something about a (larger) **population**.

This is **statistical inference**.



In hypothesis testing, we use data from a **sample** to assess conjectures about the **population**.

Because the sample is not the population:

- polls have a **margin of error**
- regression coefficients have **standard errors**
- our conclusions in hypothesis testing are guesses, with confidence summarized by **p-values**

Understanding margin of error in a poll

Understanding margin of error in a poll

Recall from the measurement lecture:

measured value = true value + bias + random error

Understanding margin of error in a poll

Recall from the measurement lecture:

measured value = true value + bias + random error

To get rid of bias:

- in measuring concepts (week 2), we sought **valid** measures
- in selecting cases (week 4), we used **random sampling** or other approaches in which "criteria determining selection are not correlated with the outcome of interest"

Understanding margin of error in a poll

Recall from the measurement lecture:

measured value = true value + bias + random error

To get rid of bias:

- in measuring concepts (week 2), we sought **valid** measures
- in selecting cases (week 4), we used **random sampling** or other approaches in which "criteria determining selection are not correlated with the outcome of interest"

"Margin of error" tries to summarize the **magnitude of random error due to sampling**.

Thought experiment

Thought experiment

measured value = true value + bias + random error

Thought experiment

measured value = true value + bias + random error

Imagine that on June 22, 2016, you take a random sample of people who will vote in the EU referendum and ask whether they support "Leave".

Will the level of support in your poll be close to the true average support?

Thought experiment

measured value = true value + bias + random error

Imagine that on June 22, 2016, you take a random sample of people who will vote in the EU referendum and ask whether they support "Leave".

Will the level of support in your poll be close to the true average support?

If truly a random sample, there is no **bias**: you should expect to get the true value on average.

What would the magnitude of the **random error** depend on?

Thought experiment

measured value = true value + bias + random error

Imagine that on June 22, 2016, you take a random sample of people who will vote in the EU referendum and ask whether they support "Leave".

Will the level of support in your poll be close to the true average support?

If truly a random sample, there is no **bias**: you should expect to get the true value on average.

What would the magnitude of the **random error** depend on?

- size of sample (1,006 GB adults vs. 10,000,000)

Thought experiment

measured value = true value + bias + random error

Imagine that on June 22, 2016, you take a random sample of people who will vote in the EU referendum and ask whether they support "Leave".

Will the level of support in your poll be close to the true average support?

If truly a random sample, there is no **bias**: you should expect to get the true value on average.

What would the magnitude of the **random error** depend on?

- size of sample (1,006 GB adults vs. 10,000,000)
- true level of support (what if 100% supported remaining in EU?)

Simulating the thought experiment in R

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 0 0 0

> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 1 1 0

> sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
   [1] 0 0 0 0 0 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 1
  [39] 0 0 0 1 1 1 1 0 0 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1
  [77] 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1
 [115] 1 0 1 1 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 0 1 0 0 1 0
 [153] 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 1 1 0 1
 [191] 1 0 0 1 1 1 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 1 0 1 1
```

Simulating the thought experiment in R

We know that 52% of all voters supported **Leave. We want to know** how much the result of a poll might deviate from the true level of support.

Let's find out using R!

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 0 0 0
```

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 1 1 0
```

```
> sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
   [1] 0 0 0 0 0 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 1
  [39] 0 0 0 1 1 1 1 0 0 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1
  [77] 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1
 [115] 1 0 1 1 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 0 1 0 0 1 0
 [153] 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 1 1 0 1
 [191] 1 0 0 1 1 1 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 1 0 1 1
```

Simulating the thought experiment in R

We know that 52% of all voters supported **Leave. We want to know** how much the result of a poll might deviate from the true level of support.

Let's find out using R!

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 0 0 0
```

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 1 1 0
```

```
> sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
   [1] 0 0 0 0 0 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 1
  [39] 0 0 0 1 1 1 1 0 0 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1
  [77] 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1
 [115] 1 0 1 1 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 0 1 0 0 1 0
 [153] 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 1 1 0 1
 [191] 1 0 0 1 1 1 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 1 0 1 1
```

Simulating the thought experiment in R

We know that 52% of all voters supported **Leave. We want to know** how much the result of a poll might deviate from the true level of support.

Let's find out using R!

Using R, I can randomly draw 10 ones and zeros, where the probability of drawing a one is 0.52:

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 0 0 0
```

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 1 1 0
```

```
> sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
   [1] 0 0 0 0 0 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 1
  [39] 0 0 0 1 1 1 1 0 0 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1
  [77] 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1
 [115] 1 0 1 1 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 0 1 0 0 1 0
 [153] 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 1 1 0 1
 [191] 1 0 0 1 1 1 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 1 0 1 1
```

Simulating the thought experiment in R

We know that 52% of all voters supported **Leave. We want to know** how much the result of a poll might deviate from the true level of support.

Let's find out using R!

Using R, I can randomly draw 10 ones and zeros, where the probability of drawing a one is 0.52:

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 0 0 0
```

I can do it again:

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 1 1 0
```

```
> sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
   [1] 0 0 0 0 0 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 1
  [39] 0 0 0 1 1 1 1 0 0 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1
  [77] 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1
 [115] 1 0 1 1 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 0 1 0 0 1 0
 [153] 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 1 1 0 1
 [191] 1 0 0 1 1 1 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 1 0 1 1
```

Simulating the thought experiment in R

We know that 52% of all voters supported **Leave.** **We want to know** how much the result of a poll might deviate from the true level of support.

Let's find out using R!

Using R, I can randomly draw 10 ones and zeros, where the probability of drawing a one is 0.52:

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 0 0 0
```

I can do it again:

```
> sample(x = c(0,1), size = 10, replace = T, prob = c(.48, .52))
 [1] 1 1 0 0 1 1 0 1 1 0
```

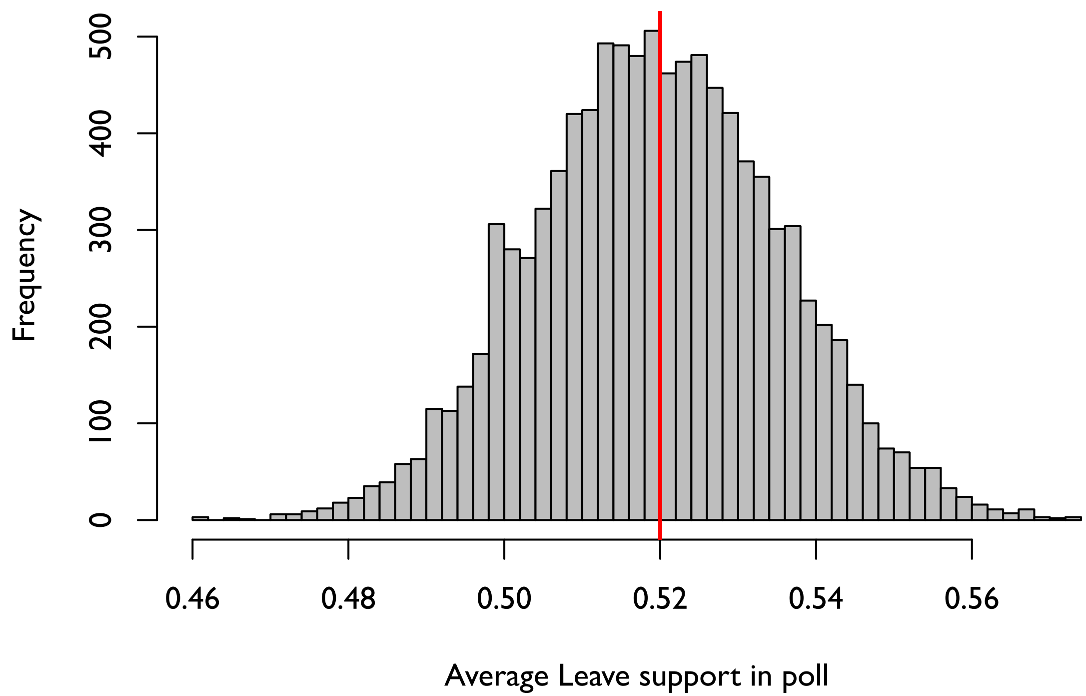
I can increase the number of "respondents" to 1,006:

```
> sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
   [1] 0 0 0 0 0 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 1
  [39] 0 0 0 1 1 1 1 0 0 0 1 0 0 1 0 1 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1
  [77] 1 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 1 0 0 0 0 0 1 1 0 0 1
 [115] 1 0 1 1 0 1 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 1 1 0 1 0 1 0 1 0 0 1 0
 [153] 0 1 1 0 0 1 1 1 0 0 1 0 1 0 1 1 1 1 0 0 0 1 0 1 0 1 1 1 1 1 1 0 1
 [191] 1 0 0 1 1 1 1 0 1 0 1 0 0 0 1 0 0 1 0 0 0 1 1 1 0 0 0 0 0 1 0 1 1
```

Simulating the thought experiment (2)

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5318091
```

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5119284
```

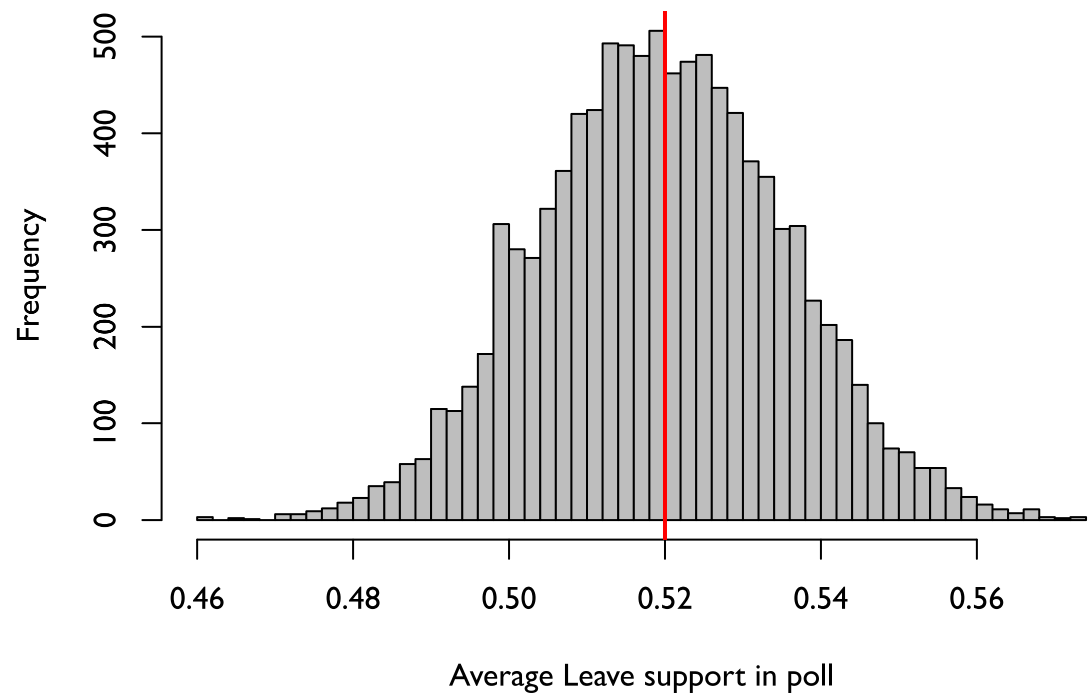


Simulating the thought experiment (2)

I can store the sample and take the mean:

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5318091
```

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5119284
```



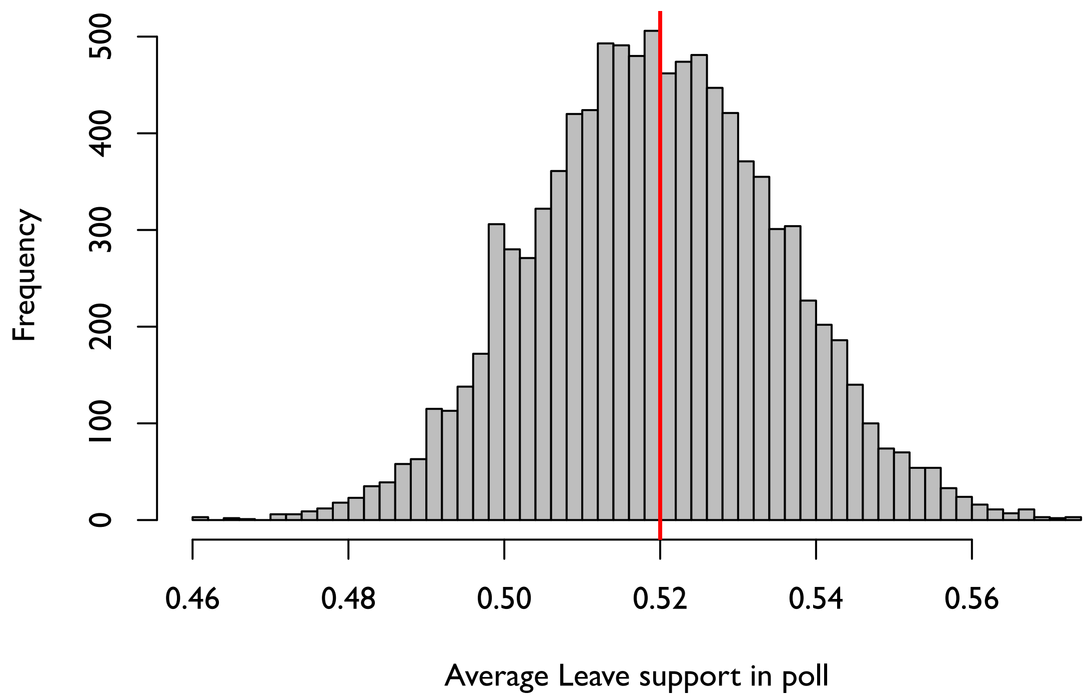
Simulating the thought experiment (2)

I can store the sample and take the mean:

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5318091
```

I can do it again:

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5119284
```



Simulating the thought experiment (2)

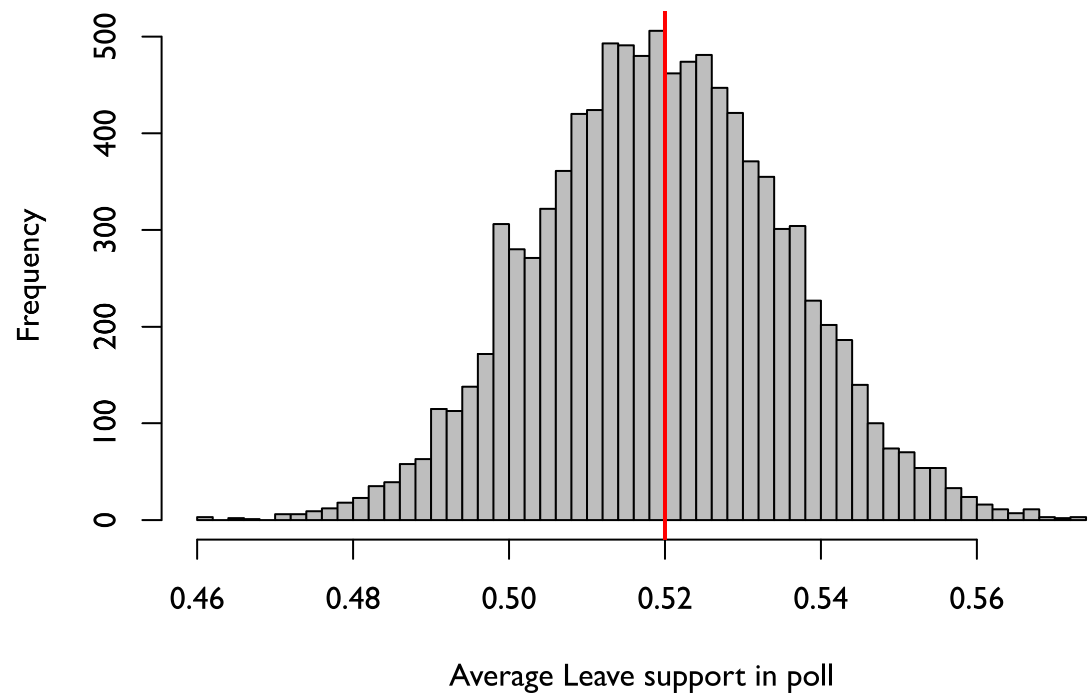
I can store the sample and take the mean:

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5318091
```

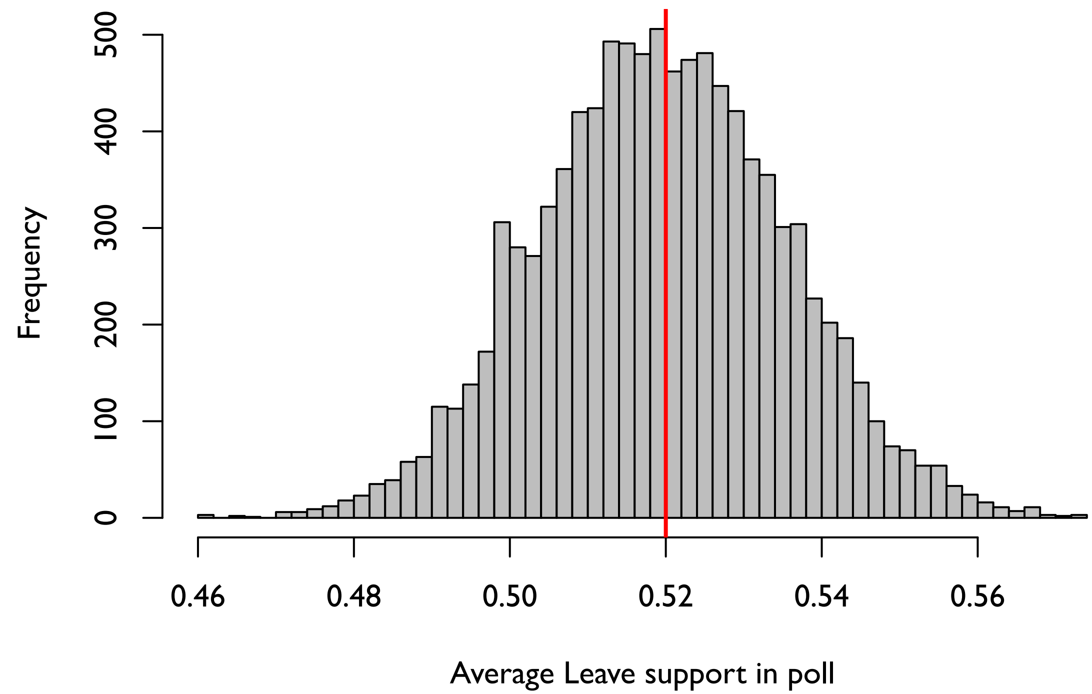
I can do it again:

```
> samp = sample(x = c(0,1), size = 1006, replace = T, prob = c(.48, .52))
> mean(samp)
[1] 0.5119284
```

I can do it 10,000 times and look at the histogram of support:



Simulating the thought experiment (3)

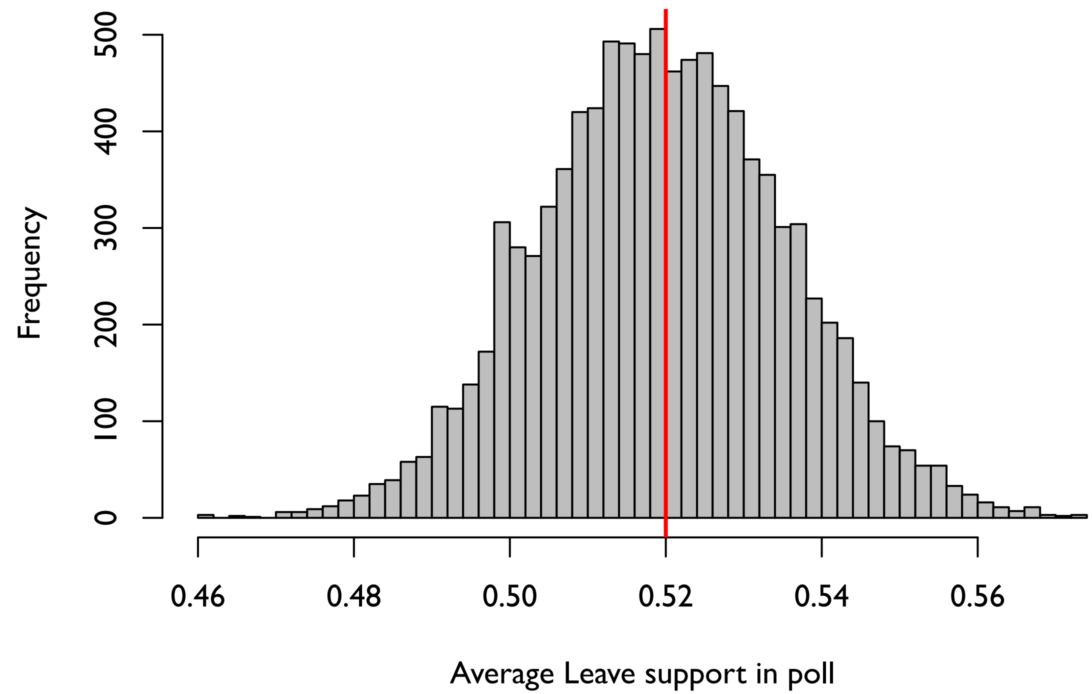



```
> sd(poll.results)
[1] 0.01572411
```

```
> quantile(poll.results, c(.025, .975))
     2.5%     97.5%
0.4890656 0.5497018
```

Simulating the thought experiment (3)

The results vary across our 10,000 "surveys" because of **sampling error.**



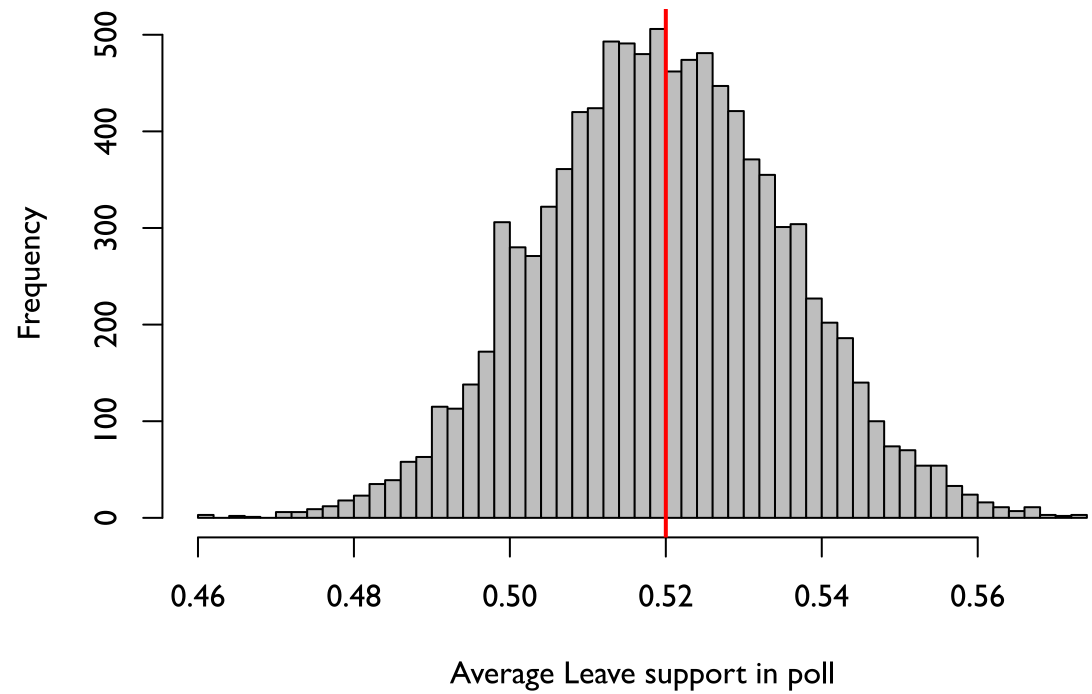

```
> sd(poll.results)
[1] 0.01572411
```

```
> quantile(poll.results, c(.025, .975))
     2.5%     97.5%
0.4890656 0.5497018
```

Simulating the thought experiment (3)

The results vary across our 10,000 "surveys" because of **sampling error.**

How much sampling error is there in our simulation?



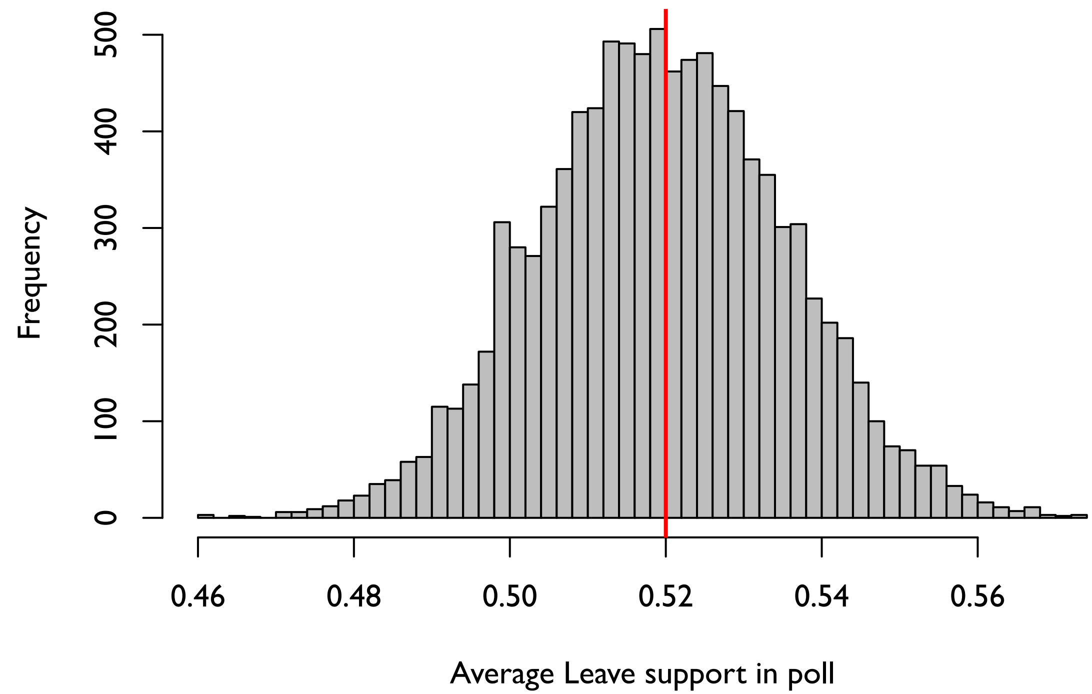

```
> sd(poll.results)
[1] 0.01572411
```

```
> quantile(poll.results, c(.025, .975))
     2.5%     97.5%
0.4890656 0.5497018
```

Simulating the thought experiment (3)

The results vary across our 10,000 "surveys" because of **sampling error.**

How much sampling error is there in our simulation?



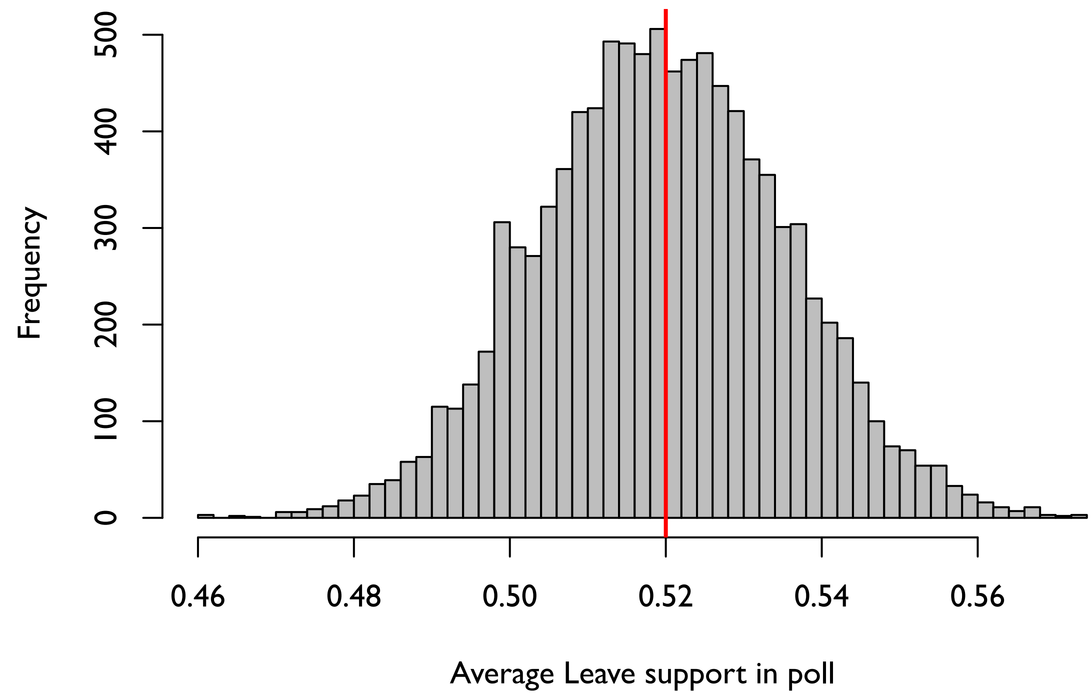
The standard deviation:

```
> sd(poll.results)
[1] 0.01572411
```

```
> quantile(poll.results, c(.025, .975))
     2.5%     97.5%
0.4890656 0.5497018
```

Simulating the thought experiment (3)

The results vary across our 10,000 "surveys" because of **sampling error.**



How much sampling error is there in our simulation?

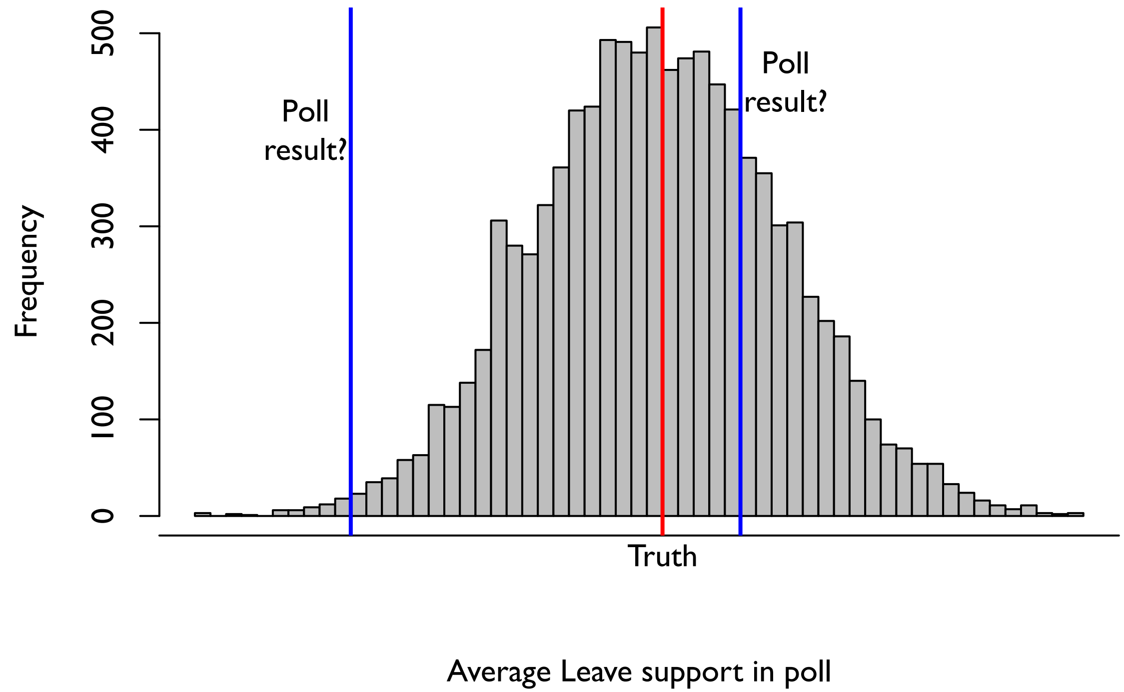
The standard deviation:

```
> sd(poll.results)
[1] 0.01572411
```

95% of the samples had a mean between 0.49 and 0.55:

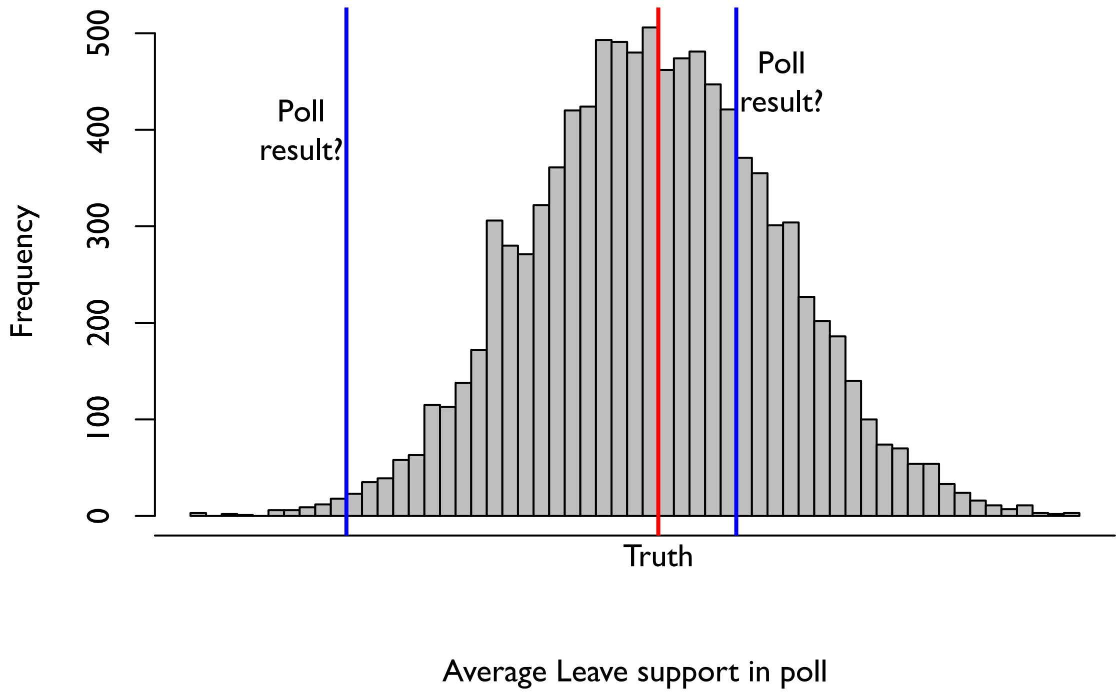
```
> quantile(poll.results, c(.025, .975))
     2.5%     97.5%
0.4890656 0.5497018
```

From thought experiment to margin of error



From thought experiment to margin of error

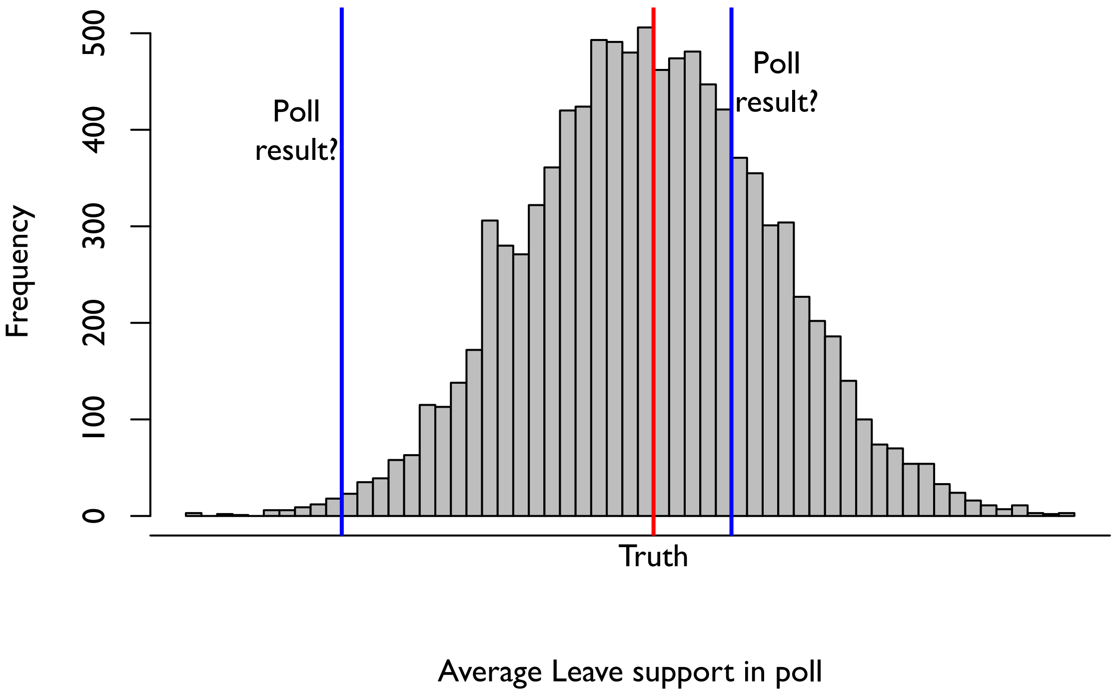
In a real survey, you don't know the answer; all you get is a **single number,** i.e. your poll result.



From thought experiment to margin of error

In a real survey, you don't know the answer; all you get is a **single number,** i.e. your poll result.

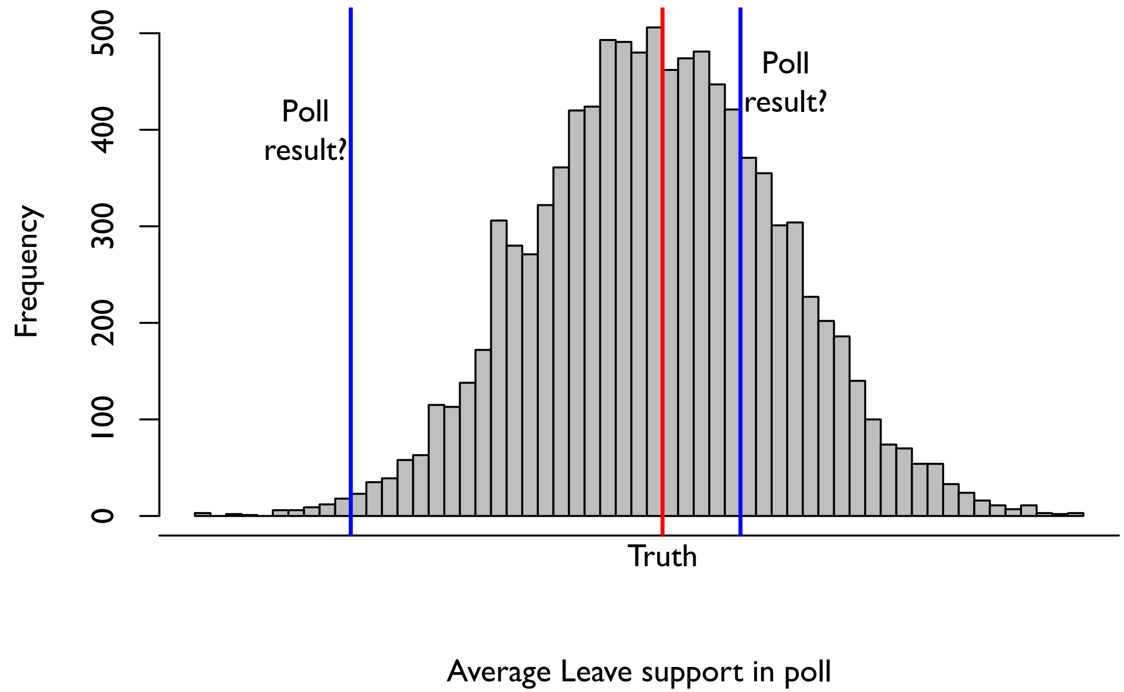
But the histogram from the thought experiment gives you a clue how close your number is to the "Truth".



From thought experiment to margin of error

In a real survey, you don't know the answer; all you get is a **single number,** i.e. your poll result.

But the histogram from the thought experiment gives you a clue how close your number is to the "Truth".

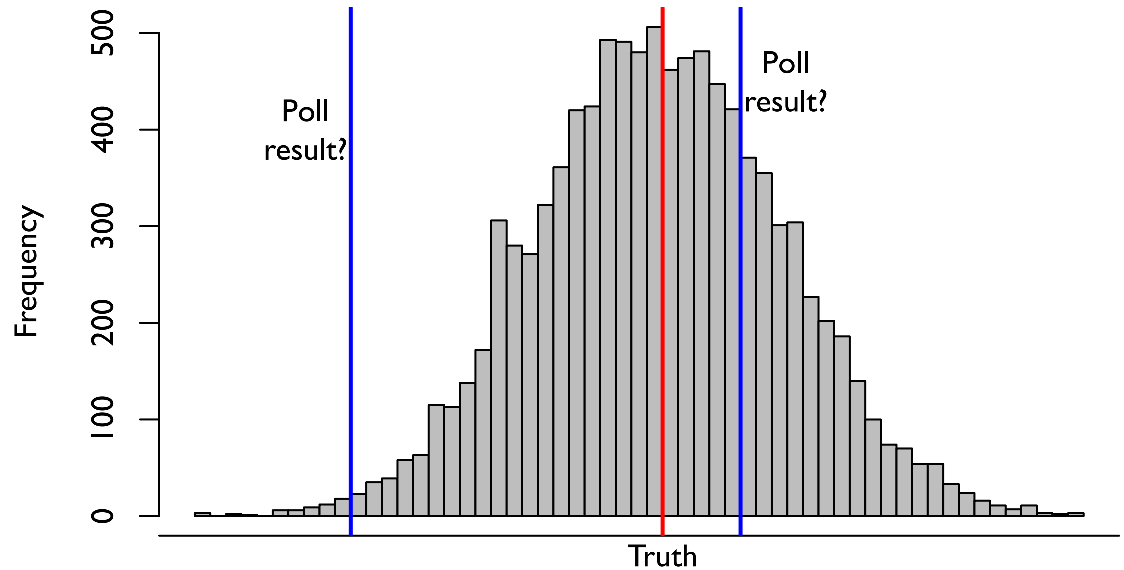


In our thought experiment (where we **know** the truth), 95% of the samples were within 0.031 of the truth.

From thought experiment to margin of error

In a real survey, you don't know the answer; all you get is a **single number,** i.e. your poll result.

But the histogram from the thought experiment gives you a clue how close your number is to the "Truth".



Average Leave support in poll

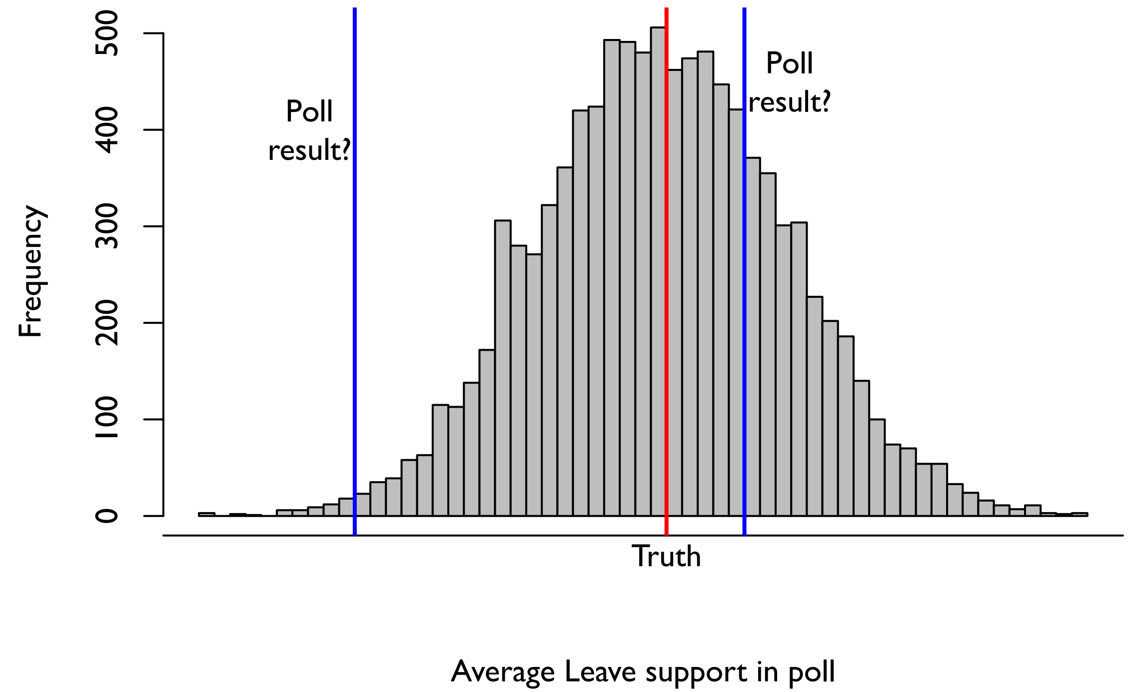
In our thought experiment (where we **know** the truth), 95% of the samples were within 0.031 of the truth.

In an actual survey (where **don't know** the truth), we have 95% confidence that our estimate is within 0.031 of the truth.

From thought experiment to margin of error

In a real survey, you don't know the answer; all you get is a **single number,** i.e. your poll result.

But the histogram from the thought experiment gives you a clue how close your number is to the "Truth".



In our thought experiment (where we **know** the truth), 95% of the samples were within 0.031 of the truth.

In an actual survey (where **don't know** the truth), we have 95% confidence that our estimate is within 0.031 of the truth.

Margin of error

So when we do a survey, we get:

So when we do a survey, we get:
An estimate for Leave support (e.g. 49%)

So when we do a survey, we get:

An estimate for Leave support (e.g. 49%)

(From the thought experiment:) An estimate of the standard deviation of poll results across samples: 0.157 (called the **standard error** of the poll)

So when we do a survey, we get:

An estimate for Leave support (e.g. 49%)

(From the thought experiment:) An estimate of the standard deviation of poll results across samples: 0.157 (called the **standard error** of the poll)

(Combining the two:) A 95% confidence interval, which we expect to include the truth in 95% of samples: e.g. 49% ± 3.1% (3.1% is the **margin of error** of the poll)

From thought experiment to margin of error (2)

So when we do a survey, we get:

An estimate for Leave support (e.g. 49%)

(From the thought experiment:) An estimate of the standard deviation of poll results across samples: 0.157 (called the **standard error** of the poll)

(Combining the two:) A 95% confidence interval, which we expect to include the truth in 95% of samples: e.g. 49% ± 3.1% (3.1% is the **margin of error** of the poll)

Another way to get the margin of error (1)

$$\sqrt{\frac{.248}{1006}} = 0.157$$

Another way to get the margin of error (1)

Another way to get the margin of error from a single sample:

$$\sqrt{\frac{.248}{1006}} = 0.157$$

Another way to get the margin of error (1)

Another way to get the margin of error from a single sample:

The **central limit theorem** says that the proportion of support in samples of size n will follow a Normal distribution centered on the truth with approximate standard deviation:

$$\sqrt{\frac{.248}{1006}} = 0.157$$

Another way to get the margin of error (1)

Another way to get the margin of error from a single sample:

The **central limit theorem** says that the proportion of support in samples of size n will follow a Normal distribution centered on the truth with approximate standard deviation:

$$\sqrt{\frac{\text{Variance of sample}}{n}}$$

$$\sqrt{\frac{.248}{1006}} = 0.157$$

Another way to get the margin of error (1)

Another way to get the margin of error from a single sample:

The **central limit theorem** says that the proportion of support in samples of size n will follow a Normal distribution centered on the truth with approximate standard deviation:

$$\sqrt{\frac{\text{Variance of sample}}{n}}$$

Our sample of e.g. 546 “Leaves” and 460 “Remains” has a variance of .248.

$$\sqrt{\frac{.248}{1006}} = 0.157$$

Another way to get the margin of error (1)

Another way to get the margin of error from a single sample:

The **central limit theorem** says that the proportion of support in samples of size n will follow a Normal distribution centered on the truth with approximate standard deviation:

$$\sqrt{\frac{\text{Variance of sample}}{n}}$$

Our sample of e.g. 546 “Leaves” and 460 “Remains” has a variance of .248.

So the estimated standard deviation (**standard error**) of our estimate is:

$$\sqrt{\frac{.248}{1006}} = 0.157$$

Another way to get the margin of error (1)

Another way to get the margin of error from a single sample:

The **central limit theorem** says that the proportion of support in samples of size n will follow a Normal distribution centered on the truth with approximate standard deviation:

$$\sqrt{\frac{\text{Variance of sample}}{n}}$$

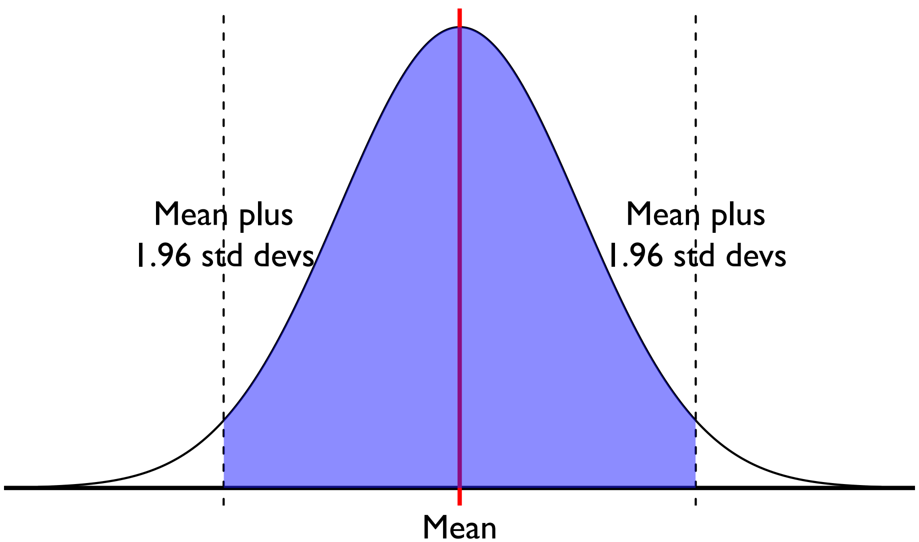
Our sample of e.g. 546 "Leaves" and 460 "Remains" has a variance of .248.

So the estimated standard deviation (**standard error**) of our estimate is:

$$\sqrt{\frac{.248}{1006}} = 0.157$$

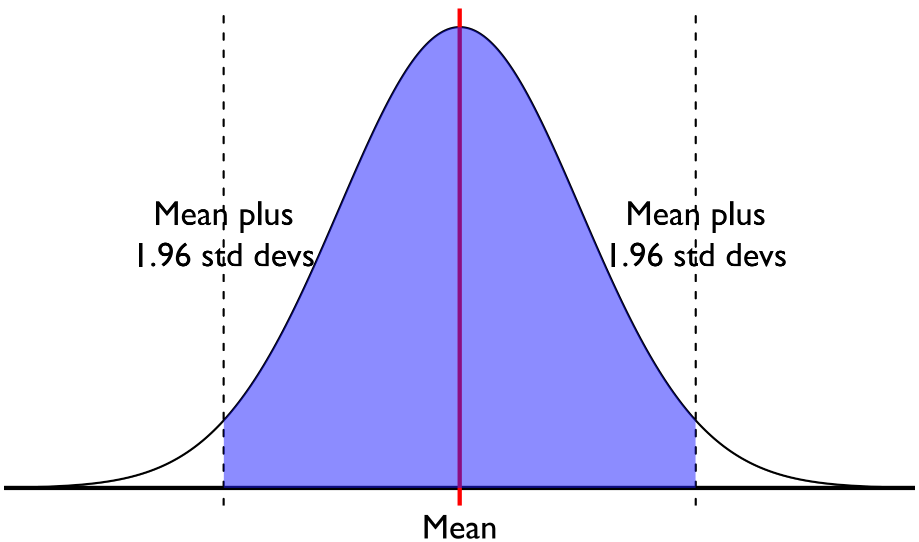
Compare: the standard deviation of our simulations was 0.0157

Another way to get the margin of error (2)



Another way to get the margin of error (2)

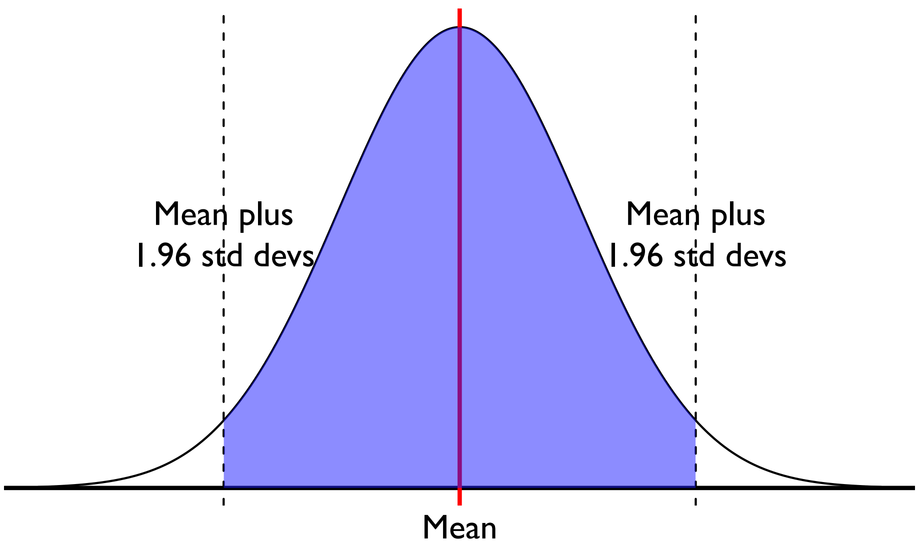
In a Normal distribution, about 95% of the draws are within 1.96 standard deviations of the mean.



Another way to get the margin of error (2)

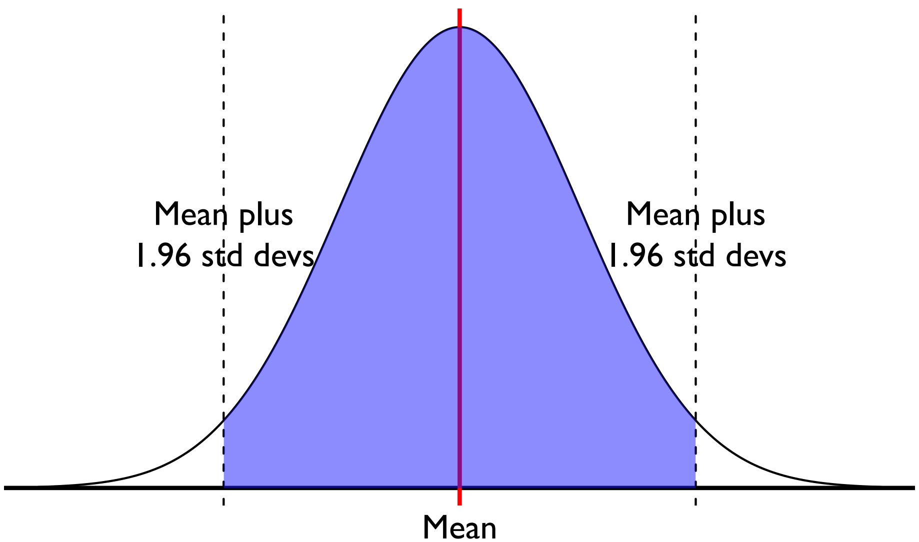
In a Normal distribution, about 95% of the draws are within 1.96 standard deviations of the mean.

This indicates that in 95% of surveys we run, our answer should be within 1.96 standard deviations of the truth.



Another way to get the margin of error (2)

In a Normal distribution, about 95% of the draws are within 1.96 standard deviations of the mean.

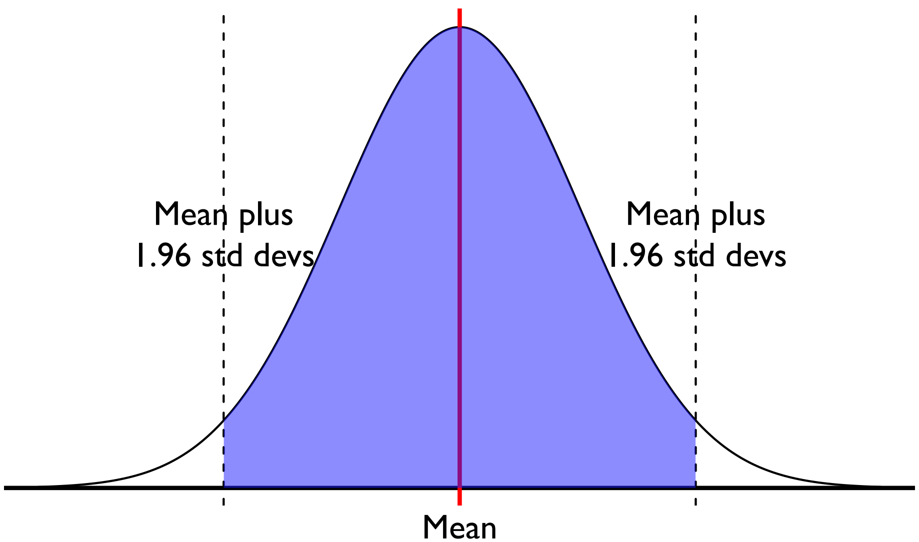


This indicates that in 95% of surveys we run, our answer should be within 1.96 standard deviations of the truth.

Given estimated standard deviation (standard error) of 0.0157, we have a **margin of error** (1.96 times standard error) of .031.

Another way to get the margin of error (2)

In a Normal distribution, about 95% of the draws are within 1.96 standard deviations of the mean.



This indicates that in 95% of surveys we run, our answer should be within 1.96 standard deviations of the truth.

Given estimated standard deviation (standard error) of 0.0157, we have a **margin of error** (1.96 times standard error) of .031.

Compare: our simulations implied a margin of error of 0.031.

Quick recap: survey part

Quick recap: survey part

- **Standard error of poll**: an estimate of how much the estimate might vary due to random error (**sampling error**)

Quick recap: survey part

- **Standard error of poll**: an estimate of how much the estimate might vary due to random error (**sampling error**)
- In 95% of polls, the true value should be within **margin of error** ($\approx 2 \times$ standard error) of the estimate (assuming no bias)

Quick recap: survey part

- **Standard error of poll**: an estimate of how much the estimate might vary due to random error (**sampling error**)
- In 95% of polls, the true value should be within **margin of error** (≈ 2 × standard error) of the estimate (assuming no bias)
- Two ways we got the margin of error:

Quick recap: survey part

- **Standard error of poll**: an estimate of how much the estimate might vary due to random error (**sampling error**)
- In 95% of polls, the true value should be within **margin of error** ($\approx 2 \times$ standard error) of the estimate (assuming no bias)
- Two ways we got the margin of error:
  - **Simulation** in R of 10,000 random samples of size 1,006 given a known level of support for "Leave"

Quick recap: survey part

- **Standard error of poll**: an estimate of how much the estimate might vary due to random error (**sampling error**)
- In 95% of polls, the true value should be within **margin of error** (≈ 2 × standard error) of the estimate (assuming no bias)
- Two ways we got the margin of error:
  - **Simulation** in R of 10,000 random samples of size 1,006 given a known level of support for "Leave"
  - **Central limit theorem:** approximation to a normal distribution

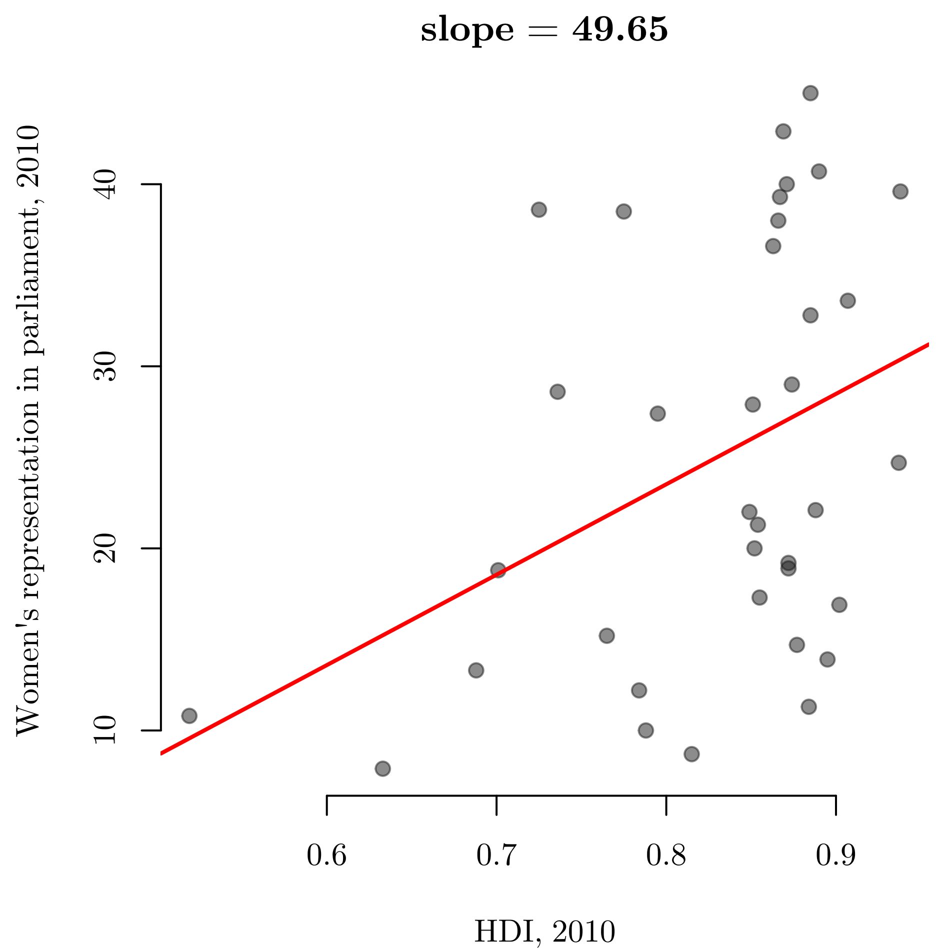
Inference in regression

Inference in regression

Recall from Lab 2: in Lijphart's data, positive relationship between development and women's representation in parliament:

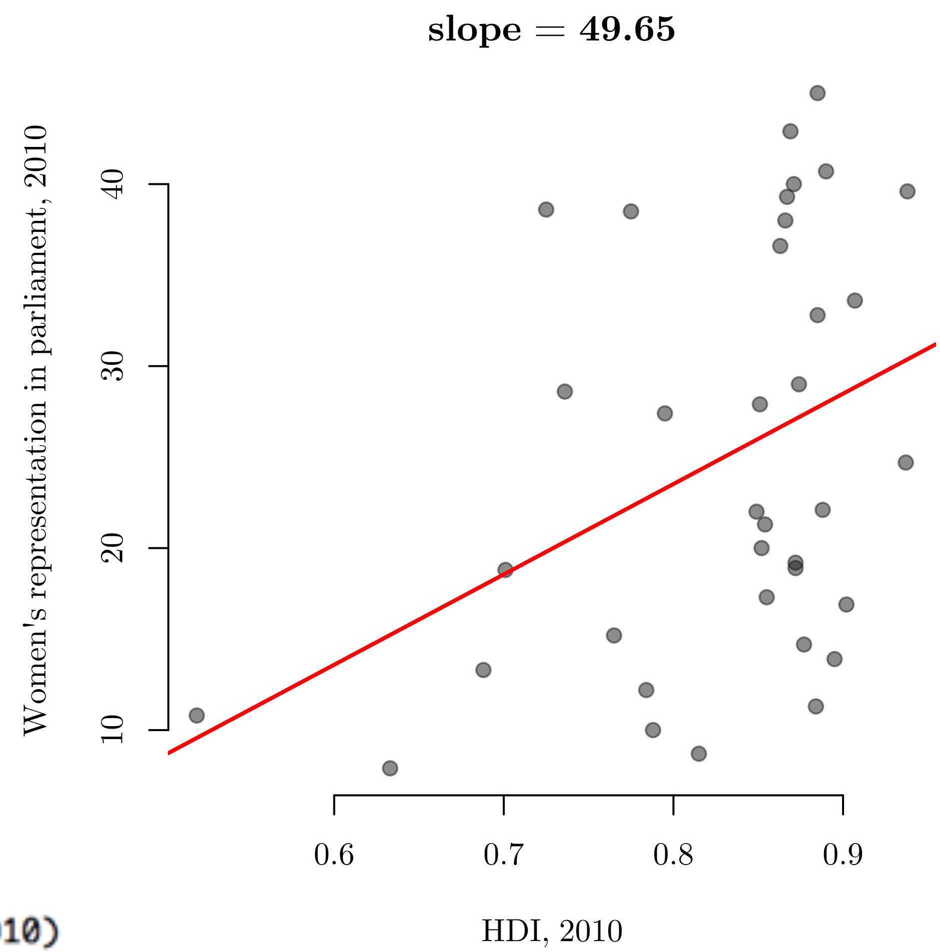
Inference in regression

Recall from Lab 2: in Lijphart's data, positive relationship between development and women's representation in parliament:



Inference in regression

Recall from Lab 2: in Lijphart's data, positive relationship between development and women's representation in parliament:




```
> lm(data$women2010 ~ data$hdi_2010)

Call:
lm(formula = data$women2010 ~ data$hdi_2010)

Coefficients:
  (Intercept)  data$hdi_2010
       -16.20          49.65
```

Inference in regression (2)

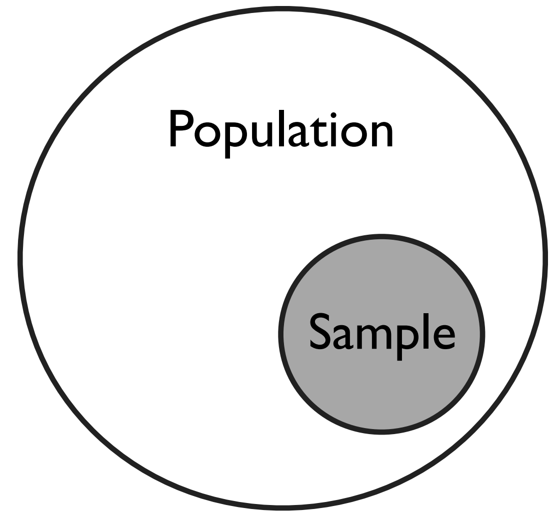
Looking at summary(), R gives us standard errors for the coefficients:

Inference in regression (2)

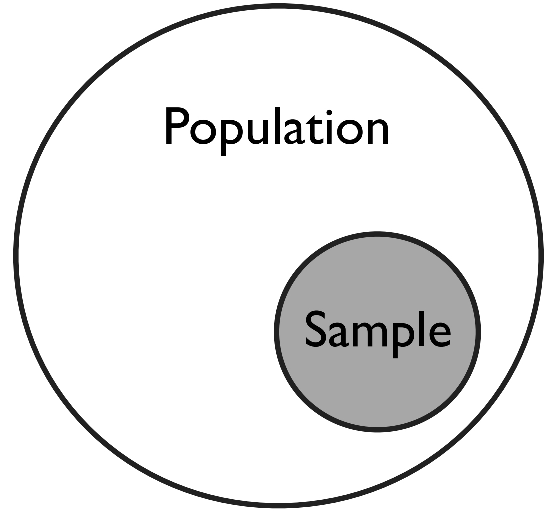
Looking at summary(), R gives us standard errors for the coefficients:

```
> model1 = lm(data$women2010 ~ data$hdi_2010)
> summary(model1)

Call:
lm(formula = data$women2010 ~ data$hdi_2010)

Residuals:
    Min      1Q  Median      3Q     Max 
-16.390  -7.970  -1.879   9.410  18.804 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)     -16.20      16.90  -0.958   0.3447  
data$hdi_2010    49.65      20.29   2.447   0.0197 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.66 on 34 degrees of freedom
Multiple R-squared:  0.1497,	Adjusted R-squared:  0.1247 
F-statistic: 5.988 on 1 and 34 DF,  p-value: 0.01973
```

Two ways to think about the standard errors in a regression

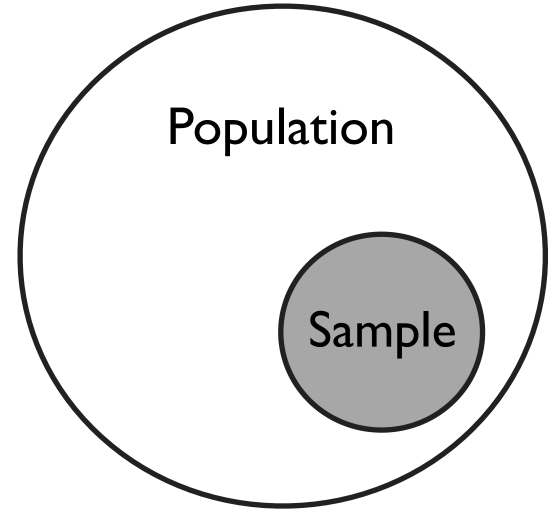


Two ways to think about the standard errors in a regression



Our estimates have standard errors because we view our data as a sample, and we want to characterize the population.

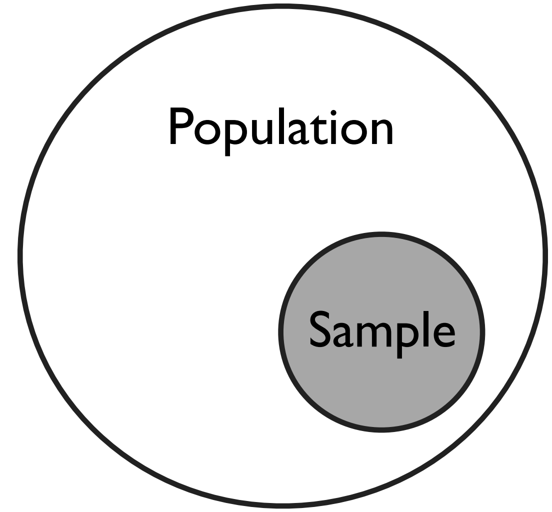
Two ways to think about the standard errors in a regression



Our estimates have standard errors because we view our data as a sample, and we want to characterize the population.

Two ways of thinking about this:

Two ways to think about the standard errors in a regression

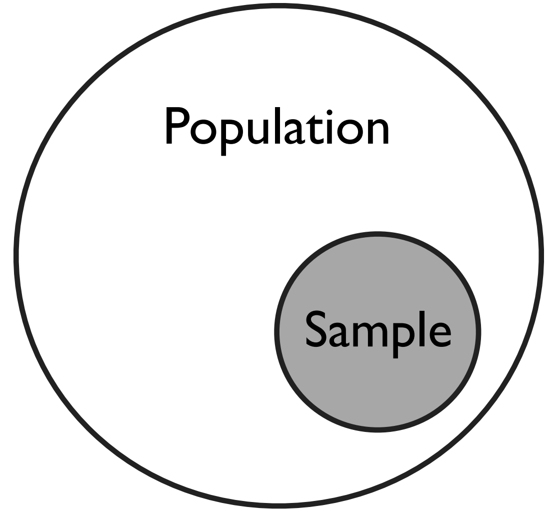


Our estimates have standard errors because we view our data as a sample, and we want to characterize the population.

Two ways of thinking about this:

1. **Just like in the survey:** Our *units* are a random sample from a population, so the coefficients vary across samples.

Two ways to think about the standard errors in a regression



Our estimates have standard errors because we view our data as a sample, and we want to characterize the population.

Two ways of thinking about this:

1. **Just like in the survey:** Our *units* are a random sample from a population, so the coefficients vary across samples.
2. **Slightly differently:** Our *residuals* are a random sample from a population, so the coefficients vary across samples.

Illustration via "re-sampling"

Illustration via "re-sampling"

We have a sample, but we want to speak about the population from which the sample is drawn.

Illustration via "re-sampling"

We have a sample, but we want to speak about the population from which the sample is drawn.

One way to do this: *sample with replacement* from the units in our sample, and do the regression in each "re-sample".

Illustration via "re-sampling"

We have a sample, but we want to speak about the population from which the sample is drawn.

One way to do this: *sample with replacement* from the units in our sample, and do the regression in each "re-sample".

Simple example:

Illustration via "re-sampling"

We have a sample, but we want to speak about the population from which the sample is drawn.

One way to do this: *sample with replacement* from the units in our sample, and do the regression in each "re-sample".

Simple example:

If our **sample** is "Andy, Andrea, Robin",

Illustration via "re-sampling"

We have a sample, but we want to speak about the population from which the sample is drawn.

One way to do this: *sample with replacement* from the units in our sample, and do the regression in each "re-sample".

Simple example:

If our **sample** is "Andy, Andrea, Robin",

three **re-samples** might be

Illustration via “re-sampling”

We have a sample, but we want to speak about the population from which the sample is drawn.

One way to do this: *sample with replacement* from the units in our sample, and do the regression in each “re-sample”.

Simple example:

If our **sample** is “Andy, Andrea, Robin”,

three **re-samples** might be

“Andrea, Robin, Robin”

Illustration via “re-sampling”

We have a sample, but we want to speak about the population from which the sample is drawn.

One way to do this: *sample with replacement* from the units in our sample, and do the regression in each “re-sample”.

Simple example:

If our **sample** is “Andy, Andrea, Robin”,

three **re-samples** might be

“Andrea, Robin, Robin”

“Andy, Andy, Robin”

Illustration via "re-sampling"

We have a sample, but we want to speak about the population from which the sample is drawn.

One way to do this: *sample with replacement* from the units in our sample, and do the regression in each "re-sample".

Simple example:

If our **sample** is "Andy, Andrea, Robin",

three **re-samples** might be

"Andrea, Robin, Robin"

"Andy, Andy, Robin"

"Andy, Andrea, Robin"

Illustration via “re-sampling”: scatterplot

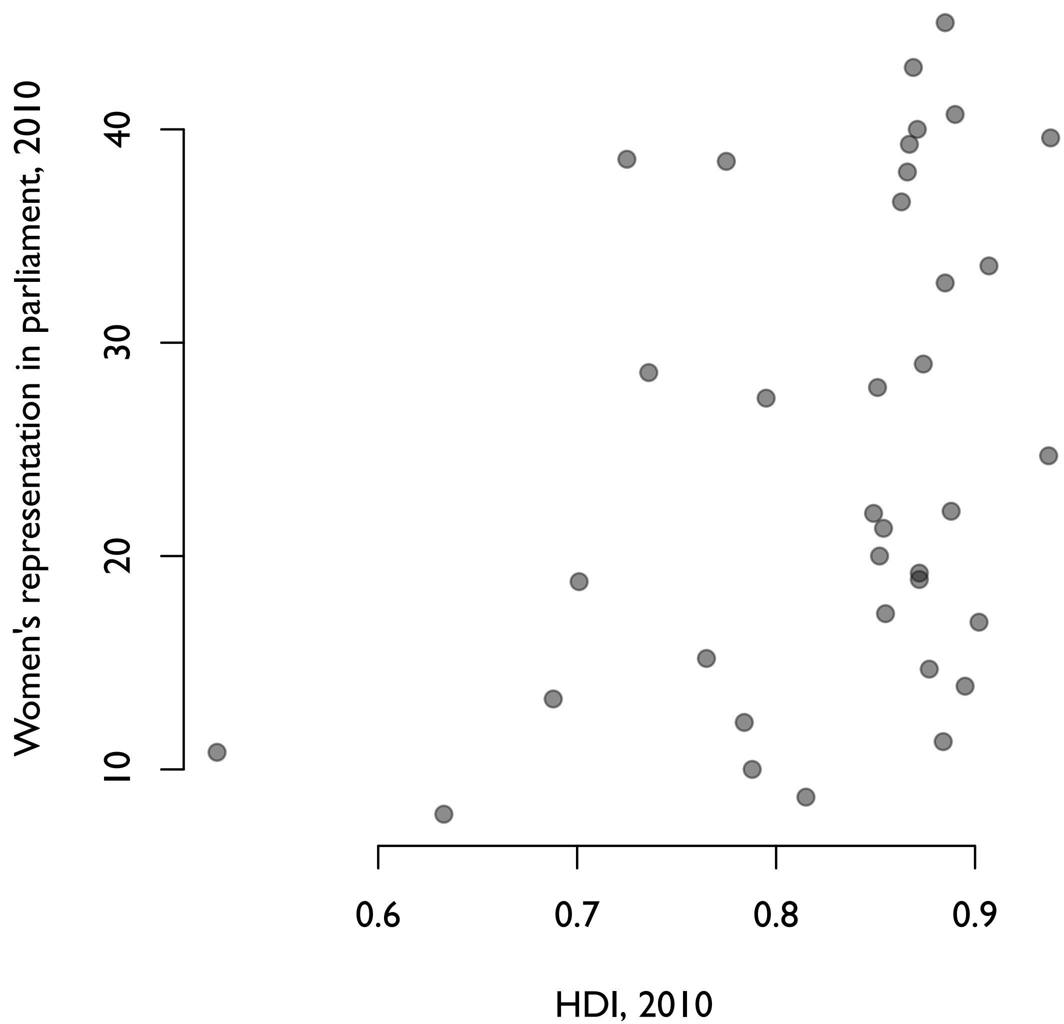


Illustration via “re-sampling”: regression in original sample

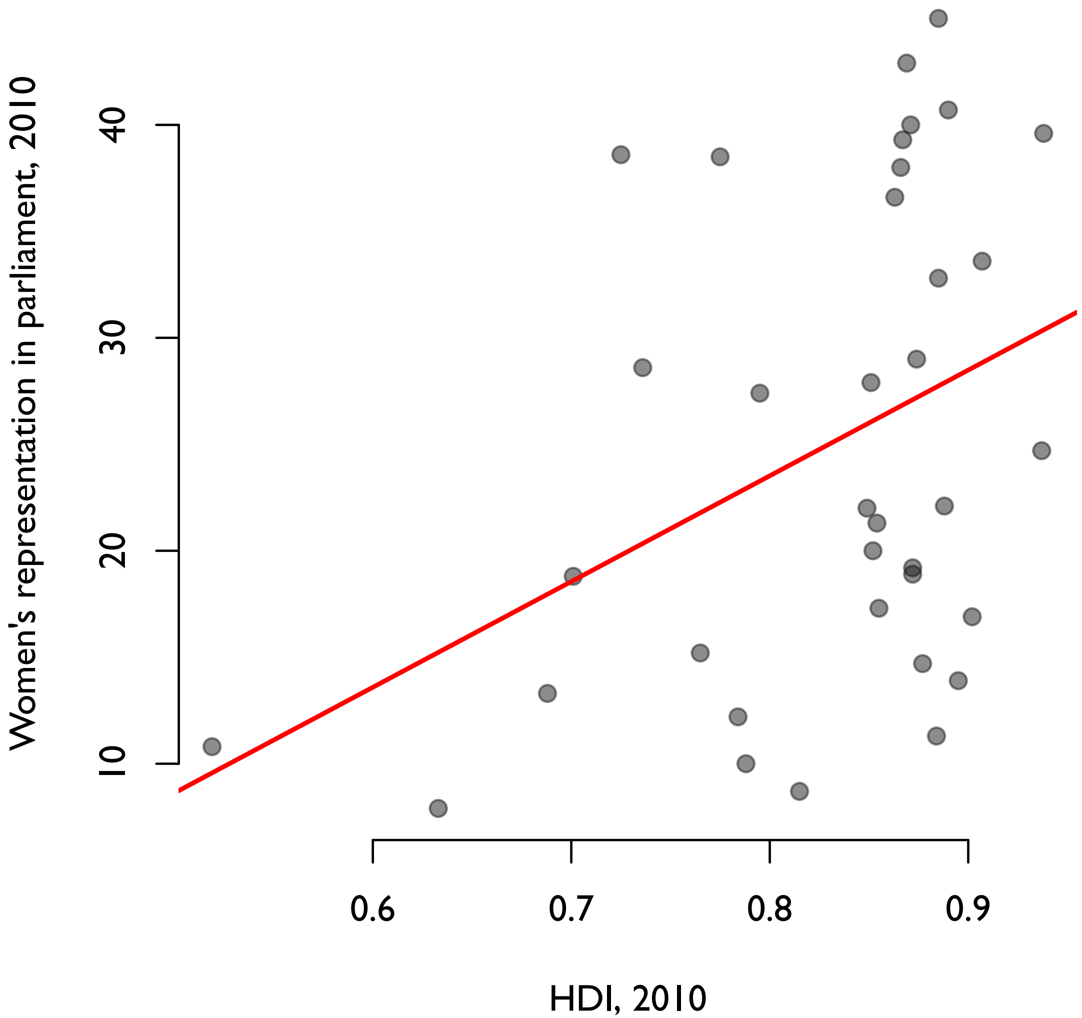


Illustration via “re-sampling”: regression in one re-sample

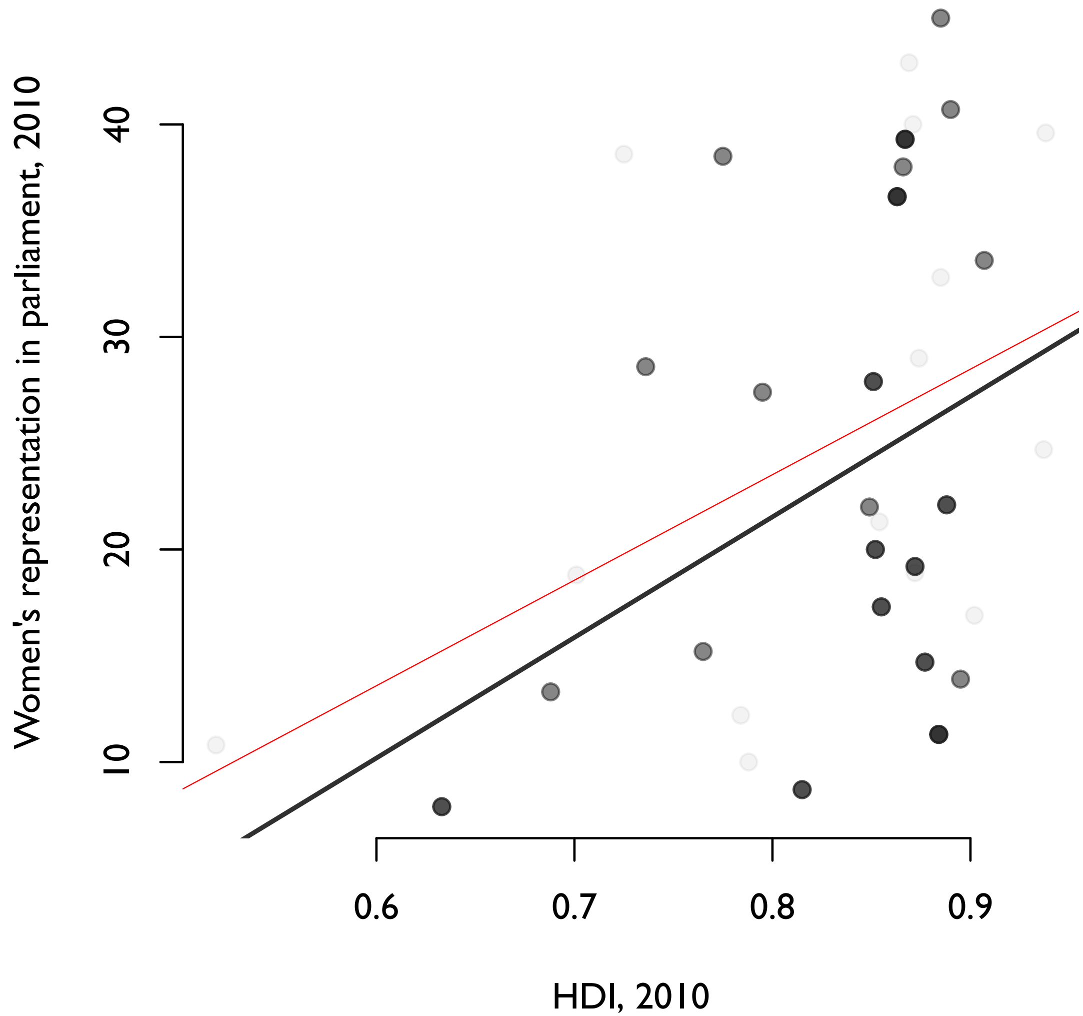


Illustration via "re-sampling": regression in 1000 re-samples

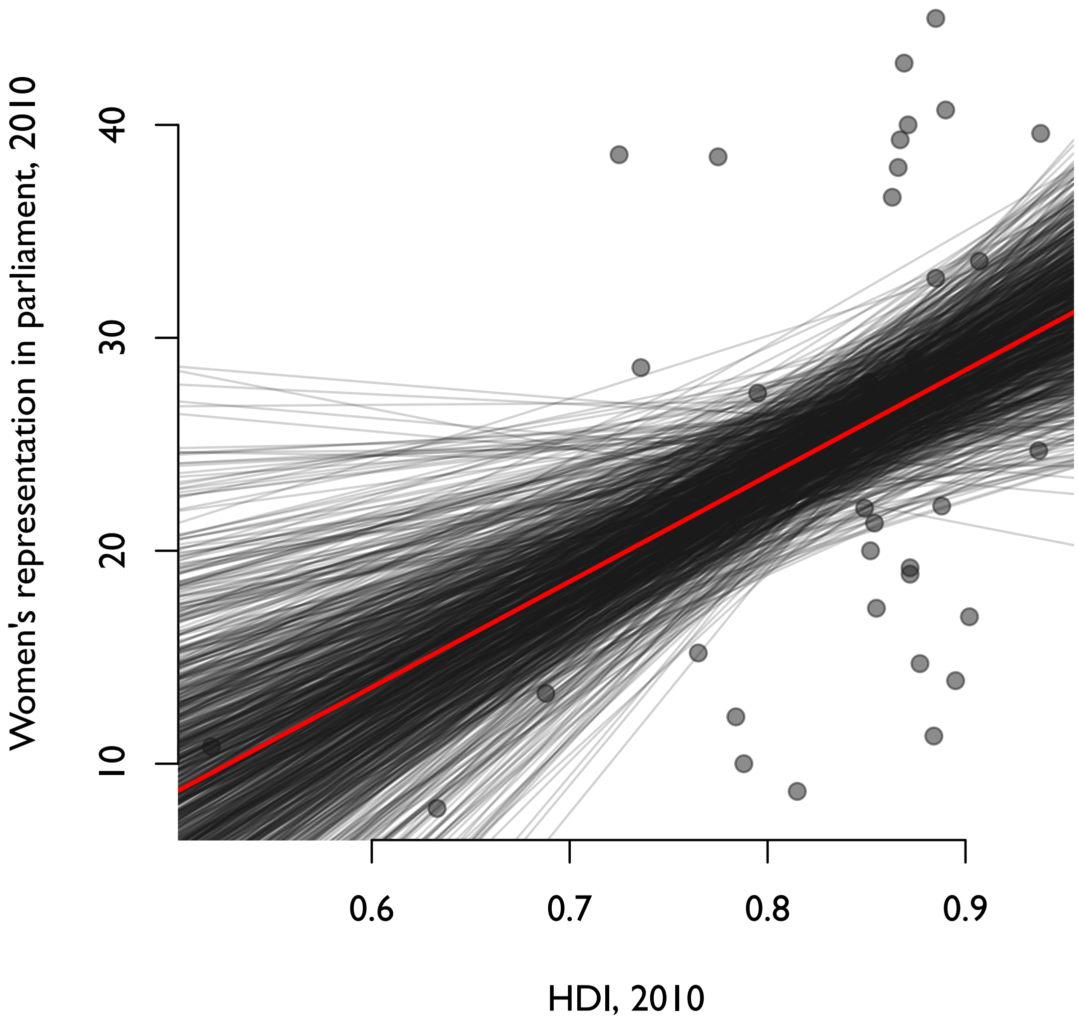
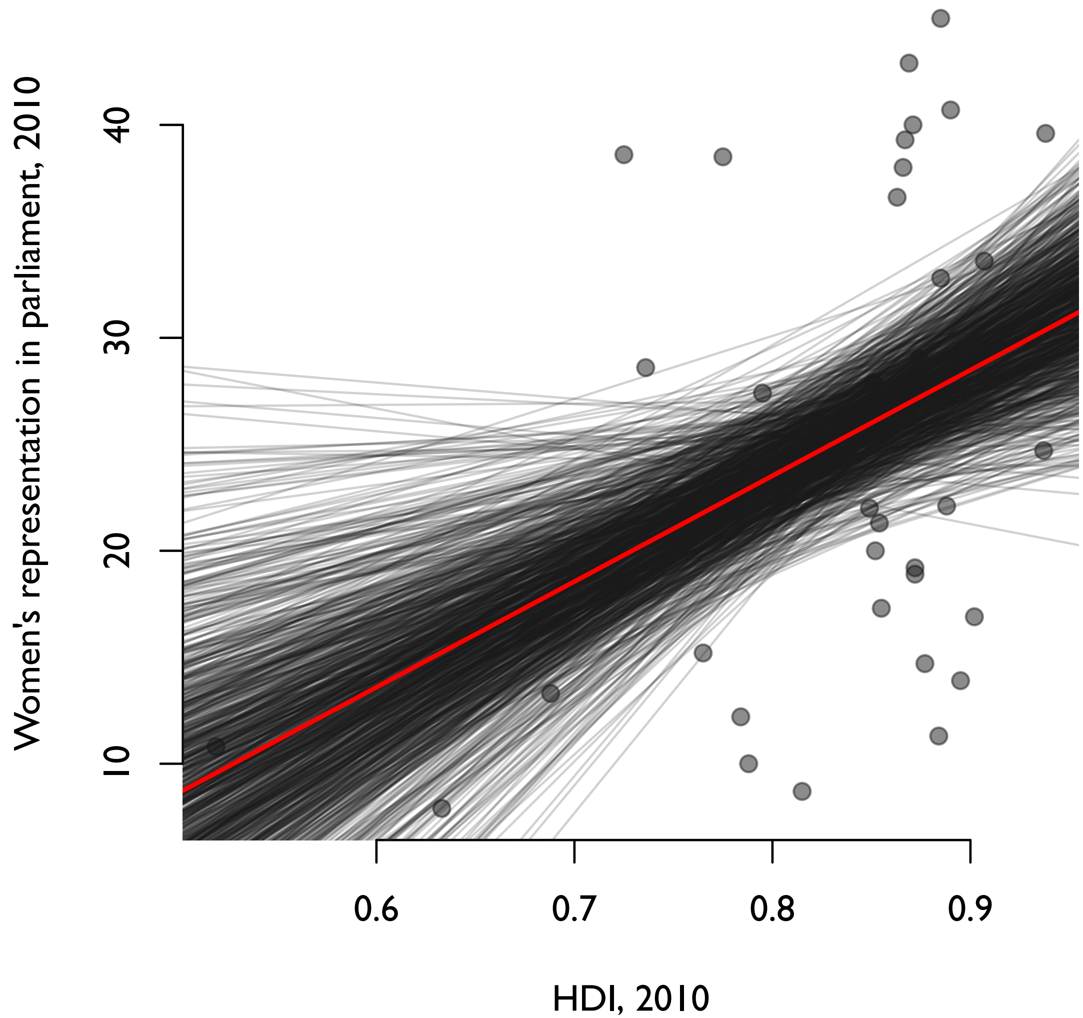


Illustration via "re-sampling": regression in 1000 re-samples



Standard deviation of slope, intercept across samples gives an approximation of the standard errors.

Inference in regression (recap)

Inference in regression (recap)

```
> model1 = lm(data$women2010 ~ data$hdi_2010)
> summary(model1)

Call:
lm(formula = data$women2010 ~ data$hdi_2010)

Residuals:
    Min      1Q  Median      3Q     Max 
-16.390  -7.970  -1.879   9.410  18.804 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)     -16.20      16.90  -0.958   0.3447  
data$hdi_2010    49.65      20.29   2.447   0.0197 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.66 on 34 degrees of freedom
Multiple R-squared:  0.1497,	Adjusted R-squared:  0.1247 
F-statistic: 5.988 on 1 and 34 DF,  p-value: 0.01973
```

Finally: hypothesis testing

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

The usual question we ask in hypothesis testing:

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

The usual question we ask in hypothesis testing:

"In my **sample** there is a relationship between two variables. Is it plausible that we would observe such a strong relationship in the **sample** if there is no relationship in the **population**?"

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

The usual question we ask in hypothesis testing:

"In my **sample** there is a relationship between two variables. Is it plausible that we would observe such a strong relationship in the **sample** if there is no relationship in the **population**?"

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

The usual question we ask in hypothesis testing:

"In my **sample** there is a relationship between two variables. Is it plausible that we would observe such a strong relationship in the **sample** if there is no relationship in the **population**?"

The **p-value** measures how likely it is that we would observe such a strong relationship in a sample if in fact there were no relationship in the population. A relationship is **statistically significant** if this p-value is below a (low) threshold.

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

The usual question we ask in hypothesis testing:

"In my **sample** there is a relationship between two variables. Is it plausible that we would observe such a strong relationship in the **sample** if there is no relationship in the **population**?"

The **p-value** measures how likely it is that we would observe such a strong relationship in a sample if in fact there were no relationship in the population. A relationship is **statistically significant** if this p-value is below a (low) threshold.

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

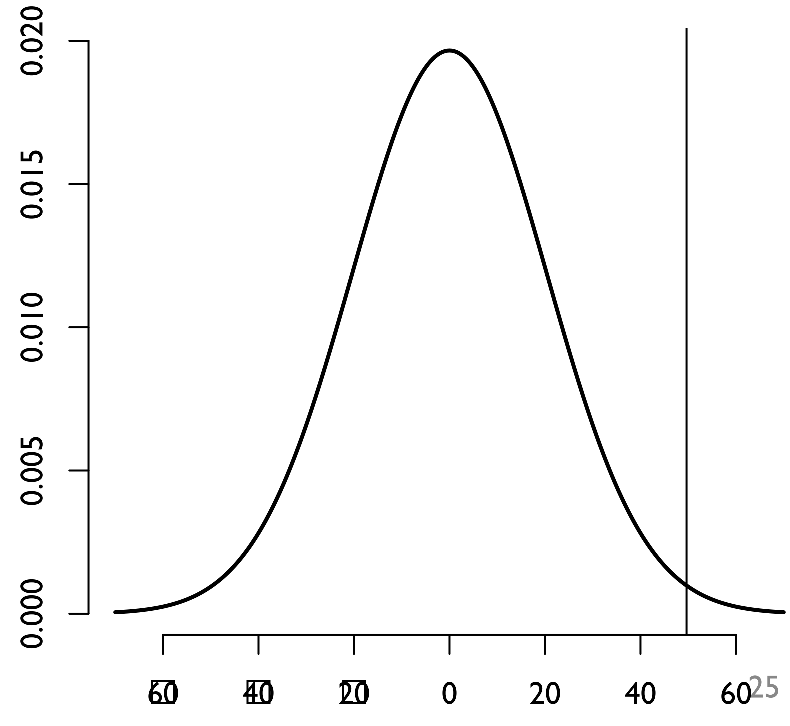
The usual question we ask in hypothesis testing:

"In my **sample** there is a relationship between two variables. Is it plausible that we would observe such a strong relationship in the **sample** if there is no relationship in the **population**?"

The **p-value** measures how likely it is that we would observe such a strong relationship in a sample if in fact there were no relationship in the population. A relationship is **statistically significant** if this p-value is below a (low) threshold.

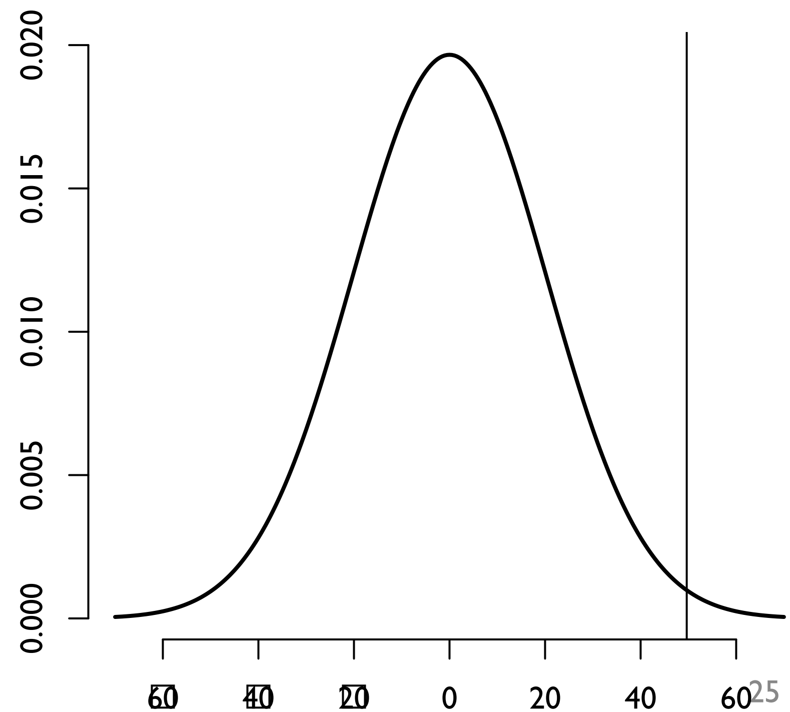
We use standard errors to do hypothesis testing.

Hypothesis testing (2)



Hypothesis testing (2)

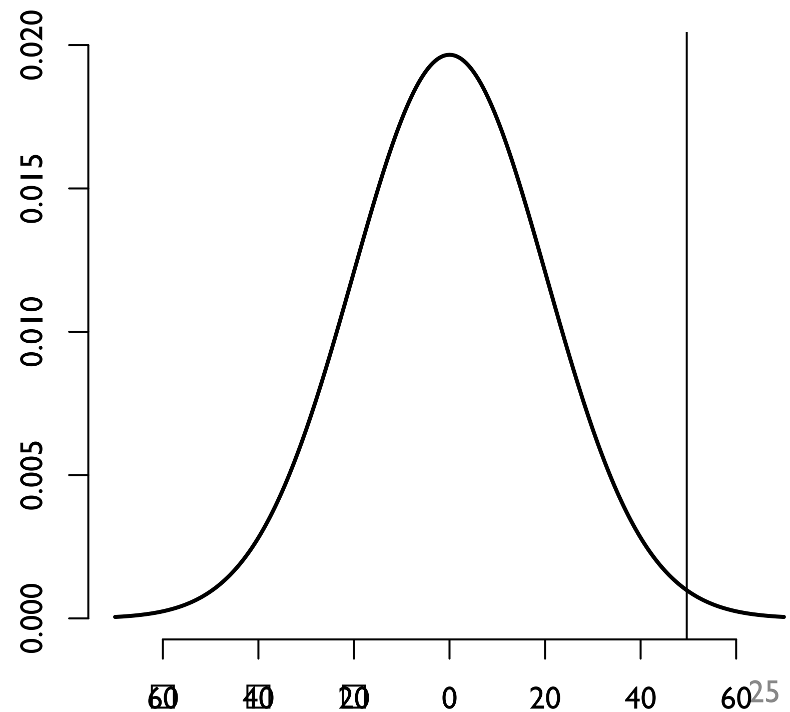
Our standard errors tell us how much estimates might vary across samples.



Hypothesis testing (2)

Our standard errors tell us how much estimates might vary across samples.

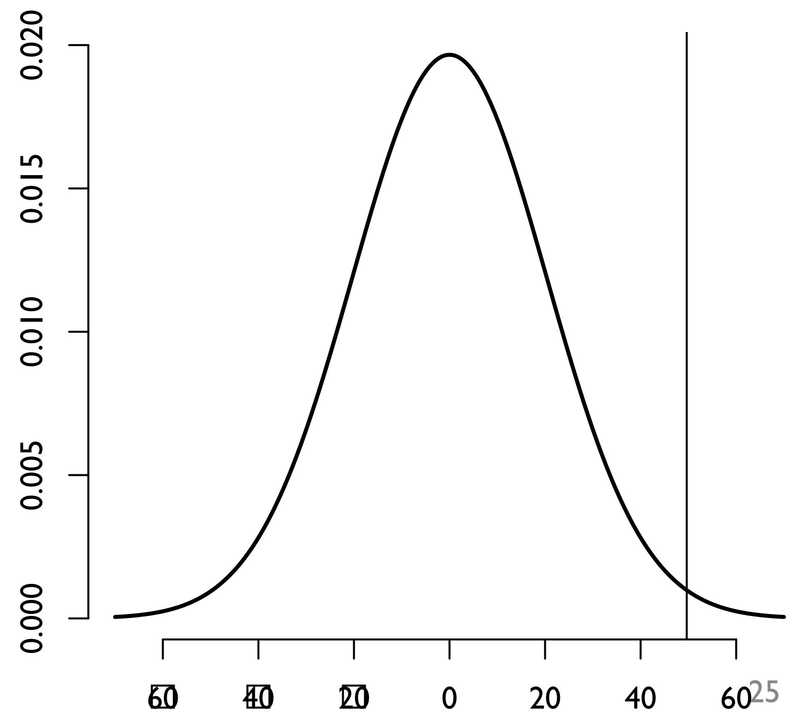
Suppose our estimated coefficient is 49.65, and our estimated standard error is 20.29. How likely is it that the regression coefficient in the population is 0?



Hypothesis testing (2)

Our standard errors tell us how much estimates might vary across samples.

Suppose our estimated coefficient is 49.65, and our estimated standard error is 20.29. How likely is it that the regression coefficient in the population is 0?

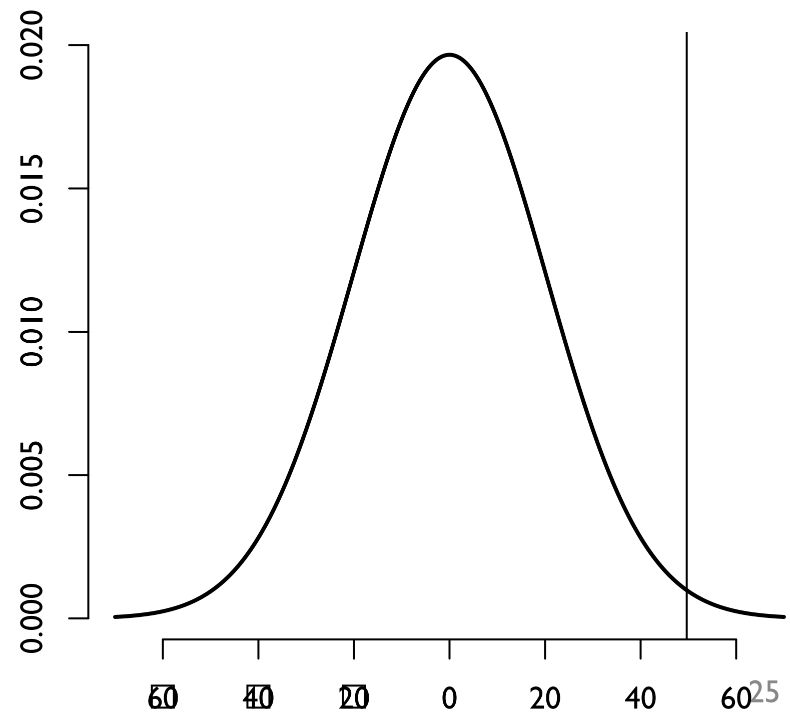


Hypothesis testing (2)

Our standard errors tell us how much estimates might vary across samples.

Suppose our estimated coefficient is 49.65, and our estimated standard error is 20.29. How likely is it that the regression coefficient in the population is 0?

Unlikely!



Inference in regression (recap)

Inference in regression (recap)

```
> model1 = lm(data$women2010 ~ data$hdi_2010)
> summary(model1)

Call:
lm(formula = data$women2010 ~ data$hdi_2010)

Residuals:
    Min      1Q  Median      3Q     Max 
-16.390  -7.970  -1.879   9.410  18.804 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)     -16.20      16.90  -0.958   0.3447  
data$hdi_2010    49.65      20.29   2.447   0.0197 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.66 on 34 degrees of freedom
Multiple R-squared:  0.1497,	Adjusted R-squared:  0.1247 
F-statistic: 5.988 on 1 and 34 DF,  p-value: 0.01973
```

Illustration via "re-sampling": regression in 1000 re-samples

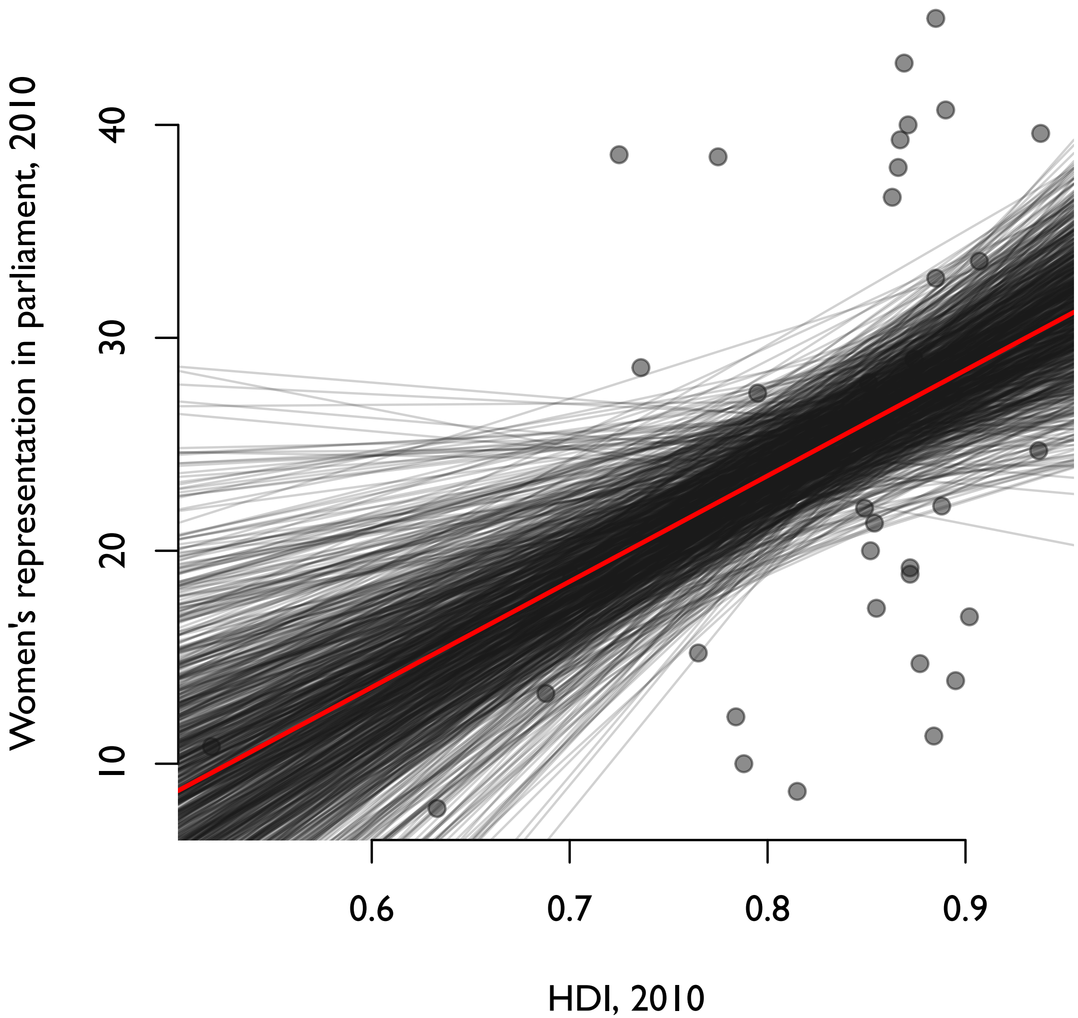
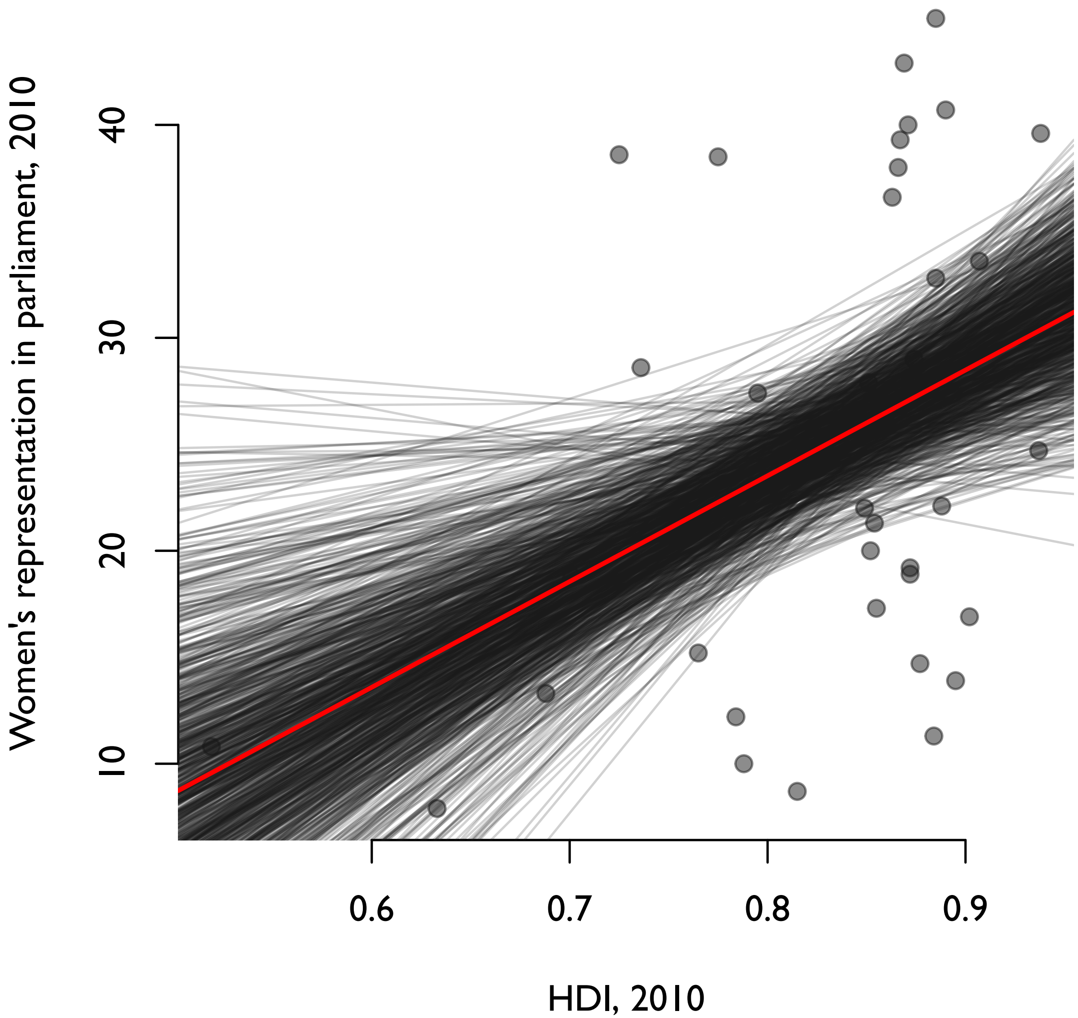


Illustration via "re-sampling": regression in 1000 re-samples



Standard deviation of slope, intercept across samples gives an approximation of the standard errors.

Finally: hypothesis testing

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

Usually, the logic of hypothesis testing is to check whether an observed fact about a sample could have happened by chance, i.e. if nothing is actually happening:

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

Usually, the logic of hypothesis testing is to check whether an observed fact about a sample could have happened by chance, i.e. if nothing is actually happening:

1. Calculate your statistic in the sample/data (e.g. support level, correlation, regression coefficient)

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "**statistically significant**" mean? What is a "**p-value**"?

Usually, the logic of hypothesis testing is to check whether an observed fact about a sample could have happened by chance, i.e. if nothing is actually happening:

1. Calculate your statistic in the sample/data (e.g. support level, correlation, regression coefficient)
2. Define a "**null hypothesis**" that indicates "nothing happening" (e.g. support is 50%, correlation is 0, regression coefficient is 0)

Finally: hypothesis testing

How do we assess whether we have found something in the data?
What does "**statistically significant**" mean? What is a "**p-value**"?

Usually, the logic of hypothesis testing is to check whether an observed fact about a sample could have happened by chance, i.e. if nothing is actually happening:

1. Calculate your statistic in the sample/data (e.g. support level, correlation, regression coefficient)
2. Define a "**null hypothesis**" that indicates "nothing happening" (e.g. support is 50%, correlation is 0, regression coefficient is 0)
3. Calculate the **p-value**: probability of getting a statistic as large as yours if the null hypothesis were true (e.g. p=0.2, p=.002)

Finally: hypothesis testing

How do we assess whether we have found something in the data?

What does "statistically significant" mean? What is a "**p-value**"?

Usually, the logic of hypothesis testing is to check whether an observed fact about a sample could have happened by chance, i.e. if nothing is actually happening:

1. Calculate your statistic in the sample/data (e.g. support level, correlation, regression coefficient)
2. Define a "**null hypothesis**" that indicates "nothing happening" (e.g. support is 50%, correlation is 0, regression coefficient is 0)
3. Calculate the p-value: probability of getting a statistic as large as yours if the null hypothesis were true (e.g. p=0.2, p=.002)
4. If p-value is low enough, reject null hypothesis, and say the correlation or regression coefficient is "statistically significant"

To discuss: why have recent polls been wrong?

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

Random error?

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

Random error?

In general election 2015, no; it was **bias** (in the statistical sense):

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

Random error?

In general election 2015, no; it was **bias** (in the statistical sense):

- Conservative voters under-represented in surveys, Labour voters over-represented.

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

Random error?

In general election 2015, no; it was **bias** (in the statistical sense):

- Conservative voters under-represented in surveys, Labour voters over-represented.
- Politically engaged over-represented.

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

Random error?

In general election 2015, no; it was **bias** (in the statistical sense):

- Conservative voters under-represented in surveys, Labour voters over-represented.
- Politically engaged over-represented.

Extremely difficult to get truly representative random sample.

To discuss: why have recent polls been wrong?

measured value = true value + bias + random error

Random error?

In general election 2015, no; it was **bias** (in the statistical sense):

- Conservative voters under-represented in surveys, Labour voters over-represented.
- Politically engaged over-represented.

Extremely difficult to get truly representative random sample.

Important: Margin of error captures random error (i.e. sampling error), not bias.

Now you should understand:

Dependent variable: Nobel Prizes awarded per capita (in log scale)

	(1)	(2)	(3)
Intercept	-1.629* (0.509)	-3.166* (0.511)	-2.982* (0.527)
Chocolate consumption per capita (log scale)	2.092* (0.298)	1.026* (0.326)	0.709 (0.415)
GDP/capita (thousands of USD)		0.105* (0.024)	0.106* (0.024)
NW Europe			0.549 (0.452)
R^2	0.70	0.85	0.86
N	34	34	34

Standard errors in parentheses. * Indicates $p<0.05$

- what a dependent variable is
- what an independent variable is
- what the coefficients mean (intercept, slopes)
- what the stars mean (i.e. what $p<0.05$ means)
- what the standard errors mean

And this too!

TABLE 15.2

Multivariate regression analyses of the effect of consensus democracy (executives-parties dimension) on five indicators of violence, with controls for the effects of the level of economic development, logged population size, and degree of societal division, and with extreme outliers removed

Performance variables	Estimated regression coefficient	Absolute t-value	Countries (N)
Political stability and absence of violence (1996–2009)	0.189***	3.360	34
Internal conflict risk (1990–2004)	0.346**	2.097	32
Weighted domestic conflict index (1981–2009)	−105.0*	1.611	30
Weighted domestic conflict index (1990–2009)	−119.7**	2.177	33
Deaths from domestic terrorism (1985–2010)	−2.357**	1.728	33

* Statistically significant at the 10 percent level (one-tailed test)

** Statistically significant at the 5 percent level (one-tailed test)

*** Statistically significant at the 1 percent level (one-tailed test)

Source: Based on data in Kaufmann, Kraay, and Mastruzzi 2010; PRS Group 2004; Banks, 2010; and GTD Team 2010

- what the dependent and independent variables are
- what Lijphart means by "controlling for" three other variables
- what the stars mean
- t-values: estimate divided by standard error