Frequentist Inference, pt. III

ECON 3640–001

Marcio Santetti
Spring 2022

Motivation

Housekeeping

Notes based on `Keller (2009)`:

- Chapter **12**, sections `12.1`.

Motivation

So far, we have assumed that the **population standard deviation** (σ) was known when computing confidence intervals and hypothesis testing.

This assumption, however, is **unrealistic**.

Now, we **relax** such belief, and move on using a very similar approach.

The Student t distribution

The Student t distribution

Recall the formula for the z test:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Now that the population standard deviation is *unknown*, what is the **best move**?

Replace it by its **sample estimator**, s!

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

The Student t distribution

Now, we do not know the population standard deviation anymore.

However, frequentist methods still assume that the population mean (μ) is **normally distributed**.

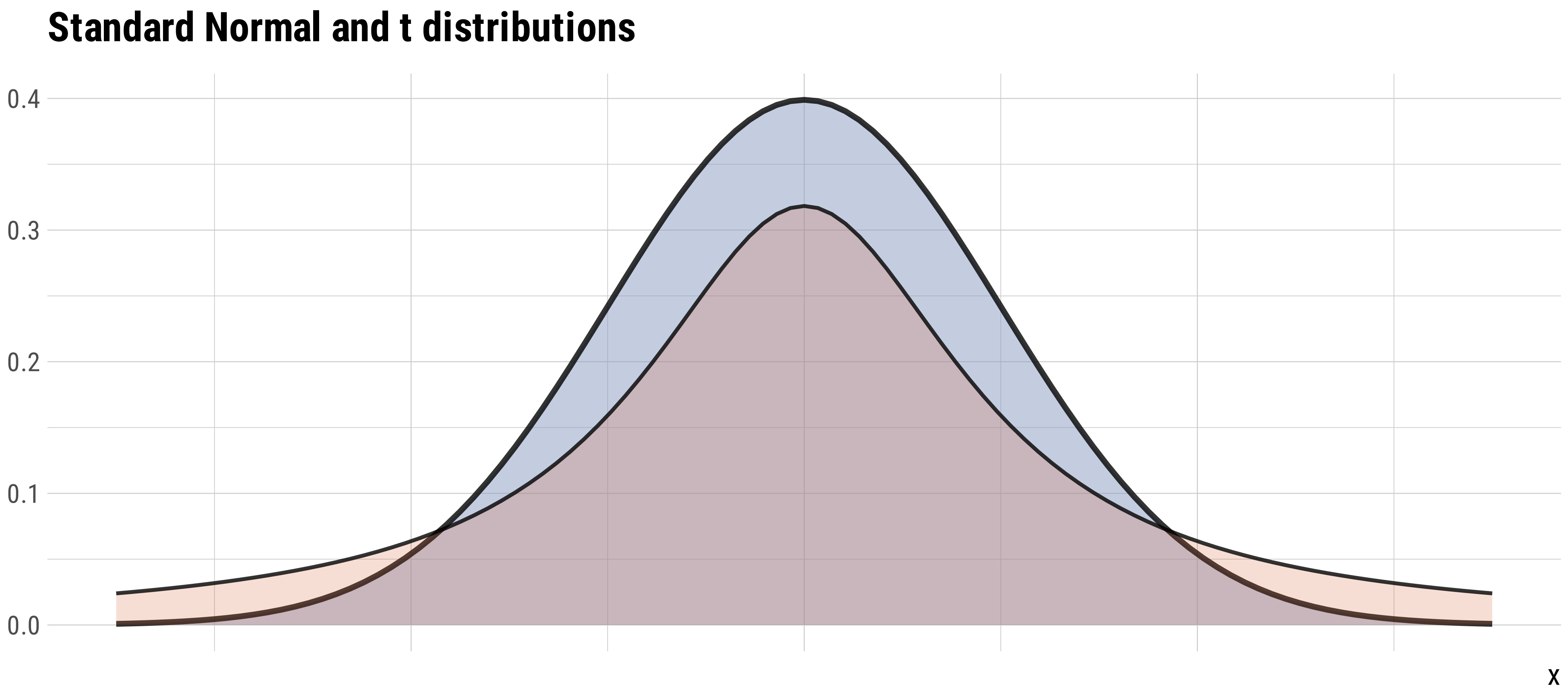
In 1908, William S. Gosset came up with the **Student t** distribution, whose mean and variance are

- $E(X) = 0$
- $\mathrm{Var}(X) = \dfrac{\nu}{\nu - 2}$

where $\nu = n - 1$.

The Student *t* distribution

The *t* distribution has a "*mound-shaped*" density curve, while the Normal's is bell-shaped.

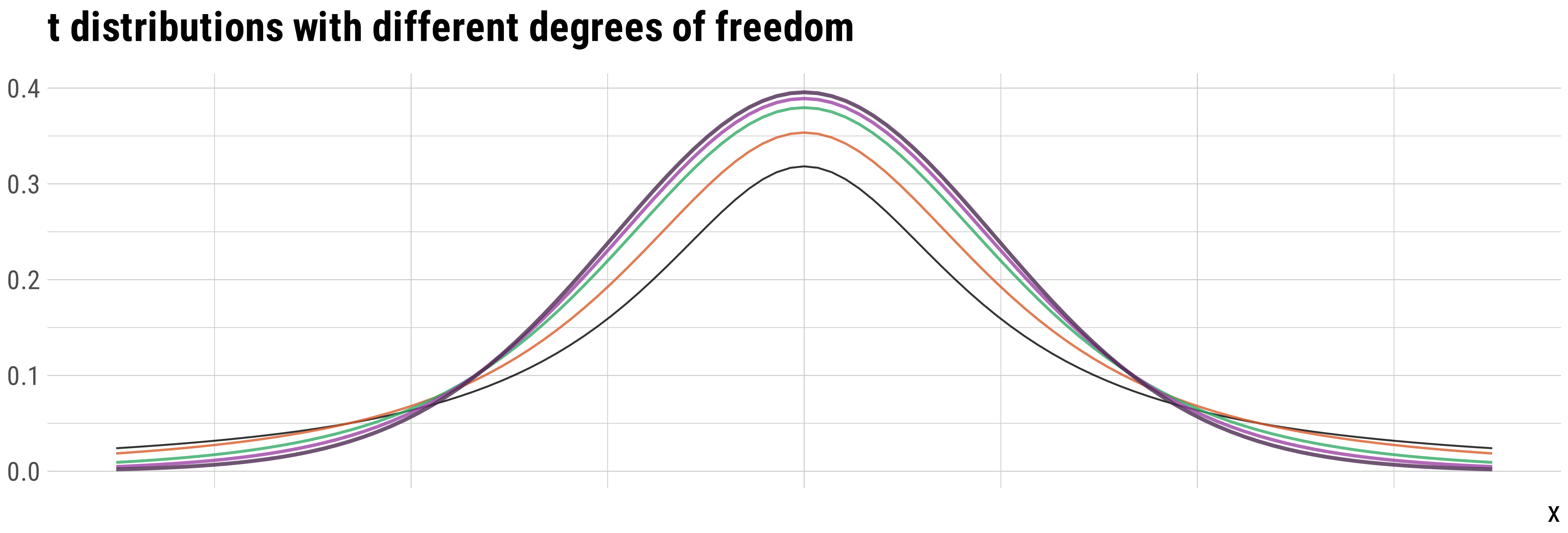


The Student *t* distribution

Its only parameter is ν, the **degrees of freedom** of the distribution:

$$X \sim t(\nu)$$

The larger the value of ν, the more **similar** the *t* distribution is to the Normal.



The Student t distribution

When σ is unknown, we can define the **confidence interval** for the population mean (μ) as:

$$\bar{x} \pm t_{\alpha/2,\,\nu}\left(\frac{s}{\sqrt{n}}\right)$$

Notice that, in addition to the significance level (α), we also must take the number of degrees of freedom ($\nu = n - 1$) into account to find the **standard error** for $\bar{x}$.

Furthermore, for **hypothesis testing**, the *t-statistic* is obtained with:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

The Student *t* distribution

We **cannot** use the sampling distribution of μ anymore, since the population standard deviation is unknown.

Such assumption considers an **infinitely large** sample, which is almost never the case in practice.

For **smaller** sample sizes, the t distribution is extremely useful, and its shape is conditional on this sample size.

But why `(n-1)` degrees of freedom?

The Student t distribution

An example:

Assuming that it can be profitable to recycle newspapers, a company's financial analyst has computed that the firm would make a profit if the mean weekly newspaper collection from each household exceeded 2 lbs. His study collected data from a sample of 148 households. The calculated sample average weight was 2.18 lbs.

Do these data provide sufficient evidence to allow the analyst to conclude that a recycling plant would be profitable? Assume a significance level of 1%, and a sample variance of .962 lbs^2.

Next time: Review & practice problems