Final Review Problems

1. For each of the following situations, state whether the parameter of interest is the mean or the proportion. Additionally, state the type of test statistic you would use (z-score or t-stat).

   (a) In a survey, one hundred college students are asked how many hours per week they spend on the Internet.

   mean, # of hours (t-test)

   (b) In a survey, one hundred college students are asked: "What percentage of the time you spend on the Internet is as part of your course work?"

   mean, although it is a % it is still numerical (t-test)

   (c) In a survey, one hundred college students are asked whether or not they cited information from Wikipedia on their papers.

   proportion, yes they did/no they didn't (z-score)

   (d) In a survey, one hundred college students are asked what percentage of their total weekly spending is on alcoholic beverages.

   mean, although it is a % it is still numerical (t-test)

   (e) In a sample of one hundred recent college graduates, it is found that 85 percent expect to get a job within one year of their graduation date.

   proportion, get a job or not (z-score)

2. Hewes and Associates, a law firm in Manhattan, is investigating whether or not a chemical called aerocyte, used in mining operations by Ultima National Resources (UNR), increases the risk of certain types of cancer. Ellen Parsons, an associate at this firm, interviewed 300 randomly selected residents of counties where UNR has operations using aerocyte. Among these 27 had been diagnosed with stomach cancer. Based on research by the Centers for Disease Control it is known that among at risk populations 5% will develop stomach cancer. Would you advise Hewes and Associates to proceed with the case, i.e. is there sufficient evident to suggest that arocyte is associated with a higher risk of stomach cancer?

(a) What are the hypotheses?

$H_0: p = 0.05$ (aerocyte has no effect on risk of cancer)

$H_A: p > 0.05$ (aerocyte is associated w/ a higher risk of cancer)

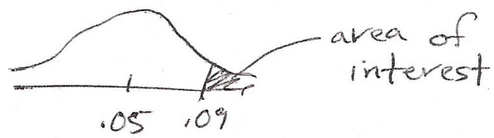
(b) Are the assumptions/conditions for inference satisfied?

1. Random
2. $300 \times 10 = 3000$. There are more than 3000 people who have surgery w/ aerocyte ✓
3. Success / Failure $np = 300(.05) = 15$ ✓ $nq = 300(.95) = 285$ ✓

(c) Calculate the test statistic.

$\hat{p} = \frac{27}{300} = .09$

$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{.09 - .05}{\sqrt{\frac{.05(.95)}{300}}} = \frac{.04}{0.0126} = 3.18$



(d) Find the p-value.

Look-up $z = 3.18$ on z-table $= .9993$

So p-value $= 1 - .9993 = .0007$

(e) What do you conclude? Interpret your conclusion in context.

Since our p-value of .0007 is less than $\alpha = 0.05$, we reject the null hypothesis. The sample provides evidence to suggest that aerocyte is associated w/ a higher risk of cancer.

(f) What type of error might you have committed? Choose only one answer.

(a) Type I (b) Type II

(g) Would you expect a confidence interval with an equivalent confidence level for the proportion of at risk populations who will develop stomach cancer to the test to include 0.05?

Since we have rejected the null hypothesis and there is evidence of the population proportion being higher than 0.05 we would not expect 0.05 to be in our interval.

(h) Calculate a 90% confidence interval?

$\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

$.09 \pm 1.645 \sqrt{\frac{.09(.91)}{300}}$

$.09 \pm 1.645(0.0165)$

$.09 \pm .0271$

$(0.063, 0.1171)$

3. A CBS News article (*Poll: 4 in 5 Support Full-Body Airport Scanners*, November 15, 2010) report that "Americans have differing views on two potentially inconvenient and invasive practices that airports could implement to uncover potential terrorist attacks." This news piece was based on a survey conducted among among a random sample of 1,137 adults nationwide, interviewed by telephone between November 7-10, 2010, where one of the questions on the survey was

> Some airports are now using "full-body" digital x-ray machines to electronically screen passengers in airport security lines. Do you think these new x-ray machines should or should not be used at airports?

Below is a breakdown of answers based on party affiliation:

		Party Affiliation		
		Republican	Democrat	Independent
Answer	Should	264	299	351
	Should not	38	55	77
	Don't know/No answer	16	15	22
	Total	318	369	450

Based on these data, can we conclude that there is a significant difference in opinion on the use of full-body scans between republicans and democrats, i.e. are proportions of republicans and democrats who think full-body scans *should* be used at airports are significantly different?

(a) Write the hypotheses in words and in symbols.

H_o: $P_{rep} = P_{Dem}$ The proportion of Reps + Dems who support full body scans is equal.

H_a: $P_{rep} \neq P_{Dem}$ The proportion of Reps + Dems who support full body scans is not equal.

(b) Calculate the test statistic.

$\hat{P}_{rep} = \frac{264}{318} = .8302$ $\quad SD = \sqrt{\frac{.8195(.1805)}{318} + \frac{.8195(.1805)}{369}} = \sqrt{.0009} = .03$

$\hat{P}_{Dem} = \frac{299}{369} = .8103$

pool b/c 2 proportions hypothesis test $\rightarrow \hat{P}_{Pool} = \frac{264+299}{318+369} = \frac{563}{687} = .8195$ $\quad z = \frac{(.8302 - .8103) - 0}{.03} = \frac{.0199}{.03} = .67$

(c) Find the p-value.

Look-up $z = .67$ on z-table to get .7486, then $1 - .7486 = .2514$

We must multiply by 2 b/c Ha is $\neq$ so $2 \times .2514 = .5028$

(d) What do you conclude?

Since our p-value of .5028 is greater than $\alpha = 0.05$ we fail to reject the null hypothesis. This suggests there is no difference in proportions of Reps + Dems who support full body scans

(e) What type of error might you have committed? Choose only one answer.

(a) Type I $\qquad$ (b) Type II

(f) Would you expect a confidence interval with an equivalent confidence level for the proportion of at risk populations who will develop stomach cancer to the test to include 0?

Since we have failed to reject to H_0 saying there is no difference in the proportions of Dems & Reps then the proportions can be equal meaning 0 will be in our interval.

(g) Calculate a 90% confidence interval for the difference between $(p_R - p_D)$.

$$.8302 - .8103 \pm 1.645\sqrt{\frac{.8302(.1698)}{318} + \frac{.8103(.1897)}{369}}$$

$$0.0199 \pm 1.645(0.0283)$$

$$0.0199 \pm 0.0466$$

$$(-0.0267,\ 0.0665)$$

(h) Interpret the confidence interval in context.

We are 90% confident that the true difference in proportions between Republicans who support full body scans is from 2.65% lower to 6.67% higher than Democrats.

(i) Does this prove that a there is no difference in opinion on the use of full-body scans between republicans and democrats? Explain.

No, this doesn't prove it. We can't make casual statements b/c this was an observational study.

4. To this point in his career Blake Griffin has averaged 21.3 points per game. During this season so far he has played 17 games averaging only 17.5 points per game with a standard deviation of 3.9 points per game. Assume that Griffin's points per game is distributed nearly normal. Are these 17 games evidence, that he has become a worse scorer in his third full season as a professional basketball player? Test the hypothesis using a significance level of 0.05.

(a) Perform a hypothesis test.

1. $H_0: \mu = 21.3$
   $H_a: \mu < 21.3$

2. Assume conditions are met.

3. $t^*_{df} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \cancel{} = \frac{17.5 - 21.3}{\frac{3.9}{\sqrt{17}}} = -4.0173$

   look-up on t-table w/ df $17 - 1 = 16$

4. p-value < 0.0005

5. Since p-value is less than $\alpha = 0.05$ we reject. H_0. There is evidence that Griffin has become a worse scorer.

(b) Calculate a 95% confidence interval for Blake Griffin's average points per game so far during the 2012 season.

$\bar{x} \pm t^*_{df} \frac{s}{\sqrt{n}}$

$17.5 \pm 2.120 \frac{3.9}{\sqrt{17}}$

$17.5 \pm 2.120(.9459)$

17.5 ± 2.0053

$(15.4947, 19.5053)$

We are 95% confident that Blake Griffin's true points per game is between 15.5 and 19.5 points

5. A Washington Post article (*Public option gains support*, October 20, 2009) reports that "a new Washington Post-ABC News poll shows that support for a government-run health-care plan to compete with private insurers has rebounded from its summertime lows and wins clear majority support from the public." More specifically the article says "seven in 10 Democrats back the plan, while almost nine in 10 Republicans oppose it. Independents divide 52 percent against, 42 percent in favor of the legislation." There were were 819 democrats, 566 republicans and 783 independents surveyed.

Is there significant evidence to suggest that a higher proportion of democrats than independents support the public option plan?

(a) Write the hypotheses in words and in symbols.

$H_0: P_{Dem} = P_{IND}$ (proportion of Dems & Ind who support the plan is equal)

$H_a: P_{Dem} > P_{IND}$ (proportion of Dems who support the plan is higher than Ind)

(b) Calculate the test statistic.

$\hat{p}_{Dem} = .70$ $\hat{p}_{IND} = .42$

$x_{DEM} = .7(819) = 573$ $x_{IND} = .42(783) = 328$

$\hat{p}_{Pool} = \frac{573+328}{819+783} = .5624$

$SD = \sqrt{\frac{.5624(.4376)}{819} + \frac{.5624(.4376)}{783}} = .0245$

$z = \frac{(.70-.42) - 0}{.0245} = \frac{.28}{.0245} = 11.43$

(c) Find the p-value and interpret it in context.

A z-score of 11.43 doesn't show up on your z-table. A larger z-score is very very rare so the corresponding value is ≈ 0

(d) What do you conclude?

Since our p-value is basically 0 this is smaller than $\alpha = 0.05$ so we reject the null hypothesis. There is evidence that the proportion of democrats who support the plan is higher than Independents.

(e) Calculate a 95% confidence interval for the difference between $(p_D - p_I)$.

$.70 - .42 \pm 1.96\sqrt{\frac{.7(.3)}{819} + \frac{.42(.58)}{783}}$

$.28 \pm 1.96(.0245)$

$.28 \pm 0.0480$

$(0.232, 0.328)$

(f) Interpret the confidence interval in context.

We are 95% confident that the true proportion of Dems support the plan is from 23.2% to 32.8% higher than Ind.

(g) Is there evidence of a higher proportion of democrats than independents support the plan?

Yes because 0 is not within our interval.

(h) Does this prove that a higher proportion of democrats support the plan than independents? Explain.

Once again we can't prove this because it an observational Study. We can't make casual conclusions from observational studies.

6. Many office "coffee stations" collect voluntary payments for the food consumed. Researchers at the University of Newcastle upon Tyne performed an experiment to see whether the image of eyes watching would change employee behavior. They alternated pictures of eyes looking at the viewer with pictures of flowers every other weekend on the cupboard behind the "donation" box. They recorded the amount of donations in pounds each week per estimated kilogram of food consumed.

	Eyes	Flowers
n (# of weeks)	5	5
$\bar{y}$	0.417	0.151
s	0.1811	0.067

(a) Do these results provide evidence that there really is a difference in honesty even when it's only photographs of eyes that are "watching"? Perform the hypothesis test. First, state the hypothesis. Assume the conditions are met.

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

$$t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{SD} \qquad SD = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(b) Calculate the test statistic.

$$SD = \sqrt{\frac{0.1811^2}{5} + \frac{0.067^2}{5}} = \sqrt{.0074} = 0.0860$$

$$t_4 = \frac{(0.417 - 0.151) - 0}{0.0860} = \frac{0.266}{.0860} = 3.0930$$

(c) Find the p-value and interpret it in context.

Look-up $t = 3.0930$ with $df = \min(5-1, 5-1) = 4$ and two-tails

The p-value is $0.02 < \text{p-value} < 0.05$

(d) What do you conclude?

Since our p-value is less than $\alpha = 0.05$ we reject the null hypothesis. There is evidence that there really is a difference in ~~[illegible]~~ honesty.

(e) The Cherry Blossom Run is a ten mile race that takes place in Washington, DC. In 2012, a sample of 100 runners had a mean age of 35.22 years old with a standard deviation of 5.5 years old. Construct a 99% confidence interval based on the 2012 sample of runners.

(Next Page)

$\bar{x} \pm t^*_{df} \frac{s}{\sqrt{n}}$ find t^*_{df} with $df = n-1 = 100-1 = 99$

but b/c no 99 round down to 80

$t^*_{80} = 2.639$

$35.22 \pm 2.639 \frac{5.5}{\sqrt{100}}$

$35.22 \pm 2.639(.55)$

35.22 ± 1.4515

$(33.7685, 36.6715)$