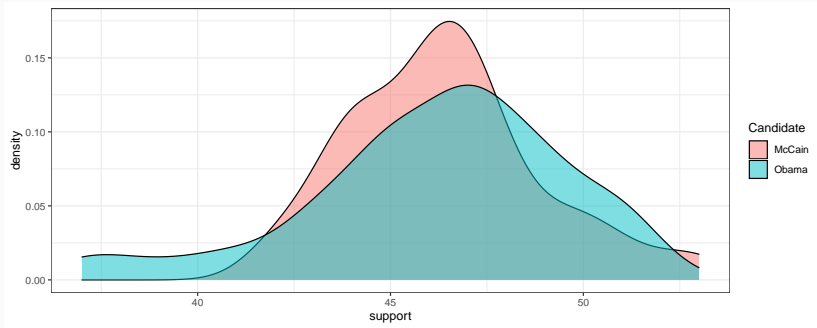


Uncertainty, 1

Frank Edwards

The basic challenge: Will Obama win Florida in '08?

##	state	Pollster	Obama	McCain	middate
## 1	FL	Zogby-4	48	46	2008-10-31
## 2	FL	ARG-3	50	46	2008-10-30
## 3	FL	Mason-Dixon-2	47	45	2008-10-29
## 4	FL	Datamar-2	47	47	2008-10-29
## 5	FL	Rasmussen-1	51	47	2008-10-26
## 6	FL	LA Times-3	50	43	2008-10-26



To estimate a proportion of voters supporting a candidate

- Take a simple random sample of n voters

To estimate a proportion of voters supporting a candidate

- Take a simple random sample of n voters
- Ask them their preferences (support for Obama $x = [0, 1]$)

To estimate a proportion of voters supporting a candidate

- Take a simple random sample of n voters
- Ask them their preferences (support for Obama $x = [0, 1]$)
- Compute the proportion of voters supporting Obama in the sample

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

To estimate a proportion of voters supporting a candidate

- Take a simple random sample of n voters
- Ask them their preferences (support for Obama $x = [0, 1]$)
- Compute the proportion of voters supporting Obama in the sample
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$
- Use \bar{x} to inform our understanding of p , the true proportion of voters supporting Obama

Is our estimate biased?

$$\text{estimation error} = \text{estimate} - \text{truth} = \bar{x} - p$$

Is our estimate biased?

$$\text{estimation error} = \text{estimate} - \text{truth} = \bar{x} - p$$

Problems?

Obtain average error, when exact error is unobservable

The data:

$$x \sim \text{Bernoulli}(p)$$

A single sample

```
sample<-rbinom(size=1, n=10, prob=0.55)
### x_i
sample
```

```
## [1] 0 0 0 1 0 1 0 1 0 1
```

```
###  $\bar{x}$ 
sum(sample)/length(sample)
```

```
## [1] 0.4
```

What is the estimation error?

Treat the sample mean \bar{x} as a random variable

Recall that a Binomial variable represents the number of successes from n Bernoulli trials with a probability of success p

Treat the sample mean \bar{x} as a random variable

Recall that a Binomial variable represents the number of successes from n Bernoulli trials with a probability of success p

We can represent the average share of Obama supporters in a sample of voters as a Binomial random variable:

$$\bar{x} \times n \sim \text{Binomial}(p, n)$$

Treat the sample mean \bar{x} as a random variable

Recall that a Binomial variable represents the number of successes from n Bernoulli trials with a probability of success p

We can represent the average share of Obama supporters in a sample of voters as a Binomial random variable:

$$\bar{x} \times n \sim \text{Binomial}(p, n)$$

What does this probability distribution mean?

Repeat our sample of voters

```
sample1<-rbinom(size=1, n=10, prob=0.55)
sample2<-rbinom(size=1, n=10, prob=0.55)
sample3<-rbinom(size=1, n=10, prob=0.55)
mean(sample1); mean(sample2); mean(sample3)
```

```
## [1] 0.6
```

```
## [1] 0.6
```

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

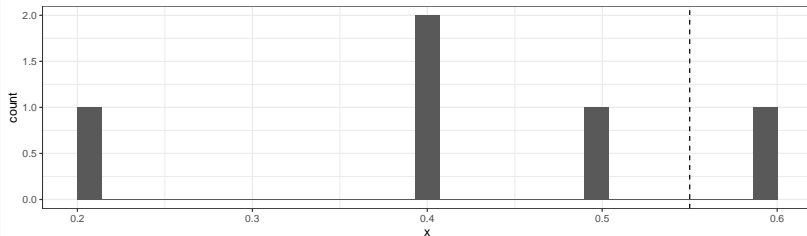
We can repeat this sampling a large number of times

Recall the Law of Large Numbers: as a sample of draws from a random variable increases, the sample mean converges to the population mean $E(X)$

$$\bar{x}_n \rightarrow E(X)$$

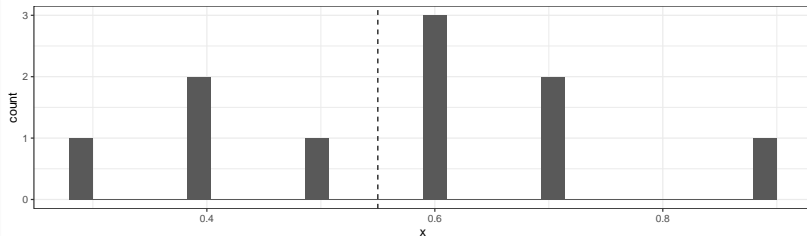
n = 5

```
sim_dat<-data.frame(x = rep(NA, 5))  
for(i in 1:5){  
  sample<-rbinom(size=1,n=10,prob=0.55)  
  sim_dat$x[i]<-mean(sample)  
}  
ggplot(sim_dat, aes(x=x)) +  
  geom_histogram(bins=30) +  
  geom_vline(xintercept = 0.55, lty=2)
```



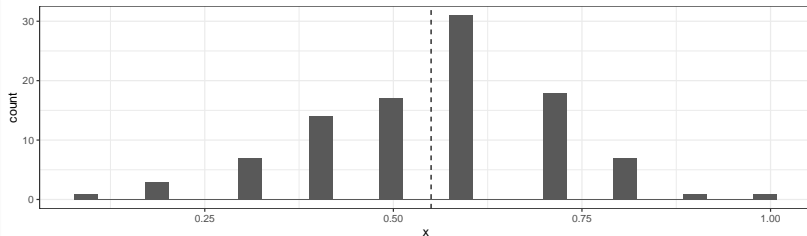
n = 10

```
sim_dat<-data.frame(x = rep(NA, 10))  
for(i in 1:10){  
  sample<-rbinom(size=1,n=10,prob=0.55)  
  sim_dat$x[i]<-mean(sample)  
}  
ggplot(sim_dat, aes(x=x)) +  
  geom_histogram(bins=30) +  
  geom_vline(xintercept = 0.55, lty=2)
```



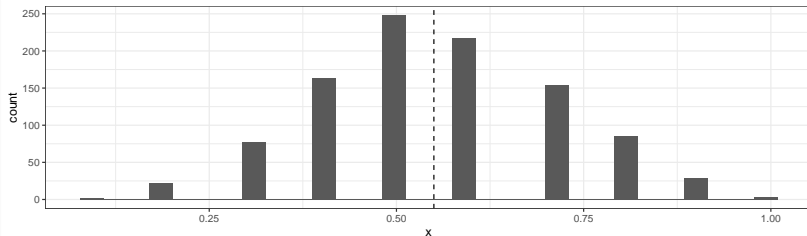
n = 100

```
sim_dat<-data.frame(x = rep(NA, 100))  
for(i in 1:100){  
  sample<-rbinom(size=1,n=10,prob=0.55)  
  sim_dat$x[i]<-mean(sample)  
}  
ggplot(sim_dat, aes(x=x)) +  
  geom_histogram(bins=30) +  
  geom_vline(xintercept = 0.55, lty=2)
```



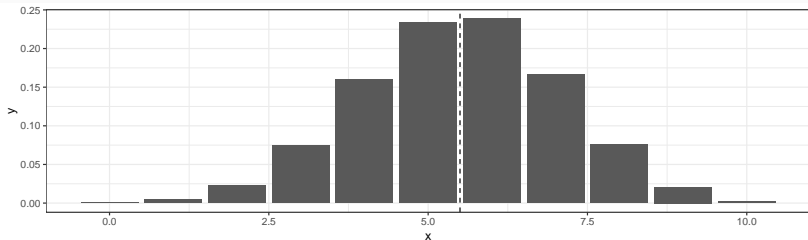
n = 1000

```
sim_dat<-data.frame(x = rep(NA, 1000))
for(i in 1:1000){
  sample<-rbinom(size=1,n=10,prob=0.55)
  sim_dat$x[i]<-mean(sample)
}
ggplot(sim_dat, aes(x=x)) +
  geom_histogram(bins=30) +
  geom_vline(xintercept = 0.55, lty=2)
```



Compare this to Binomial(n,p)

```
x<-data.frame(x=0:10,  
              y=dbinom(0:10,size=10,prob=0.55))  
ggplot(x, aes(x=x,  
              y=y)) +  
  geom_col() +  
  geom_vline(xintercept = 0.55 * 10, lty=2)
```



$$\bar{x}_n \rightarrow E(X)$$

- So if we resample x from the same population a large number of times, then compute \bar{x} , the distribution of sampled values of \bar{x} approximates the *sampling distribution* of \bar{x}

$$\bar{x}_n \rightarrow E(X)$$

- So if we resample x from the same population a large number of times, then compute \bar{x} , the distribution of sampled values of \bar{x} approximates the *sampling distribution* of \bar{x}
- This holds under simple random sampling, where each unit has an equal probability of being selected

So is our sample biased?

$$\text{estimation error} = \text{estimate} - \text{truth} = \bar{x} - p$$

So is our sample biased?

$$\text{estimation error} = \text{estimate} - \text{truth} = \bar{x} - p$$

$$\text{bias} = E(\text{estimation error}) = E(\text{estimate} - \text{truth}) = E(\bar{x}) - p$$

So is our sample biased?

$$\text{estimation error} = \text{estimate} - \text{truth} = \bar{x} - p$$

$$\text{bias} = E(\text{estimation error}) = E(\text{estimate} - \text{truth}) = E(\bar{x}) - p$$

Because $E(\bar{x}) = p$ under simple random sampling

$$\text{bias} = E(\bar{x}) - p = p - p = 0$$

So is our sample biased?

$$\text{estimation error} = \text{estimate} - \text{truth} = \bar{x} - p$$

$$\text{bias} = E(\text{estimation error}) = E(\text{estimate} - \text{truth}) = E(\bar{x}) - p$$

Because $E(\bar{x}) = p$ under simple random sampling

$$\text{bias} = E(\bar{x}) - p = p - p = 0$$

The sample mean \bar{x} is, on average, an *unbiased* estimator of the population average p as long as we've obtained a simple random sample.

So is our sample biased?

$$\text{estimation error} = \text{estimate} - \text{truth} = \bar{x} - p$$

$$\text{bias} = E(\text{estimation error}) = E(\text{estimate} - \text{truth}) = E(\bar{x}) - p$$

Because $E(\bar{x}) = p$ under simple random sampling

$$\text{bias} = E(\bar{x}) - p = p - p = 0$$

The sample mean \bar{x} is, on average, an *unbiased* estimator of the population average p as long as we've obtained a simple random sample.

However: any individual sample has unobservable error $\bar{x}_n - p$

An estimator is biased when $E(\bar{x}) \neq p$

- Bias means that the expectation is systematically different from the true value
- Positive bias means that $E(\bar{x}) > p$, negative bias means $E(\bar{x}) < p$

An estimator is consistent if it converges to the parameter as n increases, or

$$\bar{x}_n \rightarrow E(X)$$

- Simple random sampling yields unbiased and consistent estimators of a parameter

- Simple random sampling yields unbiased and consistent estimators of a parameter
- In experimental settings: Random assignment yields an unbiased estimator of the Sample Average Treatment Effect

- Simple random sampling yields unbiased and consistent estimators of a parameter
- In experimental settings: Random assignment yields an unbiased estimator of the Sample Average Treatment Effect
- In experimental settings: Simple random sampling combined with random assignment yields an unbiased and consistent estimator of the Population Average Treatment Effect

- Simple random sampling yields unbiased and consistent estimators of a parameter
- In experimental settings: Random assignment yields an unbiased estimator of the Sample Average Treatment Effect
- In experimental settings: Simple random sampling combined with random assignment yields an unbiased and consistent estimator of the Population Average Treatment Effect
- See pages 317-322 for more discussion of RCTs and bias in estimators

How wrong are we, on average?

For our vote share example, pretend that $p = 0.55$, and $n = 10$. For a Binomial random variable expressed as a proportion, the variance is:

$$\sigma^2 = \frac{p(1-p)}{n} = \frac{0.55(1-0.55)}{10} \approx 0.025$$

How wrong are we, on average?

For our vote share example, pretend that $p = 0.55$, and $n = 10$. For a Binomial random variable expressed as a proportion, the variance is:

$$\sigma^2 = \frac{p(1-p)}{n} = \frac{0.55(1-0.55)}{10} \approx 0.025$$

We define a standard error as the standard deviation of a sampling distribution

$$SE = \sigma = \sqrt{0.025} \approx 0.16$$

How wrong are we, on average?

For our vote share example, pretend that $p = 0.55$, and $n = 10$. For a Binomial random variable expressed as a proportion, the variance is:

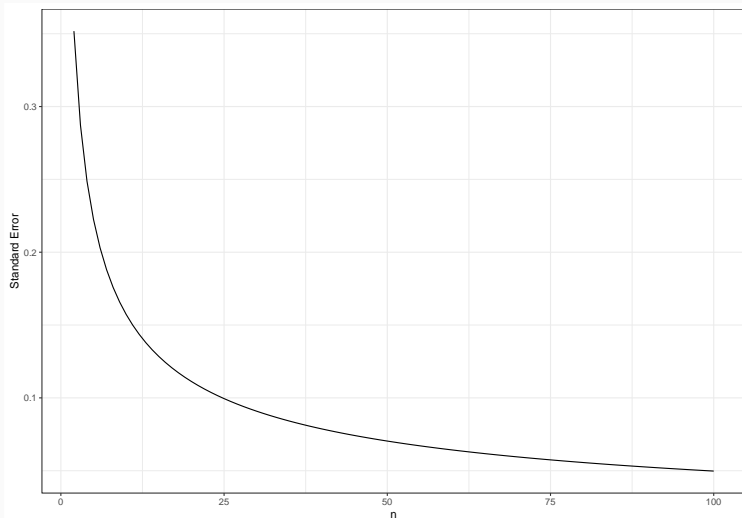
$$\sigma^2 = \frac{p(1-p)}{n} = \frac{0.55(1-0.55)}{10} \approx 0.025$$

We define a standard error as the standard deviation of a sampling distribution

$$SE = \sigma = \sqrt{0.025} \approx 0.16$$

When we sample 10 voters, our estimate of the share of the population that supports Obama will have an average error of about 0.16, or 16 percentage points.

The impact of n on a standard error for $x \sim \text{Binomial}(n, 0.55)$



The Standard Error

In general, the standard error of a sample mean \bar{x}_n where x is a sample of n i.i.d random variables is:

$$SE(\bar{x}) = \frac{sd_x}{\sqrt{n}}$$

The Standard Error

In general, the standard error of a sample mean \bar{x}_n where x is a sample of n i.i.d random variables is:

$$SE(\bar{x}) = \frac{sd_x}{\sqrt{n}}$$

We use the observed data to approximate the standard deviation of the sampling distribution for a population parameter, such as a mean.

The Standard Error

In general, the standard error of a sample mean \bar{x}_n where x is a sample of n i.i.d random variables is:

$$SE(\bar{x}) = \frac{sd_x}{\sqrt{n}}$$

We use the observed data to approximate the standard deviation of the sampling distribution for a population parameter, such as a mean.

It tells us how far away \bar{x} is from $E(x)$ on average

Describing a distribution

Let's return to our voter polls. Assume each of these polls is a simple random sample of voters.

```
data(polls08)
fl<-polls08 %>%
  filter(state=="FL") %>%
  select(Obama)
```

```
head(fl)
```

```
##   Obama
## 1    48
## 2    50
## 3    47
## 4    47
## 5    51
## 6    50
```


Describing a distribution

Let's return to our voter polls. Assume each of these polls is a simple random sample of voters.

```
data(polls08)
fl<-polls08 %>%
  filter(state=="FL") %>%
  select(Obama)
```

```
head(fl)
```

```
##   Obama
## 1    48
## 2    50
## 3    47
## 4    47
## 5    51
## 6    50
```

Each of these represents a draw from the sampling distribution of voters in Florida. The vote share estimate is the mean of a group of sampled voters, whose preference for Obama is a Bernoulli variable with probability p

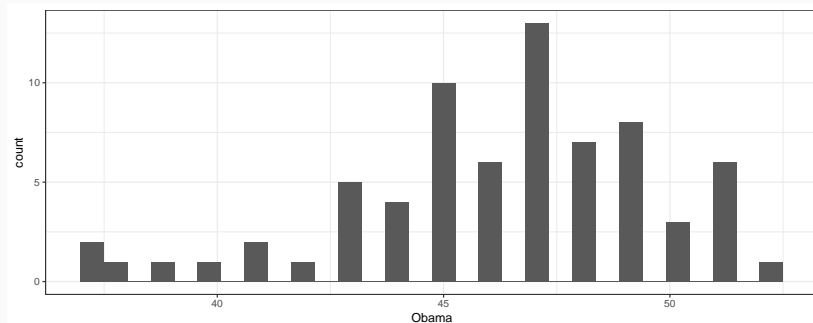
The Central Limit Theorem

The sum of i.i.d. Bernoulli trials is Binomial(n, p). When we add i.i.d. Binomial random variables together, the Central Limit Theorem shows that they converge toward a Normal distribution as the number of variables increases.

The Central Limit Theorem

The sum of i.i.d. Bernoulli trials is Binomial(n, p). When we add i.i.d. Binomial random variables together, the Central Limit Theorem shows that they converge toward a Normal distribution as the number of variables increases.

```
ggplot(fl, aes(x=Obama)) +  
  geom_histogram()
```



The Central Limit Theorem

More formally:

$$\bar{x}_n \sim N \left(E(x), \frac{V(x)}{n} \right) \text{ as } n \rightarrow \infty$$

The sampling distribution of \bar{x} is approximately Normal, if we resample x many times.

The Central Limit Theorem

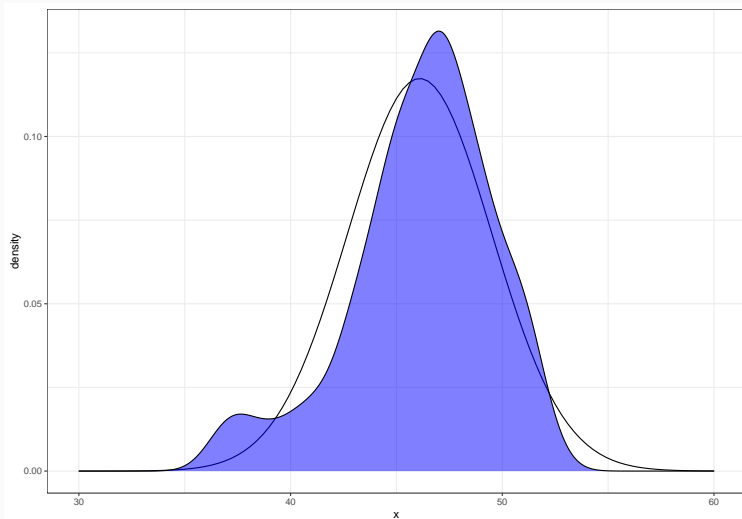
More formally:

$$\bar{x}_n \sim N\left(E(x), \frac{V(x)}{n}\right) \text{ as } n \rightarrow \infty$$

The sampling distribution of \bar{x} is approximately Normal, if we resample x many times.

Using our polling data, we obtained a distribution of \bar{x} with a mean of 46.1 and standard error of 3.4

Observed sample means (blue), and $N(46.1, 3.4)$



How certain are we about our estimate?

To provide an estimate of how certain we are about our estimate, we can compute a *confidence interval*.

1. Choose a confidence level (e.g. 0.9, 0.95, 0.99)

How certain are we about our estimate?

To provide an estimate of how certain we are about our estimate, we can compute a *confidence interval*.

1. Choose a confidence level (e.g. 0.9, 0.95, 0.99)
2. Compute the quantile of the Normal distribution that corresponds to that confidence level (symmetric)

How certain are we about our estimate?

To provide an estimate of how certain we are about our estimate, we can compute a *confidence interval*.

1. Choose a confidence level (e.g. 0.9, 0.95, 0.99)
2. Compute the quantile of the Normal distribution that corresponds to that confidence level (symmetric)
3. Multiply the value of this quantile by the standard error of the quantity, then add and subtract this value from the sample mean

Confidence interval example

Obtain a 90 percent confidence interval of Obama vote share in Florida (2008)

$$\alpha = 1 - 0.9 = 0.1$$

Confidence interval example

Obtain a 90 percent confidence interval of Obama vote share in Florida (2008)

$$\alpha = 1 - 0.9 = 0.1$$

Because a Normal is symmetric, we want to obtain all of the probability mass of the Normal curve above the 0.05 quantile and below the 0.95 quantile, or $0 + \alpha/2$ and $1 - \alpha/2$.

Confidence interval example

Obtain a 90 percent confidence interval of Obama vote share in Florida (2008)

$$\alpha = 1 - 0.9 = 0.1$$

Because a Normal is symmetric, we want to obtain all of the probability mass of the Normal curve above the 0.05 quantile and below the 0.95 quantile, or $0 + \alpha/2$ and $1 - \alpha/2$.

On the z score $N(0, 1)$ scale, this corresponds to

```
qnorm(0.05, 0, 1)
```

```
## [1] -1.644854
```

```
qnorm(0.95, 0, 1)
```

```
## [1] 1.644854
```

Confidence interval example

Our 90% confidence interval is equal to:

$$\bar{x} \pm z_{\alpha/2} \times SE$$

```
mean(fl$Obama) - 1.64*sd(fl$Obama)
```

```
## [1] 40.66452
```

```
## and
```

```
mean(fl$Obama) + 1.64*sd(fl$Obama)
```

```
## [1] 51.70168
```

Increase the confidence level to 0.95

Now we need the z-score for a Normal CDF at 0.975 and 0.025, and obtain our interval as before

```
## 90 percent interval
```

```
probs<-qnorm(c(0.05, 0.95), 0, 1)  
mean(fl$Obama) + probs*sd(fl$Obama)
```

```
## [1] 40.64819 51.71801
```

```
## 95 percent interval
```

```
probs<-qnorm(c(0.025, 0.975), 0, 1)  
mean(fl$Obama) + probs*sd(fl$Obama)
```

```
## [1] 39.58784 52.77835
```

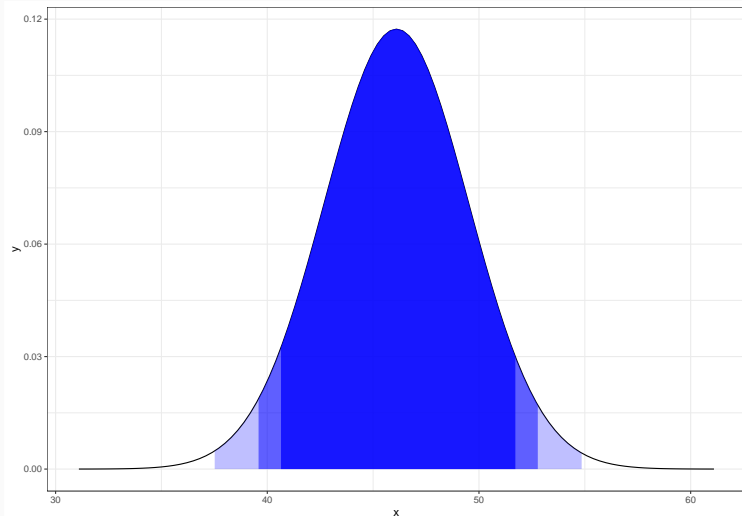
```
## 99 percent interval
```

```
probs<-qnorm(c(0.005, 0.995), 0, 1)  
mean(fl$Obama) + probs*sd(fl$Obama)
```

```
## [1] 37.51547 54.85073
```

What do you notice?

Confidence intervals for $\alpha = 0.1, 0.05, 0.01$



Interpreting confidence intervals

What does this interval mean?

```
## 95 percent interval
```

```
probs<-qnorm(c(0.025, 0.975), 0, 1)  
mean(fl$Obama) + probs*sd(fl$Obama)
```

```
## [1] 39.58784 52.77835
```


Interpreting confidence intervals

What does this interval mean?

```
## 95 percent interval
```

```
probs<-qnorm(c(0.025, 0.975), 0, 1)  
mean(fl$Obama) + probs*sd(fl$Obama)
```

```
## [1] 39.58784 52.77835
```

It means that if we draw another sample from this same population, the sampled value will fall within the computed interval 95 percent of the time.

Interpreting confidence intervals

What does this interval mean?

```
## 95 percent interval
```

```
probs<-qnorm(c(0.025, 0.975), 0, 1)  
mean(fl$Obama) + probs*sd(fl$Obama)
```

```
## [1] 39.58784 52.77835
```

It means that if we draw another sample from this same population, the sampled value will fall within the computed interval 95 percent of the time.

It does *not* mean that probability that the true value of the population mean falls within the interval is 0.95.

Interpreting confidence intervals

What does this interval mean?

```
## 95 percent interval
```

```
probs<-qnorm(c(0.025, 0.975), 0, 1)  
mean(fl$Obama) + probs*sd(fl$Obama)
```

```
## [1] 39.58784 52.77835
```

It means that if we draw another sample from this same population, the sampled value will fall within the computed interval 95 percent of the time.

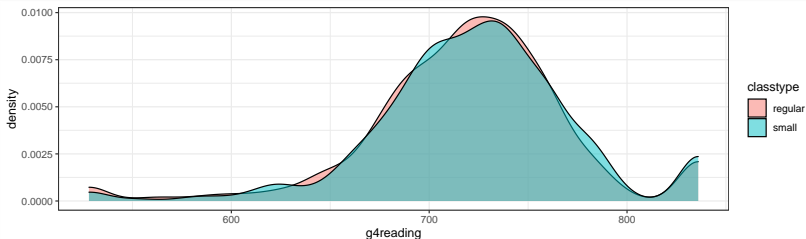
It does *not* mean that probability that the true value of the population mean falls within the interval is 0.95.

It does mean if we draw a large number of samples and construct 95 percent intervals, they will contain the population mean 95 percent of the time.

Confidence intervals for effect sizes in experiments

Let's return to the class-size experiment. We can use a confidence interval to show how uncertain our effect estimates are.

```
data(STAR)
STAR<-STAR %>%
  mutate(classtype = ifelse(classtype==1, "small",
                             ifelse(classtype==2, "regular",
                                     NA))) %>%
  filter(!is.na(classtype), !is.na(g4reading))
ggplot(STAR, aes(x=g4reading, fill = classtype)) +
  geom_density(alpha =0.5)
```



Confidence intervals for each group

```
STAR_groups<-STAR %>%  
  group_by(classtype) %>%  
  summarise(xbar_g4reading = mean(g4reading),  
            students = n(),  
            sd = sd(g4reading),  
            se_g4reading = sd(g4reading)/sqrt(n()))
```

And 95% Confidence Intervals for each group

```
alpha = 0.05  
STAR_groups %>%  
  mutate(ci_lower = xbar_g4reading - qnorm(1-alpha/2, 0, 1)*se_g4reading,  
         ci_upper = xbar_g4reading + qnorm(1-alpha/2, 0, 1)*se_g4reading)
```



```
## # A tibble: 2 x 7  
##   classtype xbar_g4reading students    sd se_g4reading ci_lower ci_upper  
##   <chr>      <dbl>      <int> <dbl>    <dbl>    <dbl>    <dbl>  
## 1 regular      720.        836  53.2     1.84     716.     723.  
## 2 small       723.        726  51.5     1.91     720.     727.
```

Confidence interval for the SATE

```
## sample average treatment effect
```

```
SATE<-STAR_groups$xbar_g4reading[2]-STAR_groups$xbar_g4reading[1]
```

```
SATE
```

```
## [1] 3.501232
```

Confidence interval for the SATE

```
## sample average treatment effect
```

```
SATE<-STAR_groups$xbar_g4reading[2]-STAR_groups$xbar_g4reading[1]
```

```
SATE
```

```
## [1] 3.501232
```

```
## SE of the SATE
```

```
SATE_se<-sqrt(STAR_groups$se_g4reading[1]^2 + STAR_groups$se_g4reading[2]^2)
```

```
SATE_se
```

```
## [1] 2.653485
```

Confidence interval for the SATE

```
## sample average treatment effect
SATE<-STAR_groups$xbar_g4reading[2]-STAR_groups$xbar_g4reading[1]
SATE

## [1] 3.501232

## SE of the SATE
SATE_se<-sqrt(STAR_groups$se_g4reading[1]^2 + STAR_groups$se_g4reading[2]^2)
SATE_se

## [1] 2.653485

### 95 percent CI
SATE+qnorm(1-alpha/2, 0, 1) *SATE_se

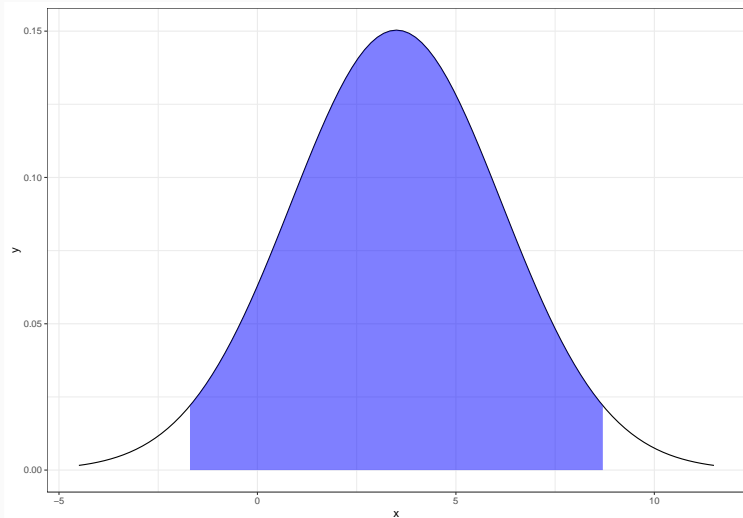
## [1] 8.701968

SATE-qnorm(1-alpha/2, 0, 1) *SATE_se

## [1] -1.699503
```

This interval is pretty wide. It suggests we don't have much certainty about whether the treatment has an effect. $[-1.7, 8.7]$.

The estimated sampling distribution of the SATE, with 95 percent interval



We would like to know whether assignment to a small class in Kindergarten improves reading scores in fourth grade.

We would like to know whether assignment to a small class in Kindergarten improves reading scores in fourth grade.

We start by proposing two hypotheses:

- H_0 : Kindergarten class assignment has no effect on fourth grade reading scores
- H_1 : Kindergarten class assignment has an effect on fourth grade reading scores

We will use the sampling distribution of the SATE to test this hypothesis

1. Specify a null and alternative hypothesis

General approach

1. Specify a null and alternative hypothesis
2. Choose a test statistic and level of α (conventionally, 0.05)

General approach

1. Specify a null and alternative hypothesis
2. Choose a test statistic and level of α (conventionally, 0.05)
3. Derive the sampling distribution of the test statistic under the null hypothesis

General approach

1. Specify a null and alternative hypothesis
2. Choose a test statistic and level of α (conventionally, 0.05)
3. Derive the sampling distribution of the test statistic under the null hypothesis
4. Compute a p-value (the probability that, under the null hypothesis, we would observe a result as extreme as we observed in the data)

General approach

1. Specify a null and alternative hypothesis
2. Choose a test statistic and level of α (conventionally, 0.05)
3. Derive the sampling distribution of the test statistic under the null hypothesis
4. Compute a p-value (the probability that, under the null hypothesis, we would observe a result as extreme as we observed in the data)
5. Reject the null hypothesis if $p \leq \alpha$, otherwise retain the null hypothesis

For the STAR experiment

$$H_0 : \mu_{small} - \mu_{regular} = 0$$

For the STAR experiment

$$H_0 : \mu_{\text{small}} - \mu_{\text{regular}} = 0$$

$$H_1 : \mu_{\text{small}} - \mu_{\text{regular}} \neq 0$$

For the STAR experiment

$$H_0 : \mu_{small} - \mu_{regular} = 0$$

$$H_1 : \mu_{small} - \mu_{regular} \neq 0$$

- Under the null hypothesis, we assume that the SATE has a Normal distribution with mean 0 and standard deviation equal to the observed standard error of the SATE

For the STAR experiment

$$H_0 : \mu_{small} - \mu_{regular} = 0$$

$$H_1 : \mu_{small} - \mu_{regular} \neq 0$$

- Under the null hypothesis, we assume that the SATE has a Normal distribution with mean 0 and standard deviation equal to the observed standard error of the SATE
- Let's go with the conventional $\alpha = 0.05$ threshold (there's nothing special about it though!)

For the STAR experiment

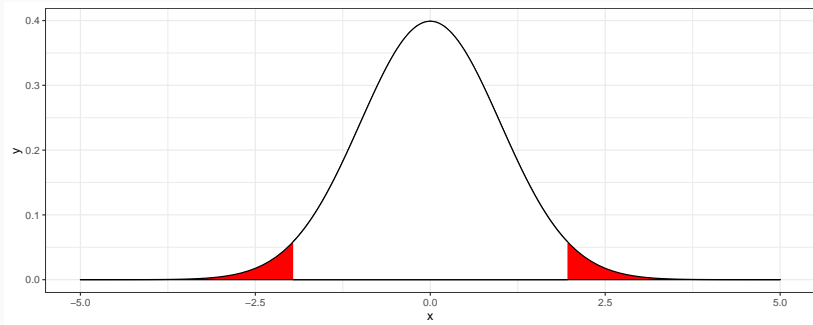
$$H_0 : \mu_{small} - \mu_{regular} = 0$$

$$H_1 : \mu_{small} - \mu_{regular} \neq 0$$

- Under the null hypothesis, we assume that the SATE has a Normal distribution with mean 0 and standard deviation equal to the observed standard error of the SATE
- Let's go with the conventional $\alpha = 0.05$ threshold (there's nothing special about it though!)
- Then we estimate the probability that we could observe the data under the null hypothesis using the Normal Cumulative Distribution Function (pnorm in R)

Rejection region: The sampling distribution of the SATE under the null hypothesis

We will reject the null hypothesis if our test statistic falls in either of the red areas (two-tailed test):



Let's calculate our p-value

Because the Normal CDF is symmetric, we use the left-tail to estimate the probability that it falls within the rejection zone on either side

```
2* pnorm(-abs(SATE), 0, SATE_se)
```

```
## [1] 0.1870072
```

Can we reject H_0 ?

Making it easier on ourselves

R has built in functions for significance tests (t-tests)

```
STAR_small<-STAR %>%
  filter(classtype=="small")
STAR_regular<-STAR %>%
  filter(classtype=="regular")
t.test(STAR_small$g4reading,
       STAR_regular$g4reading)

##
##  Welch Two Sample t-test
##
## data: STAR_small$g4reading and STAR_regular$g4reading
## t = 1.3195, df = 1541.2, p-value = 0.1872
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.703591  8.706055
## sample estimates:
## mean of x mean of y
##  723.3912  719.8900
```


- Hypothesis tests and confidence intervals are tricky to interpret, and are often interpreted incorrectly

Words of warning

- Hypothesis tests and confidence intervals are tricky to interpret, and are often interpreted incorrectly
- Hypothesis testing has become a *very* problematic feature of modern social science

Words of warning

- Hypothesis tests and confidence intervals are tricky to interpret, and are often interpreted incorrectly
- Hypothesis testing has become a *very* problematic feature of modern social science
- Just because a result has a p-value of less than 0.05 does not mean that a finding is true!

Words of warning

- Hypothesis tests and confidence intervals are tricky to interpret, and are often interpreted incorrectly
- Hypothesis testing has become a *very* problematic feature of modern social science
- Just because a result has a p-value of less than 0.05 does not mean that a finding is true!
- The routines of modern science exacerbate these problems

Words of warning

- Hypothesis tests and confidence intervals are tricky to interpret, and are often interpreted incorrectly
- Hypothesis testing has become a *very* problematic feature of modern social science
- Just because a result has a p-value of less than 0.05 does not mean that a finding is true!
- The routines of modern science exacerbate these problems

If we test 10 hypotheses in an analysis, the likelihood that at least one will be significant is:

```
1 - 0.95^10
```

```
## [1] 0.4012631
```

- For homework, you will compute confidence intervals and conduct significance tests using the polls08 data. The assignment will be posted in HW9 on slack
- Lab will review how to compute standard errors, confidence intervals, and significance tests

Lab

Returning to the Pager data

```
dat<-read_csv("../data/criminalrecord.csv")
head(dat)
```

```
## # A tibble: 6 x 9
##   jobid callback black crimrec interact city distance custserv manualskill
##   <dbl>    <dbl> <dbl>   <dbl>   <dbl> <dbl>    <dbl>    <dbl>    <dbl>
## 1   108        1     0       1       1     0      15        1        0
## 2   113        0     0       0       1     0      20        0        1
## 3   101        1     0       0       0     0      15        1        0
## 4    64        1     0       0       0     1       7        1        0
## 5    33        0     0       1       0     1       5        1        0
## 6    73        0     0       1       0     1      10        0        1
```


Compute the sample average callback rate

- For Black and white testers

```
bw<-dat %>%  
  group_by(black) %>%  
  summarise(callback = mean(callback))
```

- For those assigned to criminal record treatments

```
cr<-dat %>%  
  group_by(crimrec) %>%  
  summarise(xbar = mean(callback))
```

- For criminal record treatments for both Black and white testers

```
bw_cr<-dat %>%  
  group_by(black, crimrec) %>%  
  summarise(xbar = mean(callback))
```

Now compute the standard error for each group

- For Black and white testers

```
bw<-dat %>%  
  group_by(black) %>%  
  summarise(xbar = mean(callback),  
            se = sd(callback)/sqrt(n()))
```

- For those assigned to criminal record treatments

```
cr<-dat %>%  
  group_by(crimrec) %>%  
  summarise(xbar = mean(callback),  
            se = sd(callback)/sqrt(n()))
```

- For criminal record treatments for both Black and white testers

```
bw_cr<-dat %>%  
  group_by(black, crimrec) %>%  
  summarise(xbar = mean(callback),  
            se = sd(callback)/sqrt(n()))
```

Now compute critical values for confidence intervals for each group

- At the 90% level

```
## critical values are (1 - alpha)/2
## 95 percent (1 - 0.95)/2
q_90<-qnorm(0.95, mean = 0, sd = 1)
q_95<-qnorm(0.975, mean = 0, sd = 1)
q_99<-qnorm(0.995, 0, 1)

bw<-dat %>%
  group_by(black) %>%
  summarise(xbar = mean(callback),
            se = sd(callback)/sqrt(n()))

bw %>%
  mutate(ci_90_lwr = xbar - se * q_90,
         ci_90_upr = xbar + se * q_90,
         ci_95_lwr = xbar - se * q_95,
         ci_95_upr = xbar + se * q_95,
         ci_99_lwr = xbar - se * q_99,
         ci_99_upr = xbar + se * q_99)

## # A tibble: 2 x 9
##   black  xbar      se ci_90_lwr ci_90_upr ci_95_lwr ci_95_upr ci_99_lwr
##   <dbl> <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1     0 0.253  0.0252    0.212    0.295    0.204    0.303    0.189
## 2     1 0.0960 0.0148    0.0716   0.120   0.0669   0.125   0.0578
## # ... with 1 more variable: ci_99_upr <dbl>
```

Now compute a significance test at the 0.05 level for this hypothesis

Where w indicates white testers and b indicates Black testers

$$\bullet H_0 : \mu_b = \mu_w$$

note, this is equivalent to $H_0 : \mu_b - \mu_w = 0$

```
dat_w<-dat %>%  
  filter(black==0)  
dat_b <- dat %>%  
  filter(black==1)  
  
t_test<-t.test(dat_w$callback, dat_b$callback)  
  
sate <- bw$xbar[1] - bw$xbar[2]  
se <- sqrt(bw$se[1]^2 - bw$se[2]^2)  
  
2 * pnorm(-abs(sate), 0, se)  
  
## [1] 9.647234e-15
```