Hypothesis Testing

Grinnell College

October 27, 2025

Warm-up

1. What is a sampling distribution, and why does it make sense for us to ask about one?
2. What do the critical values of sampling distribution represent?
3. The critical value for 90% confidence for a t distribution with 15 degrees of freedom is $C = 1.75$. What does this mean?
4. What is a t-statistic? How does a t-distribution relate to a normal distribution?
5. How do I use my t-statistic with a critical value?

Johns Hopkins Example (Review)

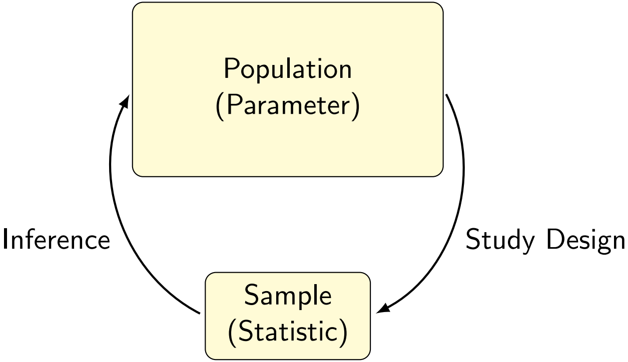
In a study conducted by Johns Hopkins University researchers investigated the survival of babies born prematurely. They searched their hospital's medical records and found 39 babies born at 25 weeks gestation (15 weeks early), 31 of these babies went on to survive at least 6 months. With your group:

1. Use the appropriate t-distribution to construct a 90% confidence interval estimate for the true proportions of babies born at 25 weeks gestation that are expected to survive
2. An article on Wikipedia suggests that 70% of babies born at a gestation period of 25 weeks survive. Is the Johns Hopkins study consistent with this claim?

Goals for Today

1. Introduce mechanics of hypothesis testing
2. Identify null distribution for a given hypothesis
3. Learn to use test statistics to evaluate plausibility of a given hypothesis

The Statistical Framework



Hypothesis Testing

Hypothesis testing involves:

1. Formulating an *unambiguous* statement about a population parameter, called our **null hypothesis**
2. Collecting observational or experimental data
3. Determining if the data collected is consistent with our hypothesis
4. Either *rejecting* or *failing to reject* our hypothesis based on the *strength* of the evidence

Null Hypothesis

Our hypothesis about a parameter prior to seeing any data is called our **null hypothesis**, typically expressed in the form

$$H_0 : \mu = \mu_0$$

where μ_0 ("mew-naught") represents a specific quantity.

For example, when considering the Wikipedia claim that the 6 month survival rate of preterm children is 70%. We could express this claim as

$$H_0 : p = p_0 = 0.7$$

Test statistics

We relate the data that we have observed (i.e., $\overline{x}$, $\hat{\sigma}$) with our null hypothesis with the use of **test statistics**

For example, if $H_0 : \mu = \mu_0$ is correct, then

$$t = \frac{\overline{x} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

will be centered at zero and will follow a t-distribution. If H_0 is *not* correct, the distribution of t statistics will be centered at $\overline{x} - \mu_0$ instead

The **null distribution** describes the distribution that our test statistics will follow if the null hypothesis is true

Hypothesis Testing and Confidence

How do we go about using a null distribution?

Assuming the null hypothesis is true, we know what distribution our statistic should follow and, accordingly, the *critical values* associated with the bounds of our distribution

Test statistics that fall sufficiently outside of the bounds of where we expect our data to fall may be considered evidence *against* the null hypothesis

The General Idea

The general idea is this:

1. We specify some null hypothesis $H_0 : \mu = \mu_0$
2. *Assuming that the null hypothesis is true*, the test statistic

$$t = \frac{\overline{x} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

   will follow a t-distribution, centered at 0
3. From our data, we will compute a test statistic using both $\overline{x}$ and μ_0
4. We will then check our test statistic t against critical values C
5. If $C < |t|$, our test statistically is sufficiently far from what we expect and we reject our null hypothesis

Example

From our Johns Hopkins study, we found that

$$\hat{p} = \frac{31}{39} = 0.795, \qquad SE = \sqrt{\frac{0.795(1 - 0.795)}{39}} = 0.065$$

Consider two competing hypotheses for the true proportion of babies expected to survive at 6 months:

1. $H_{0_1} : p_{0_1} = 0.7$
2. $H_{0_2} : p_{0_2} = 0.95$

We can construct test statistics to determine how our observed data relates to each hypothesis

Hypothesis 1

If the first null hypothesis, H_{0_1} were true and we were to repeatedly collect samples to find $\hat{p}$, the statistic

$$\frac{\hat{p} - p_{0_1}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

would follow a t-distribution and would be centered at 0. For the data we *actually* observed, we find a statistic of

$$t = \frac{\hat{p} - p_{0_1}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{0.795 - 0.7}{0.065} = 1.4615$$

indicating that our observed data is about 1.4615 “standard deviations” (kinda) away from the expected mean

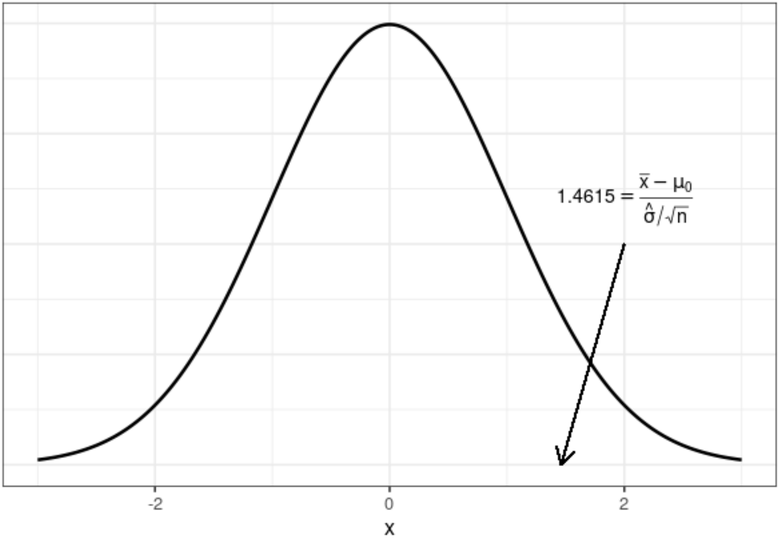
Hypothesis 1

For $t = 1.4615$, we can assess at which confidence levels we would *reject* our null hypothesis and at which levels we would fail to reject.

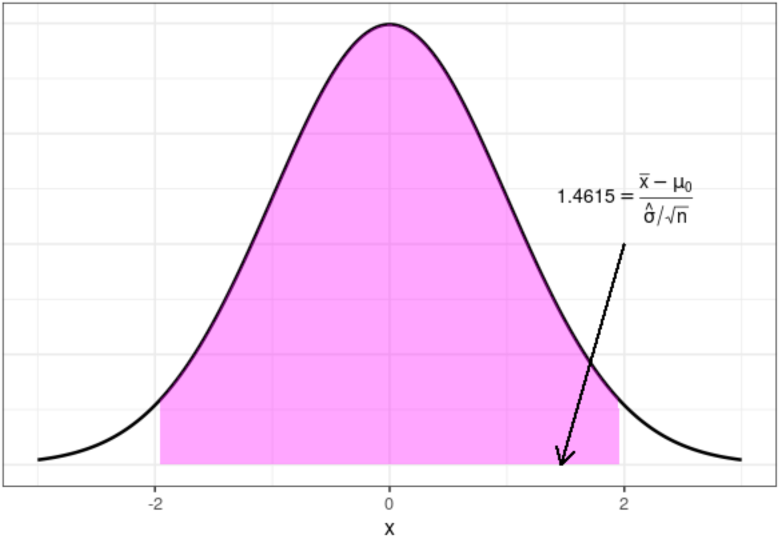
```
> qt(0.9, df = 38)
[1] 1.3042
> qt(0.95, df = 38)
[1] 1.686
> qt(0.975, df = 38)
[1] 2.0244
```

We see that if our confidence level was 80%, $t > C$, causing us to reject. However for 90% and 95% we have $t < C$, leading us to fail to reject

t distribution when $\mu_0 = 0.7$



t distribution when $\mu_0 = 0.7$



Hypothesis 2

It is absolutely critical to remember that our t-statistics *and* the results of our test depend entirely on the null hypothesis we wish to investigate.

For example, if instead we were testing the second hypothesis, $H_{0_2} : p_{0_2} = 0.95$, we would find a t-statistic of

$$t = \frac{0.795 - 0.95}{0.065} = -2.385$$

Relative to this second hypothesis, our t-statistic is larger, indicating that our observed sample mean is further from the mean we would expect if the null hypothesis were true

For our second hypothesis with $t = -2.385$, we see that the thresholds are the same but the decisions we come to may be quite different

```
> qt(0.9, df = 38)
[1] 1.3042
> qt(0.95, df = 38)
[1] 1.686
> qt(0.975, df = 38)
[1] 2.0244
> qt(0.995, df = 38)
[1] 2.7116
```

In fact, for all confidence intervals less than 99% we would reject our null hypothesis

Review Steps

- Formulate a null hypothesis H_0
- Use this and your sample data to construct a test statistic (i.e., t-statistic)
- If the null hypothesis is true, the t-statistic will follow a t-distribution centered at zero. This is our *null distribution*
- Find the critical values of the null distribution, i.e., *if the null hypothesis were true*, what are the bounds of where we would expect to see our data?
- Compare your statistic to the critical values. If our statistic exceeds those bounds, we reject H_0