MAT 135, Discussion Questions 4.20

1. What does it mean for a hypothesis test to be statistically significant? How is this different from our everyday notions of "significant"?

It means the result is unlikely to occur by chance

it does not mean "large" or "meaningful"

2. What is the most commonly used level of significance?

$\alpha = .05$

3. Why might we want to raise or lower it?

we might lower it to be more certain the evidence is strong
we might lower it if the consequences of being incorrect are small or none

4. Why do we use the statements "reject the null" and "fail to reject the null" rather than "reject the null" and " accept the null" or some other formulation? Why does it matter? (Compare to the legal case: why do we find someone "not guilty" instead of "innocent"?)

we don't want to say "accept the null" as this is too strong. we are saying we don't have strong evidence to think otherwise. Just as in court we find people not guilty and not innocent.

5. What are the steps in testing hypotheses?

State null & alternative
Calculate the z- or t-test statistic
Compare to critical value or calculate p-value & compare to significance level

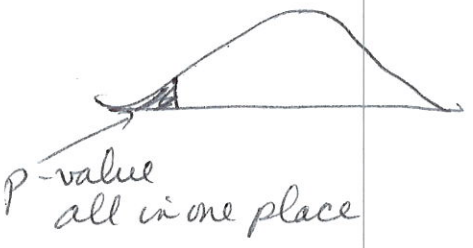
accept or reject the null

6. When are hypothesis tests one-tailed or two-tailed?

one tailed when Ha has $>$ or $<$

two tailed when Ha has $\neq$

7. How do the procedures for one-tailed and two-tailed tests differ? Draw a graph to illustrate.



p-value all in one place

calculate % on one side & multiply by two to get other side

8. What is a p-value?

it's the chance the null is true & we obtained results only by chance

9. How do we use p-values to test hypotheses?

Compare p-value to α (significance level)

if p-value $< \alpha$ reject the null since chance we obtained result but null is true is very low

10. For small samples, why do we use t-tests instead of z-tests?

t has bigger tails

& so extreme values are more likely

which better reflects variability of small sample sizes

11. An engineer designs an improved light bulb. The previous design had an average lifetime of 1200 hours. The new bulb had a lifetime of 1200.2 hours, using a sample of 40,000 bulbs. Although the difference is quite small, the effect was statistically significant. The most likely explanation for this result is what?

very large sample size

12. What conditions must hold for using a z-test for proportions in a hypothesis test?

$np \geq 5$ & $n(1-p) \geq 5$

better ($np \geq 10$ & $n(1-p) \geq 10$)

13. A city ordinance requires that more than 75% of its residents must agree to the construction of new public buildings (using tax dollars) before any such structures can be built. A proposal has been made to build a new recreational facility in the city, and sponsors of the proposal want to conduct a small survey to see if it would be approved if put to an official vote of all residents. A simple random sample of 150 residents revealed that 123 supported a change (and 27 did not).

$H_0: p = .75$ or $\leq .75$ $p_0 = .75$

$H_a: p > .75$

$x = 123$

$n = 150$

1PropZTest

prop > p_0

$z = 1.9798...$

$p = .0238 < .05$

reject H_0

there is enough support to pass/hold election

14. Scientists think that robots will play a crucial role in factories in the next several decades. Suppose that in an experiment to determine whether the use of robots to weave computer cables is feasible, a robot was used to assemble 500 cables. The cables were examined and there were 15 defectives. If human assemblers have a defect rate of 0.035, does this data support the hypothesis that the proportion of defectives is lower for robots than for humans? Use a 0.01 significance level.

$H_0: p = .035$

$H_a: p < .035$

$p_0 = .035$

$x = 15$

$n = 500$

1PropZTest

prop < p_0

$z = -1.608$

$p = .27 > .01$

fail to reject H_0

there is not enough evidence to think defect rate is lower than humans

15. Read and comment on the article at http://sciblogs.co.nz/kidney-punch/2015/01/15/beyond-reasonable-doubt-a-significant-improvement/.