Probability and Statistics I

26. The χ^2, t, and F Distributions & Distributions Based on a Normal Random Sample

Revised 02/12/24

6.3 The χ^2, t, and F Distributions
6.4 Distributions Based on a Normal Random Sample

Assessing the Mean of a Distribution

Example 26.1: Do exams cause stress?

The normal resting heart rate of people between the ages of 18 and 25 is normally distributed with a mean of 70 beats per minute (bpm).

A professor measures the heart rates of 15 of 250 students as they leave their exam. The sample mean is $\bar{x} = 74$ bpm with a standard deviation of $s = 7$ bpm.

Can the professor conclude that the mean heart rate of students leaving tests is above the normal resting heart rate?

Distribution of the Sample Mean for a Normal Population

Let $X_1, \ldots, X_N$ be a random sample from a normal distribution with mean μ and variance $\sigma^2 < \infty$.

Let $\bar{X}_n$ be the sample mean. Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \text{Normal}(0, 1).$$

t-Distribution

Let $X_1, \ldots, X_N$ be a random sample from *any* distribution with mean μ and variance $\sigma^2 < \infty$.

Let $\bar{X}_n$ be the sample mean and S_n the sample standard deviation. Then

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{(n-1)}.$$

t Distribution

A continuous random variable, T, has a t distribution with ν degrees of freedom if the pdf of T is

$$f(t) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < t < \infty.$$

Mathematically, we write $T \sim t_\nu$.

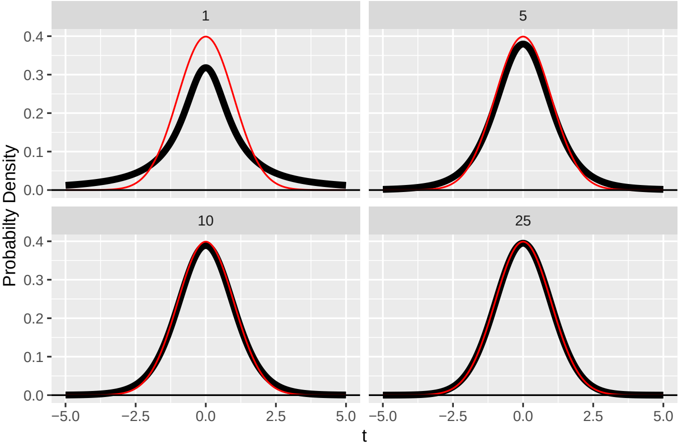
Properties

- CDF: No closed form
- Mean: $E(T) = 0$
- Variance: $V(T) = \frac{\nu}{\nu-2}, \quad \nu > 2$

Calculator

https://stattrek.com/online-calculator/t-distribution

t Distribution (black) vs Standard Normal (red)



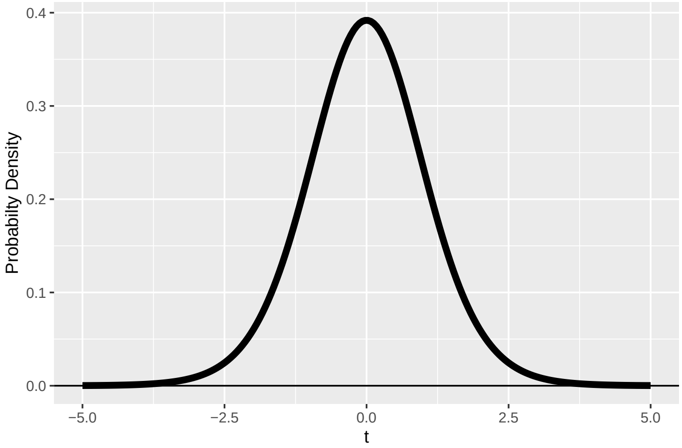
Example 26.1 ctd: Do exams cause stress?

The normal resting heart rate of people between the ages of 18 and 25 is normally distributed with a mean of 70 beats per minute (bpm).

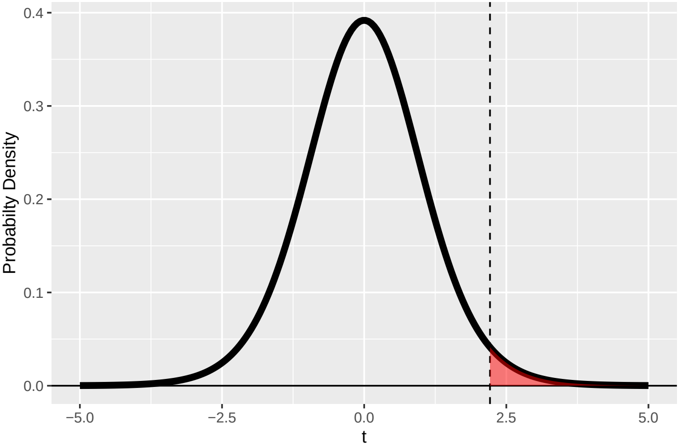
A professor measures the heart rates of 15 of 250 students as they leave their exam. The sample mean is $\bar{x} = 74$ bpm with a standard deviation of $s = 7$ bpm.

Can the professor conclude that the mean heart rate of students leaving tests is above the normal resting heart rate?

Example 26.1 ctd: Do exams cause stress?



Example 26.1 ctd: Do exams cause stress?



Assessing the Variance of a Distribution

Example 26.2

The amount of string on a spool produced by the Acme is supposed to be normally distributed with a mean of 50 m and a standard deviation of .1 *m*.

Each day the company tests 20 randomly selected spools. Suppose they find that the observed standard deviation is .11 m.

Can they conclude that the standard deviation is too high?

Distribution of the Sample Variance for a Normal Population

Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance $\sigma^2 < \infty$.

Let

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

be the sample variance. Then

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

Chi-Squared Distribution (Lecture 17)

If $X \sim \text{Gamma}(\nu/2, 2)$ then we say that X follows a chi-squared distribution with ν degrees of freedom:

$$X \sim \chi^2_\nu.$$

The pdf of the chi-squared distribution is

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2}, \quad x \geq 0.$$

Properties

- CDF: No closed form
- Mean: $E(X) = \nu$
- Variance: $V(X) = 2\nu$

Calculator

`https://stattrek.com/online-calculator/chi-square`

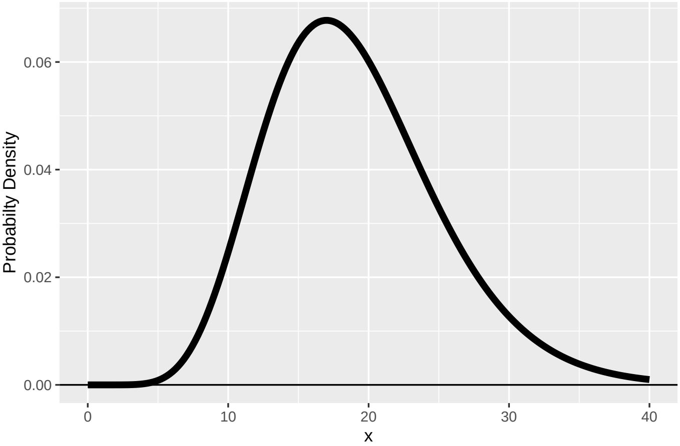
Example 26.2 ctd

The amount of string on a spool produced by the Acme is supposed to be normally distributed with a mean of 50 m and a standard deviation of .1 m.

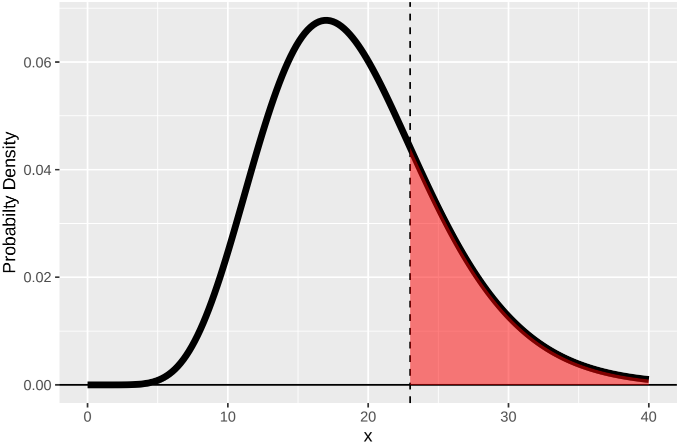
Each day the company tests 20 randomly selected spools. Suppose that the observed standard deviation on one day is .11 m.

Can they conclude that the standard deviation is too high?

Example 26.2 ctd



Example 26.2 ctd



Comparing the Variances of Two Distributions

Example 26.3

The amount of string on a spool produced by the Acme is supposed to be normally distributed with a mean of 50 m.

Each day the company tests 20 randomly selected spools. Suppose that the observed standard deviation is .11 m and .09 m the next.

Can they conclude that the standard deviation is higher on the first day?

Distribution of the Ratio of Sample Variances for two Normal Populations

Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ_1 and variance $\sigma_1^2 < \infty$ and $Y_1, \ldots, Y_m$ be a random sample from a normal distribution with mean mean μ_2 and variance $\sigma_2^2 < \infty$.

Let

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

and

$$S_2^2 = \frac{1}{m-1}\sum_{i=1}^{m}(Y_i - \bar{Y})^2.$$

Then

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n-1,m-1}$$

Properties

- CDF: No closed form
- Mean: $E(X) = \frac{\nu_2}{(\nu_2 - 2)}$, $\nu_2 > 0$
- Variance: $V(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 + 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$, $\nu_2 > 4$

Calculator

`https://stattrek.com/online-calculator/f-distribution`

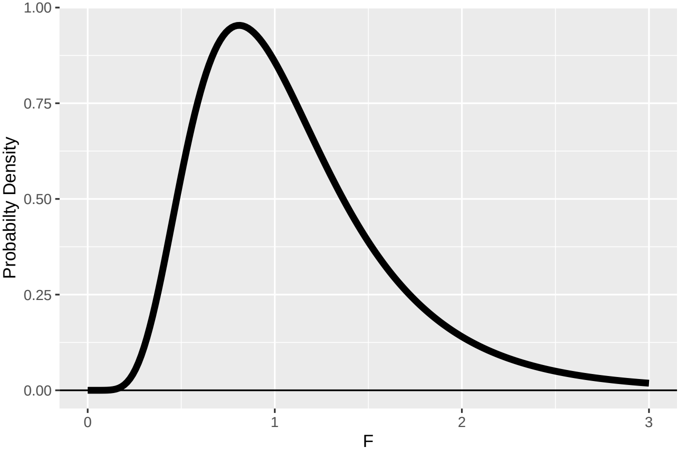
Example 26.3 ctd

The amount of string on a spool produced by the Acme is supposed to be normally distributed with a mean of 50 m.

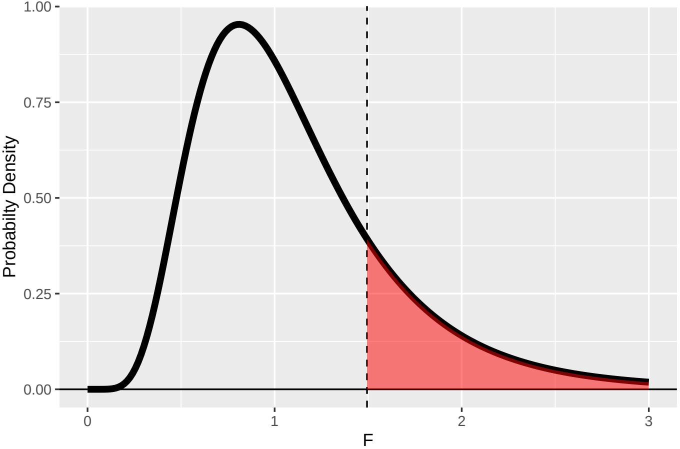
Each day the company tests 20 randomly selected spools. Suppose that the observed standard deviation is .11 m and .09 m the next.

Can they conclude that the standard deviation is different?

Example 26.3 ctd



Example 26.3 ctd



Questions?

Exercise 26.1

The historical average maximum daily temperature in London in October is normally distributed with a mean of 15C. Suppose that the temperatures on each of the 31 days are mutually independent[1]

a) The observed mean in October of this year was 17.76C with a standard deviation of 4.59C. Is it reasonable to believe that the mean is still 15C?

b) The historical standard deviation is 4C. Is it reasonable to believe that this is still true?

c) The standard deviation in September of this year was 3.00C. Can we conclude that daily maximum temperatures in October are more variable than in September?

[1] This is a highly questionable assumption.