Social Statistics

Introducing Hypothesis Testing

October 27, 2021

Introducing Significance Testing

Confidence Intervals: Based on our sample estimate and confidence level, what is the range of possible values for population parameter?

Shift to significance: How unusual - how unlikely - is our sample estimate?

More specifically: What is the probability (.1, .05, .01, etc.) of getting a value at least as extreme as our estimate?

Can we be confident at a certain level of probability that the difference between our estimate and another value is *statistically significant*, meaning not just from random chance?

Parts of the Significance Test

Assumptions: Random, Normal, Large, Increasing Validity

Build two hypotheses covering the entire range

Null Hypothesis: What we are testing *against*

- Null is usually 0 but can be any value
- $H_0 : \hat{\mu} = 0$

Alternative Hypothesis: What we estimate from our data

- Estimated mean is not 0
- $H_A : \hat{\mu} \neq 0$

Parts of the Significance Test

Compute a test statistic for our estimate

Previously we wanted distance from the mean in SDs

- That formula: $z = \frac{x - \mu}{\sigma}$

Now we want distance from the null hypothesis' value in SEs

- This formula: $t = \frac{\mu_x - \mu_0}{SD_x / \sqrt{n}}$
- t adjusts for *degrees of freedom*
- Denominator = standard error
- Fill in μ_0 from null hypothesis (0)

From Z to T

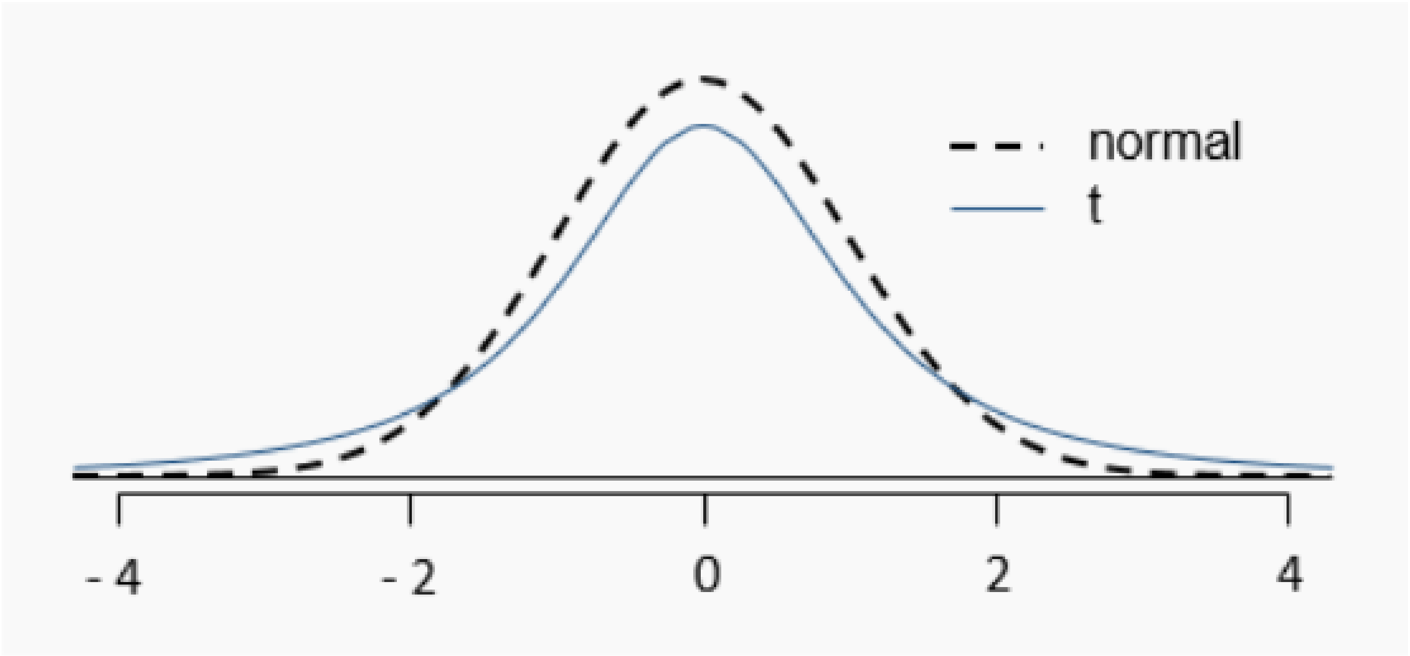
When the population standard deviation is unknown (almost always), the uncertainty of the standard error estimate is addressed by using a new distribution: the t distribution.

This distribution also has a bell shape, but its tails are thicker than the normal model's

Therefore observations are more likely to fall beyond two SDs from the mean than under the normal distribution

These extra thick tails are also helpful for resolving our problem with a less reliable estimate and a bigger standard error (if n is small)

From Z to T



From Z to T

When the sample size is big enough, z and t are the same.

When in doubt, use t and assert the degrees of freedom. It will always work.

```r
qnorm(.975)
```

```
## [1] 1.959964
```

```r
qt(.975, df = 1000)
```

```
## [1] 1.962339
```

```r
qt(.975, df = 100)
```

```
## [1] 1.983972
```

From Z to T

Other `norm()` functions also work with `t()`:

```
pt(1.962339, df = 1000)
```

```
## [1] 0.975
```

```
1 - pt(1.983972, df = 100)
```

```
## [1] 0.02499997
```

Back to the Significance Test

Calculate the test statistic, and convert it to a probability (a p-value) using `pt()` if negative or `1 - pt()` if positive.

- Multiply this probability by two to get the probability for each tail.

That p-value is probability of observing a value of the test statistic at least as extreme as the absolute value of our estimate

- A smaller p-value presents stronger evidence against H_0.

Careful: not the probability that the null hypothesis is true

Parts of the Significance Test

Choose significance level

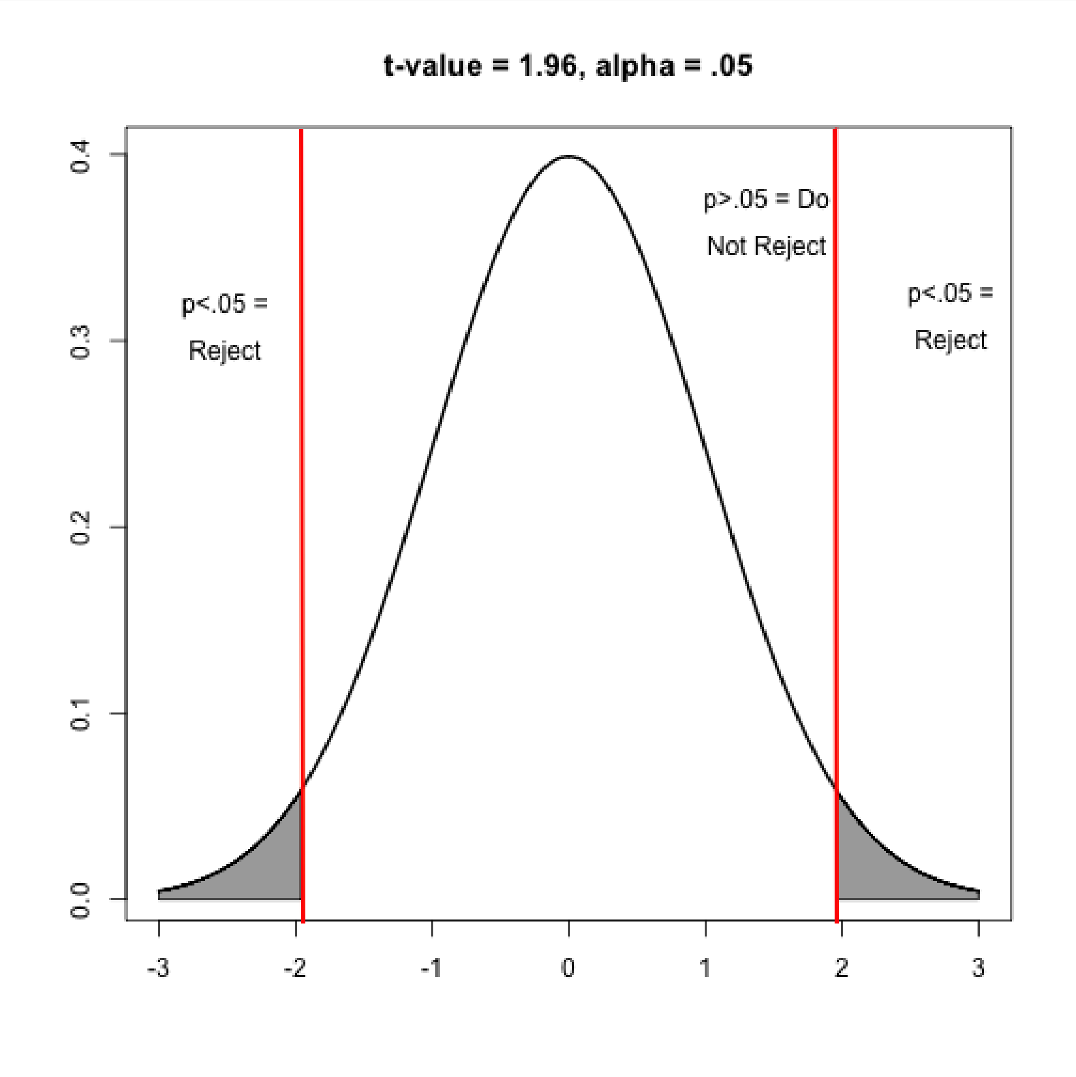
- Alpha α is significance level, or level of uncertainty, against which we want to evaluate the test statistic
- Also known as rejection region or critical region

We will reject the null hypothesis if test statistic falls within this rejection/critical region

- But we will not *accept* the null hypothesis if test statistic falls outside this region

Described in terms of probability of being *beyond* the range of values that are more likely, so we say .05 rather than .95

Critical Region



Significance Test for Mean - Example

Imagine we want to test if the average number of children a GSS respondent has in 2018 is different from the average from 1984. Why might we expect people have different numbers of children in these two periods?

Using `gss_week7.csv` file on Canvas, let's test if the mean value of `childs` in 2018 differed from the 1984 mean of 2.2 (which we know from previous research).

What was the mean for `childs` in 2018?

```
mean(gss_week7$childs[gss_week7$year==2018])
```

```
## [1] 1.998255
```

Significance Test for Mean - Example

Null Hypothesis = Actual mean is 2.2 (filled in from 1984)

Alternative Hypothesis = Actual mean is *not* 2.2

- Note, if alternative were that the mean is 2, hypotheses would not cover entire range!

Significance level = .05

Significance Test for Mean - Example

$$t = \frac{y - y_{H0}}{SD/\sqrt{n}}$$

$$t = \frac{1.998 - 2.2}{0.05951692}$$

t = -3.389708

In words: sample mean of 1.998 is 3.39 *standard errors* below the null hypothesis value.

Significance Test for Mean - Example

What is the t-value for the critical region for our sample size and an alpha level of .05?

There are 573 observations in the 2018 sample, so we use 572 for the degrees of freedom

```
qt(.975, df = 572)
```

```
## [1] 1.96412
```

For mean to be significant at .05 level, the test statistic needs to be at least 1.96412 standard errors away. Our test statistic of 3.39 is more extreme than 1.96412, so we can reject the null hypothesis that the mean number of children in 2018 differs from 2.2.

Significance Test for Mean - Example

We would reach the same conclusion using p-values. The test statistic of 3.39 is negative, so we can use `pt()` instead of `1 - pt()`:

```
pt(-3.39, df = 572)
```

```
## [1] 0.0003735504
```

```
# Multiply By 2 For Both Tails:

2 * 0.0003735504
```

```
## [1] 0.0007471008
```

The p-value for our hypothesis test is less than .05, so we can reject the null hypothesis.

Significance Test for Mean - Example

We would also reach the same conclusion using confidence intervals.

```
idealchilds_ci
```

```
## [1] 1.881356 1.998255 2.115153
```

The null hypothesis value of 2.2 falls outside our confidence interval, so we can reject the null hypothesis.

Big Shortcut: t.test()

```r
t.test(gss_week7$childs[gss_week7$year==2018],
      mu = 2.2)
```

```
##
##      One Sample t-test
##
## data:  gss_week7$childs[gss_week7$year == 2018]
## t = -3.3897, df = 572, p-value = 0.0007479
## alternative hypothesis: true mean is not equal to 2.2
## 95 percent confidence interval:
##  1.881356 2.115153
## sample estimates:
## mean of x
##  1.998255
```

Significance Test for Mean - Example

Red box gives the test statistic. Blue box gives the probability of getting a value more extreme than the test statistic. Green box is the 95% confidence interval. Orange box gives the sample mean.

```
        One Sample t-test

data:  gss_week7$childs[gss_week7$year == 2018]
t = -3.3897, df = 572, p-value = 0.0007479
alternative hypothesis: true mean is not equal to 2.2
95 percent confidence interval:
 1.881356 2.115153
sample estimates:
mean of x
 1.998255
```

Exercise

Let's look at 2018's mean ideal number of children (variable = `chldidel`)

Using `t.test()`, can we reject the null hypothesis that the 2018 mean is the same as the 1990 mean of 2.55 at the .05 alpha level?

```
t.test(gss_week7$chldidel[gss_week7$year==2018],
      mu = 2.55)
```

Exercise

```
##
##      One Sample t-test
##
## data:  gss_week7$chldidel[gss_week7$year == 2018]
## t = 1.5856, df = 572, p-value = 0.1134
## alternative hypothesis: true mean is not equal to 2.55
## 95 percent confidence interval:
##  2.535483 2.686158
## sample estimates:
## mean of x
##   2.61082
```

Significance Test for Mean - Exercise

No, do not reject the null hypothesis:

- The test statistic is not more extreme than -1.97
- The p-value is greater than .05
- The confidence interval does include null hypothesis value of 2.55

Other t.test() Options

Can we reject the null hypothesis that the 2016 mean is the same as 2.5 at the .01 alpha level? We can change the default level of .05 to .01 using the conf.level option (which requires the confidence level, so .99 for the .01 alpha level).

```
#To Change Confidence Level:

t.test(gss_week7$chldidel[gss_week7$year==2016],
      mu = 2.5, conf.level = .99)
```

Other t.test() Options

```
#To Change Confidence Level:

t.test(gss_week7$chldidel[gss_week7$year==2016],
      mu = 2.5, conf.level = .99)
```

```
##
##      One Sample t-test
##
## data:  gss_week7$chldidel[gss_week7$year == 2016]
## t = 1.5082, df = 720, p-value = 0.1319
## alternative hypothesis: true mean is not equal to 2.5
## 99 percent confidence interval:
##  2.466898 2.626029
## sample estimates:
## mean of x
##  2.546463
```

Inference For Single Proportions

- Means and proportions have different distributions, standard errors and hypothesis tests.
- For proportions, we'll use `prop.test()` rather than `t.test()`.
- Example: Does the proportion of respondents whose number of children is equal to their ideal number of children differ from .33?

Inference For Single Proportions

- First step, create binary variable capturing respondents whose number of children is equal to their ideal number of children. Call it `has_ideal`.

```
gss_week7 <- gss_week7 |>
  mutate(has_ideal = ifelse(childs == chldidel, 1, 0))
```

Inference For Single Proportions

- For the test, we will need the frequency with a 1 and the total in the sample.

```
addmargins(table(gss_week7$has_ideal))
```

```
##
##    0    1  Sum
## 2034  889 2923
```

Inference For Single Proportions

- Enter those two values in `prop.test()` along with the null hypothesis value you want to test. The function calculates the proportion and compares it to the null hypothesis value.

```
prop.test(889, 2923, p = .33)
```

Inference For Single Proportions

```r
prop.test(889, 2923, p = .33)
```

```
##
##     1-sample proportions test with continuity correction
##
## data:  889 out of 2923, null probability 0.33
## X-squared = 8.7246, df = 1, p-value = 0.003139
## alternative hypothesis: true p is not equal to 0.33
## 95 percent confidence interval:
##  0.2875599 0.3212387
## sample estimates:
##         p
## 0.3041396
```

Another Example

Does the proportion of respondents with less children than their ideal number differ from .53 at the 99% confidence level?

Another Example

```
gss_week7 <- gss_week7 |>
  mutate(less_ideal = ifelse(childs < chldidel, 1, 0))
```

```
addmargins(table(gss_week7$less_ideal))
```

```
##
##    0    1  Sum
## 1437 1486 2923
```

Another Example

```r
prop.test(1486, 2923, p = .53, conf.level = .99)
```

```
##
##      1-sample proportions test with continuity correction
##
## data:  1486 out of 2923, null probability 0.53
## X-squared = 5.3975, df = 1, p-value = 0.02017
## alternative hypothesis: true p is not equal to 0.53
## 99 percent confidence interval:
##  0.4844006 0.5323247
## sample estimates:
##         p
## 0.5083818
```