Project 7 Solutions

Statistics 401: Fall 2016, 34 points

1. (4 pts) Regarding the LA Times article on illegal immigrants voting: the language "the evidence fails to suggest that ..." or "there is no evidence that ..." are used to describe failure to conclude an alternative hypothesis. The use of this same language in the LA Times' headline "there is no evidence that thousands of non-citizens are illegally voting and swinging elections" implies that the alternative hypothesis must be H_a: *more than a thousand illegal immigrants vote.* Hence, the null and alternative hypotheses consistent with the LA Times' story are:

   (a) (i)H_0: a thousand illegal immigrants (or less) vote
   H_a: more than a thousand illegal immigrants vote

   (b) The LA Times conclusion implies that H_0 was not rejected in favor of H_a.

   (c) To test the hypotheses in 1a, one might randomly select counties in the US. Within the selected counties, one could select either a systematic sample or a random sample of voters then follow-up with these voters to determine citizenship.

2. (10 pts) In 11/17's class, we collected data from all $n = 12$ students in class to estimate the true percentage of all Montanan college students who support a ban on trapping on Montana public lands. We found that $\hat{p} = 7/12$ students supported the ban.

   (a) Unfortunately, the test and CI we calculated on 11/17 regarding the proportion of all college students is not reliable because:

   - A sample size of $n = 12$ is too small because $n\hat{p} = 12 \times \frac{7}{12} = 7 < 10$ (or we could have noticed that $n(1-\hat{p}) = 12 \times \left(1 - \frac{7}{12}\right) = 5 < 10$).
   - Choosing STAT401 students as the sample was out of convenience, so the sample is not a SRS.

   (b) Even though we realize that the CLT should not be applied (see 2a), if $\hat{p}$ was normal, then a 95% CI for the true proportion p of Montanan's represented by MSU STAT401 students who support the ban on trapping is

$$\hat{p} \pm z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{7}{12} \pm 1.96\sqrt{\frac{\frac{7}{12}(1-\frac{7}{12})}{12}}$$
$$= [0.304, 0.862]$$

   (c) Assuming that STAT401 students are representative of Montana college students, we are 95% confident that the true percentage of Montana college students who support a ban on trapping on public lands is between 30.4% and 86.2%.

   (d) On 11/8's ballot, Montanan voters recently rejected a ballot initiative seeking to ban trapping on public lands in Montana, with only 37.2% of voters supporting the ban. We will test whether the proportion of Montana college students p who support the ban is larger than 0.372:

- The statistical hypotheses are:
  $H_0 : p = 0.372$
  $H_a : p > 0.372$
- Because the CI includes 0.372, then we fail to reject H_0.
- The evidence from STAT401 students fails to suggest that the proportion of college students who support a ban on trapping is larger than the proportion of all Montana voters who supported the ban.

(e) We want to take a SRS of $n = 30$ students to test the hypotheses in 2d at a significance level of $\alpha = 0.05$. If the true proportion of college students who support banning trapping is 58.3% (i.e., we will assume that $p = 0.583$; this is a value of p consistent with H_a in 2d), we will calculate the power of the test using the following steps:

- First, we will assume that H_0 is true and determine for what values of $\hat{p}$ that a Type I Error occurs (ie., when we REJECT H_0 when we should not because H_0 is true).
- When H_0 is true, $p = 0.372$.
- Because $n = 30$ is large (i.e., np and $n(1-p)$ are both larger than 10), then $\hat{p} \dot\sim N\left(0.372, \frac{0.372(1-0.372}{30}\right)$.
- We need to calculate the value of $\hat{p}^*$ such that if we get a sample with $\hat{p}$ above $\hat{p}^*$ (which happens with probability α when H_0 is true), we will REJECT H_0. Because $\alpha = 0.05$, then, by the normal tables, $\hat{p}^*$ is 1.645 SDs above the true proportion of $p = 0.372$. That is, $\hat{p}^* = 0.372 + 1.645\sqrt{\frac{.372(1-.372)}{30}} = 0.517$. In other words, $P(\hat{p} > 0.517 \text{ when } H_0 \text{ is true}) = \alpha = 0.05$.

- Now we will assume that H_a is true and determine how often a Type II Error occurs (ie., when we Fail To Reject H_0 when we should because H_a is true).
- When H_a is true, $p > 0.372$. We need to pick one of these values bigger than p! Which one? The problem tells us to use $p = 0.583$.
- We need to calculate $P(\hat{p} < 0.517 \text{ when } H_a \text{ is true})$. Using the normal table we standardize 0.517 with respect to H_a like this: $Z = \frac{0.517-0.583}{\sqrt{\frac{0.583*(1-.583)}{30}}} = -0.733$. So

$$\beta = P(\hat{p} < 0.517 \text{ when } H_a \text{ is true}) = P(Z < -0.733) = 0.232.$$

- The power to REJECT $H_0 : p = 0.372$ when $p = 0.582$ is $1-\beta = 1-0.232 = 0.768$.

3. (10 pts) Regarding the study of middle aged Finnish men (see the abstract for *Coffee Drinking Is Dose-Dependently Related to the Risk of Acute Coronary Events in Middle-Aged Men* from the September 2004 Journal of Nutritional Epidemiology:

   (a) Since a can of *Miller High Life* admits that 12 fluid ounces is equal to 355 ml of the champagne of beers, then 950ml of coffee is equal to $950\left(\frac{12}{355}\right) = 32.11$ fluid ounces of coffee, which is equal to a little over 4 cups of coffee each day.

   I drink about 3 cups of coffee a day, so I am not a heavy coffee drinker.

   (b) A pooled procedure appears appropriate since the sample standard deviations $s_1 = 1.4$ and $s_2 = 1.7$ are pretty close (and neither more than twice the other), indicating that $\sigma_1 = \sigma_2$ is not a bad assumption.

   (c) i. The **statistical hypotheses** consistent with testing whether drinking coffee is associated with higher white blood cell counts, on the average, in middle aged men:
   $H_0 : \mu_1 = \mu_{\text{no coffee}} = \mu_{\text{coffee}} = \mu_2$
   $H_a : \mu_1 = \mu_{\text{no coffee}} < \mu_{\text{coffee}} = \mu_2$

   ii. **Assumptions:** We must assume that we have SRSs of men in the coffee drinking and non-coffee drinking groups, although this seems like a dubious assumption. We will also assume that the two samples are independent of each other. There are millions of coffee drinkers and non-coffee drinkers on the planet, so the populations are very large (and more than 5%) in comparison to the sample sizes. And lastly, the sample sizes of $n_1 = 77$ and $n_2 = 351$ are both larger than 30. Hence the Central Limit Theorem applies: both $\overline{X}_1$ and $\overline{X}_2$ are normal, which implies that $\overline{X}_1 - \overline{X}_2$ is normal.

   iii. The test statistic using a pooled procedure is given by:

$$t = \frac{5.2 - 6}{s_p\sqrt{\frac{1}{77} + \frac{1}{351}}} = -3.85$$

   where $s_p = \sqrt{\frac{76(1.4^2)+350(1.7^2)}{77+351-2}} = 1.65$

   iv. Using a t-distribution with degrees of freedom $df = n_1 + n_2 - 2 = 426$, the p-value is $P(T < -3.85) = 6.8 \times 10^{-5}$ (see Appendix).

   v. Because p-value is tiny and less than α, we REJECT H_0,

   vi. Heavy coffee drinkers have statistically significantly higher white blood cell counts, on the average, compared to non-coffee drinkers.

   (d) A Type I Error is when we find a statistically significant difference (i.e., REJECT H_0) in the mean number of white blood cells between coffee drinkers and non-coffee drinkers when in fact there is no difference.

   (e) A Type II Error is when we fail to find a statistically significant difference (i.e., Fail to Reject H_0) in the mean number of white blood cells between coffee drinkers and non-coffee drinkers when in fact there really IS difference.

   (f) This was an observational study, not an experiment. A "coffee treatment" was not randomly assigned to the individuals in the study. Thus, cause-and-effect statements can not be made on the basis of this study alone. The assertion in the abstract that "heavy coffee consumption increases the short-term risk of acute myocardial infarction or coronary death," solely based on this study, seems dubious since this is an observational study. This is highlighted by the fact that the abstract points out that the "evidence remains equivocal" due to conflicting results from many previous observational studies.

4. (10 pts) Regarding treatment of Ductal Carcinoma In Situ (DCIS):

   (a) A 2-sample test of proportions that compares the observational study of $n_1 = 136$ women to the the treatment group in the CRD of $n_2 = 408$ women who received only a partial masectomy:

      i. To test whether a partial masectomy alone reduces the risk of invasive breast cancer compared to doing nothing, the **statistical hypotheses** are:
      $H_0 : \ p_1 = p_2$
      $H_a : \ p_1 > p_2$

      ii. **Assumptions:** Because this is a trial of humans, neither of the two samples are likely SRSs. However, it is very likely that the two samples of women are independent. There are millions of women with DCIS who have never had it treated, and thousands of women with DCIS who have undergone a partial masectomy to treat their DCIS, so the population sizes are large (and more than 5%) compared to the sample sizes $n_1 = 136$ and $n_2 = 408$. The sample sizes are large enough because:
      - $n_1 p_1 = 136(0.25) = 34 \geq 10$
      - $n_1(1 - p_1) = 102 \geq 10$
      - $n_2 p_2 = 408(0.17) = 69 \geq 10$
      - $n_2(1 - p_2) = 339 \geq 10$

      iii. Because we assume that $H_0 : p_1 = p_2$ is true, then the test statistic for this problem requires that first we calculate a single estimate from both samples of the proportion of women who get invasive cancer:

$$\hat{p} = \frac{\text{count}_1 + \text{count}_2}{n_1 + n_2} = \frac{0.25 \times 136 + 0.17 \times 408}{136 + 408} = 0.19.$$

      So the test statistic is

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} = \frac{0.25 - 0.17}{\sqrt{\frac{0.19(1-0.19)}{136} + \frac{0.19(1-0.19)}{408}}} = 2.06$$

      iv. The p-value for the upper one-sided H_a is $P(Z > 2.06) = 0.0197$.

      v. Because p-value $= 0.0197 < 0.05 = \alpha$, REJECT H_0.

      vi. The evidence suggests that a partial masectomy does reduce the risk of DCIS progressing to an invasive cancer. With 95% confidence, the proportion of women who get invasive cancer decreased by at least 1.2%.

   (b) A Type I Error is finding that having a partial masectomy reduces the risk of contracting an invasive cancer compared to doing nothing (i.e., REJECTING H_0) when in fact a partial masectomy does not reduce the risk.

(c) A Type II Error is failing to find that having a partial masectomy reduces the risk of contracting an invasive cancer compared to doing nothing when in fact a partial masectomy does reduce the risk.

(d) We found that a partial masectomy and radiation statistically significantly reduces the risk of an invasive cancer compared to partial masectomy alone ($p = 4.9 \times 10^{-5}$, see the Appendix). With 95% confidence, the proportion of women who get invasive cancer decreased by at least 5.2%. We CAN conclude that the addition of radiation <u>caused</u> a decrease in the lower proportion of invasive breast cancer because the data from the two groups of women was generated by an experiment where the women were randomly assigned to the treatment groups.

1 Appendix

```
##########
# PROBLEM 2 - trappng

# 2b - A 95% CI for p
> 7/12 + c(-1,1)*qnorm(.975)*sqrt(7/12*(1-7/12)/12)
[1] 0.3043937 0.8622730

# 2e - Power calc
# Assuming H0 is true, calculate the value of p-hat above which Ho will be rejected
# It is 1.645 SDs above the true proportion of p = 0.372
> .372+1.645*sqrt(.372*(1-.372)/30)
[1] 0.5171632

# Assuming Ha is true, calculate beta = the probability that we see a sample with
# a p-hat less than p-hat = 0.517
> pnorm(0.517,mean=0.583,sd=sqrt(0.583*(1-.583)/30))
[1] 0.2317286

# Power is 1 - beta
> 1 - 0.2317286
[1] 0.7682714




########
# PROBLEM 3 - coffee

# Pooled SD
> sp=sqrt((76*1.4^2 + 350*1.7^2)/(77 + 351 - 2))
> sp
[1] 1.65048

# p-value
```

```
> pt(-3.85,426)
[1] 6.812072e-05

########
# PROBLEM 4 - DCIS

# First, test whether partial masectomy reduces the risk of invasice cancer
# compared to no treatment
#
# Checking sample size of 136 from observational study of women with DCIS with
# no treatment
> 136*c(0.25,.75)
[1] 34 102

# Checking sample size of 408 from CRD of women with DCIS with partial masectomies
> 408*c(.17,.83)
[1]  69.36 338.64

# Calculate the pooled estimate of p1=p2 under H0:
> (.25*136 + 0.17*408)/(136+408)
[1] 0.19

# Calculate the test stat
> (.25-.17)/sqrt(.19*.81/(136) + .19*.81/(408))
[1] 2.059543

# Upper 1-sided p-value
> 1-pnorm(2.059543)
[1] 0.01972113

# Use R's prop.test to verify my results:
> prop.test(c(.25*136,.17*408),c(136,408),correct=F,alternative="greater")

        2-sample test for equality of proportions without continuity
        correction

data:  c(0.25 * 136, 0.17 * 408) out of c(136, 408)
X-squared = 4.2417, df = 1, p-value = 0.01972      # <--- Get the same p-value
alternative hypothesis: greater
95 percent confidence interval:
 0.01169381 1.00000000
sample estimates:
prop 1 prop 2
  0.25   0.17

# Get the same test stat!
> sqrt(4.2417)
```

```
[1] 2.059539

#######
# Second, test whether partial masectomy + radiation reduces the risk of
# invasice cancer compared to partial masectomy alone
> prop.test(c(.17*408,0.08*410),c(408,410),correct=F,alternative="greater")

        2-sample test for equality of proportions without continuity
        correction

data:  c(0.17 * 408, 0.08 * 410) out of c(408, 410)
X-squared = 15.156, df = 1, p-value = 4.949e-05    # <--- Tiny p-value
alternative hypothesis: greater
95 percent confidence interval:
 0.05229927 1.00000000
sample estimates:
prop 1 prop 2
  0.17   0.08
```