Hypothesis Testing

Grinnell College

April 2, 2025

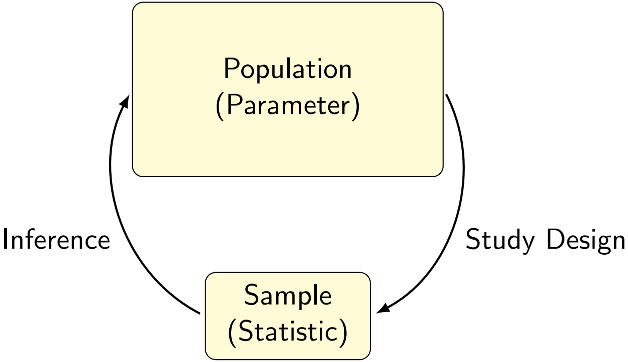
Warm-up

blah

Goals for Today

What is hypothesis test Null hypothesis/null distribution what is t statistic

The Statistical Framework



Hypothesis Testing

Hypothesis testing involves:

1. Formulating an *unambiguous* statement about a population parameter (i.e., the true survival rate at 6 months is 0.7)
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 prior to seeing any data is called the **null hypothesis**, denoted H_0. For example, the null hypothesis that the survival rates of preterm children is 0.7 can be expressed as $H_0 : p = 0.7$

When considering hypotheses regarding the relationship of two or more variables, we typically take the assumption of there being no effect, difference, change, or relationship between them. That is, we begin with the "status quo" and it becomes incumbent upon the evidence to suggest otherwise

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

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. $H_{0_1} : p_1 = 0.7$
2. $H_{0_2} : p_2 = 0.95$

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

$$\frac{\hat{p} - p_1}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

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

$$t = \frac{\hat{p} - p_1}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{0.795 - 0.7}{0.065} = 1.4615$$

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

Hypothesis 1

Is the data we collected consistent with our null hypothesis? In other words, is the statistic $t = 1.4615$ about what we would expected or is it "far away"?

Our threshold for "far away" is going to depend on the amount of *confidence* we wish to have in our decision. We can determine the cutoffs for these threshold with our *critical values*

Once we have determined our thresholds, we can say that t-statistics between our critical values are *consistent* with our null hypothesis, while those outside the bounds are inconsistent

Confidence Levels – Quick Aside

There is a one to one relationship between the the critical values we have used for CI and the confidence we have in our test

In particular, a null hypothesis can be correct yet, by chance, we observe a test statistic that is unlikely and falls outside the range of values we expect to see

This may result in us incorrectly rejecting a null hypothesis that is actually true. This is called a *Type I Error*, which we will discuss in more detail later

Confidence Level – Quick Aside

For example, suppose we specify a confidence level of 90%, indicating we want the range of where 90% of our statistics would fall *if the null hypothesis were true*

By definition, 10% of observations will fall outside of this interval. As such, *if the null hypothesis were true*, there is a 10% chance we see a statistic outside these bounds

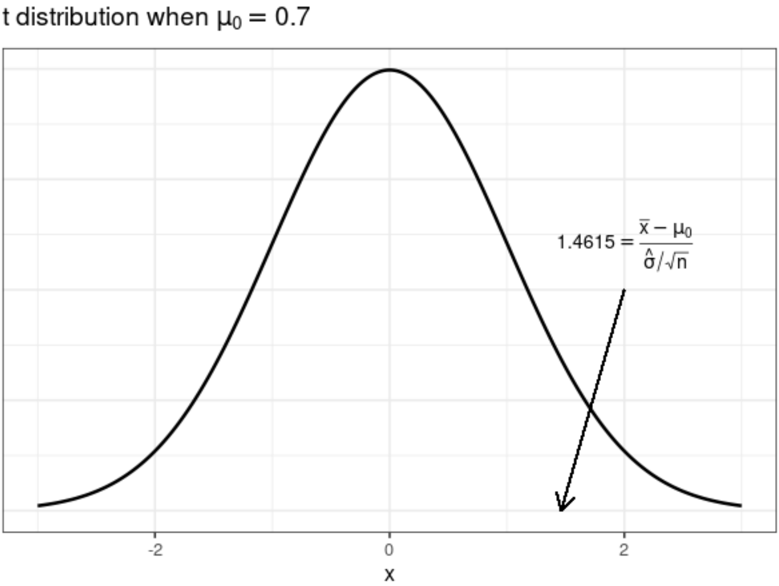
This equates to there being a 10% chance that we incorrectly reject our null. Said another way, our *Type I error rate* is 10%

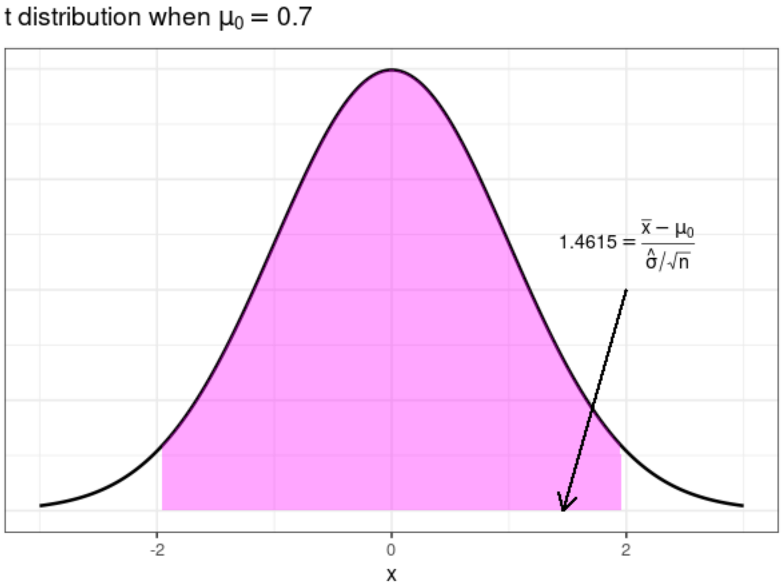
Hypothesis 1

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 μ0 = 0.7
1.4615 = (x̄ − μ0) / (σ̂/√n)
-2
0
2
x


t distribution when $\mu_0 = 0.7$
$1.4615 = \frac{\bar{x} - \mu_0}{\hat{\sigma}/\sqrt{n}}$
-2
0
2
x

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, $H_{0_2} : p_2 = 0.95$, 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