Three More Inferences

Grinnell College

November 2025

Big Picture

The last few slide decks we have been focusing on testing if a population mean is the same as BLANK or building confidence intervals

We are now going to *generalize* the tests to....

1. The difference between two means
2. One proportion
3. The difference between two proportions

Critically, they work almost the exact same as a z-test stuff....

2 Inferences: Hypothesis Test

Almost all hypothesis tests carry the same format...

- State our null hypothesis (default state of the world we want to disprove) and alternative (thing we are trying to show)
- After checking assumptions we have a sampling distribution.
  - For us it'll be normal so long as we know σ^2

Difference of Two Means

Set up: We have two distinct subpopulations and are interested in the difference of their means. Eg are Dream Island penguin's mean flipper length the same as Toregersen Island penguin's flipper length?

H_0: $\mu_1 - \mu_2 = 0$ OR $\mu_1 - \mu_2 \geq 0$ OR $\mu_1 - \mu_2 \leq 0$

H_A: $\mu_1 - \mu_2 \neq 0$ OR $\mu_1 - \mu_2 < 0$ OR $\mu_1 - \mu_2 > 0$

The alternative is below it's corresponding null hypothesis.

It's possible to add a constant to one side (eg H_0: $\mu_1 - \mu_2 = 3$) but rarely done in practice

---

$$H_0 : \mu_{Dream\ Island} = \mu_{Toregersen}$$
$$H_A : \mu_{Dream\ Island} \neq \mu_{Toregersen}$$

Difference in 2 Means: Assumptions

Assumptions:

- Randomly collected data
- Independent and identically distributed
  - We assume they have the same mean for the moment
  - We do NOT assume they have the same variance (ie this is the unequal variance t-test; also called the Welch t-test)
- Both sample sizes (from Dream Island and Toregeresen Island) are large or the populations are both normal
  - Have to check for both groups!
  - Often indicate relevant statistics with a subscript to denote the two groups
  - Eg n_D and n_T for the sample size from Dream Island and Toregeresen Island, respectively

Test Stat

$$\text{General Form:} = \frac{\text{Observed - Hypothesized}}{\text{Standard Error}}$$

$$\text{Z-test statistic} = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}}$$

$$\text{t-test statistic} = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$$

$$\text{t-test stat for 2 means} = \frac{(\bar{x_D} - \bar{x_T}) - (\mu_D - \mu_T)}{\sqrt{\frac{s_D^2}{n_D} + \frac{s_T^2}{n_T}}} \quad = \quad \frac{\bar{x_D} - \bar{x_T}}{\sqrt{\frac{s_D^2}{n_D} + \frac{s_T^2}{n_T}}}$$

NOTE: We are assuming H_0 is true so μ_D - $\mu_T = 0$

Df and P-values

Degrees of freedom is actually super complicated to calculate in this situation (see Satterthwaits approximation for the gory details)

In practice computers calculate the degrees of freedom at the same time it calculates the p-value (and honestly the test statistic) which brings us to a new R command.... t.test()

From R...

- test statistic $= 1.701$
- degrees of freedom $= 111.37$
- p-value $= 0.0916$

So we have weak evidence the mean flipper length is different between the two islands.

Difference in Means Confidence Interval

And we can give a range of what we think the difference in means might be.

$$\bar{x}_1 - \bar{x}_2 \;\pm\; t_{df}\sqrt{\frac{s_D^2}{n_D} + \frac{s_T^2}{n_T}}$$

Again, we just use R because finding the degrees of freedom (and by extension t_{df}) is complex...

(-0.309, 4.062)

We are 95% confidence the true difference between flipper lengths from Dream Island Penguins to Toregersen Island Penguins is between -.309 and 4.062 mm.

Difference in Means Summary

Has similar inference as that for 1 mean.

This is definitely the point where R needs to be used to run this test effectively.

Big picture for me is to think about the (difference of means) as

- A single variable
- with mean 0 (for hypothesis test)
- or mean $\bar{x_1} - \bar{x_2}$ (for confidence intervals)
- and a "pooled standard error"

Ex: Wool Breaks

The number of breaks in yarn given the type of yarn it was made of (type A or type B). To do this, 54 batches of wool were formed into yarn either using type A method or type B method. 27 bundles of yarn were made in both type A and type B style. The resulting bundles were tested on a loom and the total number of breaks found.

Find the

1. Factor(s) and it's levels
2. Experimental unit (= observation unit for this example)
3. A t-test to see if there is a difference in the two means
4. A 90% confidence interval around the differences in means

Switching Gears

We are now going to go *back* to the z-test.

What is the least normal distribution you can think of?

Switching Gears

We are now going to go *back* to the z-test.

What is the least normal distribution you can think of?

Probably an indicator variable that is 0/1?

Z-test Review: Sampling Distribution

Previously, given our data was...

1. randomly selected
2. independent and identically distributed
3. pop is normal or the sample is large

our sampling distribution would be.....

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

Pop Question

What's the mean for an indicator variable called?

Pop Question

What's the mean for an indicator variable called?

The mean of an indicator variable is called a **proportion**

Parameter (population proportion): p

Statistic (sample proportion): $\hat{p}$

Pop Question

What's the mean for an indicator variable called?

The mean of an indicator variable is called a **proportion**

Why on earth am I defining what a proportion is?

Pop Question

What's the mean for an indicator variable called?

The mean of an indicator variable is called a **proportion**

Why on earth am I defining what a proportion is?

Because you test a proportion the same way you test a mean with known variance, the z-test

One last background piece....

Earlier I said the normal model can be used to approximate several other distributions.

One of those distributions is the *binomial distribution* which is (number of success) our of (number of trials).

There is some weird things about the binomial distribution....

- Defined by the probability of success p and number of trials n
- Ie doesn't mention it's variance (unlike Normal)
- The variance is instead a function of p and n
  - $np(1-p)$
- Short hand is Bin(p, n) or Binom(p, n) or Binomial(p, n)

Inference for a proportion

First the sampling distribution and then we will talk about it's (not different) assumptions

---

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

So the mean of a sample (be it normal or an indicator) will follow a normal distribution centered at the population's mean and with a (known) variance

1 Prop: Assumptions Unchanged

For a sample's mean...

1. The sample was randomly collected
2. The observations are independent and identically distributed
3. The population is normal or n is large

For a sample's proportion...

1. The sample was randomly collected
2. The observations are independent and identically distributed
3. ~~The population is normal or~~ n is large
   - My variable is an indicator so clearly not normal
   - How large n should be now follows different guidelines

1 Prop: How large of n do we need *now*??

The needed size for n is dictated by p actually....

- $np \geq 10$ AND
- $n(1-p) \geq 10$

They must both be passed which will happen, for any p not 0 or 1, given n is sufficiently large.

Alternative way to think about this condition: Each category needs 10 observations (and this extends to more complicated situations, like the multinomial)

1 Prop: Hypothesis Test Statistic

For sample means, we assumed we knew μ....

$$\frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

For sample proportions, we assume we know p....

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

1 Prop: P-value and Decision

P-value calculation remains the same as for the sample mean

- Calculate the probability in the tails of a normal distribution
- Which tail (or both tails) is dependent on H_A

Decisions need to be stated in terms of a proportion

Eg: "We have strong evidence to suggest the proportion of general townsfolk of Grinnell are more religious than the student body"

1 Prop: Confidence Intervals

General Formula:

$$(estimate) \ \pm \ (distribution\ value)(st.\ error)$$

Sample Mean:

$$\bar{x} \ \pm \ z_{1-\alpha/2}(\sqrt{\frac{\sigma^2}{n}})$$

Sample Proportion:

$$\hat{p} \ \pm \ z_{1-\alpha/2}(\sqrt{\frac{p(1-p)}{n}})$$

Caveat on Assumptions for Confidence Intervals

There is one thing that is annoying when it comes to the "n is large"

- For means we just gave guidlines on how not-normal the sample looked
- For 1 proportions hypothesis test we said np and $n(1-p) > 10$
  - In HT we assume we know p

Confidence intervals don't assume we know the true p so what can we do?

Caveat on Assumptions for Confidence Intervals

There is one thing that is annoying when it comes to the "n is large"

- For means we just gave guidlines on how not-normal the sample looked
- For 1 proportions hypothesis test we said np and $n(1-p) > 10$
  - In HT we assume we know p

Confidence intervals don't assume we know the true p so what can we do?

Let's just plug in $\hat{p}$, our best guess at what p is

Large n Summary

For a sample mean...

Shape	needed n
Symmetic, not outliers	20-ish
Skewed	50-ish
Most anything	100-ish

For a sample proportion HT with a hypothesized proportion p....

Formula	size
np	≥ 10
$n(1-p)$	≥ 10

For a sample proportion CI....

Formula	size
$n\hat{p}$	≥ 10
$n(1-\hat{p})$	≥ 10

1 Prop: Example

Do NFL teams win more often when playing at home? That is, is $p_{winning\ at\ home} = .5$?

H_0:

H_A:

1 Prop: Example

Do NFL teams win more often when playing at home? That is, is $p_{winning\ at\ home} = .5$?

H_0: $p \leq .5$

H_A: $p > .5$

(Are we allowed to write μ? Technically yes but I'll advise against it...it feels odd and isn't considered standard)

1 Prop Example: Assumptions

- Random Sample:
  - Yes but all games are from 2024 which limits our conclusions
- Independent and Identically Distributed:
  - Identically Distributed: sure? No glaring reason why not
  - Independent: Yes; ignoring the idea teams have “runs” of wins
    - ⋆ Eg Bobby Fisher won 20 games straight in the 1972 Chess Candidates Tournament, his first (and only) loss was in the final against former champ Petrosian
- Large n:
  - ?
  - ?

1 Prop Example: Assumptions

- Random Sample:
  - Yes but all games are from 2024 which limits our conclusions
- Independent and Identically Distributed:
  - Identically Distributed: sure? No glaring reason why not
  - Independent: Yes; ignoring the idea teams have “runs” of wins
- Large n:
  - $np \geq 10$
  - $n(1-p) \geq 10$
  - Ie all categories/groups have 10+ observations under the null hypothesis

1 Prop Example: Assumptions

- Random Sample:
  - Yes but all games are from 2024 which limits our conclusions
- Independent and Identically Distributed:
  - Identically Distributed: sure? No glaring reason why not
  - Independent: Yes; ignoring the idea teams have "runs" of wins
- Large n:
  - $544 * .5 \geq 272$
  - $544(1 - .5) \geq 272$
  - Ie all categories/groups should have 10+ observations under the null hypothesis

1 Prop Example: Sampling Distribution

Having passed the assumptions our sampling distribution is...

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

And we know p and n so....

$$\hat{p} \sim N(.5, \frac{.5(1-.5)}{544})$$

$$\hat{p} \sim N(.5, .0004595)$$

1 Prop Example: test stat, p-value and decision

Test Statistic:

$$\frac{\hat{p} - p}{\sqrt{\frac{p*(1-p)}{n}}} = \frac{.543 - .5}{.0214} = 1.976$$

Recall H_A: $p > .5$ so we want the right tail of a normal distribution:

$P(Z > 1.976) = .024$

We have moderate evidence to suggest that the win rate for home teams is greater than .5.

1 Prop Example: 95% Confidence Interval

What changes for our assumptions?

1 Prop Example: Confidence Interval

What changes for our assumptions?

We don't assume we know p and instead have to use our estimate $\hat{p}$. BE CAREFUL! This is a common mistake/mess up.

- $n\hat{p} = 544 * .542 = 295 \geq 10$
- $n(1 - \hat{p}) = 544 * (1\text{-}.542) = 249 \geq 10$

with an estimated sampling distribution being...

$$\hat{p} \sim \; N(\hat{p}, \; \frac{\hat{p}(1-\hat{p})}{n})$$
$$\hat{p} \sim \; N(.542, \; 0.000456)$$

1 Prop Example: 95% Confidence Interval

$$estimate \pm (distributional\ value)(standard\ error)$$

$$\bar{x} \pm z_{1-\alpha/2}(\sqrt{\frac{\sigma^2}{n}})$$

$$\hat{p} \pm z_{1-\alpha/2}(\sqrt{\frac{p(1-p)}{n}})$$

but since we don't know p we use our best guess, $\hat{p}$....

$$\hat{p} \pm z_{1-\alpha/2}(\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$$

$$0.5422 \pm 1.96(.0214)$$

(.5003, .5841)

We are 95% confident the true proportion of home team wins is between .5003 and .5841.

One Last One....

We had one mean

We then had two means

We now have one proportion

Guesses?

One Last One....

We had one mean

We then had two means

We now have one proportion

Guesses?

2 proportions!

Major Changes

- Need to check large n for both groups
  - hypothesis test: use hypothesized difference in the proportion
  - Confidence intervals: use sample/estimated difference

- Pooled proportion, p_{pooled} is the proportion from the population, IGNORING GROUPS!!
  - Similarly, pooled sample proportion, $\hat{p}_{pooled}$ is the proportion from the sample, igorning groups
  - $$p_{pooled} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

- Standard Error gets even weirder
  - Function of p_1, n_1, p_2, and n_2

2 Proportions: HT

Set up: We have two distinct subpopulations and are interested in the difference of their ~~means~~ proportions. Eg Is the home game winning rate for underdogs different than the home game winning rate for the favored team?

H_0: p_1 - $p_2 = 0$ OR p_1 - $p_2 \geq 0$ OR p_1 - $p_2 \leq 0$

H_A: p_1 - $p_2 \neq 0$ OR p_1 - $p_2 < 0$ OR p_1 - $p_2 > 0$

The alternative is below it's corresponding null hypothesis.

It's possible to add a constant to one side (eg H_0: p_1 - $p_2 = .25$) but rarely done in practice

---

$$\begin{aligned} H_0 &: p_{underdogs} = p_{favorites} \\ H_A &: p_{underdogs} \neq p_{favorites} \end{aligned}$$

Assumptions

- Random Sample: Same as before
- IID: Same as before
  - NOTE: It's okay iif the two groups have different means/proportions
  - Don't want lurking variables affecting them
  - NOTE 2: home team cannot be both the favorites and underdogs so the proportions shouldn't affect each other
- Large n
  - $n_1 p_1 \ > \ 10$
  - $n_1(1 - p_1) \ > \ 10$
  - $n_2 p_2 \ > \ 10$
  - $n_2(1 - p_2) \ > \ 10$
  - Ie 10 or more obs per category
    - ⋆ Use a pooled p for HT (more in a bit)
    - ⋆ Use the sample p_1 and p_2 for CI

Inference for 2 proportions

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$$

$$\hat{p} \sim N(p, \frac{p(1-p)}{n})$$

$$\hat{p}_1 - \hat{p}_2 \sim N(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2})$$

2 Props: HT Set Up

Before we get into hypothesis testing we need to discuss a somewhat weird thing...

$\hat{p}_{pooled}$ (read p-hat-pooled) is the sample proportion, ignoring the subgroups.

$$\hat{p}_{pooled} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{\text{Total Successes}}{\text{Total Sample Size}}$$

This equation is *far* more intimidating looking than what it is conceptually

2 Props: Examples

Working with the nfl question...

- $n_F = 331$
- $n_D = 213$

- Home Wins for Favs $= 226$
- Home Wins for Underdogs $= 69$

- $\hat{p}_F = ?$
- $\hat{p}_D = ?$

- $\hat{p}_{pooled} = ?$

2 Props: Ex

Working with the nfl question...

- $n_F = 331$
- $n_D = 213$

- Home Wins for Favs = 226
- Home Wins for Underdogs = 69

- $\hat{p}_F = 226/331 = .6827$
- $\hat{p}_D = 69/213 = .3239$

- $\hat{p}_{pooled} = (226 + 69) \;/\; (331 + 213) = 295/544 = .5423$

2 Props: Ex HT Assumptions

- Random? Sure but from 1 season only

- Independent and Identically Distributed?
  - Probably? Ignoring "winning streaks" like earlier

- Large n?
  - Need to use the pooled proportion (eg pretend p is the same for both)
  - $n_1 \hat{p}_{pooled} = 179.5 > 10$
  - $n_1(1 - \hat{p}_{pooled}) = 151.5 > 10$
  - $n_2 \hat{p}_{pooled} = 115.5 > 10$
  - $n_2(1 - \hat{p}_{pooled}) = 97.49 > 10$

2 Props: HT Test Stat

For sample means, we assumed we knew μ....

$$\frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

For 1 proportion, we assume we know p....

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

For 2 proportions, we assume we know p_1 - p_2 ($= 0$ almost always)....

$$\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_1} + \frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_2}}} \sim N(0, 1)$$

2 Props: Ex Test Stat

$$z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_1} + \frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_2}}}$$

$$z = \frac{.6827 - .3239 - (0)}{\sqrt{\frac{.5423(1-.5423)}{331} + \frac{.5423(1-.5423)}{213}}}$$

$$z = \frac{.3588}{.04376}$$

$$z = 12.39$$

With p-value (from the normal distribution!!) of...

2*(pnorm(12.39, lower.tail = FALSE))
[1] 2.960434e-35

We have very strong evidence that the true proportion of home wins by the favorite team is different than the underdog team's true home win proportion.

2 Props: HT Summary

Ultimately this follows a pretty old song and dance

- State our hypothesis
- Check our assumptions
  - Random
  - Independent and Identically Distributed
  - Large n
    - Use $\hat{p}_{pooled}$
    - 4 subparts to this one!!
- Calculate our test statistic and p-value and decisision

See Wiki for z-test 3/4 of the way down....

2 Props: Confidence Intervals

If we can do a hypothesis test we can do a confidence interval (like actually, they are the same under the hood)

$$
\begin{aligned}
estimate &\pm (distributional\ value)(standard\ error) \\
\bar{x} &\pm z_{1-\alpha/2}\left(\sqrt{\frac{\sigma^2}{n}}\right) \\
\hat{p} &\pm z_{1-\alpha/2}\left(\sqrt{\frac{p(1-p)}{n}}\right) \\
\hat{p}_1 - \hat{p}_2 &\pm z_{1-\alpha/2}\left(\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)
\end{aligned}
$$

NOTE: We do NOT use $\hat{p}_{pooled}$ for a confidence interval because we don't think p_1 and p_2 are the same!

2 Props: CI Assumptions

- Random Sample: Same as before
- IID: Same as before
- Large n
  - $n_1 \hat{p}_1 > 10$
  - $n_1(1 - \hat{p}_1) > 10$
  - $n_2 \hat{p}_2 > 10$
  - $n_2(1 - \hat{p}_2) > 10$
  - Ie 10 or more obs per category
  - We pass with 69 being our lower number of obs (Underdogs win at home)

2 Props: 90% CI

$$\hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2}\left(\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

$$.6827 - .3239 \pm 1.645 * \sqrt{\frac{.6827(1-.6827)}{331} + \frac{.3239(1-.3239)}{213}}$$

$$.3588 \pm .0675$$

(.2913, .4263)

We are 90% confident the true difference in proportions between the home win rate for the favorites and the home win rate for the underdogs is between .2913 and .4263.

2 Props: CI Summary

- Formulas are a bit more complicated
- But fundmentally you've seen everything before
- Interpretations/understandings are all the same
- Eg "We are BLANK % confident the true (STATISTIC) is between LOWER and UPPER"
    - BLANK = confidence level
    - STATISTIC = Difference in the two proportions
    - LOWER and UPPER = confidence interval bounds