Chapter 9: Inference for Variances

DSCC 462

Computational Introduction to Statistics

Anson Kahng

Fall 2022

Inference for Variances

Inference for Variances

- Primarily, inference tends to be focused on means (and proportions, which we will cover next)

Inference for Variances

- Primarily, inference tends to be focused on means (and proportions, which we will cover next)
- However, variances are essential to inference of means

Inference for Variances

- Primarily, inference tends to be focused on means (and proportions, which we will cover next)
- However, variances are essential to inference of means
- We may be interested in determining:

Inference for Variances

- Primarily, inference tends to be focused on means (and proportions, which we will cover next)
- However, variances are essential to inference of means
- We may be interested in determining:
  - Whether a population variance is equal to a predetermined value

Inference for Variances

- Primarily, inference tends to be focused on means (and proportions, which we will cover next)
- However, variances are essential to inference of means
- We may be interested in determining:
  - Whether a population variance is equal to a predetermined value
  - Whether two population variances are equal

Sampling Distribution of Variance

Sampling Distribution of Variance

- Recall that the sample variance of a sample $x_1, \ldots, x_n$ is defined as $s^2 = \dfrac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2$

Sampling Distribution of Variance

- Recall that the sample variance of a sample $x_1, \ldots, x_n$ is defined as $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$

- Thought experiment: Suppose we know the population mean μ

Sampling Distribution of Variance

- Recall that the sample variance of a sample $x_1, \ldots, x_n$ is defined as $s^2 = \dfrac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$
- Thought experiment: Suppose we know the population mean μ

$\approx$

- Rewrite the sample variance as (approximately) $s^2 = \dfrac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$

Sampling Distribution of Variance

- Recall that the sample variance of a sample $x_1, \ldots, x_n$ is defined as $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$
- Thought experiment: Suppose we know the population mean μ
- Rewrite the sample variance as (approximately) $\frac{s^2}{\sigma^2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \boxed{\frac{(x_i - \mu)^2}{\sigma^2}}$

  $\frac{x_i - \mu}{\sigma} \sim$
- Now, examine s^2/σ^2

Sampling Distribution of Variance

- Recall that the sample variance of a sample $x_1, \ldots, x_n$ is defined as $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$
- Thought experiment: Suppose we know the population mean μ
- Rewrite the sample variance as (approximately) $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$
- Now, examine s^2/σ^2
- We have $\frac{s^2}{\sigma^2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2$

χ^2_n

$\frac{x_i - \mu}{\sigma} \sim N(0,1)$

z

χ^2_1

Sampling Distribution of Variance

- Recall that the sample variance of a sample $x_1, \ldots, x_n$ is defined as $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$
- Thought experiment: Suppose we know the population mean μ
- Rewrite the sample variance as (approximately) $s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \mu)^2$
- Now, examine s^2/σ^2
- We have $\frac{s^2}{\sigma^2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2$
- Therefore, $\frac{s^2}{\sigma^2} = \frac{1}{n-1} \cdot \boxed{\sum_{i=1}^{n} Z_i^2}$, where Z_i is a standard normal random variable

Chi-squared (χ^2) distribution

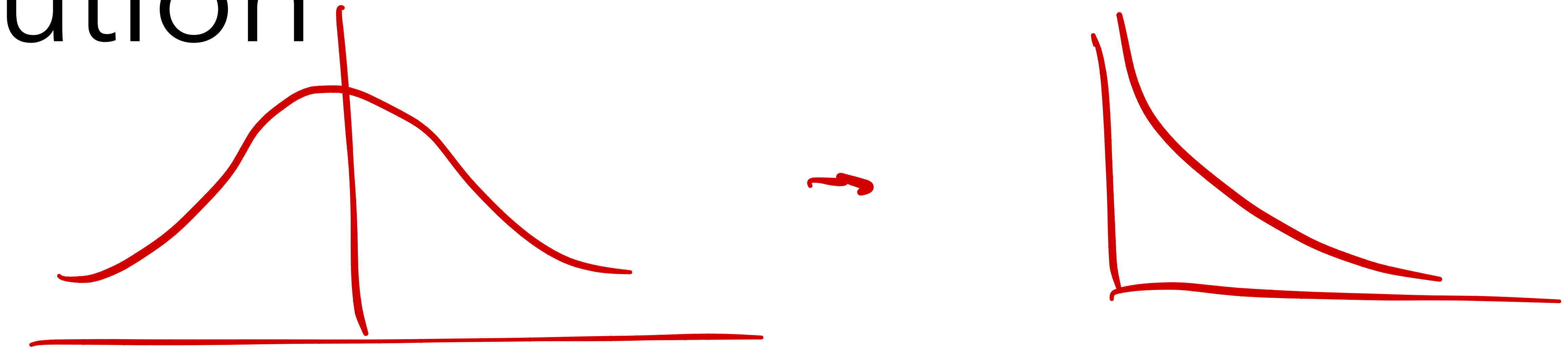
Chi-squared (χ^2) distribution

- However, $Q = \sum_{i=1}^{n} Z_i^2$ follows something called the **chi-squared distribution**

Chi-squared (χ^2) distribution

- However, $Q = \sum_{i=1}^{n} Z_i^2$ follows something called the **chi-squared distribution**
- $Q = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$ (chi-squared distribution with k degrees of freedom)

Chi-squared (χ^2) distribution

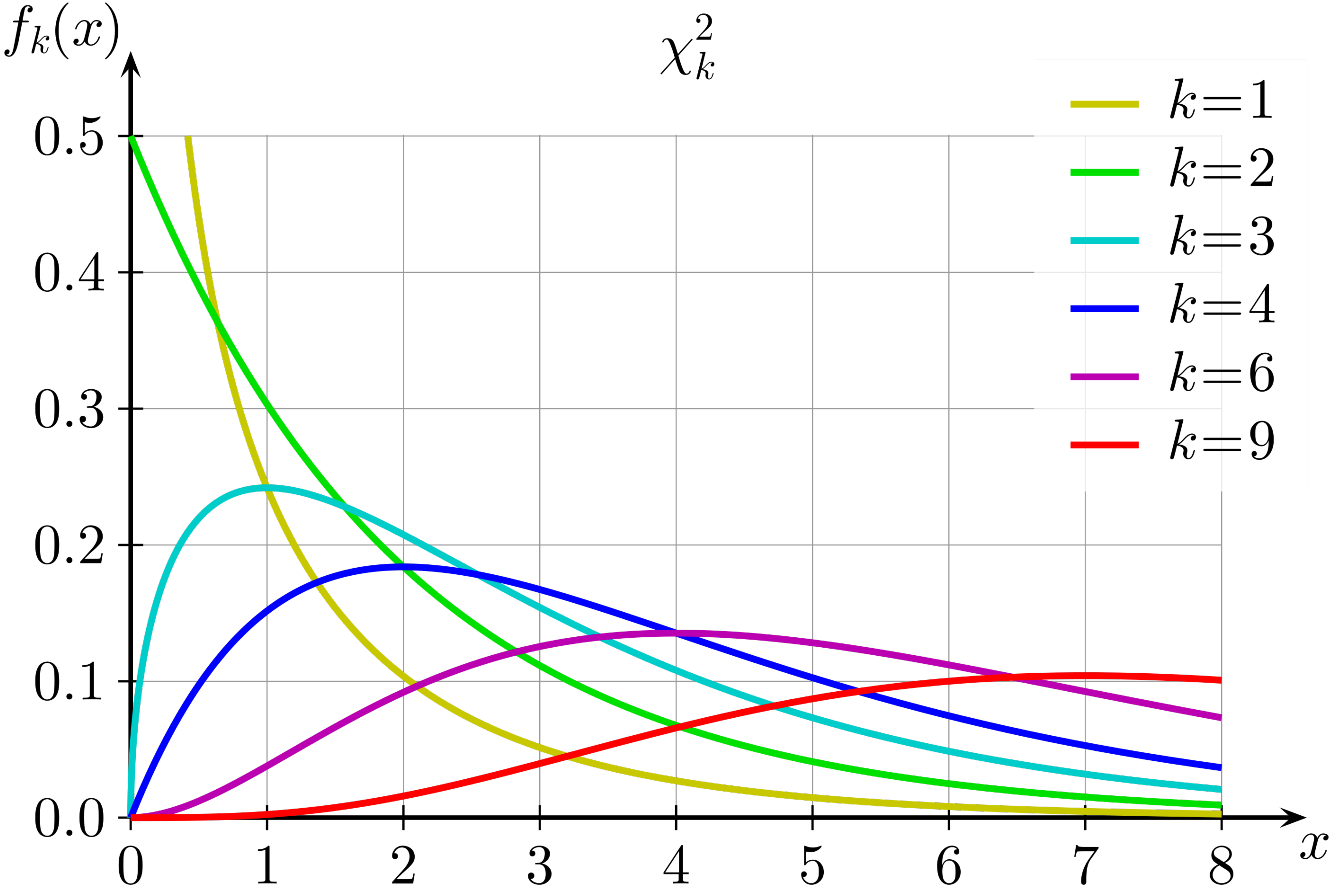


- However, $Q = \sum_{i=1}^{n} Z_i^2$ follows something called the **chi-squared distribution**
- $Q = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$ (chi-squared distribution with k degrees of freedom)
- Mean of χ_k^2 is k (degrees of freedom)

Chi-squared (χ^2) distribution

- However, $Q = \sum_{i=1}^{n} Z_i^2$ follows something called the **chi-squared distribution**
- $Q = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$ (chi-squared distribution with k degrees of freedom)
- Mean of χ_k^2 is k (degrees of freedom)
- Variance of χ_k^2 is $2k$ (twice the degrees of freedom)

Chi-squared (χ^2) distribution



Sampling Distribution of Variance

Sampling Distribution of Variance

- Returning to $\dfrac{s^2}{\sigma^2} = \dfrac{1}{n-1} \cdot \sum_{i=1}^{n} Z_i^2$, where Z_i is a standard normal random variable

Sampling Distribution of Variance $\frac{1}{n-1} \sum (x_i - \bar{x})^2$

μ

- Returning to $\dfrac{s^2}{\sigma^2} = \dfrac{1}{n-1} \cdot \sum_{i=1}^{n} Z_i^2$, where Z_i is a standard normal random variable
- Recall that we assumed we knew σ^2; we don't actually know this and have to use s^2 as an estimate

Sampling Distribution of Variance

- Returning to $\frac{s^2}{\sigma^2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} Z_i^2$, where Z_i is a standard normal random variable
- Recall that we assumed we knew σ^2; we don't actually know this and have to use s^2 as an estimate
- This means that $\boxed{(n-1) \cdot \frac{s^2}{\sigma^2}} \sim \chi^2_{n-1}$ (chi-squared distribution with $n-1$ dof)

Hypothesis Tests for One Population Variance

Hypothesis Tests for One Population Variance

- We have considered the case of comparing the *mean* of a population to a predetermined value

Hypothesis Tests for One Population Variance

- We have considered the case of comparing the *mean* of a population to a predetermined value
- We can also test whether the *variance* of a population is a specified value

Hypothesis Tests for One Population Variance

Hypothesis Tests for One Population Variance

- Can run both two-sided and one-sided tests given a null σ_0^2

Hypothesis Tests for One Population Variance

- Can run both two-sided and one-sided tests given a null σ_0^2
- One-tailed, lower hypothesis:

Hypothesis Tests for One Population Variance

- Can run both two-sided and one-sided tests given a null σ_0^2
- One-tailed, lower hypothesis:
    - $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$

Hypothesis Tests for One Population Variance

- Can run both two-sided and one-sided tests given a null σ_0^2
- One-tailed, lower hypothesis:
  - $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$
- One-tailed, upper hypothesis:

Hypothesis Tests for One Population Variance

- Can run both two-sided and one-sided tests given a null σ_0^2
- One-tailed, lower hypothesis:
  - $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$
- One-tailed, upper hypothesis:
  - $H_0 : \sigma^2 \leq \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$

Hypothesis Tests for One Population Variance

- Can run both two-sided and one-sided tests given a null σ_0^2
- One-tailed, lower hypothesis:
  - $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$
- One-tailed, upper hypothesis:
  - $H_0 : \sigma^2 \leq \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$
- Two-tailed hypothesis:

Hypothesis Tests for One Population Variance

- Can run both two-sided and one-sided tests given a null σ_0^2
- One-tailed, lower hypothesis:
  - $H_0 : \sigma^2 \geq \sigma_0^2$ vs. $H_1 : \sigma^2 < \sigma_0^2$
- One-tailed, upper hypothesis:
  - $H_0 : \sigma^2 \leq \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$
- Two-tailed hypothesis:
  - $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$

Hypothesis Tests for One Population Variance

Hypothesis Tests for One Population Variance

- What is our test statistic?

$$(n-1)\cdot\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$$

Hypothesis Tests for One Population Variance

- What is our test statistic?

- Recall that we know $(n-1)\cdot\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$, so it makes sense to define $T = \boxed{(n-1)\cdot\frac{s^2}{\sigma_0^2}}$ based on our null hypothesis

null H_0 : $\sigma^2 = \sigma_0^2$

Hypothesis Tests for One Population Variance

- What is our test statistic?
- Recall that we know $(n-1) \cdot \dfrac{s^2}{\sigma^2} \sim \chi^2_{n-1}$, so it makes sense to define $T = (n-1) \cdot \dfrac{s^2}{\sigma^2}$ based on our null hypothesis
- Under the null hypothesis, $\sigma^2 = \sigma_0^2$, we have the following:

Hypothesis Tests for One Population Variance

- What is our test statistic?

- Recall that we know $(n-1) \cdot \frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$, so it makes sense to define $T = (n-1) \cdot \frac{s^2}{\sigma^2}$ based on our null hypothesis

- Under the null hypothesis, $\sigma^2 = \sigma_0^2$, we have the following:

$$T_{obs} = (n-1) \cdot \frac{s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

Hypothesis Tests for One Population Variance

- What is our test statistic?

- Recall that we know $(n-1)\cdot\frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$, so it makes sense to define $T = (n-1)\cdot\frac{s^2}{\sigma^2}$ based on our null hypothesis

- Under the null hypothesis, $\sigma^2 = \sigma_0^2$, we have the following:

$$T_{obs} = (n-1)\cdot\frac{s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

- Once we have our test statistic, we can compare to the chi-square distribution to calculate a p-value

Hypothesis Tests for One Population Variance

Hypothesis Tests for One Population Variance

- Finding p-values:

Hypothesis Tests for One Population Variance

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(T < T_{obs})$

Hypothesis Tests for One Population Variance

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(T < T_{obs})$
    - R: `pchisq(T_obs,n-1)`

Hypothesis Tests for One Population Variance

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(T < T_{obs})$
    - R: `pchisq(T_obs,n-1)`
  - One-tailed, upper hypothesis: $p = \Pr(T > T_{obs})$

Hypothesis Tests for One Population Variance

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(T < T_{obs})$
    - R: `pchisq(`T_{obs}`,n-1)`
  - One-tailed, upper hypothesis: $p = \Pr(T > T_{obs})$
    - R: `1-pchisq(`T_{obs}`,n-1)`

Hypothesis Tests for One Population Variance

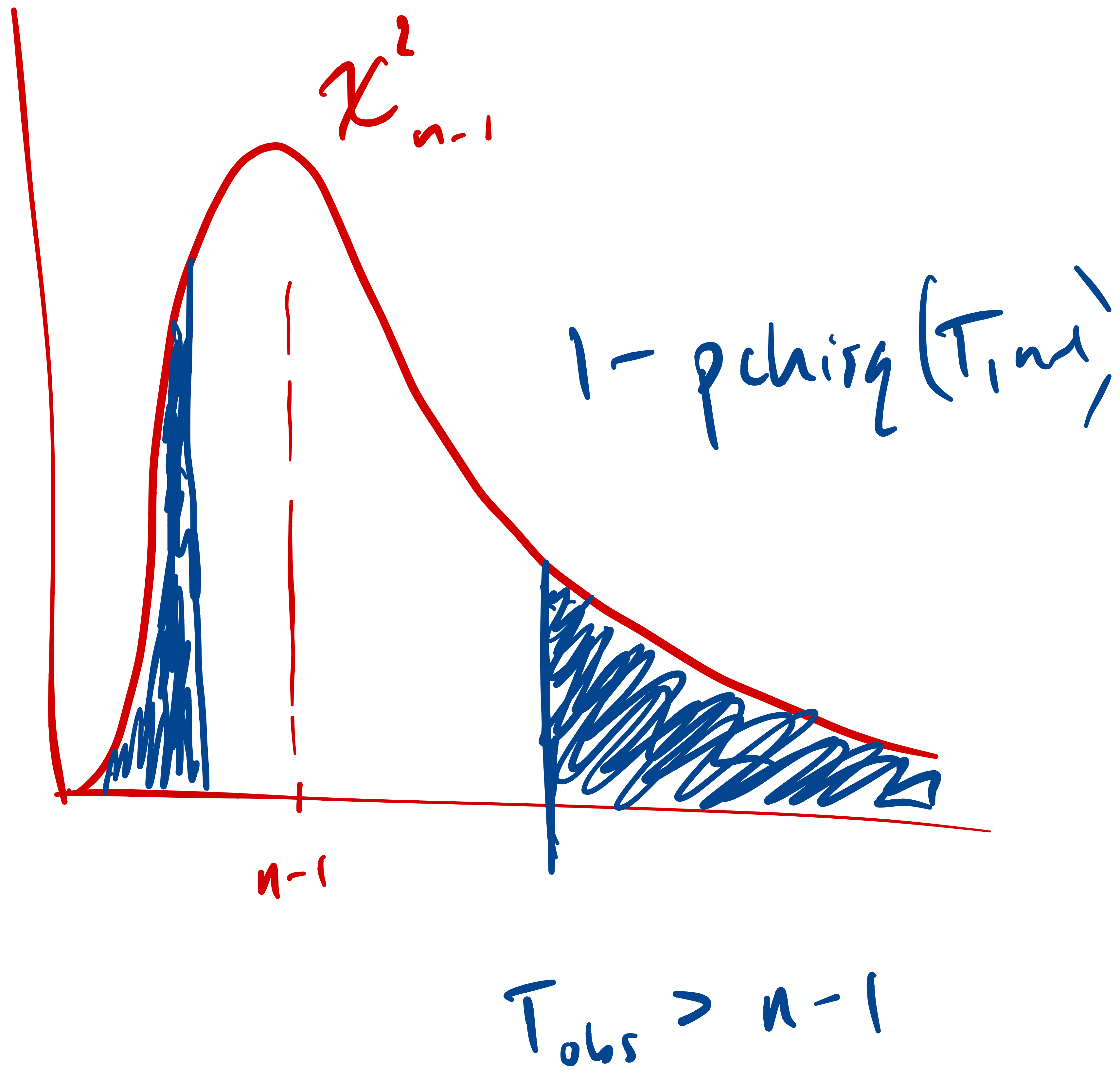
- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(T < T_{obs})$
    - R: `pchisq(T_obs,n-1)`
  - One-tailed, upper hypothesis: $p = \Pr(T > T_{obs})$
    - R: `1-pchisq(T_obs,n-1)`
  - Two-tailed hypothesis: $p = \Pr(|T| \geq |T_{obs}|)$

Hypothesis Tests for One Population Variance

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(T < T_{obs})$
    - R: `pchisq(Tobs,n-1)`
  - One-tailed, upper hypothesis: $p = \Pr(T > T_{obs})$
    - R: `1-pchisq(Tobs,n-1)`
  - Two-tailed hypothesis: $p = \Pr(|T| \geq |T_{obs}|)$
    - If $T_{obs} \leq n - 1$, we have `2*pchisq(Tobs,n-1)`

Hypothesis Tests for One Population Variance

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(T < T_{obs})$
    - R: `pchisq(Tobs,n-1)`
  - One-tailed, upper hypothesis: $p = \Pr(T > T_{obs})$
    - R: `1-pchisq(Tobs,n-1)`
  - Two-tailed hypothesis: $p = \Pr(|T| \geq |T_{obs}|)$
    - If $T_{obs} \leq n-1$, we have `2*pchisq(Tobs,n-1)`
    - If $T_{obs} > n-1$, we have `2*(1-pchisq(Tobs,n-1))`



Hypothesis Tests for One Population Variance

Hypothesis Tests for One Population Variance

- If the p-value is less than α, reject the null hypothesis and conclude that the population variances are unequal to each other

Hypothesis Tests for One Population Variance

- If the p-value is less than α, reject the null hypothesis and conclude that the population variances are unequal to each other
- If the p-value is greater than α, fail to reject the null hypothesis and conclude that the population variances are equal to each other

Hypothesis Tests for One Population Variance: Example

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes². This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes². Has the variance changed? Evaluate at $\alpha = 0.01$

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes^2. This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes^2. Has the variance changed? Evaluate at $\alpha = 0.01$

- Hypotheses:

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes². This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes². Has the variance changed? Evaluate at $\alpha = 0.01$
- Hypotheses:
- Calculate the T statistic:

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes^2. This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes^2. Has the variance changed? Evaluate at $\alpha = 0.01$

- Hypotheses:

- Calculate the T statistic:

- Calculate the p-value:

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes². This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes². Has the variance changed? Evaluate at $\alpha = 0.01$
- Hypotheses: $H_0 : \sigma^2 = 121$ vs. $H_1 : \sigma^2 \neq 121$
- Calculate the T statistic: $T = (n-1) \cdot \frac{s^2}{\sigma^2} = 44 \cdot \frac{196}{121} = 71.3 > 44$
- Calculate the p-value: $2 \times (1 - \text{pchisq}(71.3, 44)) = 0.0115$
- Conclusion: $p > \alpha$, fail to reject.

Hypothesis Tests for One Population Variance: Example

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes². This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes². Has the variance *increased*? Evaluate at $\alpha = 0.01$

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes². This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes². Has the variance *increased*? Evaluate at $\alpha = 0.01$
- Hypotheses:

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes^2. This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes^2. Has the variance *increased*? Evaluate at $\alpha = 0.01$

- Hypotheses:

- Calculate the T statistic:

Hypothesis Tests for One Population Variance: Example

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes^2. This year, we took a sample of $n = 45$ guests and found that the sample variance $s = 196$ minutes^2. Has the variance *increased*? Evaluate at $\alpha = 0.01$

- Hypotheses:

- Calculate the T statistic:

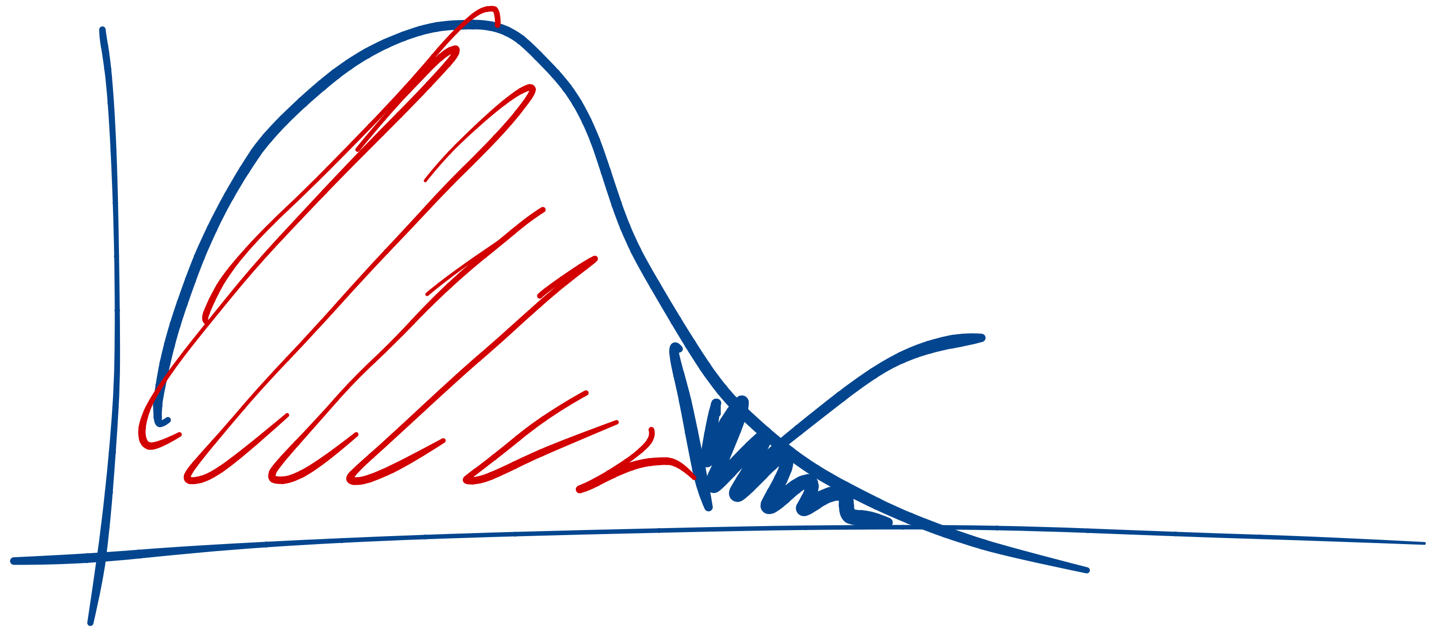
- Calculate the p-value:

$$s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \cancel{\bar{x}}\,\mu)^2 \sim \chi^2_n$$

Hypothesis Tests for One Population Variance: Example

$$\frac{1}{n-1} \sum (x_i - \bar{x})^2 \sim \chi^2_{n-1}$$

- Setup: In 2010, the distribution of the amount of time spent per table at The Cheesecake Factory is normally distributed with $\mu = 65$ minutes and $\sigma^2 = 121$ minutes². This year, we took a sample of $n = 45$ guests and found that the sample variance $s^2 = 196$ minutes². Has the variance *increased*? Evaluate at $\alpha = 0.01$
- Hypotheses: $H_0: \sigma^2 \le 121$ $\quad H_1: \sigma^2 > 121$
- Calculate the T statistic: $T = (n-1)\frac{s^2}{\sigma^2} = 71.3$



- Calculate the p-value: $1 - \text{pchisq}(71.3, 44) \approx 0.0055$
- Conclusion: $p < \alpha \rightarrow$ Reject. σ^2 Increased since 2010

Confidence Intervals for One Population Variance

Confidence Intervals for One Population Variance

- We can also calculate $(1-\alpha) \cdot 100\,\%$ (two-sided) confidence intervals for the population variance

Confidence Intervals for One Population Variance

- We can also calculate $(1-\alpha)\cdot 100\,\%$ (two-sided) confidence intervals for the population variance
- Because we have that $(n-1)\cdot\dfrac{s^2}{\sigma^2} \sim \chi^2_{n-1}$, we have

Confidence Intervals for One Population Variance

- We can also calculate $(1-\alpha)\cdot 100\,\%$ (two-sided) confidence intervals for the population variance

- Because we have that $\boxed{(n-1)\cdot\frac{s^2}{\sigma^2}} \sim \chi^2_{n-1}$, we have

qchisq(α/2, n-1)

$$\Pr\left(\chi^2_{\alpha/2,n-1} \le (n-1)\cdot\frac{s^2}{\sigma^2} \le \chi^2_{1-\alpha/2,n-1}\right) = \alpha$$

$$\implies \Pr\left(\frac{1}{\chi^2_{\alpha/2,n-1}} \ge \frac{\sigma^2}{(n-1)\cdot s^2} \ge \frac{1}{\chi^2_{1-\alpha/2,n-1}}\right) = \alpha$$

$$\implies \Pr\left(\frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,n-1}}\right) = \alpha$$

Confidence Intervals for One Population Variance

- We can also calculate $(1-\alpha)\cdot 100\,\%$ (two-sided) confidence intervals for the population variance

- Because we have that $(n-1)\cdot\dfrac{s^2}{\sigma^2}\sim\chi^2_{n-1}$, we have

$$\Pr\left(\chi^2_{\alpha/2,n-1}\le(n-1)\cdot\frac{s^2}{\sigma^2}\le\chi^2_{1-\alpha/2,n-1}\right)=\alpha$$

$$\implies\ \Pr\left(\frac{1}{\chi^2_{\alpha/2,n-1}}\ge\frac{\sigma^2}{(n-1)\cdot s^2}\ge\frac{1}{\chi^2_{1-\alpha/2,n-1}}\right)=\alpha$$

$$\implies\ \Pr\left(\frac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2,n-1}}\le\sigma^2\le\frac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,n-1}}\right)=\alpha$$

qchisq(1 - α/2, n-1) qchisq(α/2, n-1)

- Therefore, the interval $\left(\dfrac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2,n-1}},\dfrac{(n-1)\cdot s^2}{\chi^2_{\alpha/2,n-1}}\right)$ contains σ^2 with probability $1-\alpha$

Confidence Intervals for One Population Variance

Confidence Intervals for One Population Variance

- Similarly, one-sided $(1-\alpha)\cdot 100\,\%$ confidence intervals for the population variance are as follows

Confidence Intervals for One Population Variance

- Similarly, one-sided $(1 - \alpha) \cdot 100\,\%$ confidence intervals for the population variance are as follows
- One-sided, upper:

Confidence Intervals for One Population Variance

- Similarly, one-sided $(1 - \alpha) \cdot 100\,\%$ confidence intervals for the population variance are as follows

- One-sided, upper:

  - $\left(0, \dfrac{(n-1) \cdot s^2}{\chi^2_{\alpha,n-1}}\right)$

$\chi^2_{\alpha,n-1}$

Confidence Intervals for One Population Variance

- Similarly, one-sided $(1-\alpha)\cdot 100\,\%$ confidence intervals for the population variance are as follows

- One-sided, upper:
  - $\left(0,\ \dfrac{(n-1)\cdot s^2}{\chi^2_{\alpha,n-1}}\right)$

- One-sided, lower:

Confidence Intervals for One Population Variance

- Similarly, one-sided $(1-\alpha)\cdot 100\,\%$ confidence intervals for the population variance are as follows
- One-sided, upper:
  - $\left(0,\ \dfrac{(n-1)\cdot s^2}{\chi^2_{\alpha,n-1}}\right)$
- One-sided, lower:
  - $\left(\dfrac{(n-1)\cdot s^2}{\chi^2_{1-\alpha,n-1}},\ \infty\right)$

Confidence Intervals for One Population Variance: Example

Confidence Intervals for One Population Variance: Example

- Let's say we are interested in the amount of water the average person drinks per day. We sample a group of $n = 40$ people and find that the sample variance is $s^2 = 60$ oz^2. What is a two-tailed 95% confidence interval for the variance of this distribution?

Confidence Intervals for One Population Variance: Example

- Let's say we are interested in the amount of water the average person drinks per day. We sample a group of $n = 40$ people and find that the sample variance is $s^2 = 60$ oz^2. What is a two-tailed 95% confidence interval for the variance of this distribution?

- What is α?

Confidence Intervals for One Population Variance: Example

$(n-1) \cdot \frac{s^2}{\sigma^2}$

$\frac{60}{40.26}$ $\frac{98.92}{60}$

- Let's say we are interested in the amount of water the average person drinks per day. We sample a group of $n = 40$ people and find that the sample variance is $s^2 = 60$ oz². What is a two-tailed 95% confidence interval for the variance of this distribution?

$(40.26, 98.92)$

- What is α? $\alpha = 0.05$
- CI: $\left(\frac{(n-1) \cdot s^2}{\chi^2_{1-\alpha/2,\, n-1}}, \frac{(n-1) \cdot s^2}{\chi^2_{\alpha/2,\, n-1}} \right) = \left(\frac{39 \cdot 60}{\chi^2_{0.975,\, 39}}, \frac{39 \cdot 60}{\chi^2_{0.025,\, 39}} \right)$

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- Consider the case of wanting to compare two populations

Hypothesis Tests for Two Population Variances

- Consider the case of wanting to compare two populations
- We have talked about testing whether the means of the two independent populations are equal to each other

Hypothesis Tests for Two Population Variances

- Consider the case of wanting to compare two populations
- We have talked about testing whether the means of the two independent populations are equal to each other
- We can also test whether the variances of the two populations are equal to each other

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal
- We attempt to see if there is a difference in variances

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal
- We attempt to see if there is a difference in variances
- One-tailed, lower hypothesis:

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal
- We attempt to see if there is a difference in variances
- One-tailed, lower hypothesis:
  - $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs. $H_1 : \sigma_1^2 < \sigma_2^2$

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal
- We attempt to see if there is a difference in variances
- One-tailed, lower hypothesis:
  - $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs. $H_1 : \sigma_1^2 < \sigma_2^2$
- One-tailed, upper hypothesis:

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal
- We attempt to see if there is a difference in variances
- One-tailed, lower hypothesis:
  - $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs. $H_1 : \sigma_1^2 < \sigma_2^2$
- One-tailed, upper hypothesis:
  - $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs. $H_1 : \sigma_1^2 > \sigma_2^2$

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal
- We attempt to see if there is a difference in variances
- One-tailed, lower hypothesis:
  - $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs. $H_1 : \sigma_1^2 < \sigma_2^2$
- One-tailed, upper hypothesis:
  - $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs. $H_1 : \sigma_1^2 > \sigma_2^2$
- Two-tailed hypothesis:

Hypothesis Tests for Two Population Variances

- Null hypothesis: Population variances are equal
- We attempt to see if there is a difference in variances
- One-tailed, lower hypothesis:
    - $H_0 : \sigma_1^2 \geq \sigma_2^2$ vs. $H_1 : \sigma_1^2 < \sigma_2^2$
- One-tailed, upper hypothesis:
    - $H_0 : \sigma_1^2 \leq \sigma_2^2$ vs. $H_1 : \sigma_1^2 > \sigma_2^2$
- Two-tailed hypothesis:
    - $H_0 : \sigma_1^2 = \sigma_2^2$ vs. $H_1 : \sigma_1^2 \neq \sigma_2^2$

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

$(n-1) \cdot \frac{s^2}{\sigma^2} \sim \chi^2_{n-1}$

- Recall that for a single population, we have $\dfrac{s^2}{\sigma^2} \sim \dfrac{1}{n-1} \cdot \chi^2_{n-1}$

Hypothesis Tests for Two Population Variances

- Recall that for a single population, we have $\frac{s^2}{\sigma^2} \sim \frac{1}{n-1} \cdot \chi^2_{n-1}$
- When comparing two populations, it makes sense to look at the quantity $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$

Hypothesis Tests for Two Population Variances

- Recall that for a single population, we have $\frac{s^2}{\sigma^2} \sim \frac{1}{n-1} \cdot \chi^2_{n-1}$
- When comparing two populations, it makes sense to look at the quantity $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$

F distribution

- However, from the above, we have that $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim \boxed{\frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)}}$

Hypothesis Tests for Two Population Variances

- Recall that for a single population, we have $\frac{s^2}{\sigma^2} \sim \frac{1}{n-1} \cdot \chi^2_{n-1}$
- When comparing two populations, it makes sense to look at the quantity $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$
- However, from the above, we have that $\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim \frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)}$
- It turns out that this type of distribution has a name: F distribution

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- To test such hypotheses, we use an F test statistic:

Hypothesis Tests for Two Population Variances

- To test such hypotheses, we use an F test statistic:

$$F_{obs} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,\; n_2-1}$$

num

denom

Hypothesis Tests for Two Population Variances

- To test such hypotheses, we use an F test statistic:

$$F_{obs} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,\ n_2-1}$$

- This F statistic follows an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom

Hypothesis Tests for Two Population Variances

- To test such hypotheses, we use an F test statistic:

$$F_{obs} = \frac{s_1^2/\cancel{\sigma_1^2}}{s_2^2/\cancel{\sigma_2^2}} \sim F_{n_1-1,\, n_2-1}$$

- This F statistic follows an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom

- Under the null hypothesis, $\sigma_1^2 = \sigma_2^2$, we have the following:

Hypothesis Tests for Two Population Variances

- To test such hypotheses, we use an F test statistic:

$$F_{obs} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

- This F statistic follows an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom

- Under the null hypothesis, $\sigma_1^2 = \sigma_2^2$, we have the following:

$$F_{obs} = \frac{s_1^2}{s_2^2} \sim \underline{F_{n_1-1,\, n_2-1}}$$

Hypothesis Tests for Two Population Variances

- To test such hypotheses, we use an F test statistic:

$$F_{obs} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,\, n_2-1}$$

- This F statistic follows an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom

- Under the null hypothesis, $\sigma_1^2 = \sigma_2^2$, we have the following:

$$F_{obs} = \frac{s_1^2}{s_2^2} \sim F_{n_1-1,\, n_2-1}$$

- Once we have our test statistic, we can compare to the F distribution to calculate a p-value

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- Finding p-values:

Hypothesis Tests for Two Population Variances

- Finding p-values:
    - One-tailed, lower hypothesis: $p = \Pr(F < F_{obs})$

Hypothesis Tests for Two Population Variances

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(F < F_{obs})$
    - R: `pf(`F_{obs}`,`n_1`-1,`n_2`-1)`

Hypothesis Tests for Two Population Variances

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(F < F_{obs})$
    - R: `pf(`F_{obs}`,`n_1`-1,`n_2`-1)`
  - One-tailed, upper hypothesis: $p = \Pr(F > F_{obs})$

Hypothesis Tests for Two Population Variances

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(F < F_{obs})$
    - R: `pf(`F_{obs}`,`n_1`-1,`n_2`-1)`
  - One-tailed, upper hypothesis: $p = \Pr(F > F_{obs})$
    - R: `1-pf(`F_{obs}`,`n_1`-1,`n_2`-1)`

Hypothesis Tests for Two Population Variances

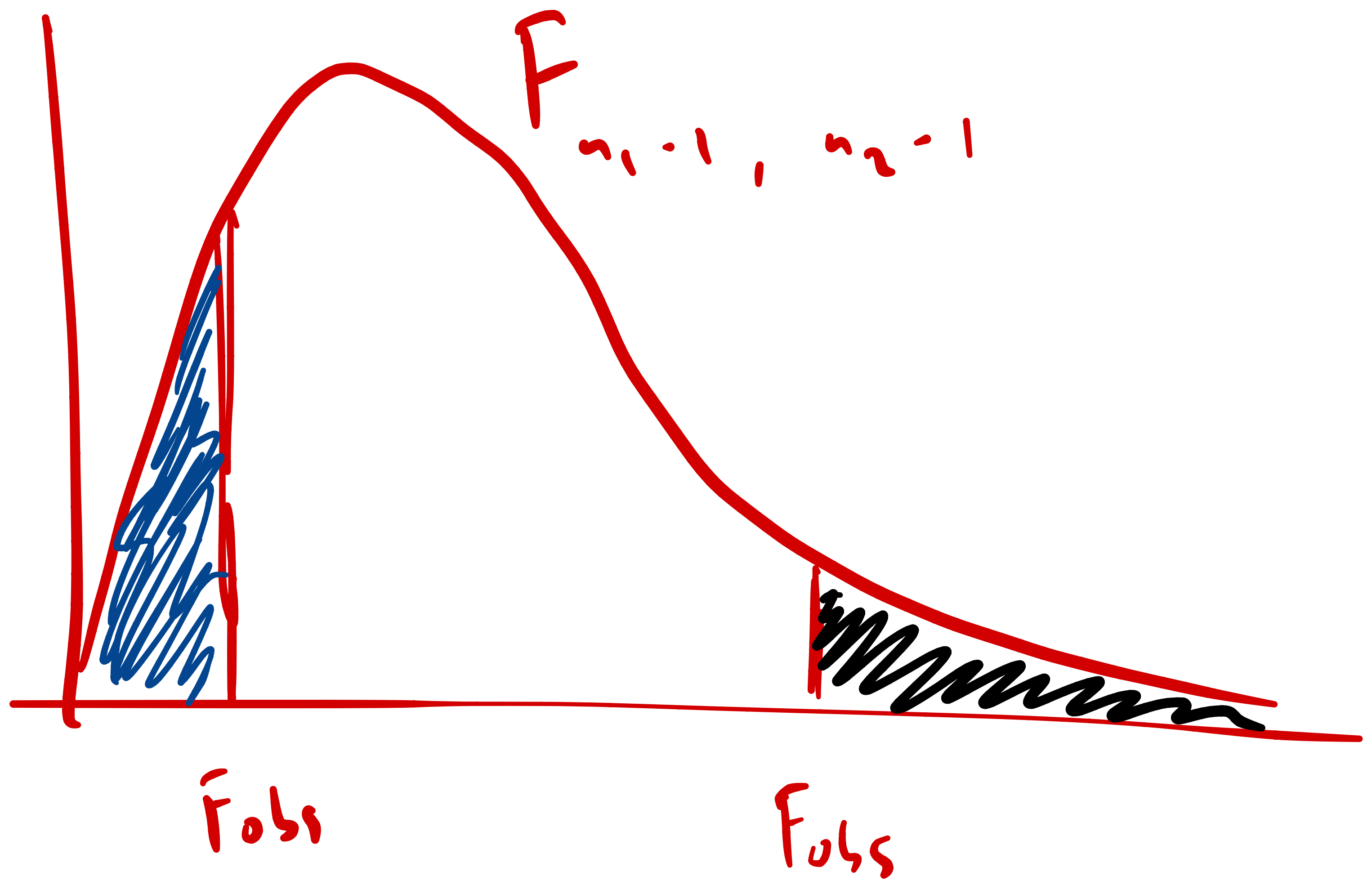
- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(F < F_{obs})$
    - R: `pf(`F_{obs}`,`n_1`-1,`n_2`-1)`
  - One-tailed, upper hypothesis: $p = \Pr(F > F_{obs})$
    - R: `1-pf(`F_{obs}`,`n_1`-1,`n_2`-1)`
  - Two-tailed hypothesis: $p = \Pr(|T| \geq |T_{obs}|)$

Hypothesis Tests for Two Population Variances

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(F < F_{obs})$
    - R: `pf(`$\texttt{F}_{\texttt{obs}}$`,`$\texttt{n}_1$`-1,`$\texttt{n}_2$`-1)`
  - One-tailed, upper hypothesis: $p = \Pr(F > F_{obs})$
    - R: `1-pf(`$\texttt{F}_{\texttt{obs}}$`,`$\texttt{n}_1$`-1,`$\texttt{n}_2$`-1)`
  - Two-tailed hypothesis: $p = \Pr(|T| \geq |T_{obs}|)$
    - If $F_{obs} \leq 1$, we have `2*pf(`$\texttt{F}_{\texttt{obs}}$`,`$\texttt{n}_1$`-1,`$\texttt{n}_2$`-1)`

Hypothesis Tests for Two Population Variances

- Finding p-values:
  - One-tailed, lower hypothesis: $p = \Pr(F < F_{obs})$
    - R: `pf(Fobs,n1-1,n2-1)`
  - One-tailed, upper hypothesis: $p = \Pr(F > F_{obs})$
    - R: `1-pf(Fobs,n1-1,n2-1)`
  - Two-tailed hypothesis: $p = \Pr(|T| \geq |T_{obs}|)$
    - If $F_{obs} \leq 1$, we have `2*pf(Fobs,n1-1,n2-1)`
    - If $F_{obs} > 1$, we have `2*(1-pf(Fobs,n1-1,n2-1))`



$$F = \frac{\chi^2_{n_1-1} / (n_1-1)}{\chi^2_{n_2-1} / (n_2-1)}$$

$\mu = 1$

$\mu = 1$

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- If the p-value is less than α, reject the null hypothesis and conclude that the population variances are unequal to each other

Hypothesis Tests for Two Population Variances

- If the p-value is less than α, reject the null hypothesis and conclude that the population variances are unequal to each other
- If the p-value is greater than α, fail to reject the null hypothesis and conclude that the population variances are equal to each other

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection
- Let $F_{n_1-1,n_2-1,\alpha}$ be the critical value for an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom and α in the lower tail

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection

- Let $F_{n_1-1,n_2-1,\alpha}$ be the critical value for an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom and α in the lower tail

- One-tailed, lower hypothesis:

$F_{n_1-1,\,n_2-1,\,\alpha}$

qf(

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection
- Let $F_{n_1-1,n_2-1,\alpha}$ be the critical value for an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom and α in the lower tail
- One-tailed, lower hypothesis:
    - Reject if $F_{obs} \leq F_{n_1-1,n_2-1,\alpha}$

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection
- Let $F_{n_1-1,n_2-1,\alpha}$ be the critical value for an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom and α in the lower tail
- One-tailed, lower hypothesis:
  - Reject if $F_{obs} \leq F_{n_1-1,n_2-1,\alpha}$
- One-tailed, upper hypothesis:

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection
- Let $F_{n_1-1,n_2-1,\alpha}$ be the critical value for an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom and α in the lower tail
- One-tailed, lower hypothesis:
  - Reject if $F_{obs} \leq F_{n_1-1,n_2-1,\alpha}$
- One-tailed, upper hypothesis:
  - Reject if $F_{obs} \geq F_{n_1-1,n_2-1,\alpha}$

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection
- Let $F_{n_1-1,n_2-1,\alpha}$ be the critical value for an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom and α in the lower tail
- One-tailed, lower hypothesis:
  - Reject if $F_{obs} \leq F_{n_1-1,n_2-1,\alpha}$
- One-tailed, upper hypothesis:
  - Reject if $F_{obs} \geq F_{n_1-1,n_2-1,\alpha}$
- Two-tailed hypothesis:

Hypothesis Tests for Two Population Variances

- Instead of p-values, we can use critical values for rejection
- Let $F_{n_1-1,n_2-1,\alpha}$ be the critical value for an F distribution with $n_1 - 1$ and $n_2 - 1$ degrees of freedom and α in the lower tail
- One-tailed, lower hypothesis:
  - Reject if $F_{obs} \leq F_{n_1-1,n_2-1,\alpha}$
- One-tailed, upper hypothesis:
  - Reject if $F_{obs} \geq F_{n_1-1,n_2-1,\alpha}$
- Two-tailed hypothesis:
  - Reject if $F_{obs} \leq F_{n_1-1,n_2-1,\alpha/2}$ or $F_{obs} \geq F_{n_1-1,n_2-1,\cancel{\alpha/2}}$ $1-\alpha/2$

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- We are interested in exploring the fuel efficiency, measured in miles per gallon (mpg), for two models of mass-produced cars: Honda Civics and Honda Accords

Hypothesis Tests for Two Population Variances

- We are interested in exploring the fuel efficiency, measured in miles per gallon (mpg), for two models of mass-produced cars: Honda Civics and Honda Accords

- In particular, we want to explore the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$

Hypothesis Tests for Two Population Variances

- We are interested in exploring the fuel efficiency, measured in miles per gallon (mpg), for two models of mass-produced cars: Honda Civics and Honda Accords

- In particular, we want to explore the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$

- Take a sample of $n_1 = 10$ Civics and $n_2 = 20$ Accords

Hypothesis Tests for Two Population Variances

- We are interested in exploring the fuel efficiency, measured in miles per gallon (mpg), for two models of mass-produced cars: Honda Civics and Honda Accords

- In particular, we want to explore the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$

- Take a sample of $n_1 = 10$ Civics and $n_2 = 20$ Accords

- Sample variances: $s_1^2 = 2.33$ and $s_2^2 = 9.97$

Hypothesis Tests for Two Population Variances

- We are interested in exploring the fuel efficiency, measured in miles per gallon (mpg), for two models of mass-produced cars: Honda Civics and Honda Accords
- In particular, we want to explore the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$
- Take a sample of $n_1 = 10$ Civics and $n_2 = 20$ Accords
- Sample variances: $s_1^2 = 2.33$ and $s_2^2 = 9.97$
- Evaluate at $\alpha = 0.05$

Hypothesis Tests for Two Population Variances

Hypothesis Tests for Two Population Variances

- Calculate the test statistic:

Hypothesis Tests for Two Population Variances

- Calculate the test statistic:

Hypothesis Tests for Two Population Variances

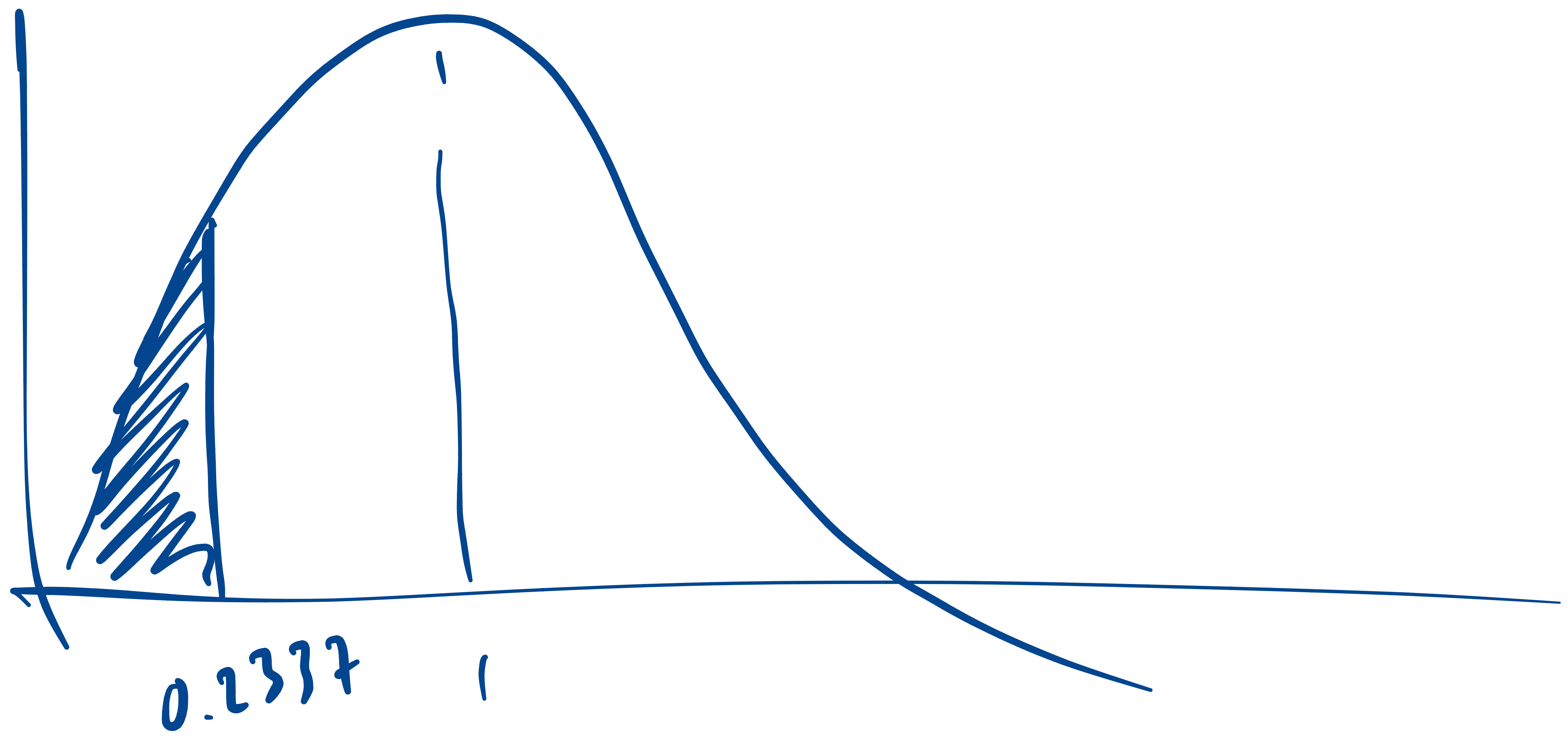
- Calculate the test statistic:

$$F = \frac{s_1^2}{s_2^2} = \frac{2.33}{9.47} \approx 0.2337$$

$$p = 2 \times pf(0.2337, 9, 19)$$

$$= \underline{0.03}$$

- Conclusion:



Hypothesis Tests for Two Population Variances

- Calculate the test statistic:

- Conclusion: Reject.

- Similarly, we could have calculated the critical values for the F-test statistic and used them to complete the test:

Hypothesis Tests for Two Population Variances

- Calculate the test statistic:

$$F = 0.2337$$

- Conclusion:

- Similarly, we could have calculated the critical values for the F-test statistic and used them to complete the test:

$$F_{n_1-1,n_2-1,\alpha/2} = \texttt{qf(0.025,9,19)} = 0.2715$$
$$F_{n_1-1,n_2-1,1-\alpha/2} = \texttt{qf(0.975,9,19)} = 2.880$$

Hypothesis Tests for Two Population Variances

- Calculate the test statistic:

- Conclusion:

- Similarly, we could have calculated the critical values for the F-test statistic and used them to complete the test:

$$F_{n_1-1,n_2-1,\alpha/2} = \texttt{qf(0.025,9,19)} = 0.2715$$
$$F_{n_1-1,n_2-1,1-\alpha/2} = \texttt{qf(0.975,9,19)} = 2.880$$

- Since $F = 0.2337 < 0.2715$, we reject the null hypothesis and conclude $\sigma_1^2 \neq \sigma_2^2$

Confidence Intervals for Two Population Variances

Confidence Intervals for Two Population Variances

- We can also calculate confidence intervals for the ratio of two population variances

Confidence Intervals for Two Population Variances

- We can also calculate confidence intervals for the ratio of two population variances
- A two-sided $(1-\alpha)\cdot 100\,\%$ confidence interval for σ_1^2/σ_2^2 is given by

$$\left(\frac{1}{F_{1-\alpha/2}}\cdot\frac{s_1^2}{s_2^2}, \frac{1}{F_{\alpha/2}}\cdot\frac{s_1^2}{s_2^2}\right)$$

$$\frac{1}{\chi^2_{n-1,\,1-\alpha/2}} \times \cdots \quad , \quad \frac{1}{\chi^2_{n-1,\,\alpha/2}} \times \cdots$$

Confidence Intervals for Two Population Variances

- We can also calculate confidence intervals for the ratio of two population variances
- A two-sided $(1-\alpha)\cdot 100\,\%$ confidence interval for σ_1^2/σ_2^2 is given by
$$\left(\frac{1}{F_{1-\alpha/2}}\cdot\frac{s_1^2}{s_2^2},\ \frac{1}{F_{\alpha/2}}\cdot\frac{s_1^2}{s_2^2}\right)$$
- A one-sided upper $(1-\alpha)\cdot 100\,\%$ confidence interval for σ_1^2/σ_2^2 is given by
$$\left(0,\ \frac{1}{F_{\alpha}}\cdot\frac{s_1^2}{s_2^2}\right)$$

Confidence Intervals for Two Population Variances

- We can also calculate confidence intervals for the ratio of two population variances
- A two-sided $(1-\alpha)\cdot 100\,\%$ confidence interval for σ_1^2/σ_2^2 is given by
$$\left(\frac{1}{F_{1-\alpha/2}}\cdot\boxed{\frac{s_1^2}{s_2^2}}, \frac{1}{F_{\alpha/2}}\cdot\frac{s_1^2}{s_2^2}\right)$$
- A one-sided upper $(1-\alpha)\cdot 100\,\%$ confidence interval for σ_1^2/σ_2^2 is given by
$$\left(0, \frac{1}{F_{\alpha}}\cdot\frac{s_1^2}{s_2^2}\right)$$
- A one-sided lower $(1-\alpha)\cdot 100\,\%$ confidence interval for σ_1^2/σ_2^2 is given by
$$\left(\frac{1}{F_{1-\alpha}}\cdot\frac{s_1^2}{s_2^2}, \infty\right)$$

Confidence Intervals for Two Population Variances: Example

Confidence Intervals for Two Population Variances: Example

- Return to the fuel efficiency example (sample of $n_1 = 10$ Civics and $n_2 = 20$ Accords, sample variances: $s_1^2 = 2.33$ and $s_2^2 = 9.97$). What is a one-tailed upper 95% confidence interval for the ratio of variances?

Confidence Intervals for Two Population Variances: Example

- Return to the fuel efficiency example (sample of $n_1 = 10$ Civics and $n_2 = 20$ Accords, sample variances: $s_1^2 = 2.33$ and $s_2^2 = 9.97$). What is a one-tailed upper 95% confidence interval for the ratio of variances?

- What is α?

Confidence Intervals for Two Population Variances: Example

- Return to the fuel efficiency example (sample of $n_1 = 10$ Civics and $n_2 = 20$ Accords, sample variances: $s_1^2 = 2.33$ and $s_2^2 = 9.97$). What is a one-tailed upper 95% confidence interval for the ratio of variances?

- What is α? $\alpha = 0.05$

  $(0, 0.689)$

- CI: $\left(0, \frac{1}{F_\alpha} \cdot \frac{s_1^2}{s_2^2}\right) = \left(0, \frac{1}{0.339} \cdot \frac{2.33}{9.97}\right)$

  $qf(0.05, 9, 19) = 0.339$