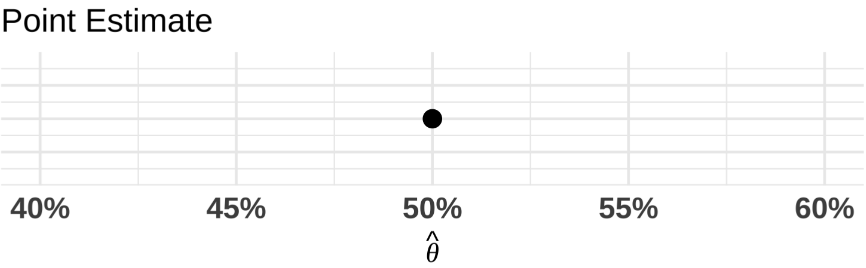
STAT 339: Statistical Theory

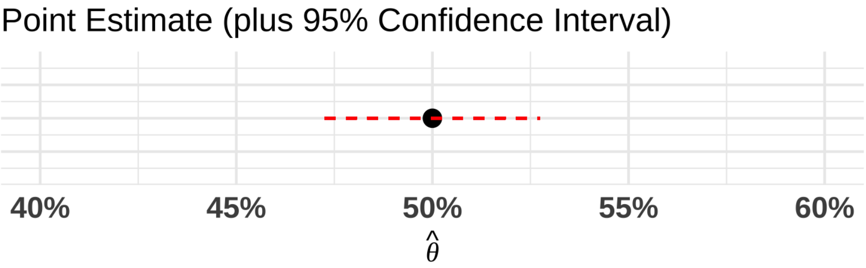
Frequentist Interval Estimation

Anthony Scotina

Pivotal Quantities

What's better than a point estimate?





Confidence Interval

Ideally, an interval estimate (or, **confidence interval**) will…

1. contain the *target parameter*, θ
2. be relatively narrow

A $(1-\alpha) \times 100\%$ **confidence interval** is an interval $[\hat{\theta}_L, \hat{\theta}_U]$ such that

$$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha,$$

where $1-\alpha$ is the **confidence coefficient**.

- If $1-\alpha$ is *high*, we can be highly *confident* that the confidence interval (based on a *single sample*) will contain θ.

🚨**Note**: $\hat{\theta}_L$ and $\hat{\theta}_U$ are *random variables* and functions of the *data* $Y_1, \ldots, Y_n$.

Pivotal Quantities

To find $\hat{\theta}_L$ and $\hat{\theta}_U$, it is useful to *first* find a **pivotal quantity**, V, for θ:

1. V is a function of the sample measurements $Y_1, \ldots, Y_n$ and the unknown parameter θ, where θ is the *only* unknown quantity.

2. Its probability distribution does *not* depend on θ.

We will primarily work with **location-scale** pivotal quantities. For example,

$$V = \frac{g(\mathbf{Y})}{\theta}$$

is a **scale** pivotal quantity, and

$$P(a \leq V \leq b) = P\left(a \leq \frac{g(\mathbf{Y})}{\theta} \leq b\right) = P\left(\frac{g(\mathbf{Y})}{b} \leq \theta \leq \frac{g(\mathbf{Y})}{a}\right) = 1 - \alpha$$

Gamma Pivot

Suppose $Y_1, \ldots, Y_n \sim iid\ Exponential(\theta)$.

Facts from STAT 338:

- $\sum_{i=1}^{n} Y_i = Y_1 + \cdots + Y_n \sim Gamma(n, \theta)$
- $V = 2(Y_1 + \cdots + Y_n)/\theta \sim \chi^2(df = 2n)$

We can use V as a **pivotal quantity**, because…

1. V is a function of the **DATA**, $Y_1, \ldots, Y_n$ and the *unknown parameter* θ.
2. The probability distribution of V, $\chi^2(2n)$, *does not depend on* θ.

Using the Gamma Pivot

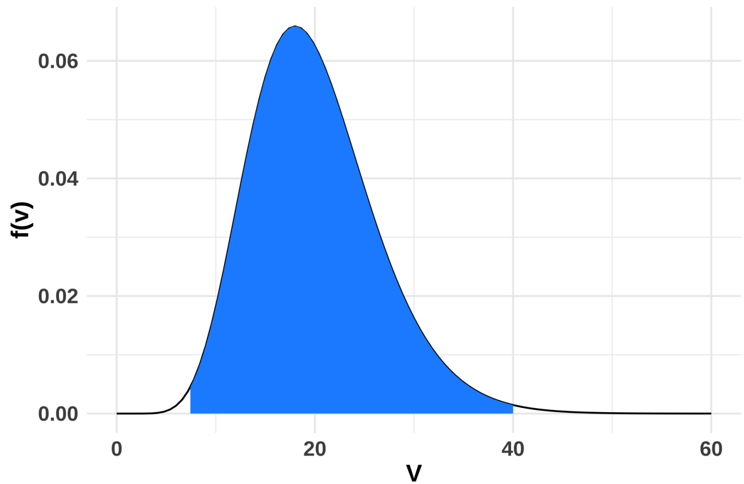
We will use $V = 2(Y_1 + \cdots + Y_n)/\theta \sim \chi^2(df = 2n)$ as a *pivotal quantity* in forming a *confidence interval* for θ.

- But now, suppose that $n = 10$. In other words, $V \sim \chi^2(20)$.

  > Let's find a 99% confidence interval for θ.

Want: $P(a \leq V \leq b) = 0.99$

- Use *quantiles* for the *middle* 99%: $\chi^2_{0.005}(20) = 7.434$; $\chi^2_{0.995}(20) = 39.997$



Using the Gamma Pivot to Form a CI!

Want: $P(a \leq V \leq b) = 0.99$

- Now we **have** $P(7.434 \leq V \leq 39.997) = 0.99$! 👀

This means we can write out

$$P\left(7.434 \leq \frac{2\sum_{i=1}^{n} Y_i}{\theta} \leq 39.997\right) = 0.99,$$

and then *rearrange* so that θ is alone in the middle.

- This gives us a 99% **confidence interval** *estimate* for θ.

Large Sample Confidence Intervals

Common Unbiased Point Estimators

(From WMS, page 397)

Target Parameter θ	Sample Size(s)	Point Estimator $\hat{\theta}$	$E(\hat{\theta})$	Standard Error $\sigma_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\dfrac{\sigma}{\sqrt{n}}$
p	n	$\hat{p} = \dfrac{Y}{n}$	p	$\sqrt{\dfrac{pq}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}^{*\dagger}$
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}^{\dagger}$

🚨**Note**: The standard deviation of the *sampling distribution* for $\hat{\theta}$ is usually called the **standard error** of $\hat{\theta}$, $SE(\hat{\theta}) = \sigma_{\hat{\theta}}$.

Revisiting the CLT

The **Central Limit Theorem** can be used to show that each of these four point estimators have *approximately Normal* distributions for large samples.

- In other words, for each of these four point estimators,

$$\hat{\theta} \sim N(\theta, \sigma^2_{\hat{\theta}}) \quad \text{or} \quad \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \sim N(0, 1)$$

Because $Z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}$ has these characteristics:

1. It is a function of the sample measurements $Y_1, \ldots, Y_n$ and the unknown parameter θ, and...

2. Its probability distribution, $N(0, 1)$, does *not* depend on θ,

it follows that $Z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}}$ is a **pivotal quantity** and we can use it to build **confidence intervals** for θ.

- (when the sample is *large*)

Building a Large Sample CI

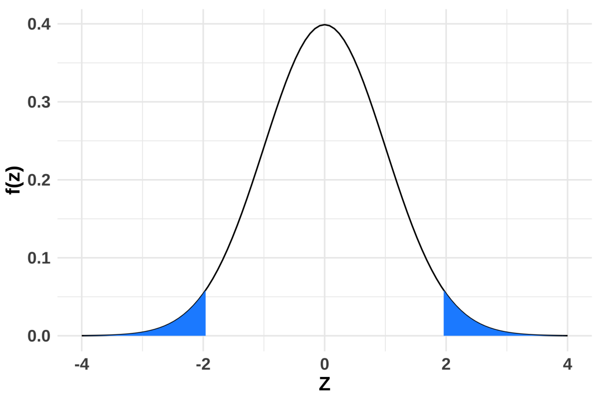
Want: $P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U) = 1 - \alpha$

- Because $Z = \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \sim N(0,1)$, and the Normal distribution is **symmetric**, we can use the following:

$$P\left(-z_{\alpha/2} \leq \frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}\right) = 1 - \alpha,$$

  where $z_{\alpha/2}$ is a Normal distribution *quantile*.

Example: If $1-\alpha = 0.95$, $\alpha = 0.05$ and $z_{\alpha/2} = 1.96$.



A Large Sample CI

The endpoints for a $100(1-\alpha)\%$ **confidence interval** for θ are given by:

- $\hat{\theta}_L = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}$
- $\hat{\theta}_U = \hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}$

🤔How can we find $z_{\alpha/2}$? Use ~~table in back of textbook~~ **R**!

- Specifically, `qnorm(...)`

Example: 90% CI $\implies 1-\alpha = 0.90 \implies \alpha/2 = 0.05$

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

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

Back to Piéchart Emporium (yet again)

Piéchart
Emporium

Goal 👇👇👇

> Construct a 90% confidence interval for μ, the true average shopping time (in mins) per customer.

The Data 📊

- $n = 144$ randomly selected customers
  - $\bar{y} = 30$ mins
  - $s = 12$ mins

A Large-Sample CI for μ

The endpoints for the $100(1-\alpha)\%$ **confidence interval** for θ can be obtained by:

$$\hat{\theta} \pm z_{\alpha/2}\sigma_{\hat{\theta}}$$

In this example...

- $\hat{\theta} = \bar{y} = 30$
- $1 - \alpha = 0.90 \implies \alpha/2 = 0.05$
- $\sigma_{\hat{\theta}} = \sigma/\sqrt{n} = \sigma/\sqrt{144} =$ 🤔

The *sample standard deviation*, S, is **consistent** for σ. So in **large samples**,

$$\bar{y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \approx \bar{y} \pm z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right).$$

One-sided CIs

Using similar derivations, we can also determine $100(1-\alpha)\%$ **one-sided confidence limits**:

- $100(1-\alpha)\%$ *lower bound* for θ is $\hat{\theta}_L = \hat{\theta} - z_\alpha \sigma_{\hat{\theta}}$
- $100(1-\alpha)\%$ *upper bound* for θ is $\hat{\theta}_U = \hat{\theta} + z_\alpha \sigma \hat{\theta}$

These limits satisfy the following:

- $P(\hat{\theta}_L \leq \theta) = 1 - \alpha$, with CI $[\hat{\theta}_L, \infty)$
- $P(\theta \leq \hat{\theta}_U) = 1 - \alpha$, with CI $(-\infty, \hat{\theta}_U]$

> Let's construct a 90% *lower confidence bound* for μ, the true average shopping time (in mins) per customer

A Large-Sample CI for p

In a poll of 1001 adults, 51% claim to be baseball fans.

> Find a 99% confidence interval for p, the true propotion of baseball fans.

Think of the 1001 adults as the random sample $Y_1, Y_2, \dots, Y_{1001} \sim iid\ Bernoulli(p)$.

- $\hat{\theta} = \hat{p} = \frac{1}{n}\sum_{i=1}^{1001} Y_i = 0.51$
- $Var(Y) = p(1-p) \implies Var(\hat{p}) = \frac{p(1-p)}{n} =$ 🤔

Putting this together, a large-sample $100(1-\alpha)\%$ confidence interval for p is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Interpreting Frequentist Interval Estimates

Be *very careful* when interpreting confidence intervals! There are some *tempting* interpretations that are wrong:

Wrong

> There is a probability of 0.90 that the true average shopping time μ per customer is between 28.36 mins and 31.65 mins.

- μ is either in *this* interval, or it isn't!
- In order to calculate the **probability** that μ lies in **this interval** $(28.36, 31.65)$, we need to first assign a *prior* distribution to μ, and then use the resulting *posterior distribution*.

Correct

> We are *90% confident* that thr true average shopping time μ per customer is between 28.36 mins and 31.65 mins.

- In this context, we are using the phrase "*90% confident*", rather than "*90% probability*".

Interpreting Frequentist Interval Estimates

We *can* attribute **probabilistic** interpretations to the *unobserved* endpoints,

$$(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}})$$

- Before these are *observed*, they are **random variables** and functions of $Y_1, \ldots, Y_n$.

> If the *same procedure* were used many times on different random samples $Y_1, \ldots, Y_n$, then approximately $100(1-\alpha)\%$ of the resulting intervals will contain θ.

Simulation Check

```r
p = 0.5
n = 1000

# Construct 95% CI
# and check whether it contains p = 0.5
coverage = replicate(10000, {
  # Collect sample
  sample_data = sample(0:1, size = n, replace = TRUE)
  p_hat = mean(sample_data)

  # Calculate CI
  theta_L = p_hat - 1.96*sqrt(p_hat*(1-p_hat)/n)
  theta_U = p_hat + 1.96*sqrt(p_hat*(1-p_hat)/n)

  # Check coverage
  (theta_L <= 0.5) & (theta_U >= 0.5)
})
mean(coverage)
```

```
## [1] 0.9467
```

Small-Sample Confidence Intervals

(for means)

Introduction

In the **large-sample** CIs that we constructed previously, we assumed the following:

The sample was sufficiently large!

- And we had *no restrictions* on the distribution of the sample data, $Y_1, \ldots, Y_n$.

The confidence intervals (for μ) that we will discuss now *require* that the sample has been randomly selected from a **normal** population.

- In other words, we'll **assume** that $Y_1, \ldots, Y_n \sim iid\ N(\mu, \sigma^2)$.
  - While we cannot truly *know* the true distribution of the Y_i, the following procedures will work well, *as long as departures from normality are not extreme.*

Setting

- $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$
- $\bar{Y}$ is the *sample mean*, and S^2 is the *sample variance*.

The t-distribution

Suppose we are performing inference for μ under the following scenario:

- $Var(Y_i) = \sigma^2$ is *unknown*.
- The sample size is **small**, so we can't use the *large-sample* techniques discussed previously.

Recall: If $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$, then

$$T = \frac{\bar{Y} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Notice that:

- T is a function of *only* the **sample measurements** and the **unknown parameter**, μ.
- The distribution of T *does not* depend on μ.

Therefore, we can use T as a **pivotal quantity** in constructing a *confidence interval* for μ.

A Small-Sample CI for μ

Using T as a pivotal quantity, a (small-sample) $100(1-\alpha)\%$ CI for μ is

$$\bar{Y} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$$

- Note that $t_{\alpha/2}$ also depends on the **degrees of freedom**, $n-1$.

This CI is also valid for *large samples*.

- But what happens to the *t*-distribution for **large** n?

$$P(-t_{\alpha/2} \leq T \leq t_{\alpha/2}) \approx P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$$

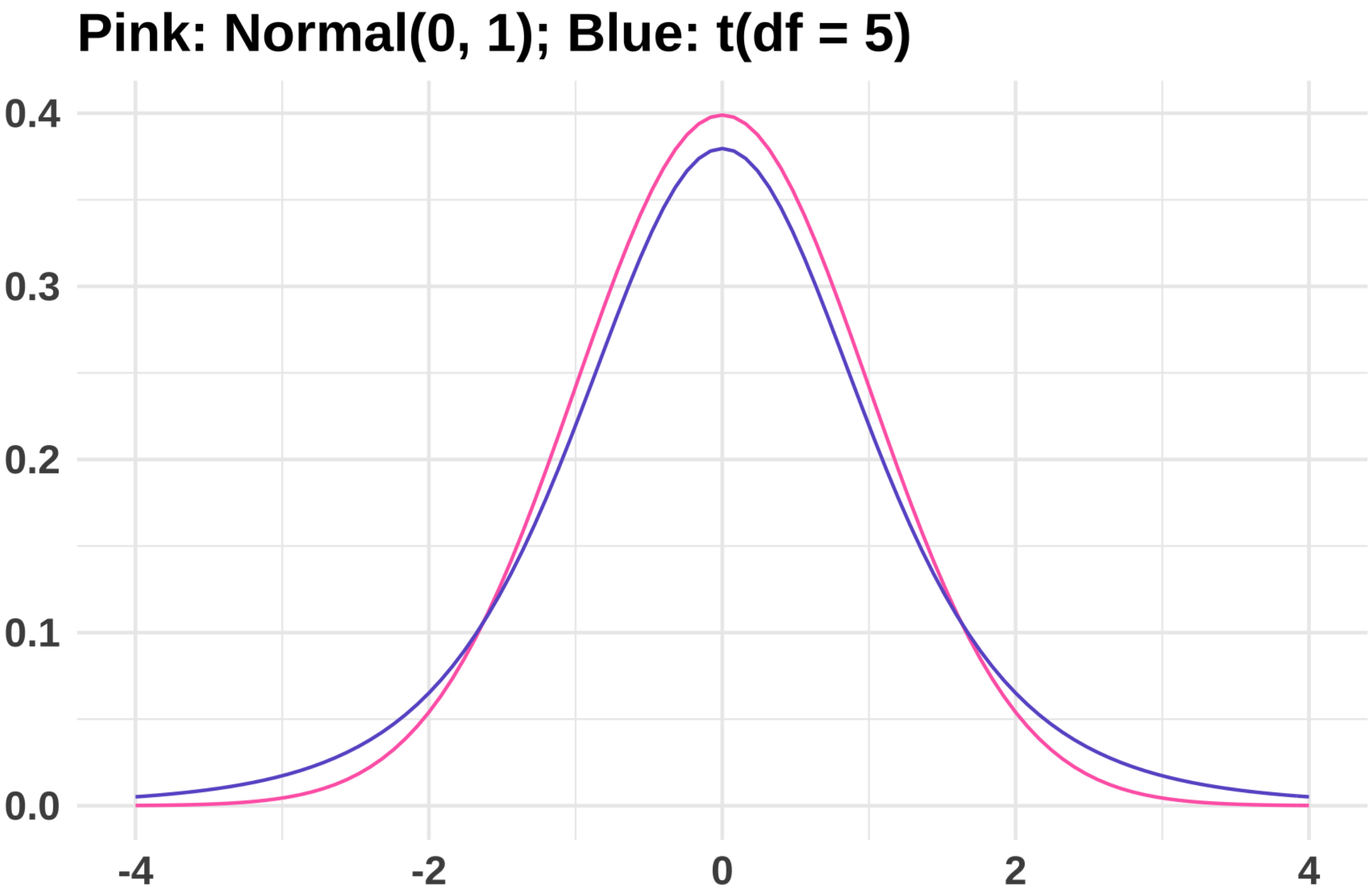
```
qnorm(0.975)
```

```
## [1] 1.959964
```

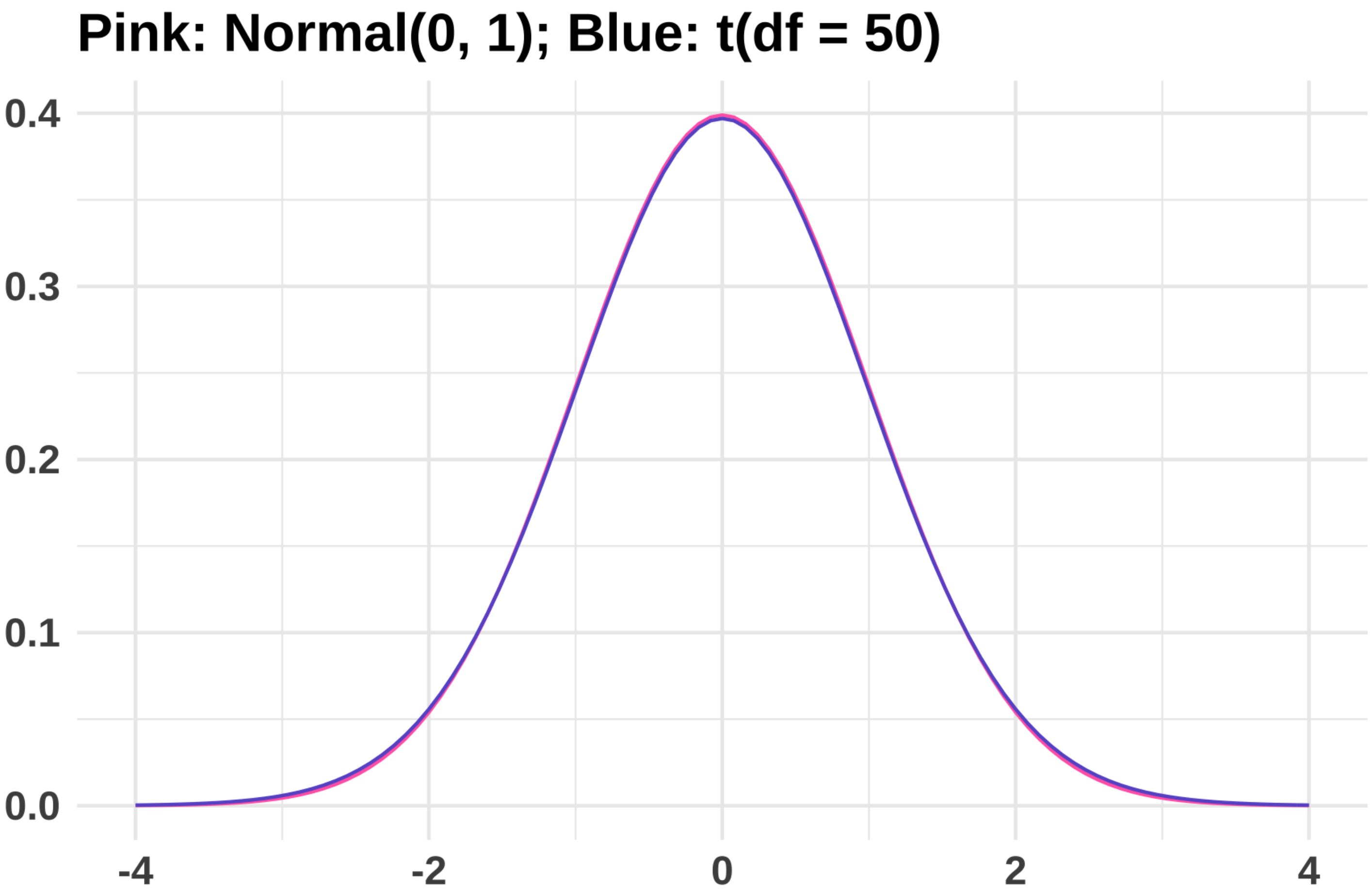
```
qt(0.975, df = c(5, 50, 500))
```

```
## [1] 2.570582 2.008559 1.964720
```

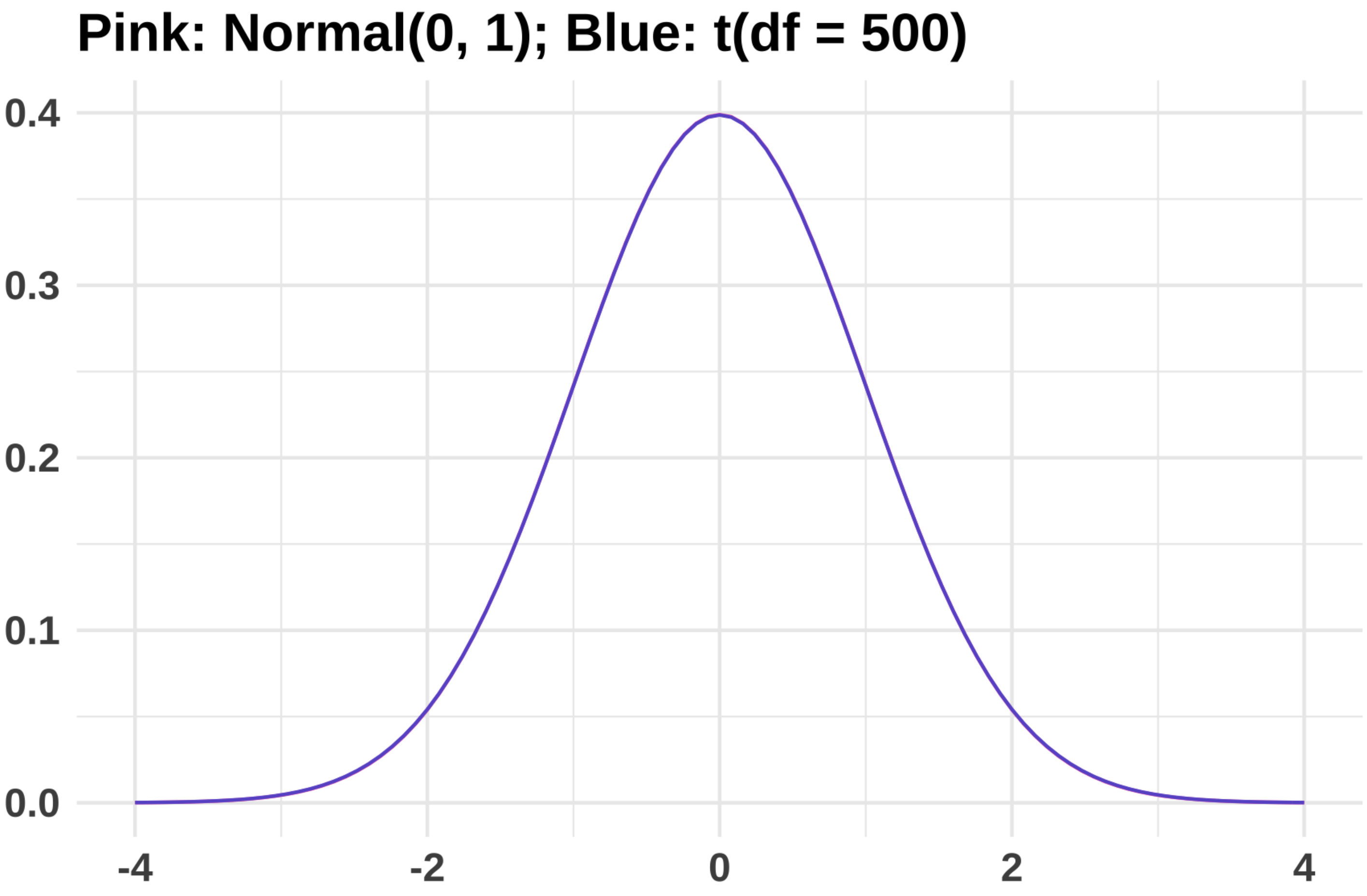
Student's t vs. Normal



Student's t vs. Normal



Student's t vs. Normal



The Stroop Test

The **Stroop Effect** describes the psychological phenomenon that occurs when the processing of one particular stimulus feature interferes with the simultaneous processing of a second stimulus feature.

- Take the **Stroop Test** here!

A random sample of $n = 8$ study participants yielded the following *reaction times* (in seconds per hundred reactions):

- 95, 99, 106, 107, 107, 114, 120, 127

Based on prior studies, it is reasonable to assume that reaction times are *normally distributed.*

> Construct a 95% CI for μ, the population mean reaction time for the Stroop Test.

The Stroop Test

First, we need to calculate some **stats**:

```
stroop = c(95, 99, 106, 107, 107, 114, 120, 127)
mean(stroop) # Y_bar
```

```
## [1] 109.375
```

```
sd(stroop) # S
```

```
## [1] 10.56865
```

⭐ How does this interval change if we use *large-sample* methods?

- We're really just changing $t_{\alpha/2}$ to $z_{\alpha/2}$:

  ```
  c(qt(0.975, df = 7), qnorm(0.975))
  ```

  ```
  ## [1] 2.364624 1.959964
  ```

A CI for μ_1 - μ_2

Suppose that we are interested in *comparing* the means of **two** normal populations:

- **Population 1**: Mean μ_1 and variance σ_1^2
- **Population 2**: Mean μ_2 and variance σ_2^2

We can construct a CI for $\mu_1 - \mu_2$ based on the *t*-distribution by making two additional assumptions:

- The two *samples* are independent from one another.
- The two *populations* have a common but *unknown* variance, $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

A CI for μ$_1$ - μ$_2$

If $Y_{11}, \ldots, Y_{1n_1} \sim N(\mu_1, \sigma^2)$ and $Y_{21}, \ldots, Y_{2n_2} \sim N(\mu_2, \sigma^2)$,

- $Var(\bar{Y}_1 - \bar{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$
- $Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1)$

🤔 But σ^2 is unknown! What do we do?

Use the **pooled variance estimator**, S_p^2:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

The Pooled Variance Estimator

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

We can think of the *pooled variance estimator*, S_p^2 as the **weighted average** of S_1^2 and S_2^2.

- If $n_1 = n_2$, then S_p^2 is simply the *average* of S_1^2 and S_2^2.
- If $n_1 \neq n_2$, then S_p^2 gives *larger weight* to the sample variance associated with the *larger sample size.*

Further,

$$W = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2} \sim \chi^2(df = n_1 + n_2 - 2),$$

so

$$T = \frac{Z}{\sqrt{W/df}} \sim t(df = n_1 +_2 -2)$$

A CI for μ1 - μ2

Because

$$T = \frac{Z}{\sqrt{W/df}} = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

is a **pivotal quantity**, we can use it to derive a *confidence interval* for $\mu_1 - \mu_2$.

A $(1 - \alpha) \times 100\%$ **confidence interval** for $\mu_1 - \mu_2$ is

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where $t_{\alpha/2}$ is determined from the *t*-distribution with $df = n_1 + n_2 - 2$.

Piéchart Emporium's Rival

Piéchart Emporium's rival on the other side of town is *Bärchart Marketplace*.

Goal 👇👇👇

> Construct a 95% confidence interval for $\mu_1 - \mu_2$, the *difference* in average daily revenue (in USD) between Piéchart Emporium and Bärchart Marketplace.

Piéchart Emporium's Rival

Goal 👇👇👇

> Construct a 95% confidence interval for $\mu_1 - \mu_2$, the *difference* in average daily revenue (in USD) between Piéchart Emporium and Bärchart Marketplace.

Piéchart Emporium😨

- $n_p = 10$ days
  - $\bar{y}_p = 875$ dollars
  - $s_p = 80$ dollars

Bärchart Marketplace📊

- $n_b = 10$ days
  - $\bar{y}_b = 960$ dollars
  - $s_p = 90$ dollars

Confidence Intervals for the Variance

Introduction

The population **variance**, σ^2 quantifies the amount of *variability* in the population.

- While σ^2 is often *unknown*, we've shown that

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

  is an **unbiased estimator** for σ^2.

We've used S^2 in the construction of confidence intervals for μ and $\mu_1 - \mu_2$, but now let's use it in the construction of confidence intervals for σ^2.

Want: $P(\chi^2_L \leq \sigma^2 \leq \chi^2_U) = 1 - \alpha$

Need: A **pivotal quantity**!

Sample Variance Sampling Distribution

Setting: $Y_1, Y_2, \ldots, Y_n \sim N(\mu, \sigma^2)$, where both μ and σ^2 are *unknown*.

- Earlier in the semester, we discussed that, when *scaled appropriately*, S^2 follows a $\chi^2(n-1)$ distribution. That is,

$$\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sim \chi^2(n-1).$$

- This is a *pivotal quantity*! It's a function of the *data* and σ^2, and its distribution does *not* depend on σ^2.

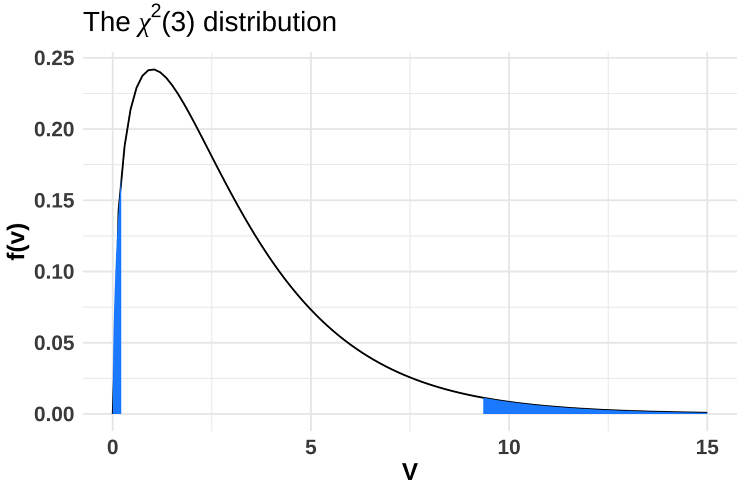
Therefore, we can derive a $(1-\alpha) \times 100\%$ **confidence interval** for σ^2 via:

$$P\left(\chi^2_L \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_U\right) = 1 - \alpha$$

A CI for σ^2

For χ^2_L and χ^2_U, we choose points that cut off *equal tail areas.*

- That is, $\chi^2_L = \chi^2_{\alpha/2}$ and $\chi^2_U = \chi^2_{1-(\alpha/2)}$.



A $(1-\alpha)\times 100\%$ **confidence interval** for σ^2 is given by

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{1-(\alpha/2)}}\right)$$

Truck Noise

(WMS 8.95)

The EPA has set a maximum noise level for heavy trucks at 83 decibels (dB). The manner in which this limit is applied will greatly affect the trucking industry and the public. One way to apply the limit is to require all trucks to conform to the noise limit.

A second but less satisfactory method is to require the truck fleet's mean noise level to be less than the limit. If the latter rule is adopted, variation in the noise level from truck to truck becomes important because a large value of σ^2 would imply that many trucks exceed the limit, even if the mean fleet level were 83 dB. A random sample of six heavy trucks produced the following noise levels (in decibels):

$$85.4 \quad 86.8 \quad 86.1 \quad 85.3 \quad 84.8 \quad 86.0$$

> Use these data to construct a 90% confidence interval for σ^2, the variance of the truck noise emission readings. Interpret your results.