Math 247: Hypothesis Testing with Proportions (Section 8.1, 8.2, 8.3)

Examples: We're getting a LOT of information right now about the coronavirus, with some of it changing almost daily, including information about the importance of social distancing and sheltering at home.

Since you all are becoming certified statistics experts, you can now take conflicting information and decide what's more likely to be true. Take for instance the recent CDC statement that "as many as 25% of people infected with the new coronavirus remain asymptomatic" which means a lot of people who continue to run around and not remain socially distant are spreading the disease. But is that 25% correct? Is it too high? Too low?

Let's look at data from Iceland where extensive testing is being done. (Their goal is to test everyone in their small island population, not just people who are sick or who have been exposed to the virus). They found that out of 180 people who tested positive, 90 of them (50 percent!) did not have any symptoms. What does this data tell us about the CDC number?

key word!

Perform a hypothesis test to see whether the true proportion of asymptomatic people in the population is higher than 25%.

1. Hypothesize

$H_0: p = .25$ — The true proportion of asymptomatic people infected in the POPULATION is 25% (CDC number)

$H_a: p > .25$ — The true proportion is higher than 25%

---

2. Prepare

- Set $\alpha = .05$ (default)
- Choose test: 1 Proportion z-Test

Parking Lot

Iceland data
$n = 180$ infected people
$X = 90$ no symptoms
$\hat{p} = \frac{90}{180} = .5$
$p_0 = .25$

→ Check conditions are met for this specific test

1. Random sample? NO! We'll proceed anyway but will address this at the end.

Independent observations? Unknown - assume

✓ 2. Large sample?

Success = no symptoms: $E = np_0 = 180(.25) = 45 \geq 10$? Yes
Failure = symptoms $E = n(1-p_0) = 180(.75) = 135 \geq 10$? Yes

✓ 3. Large population?

Be sure to describe this!

{ Population of all people (in WORLD!) infected by coronavirus } $Pop \geq 10 \cdot n = 10(180) = 1800$

At least 1800 people infected? Yes!

3. Compute – GOAL: find P-value

- sketch $\hat{p}$ curve and z-curve
- find z-score
- find P-value

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{.25(.75)}{180}}$$

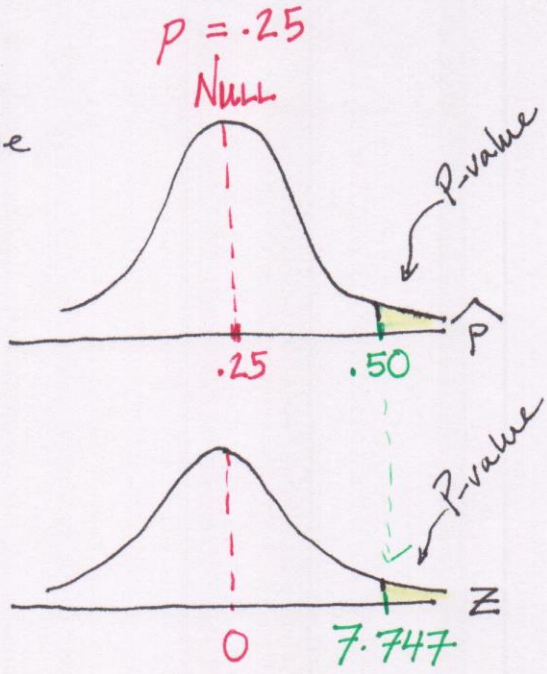
$SE = .03227\ldots$

$$z = \frac{\hat{p} - p_0}{SE} = \frac{.50 - .25}{.03227} = 7.747$$

P-value $= P(\hat{p} \geq .50)$

$= P(z \geq 7.747)$

= close to zero!



Stat Crunch: Normal Distribution Calculator

$P(z \geq 7.747) = 4.6629e-15$

What?!

$= .0000000000000046629$

Basically ZERO!

4. Interpret – Make a decision

P-value $\leq .0001 < .05 = \alpha$

so reject H_0 and accept H_a.

There is basically an almost ZERO% chance the data from Iceland would have shown 50% of COVID positive people to be asymptomatic IF the CDC number of 25% was correct. We have strong evidence the proportion of infected people who are asymptomatic is higher than 25%.

Critique of our "study":

Our sample was not random and also from a very uniform group that is not necessarily ~~reflected~~ representative of the world population.

Major confounders

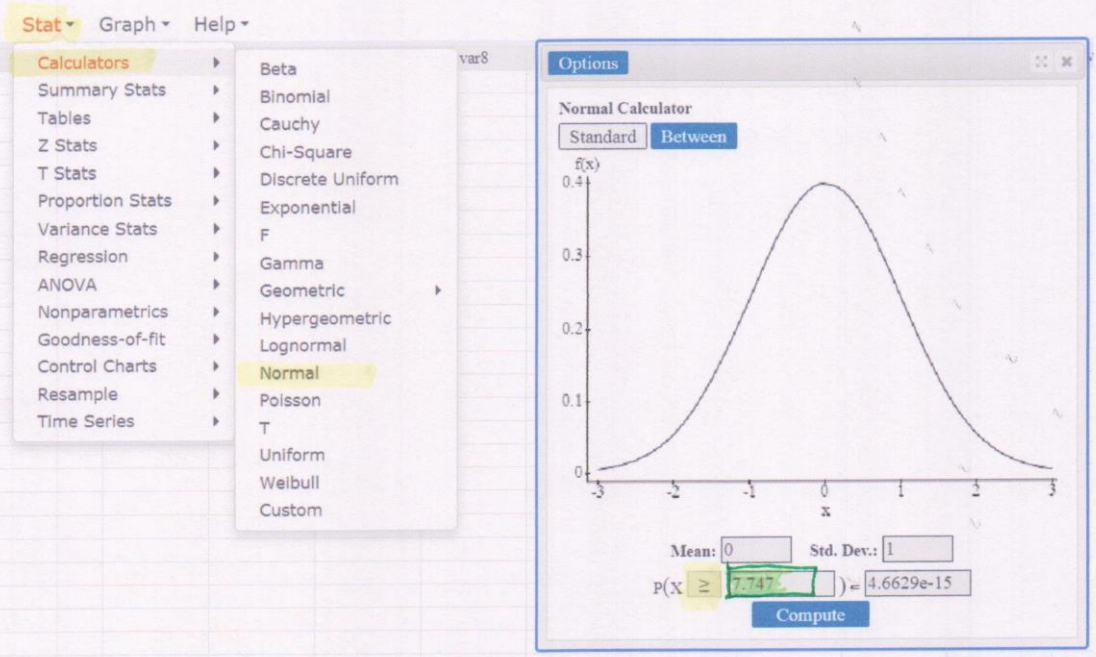
- Genetics – we don't know the genetic relationship with who gets sick and who doesn't. Icelandic people have fairly uniform genetics.

- Viral load: Some studies are showing the more virus people are exposed to initially, the sicker they get. Few people in Iceland are infected, so people aren't bumping into 10 different infected people, unlike what may have happened in Italy or New York.

among those infected

StatCrunch Results:

Normal Distribution Calculator $P(z \geq 7.747)$



Hypothesis Test: In StatCrunch, select: Stat → Proportion Stats → One Sample → With Summary