Homework 5 Soln: Sign Test, Hypothesis Tests for Mean and Proportions; due Fri 11/2 at 10 am

Total Points: 45 Completion: 10; Accuracy: 35

*Submit your homework to your TA's mailbox (and Canvas if you wish) before the due date/time. The mailboxes are to the left as you enter the Medical Science Center (1300 University Ave.) from the main University Ave. entrance.

*No late homework will be accepted for credit.

*If a problem asks you to use R, include a copy of the code and output. Please edit your code and output to be only the relevant portions.

*If a problem does not specify how to compute the answer, you may use any appropriate method but show the work that is necessary for the TA to follow and evluate your reasoning. I may ask you to use R or manual calculations on exam, so practice accordingly.

*Staple multiple papers together; present solutions to all prompts clearly; and put the discussion you attend on top to get full "Completion Points".

1. An automobile club pays for emergency road services (ERS) requested by its members. The current policy rate the automobile club pays is based on the thought that 20% of services requested will be serious mechanical problems requiring towing. However, the insurance company claims that the auto club has a higher rate of serious mechanical problems requiring towing services. Perform a hypothesis test at the 5% level (after checking assumptions) to test the insurers claim.

   Upon examining a sample of 2927 ERS calls from the club members, the club finds that 1499 calls related to starting problems, 849 calls involved serious mechanical failures requiring towing, 498 calls involved flat tires or lockouts, and 81 calls were for other reasons.

   Answer: $H_0 : \pi = .20$ *vs* $H_a : \pi > .20$ *where* π *is the percent of services requested that are serious mechanical problems requiring towing. Since* $n\pi > 5$ *and* $n(1 - \pi) > 5$*, CLT kicks in, Normal approximation will be ok. We also have to assume our sample is a SRS of the whole population of ERS calls from the club members.* $\hat{p_{obs}} = \frac{849}{2927} = .29$, $SE(\hat{p}) = \sqrt{\frac{.20*.80}{2927}} = 0.00739$ *so* $Z = \frac{.29-.20}{0.00739} = 12.18$. $P(Z > 12.18) = 0$. *We have very strong evidence against the null. It is very unlikely to see a sample percentage of .29 or higher in a sample of 2927 if the population proportion is truely .20.*

2. A pumpkin farmer weighed a simple random sample of size n= 20 pumpkins from his main patch, with these results:

   9.6, 8.8, 5.1, 9.7, 9.1, 8.9, 8, 9.2, 2.7, 9.1, 8.5, 7.3, 9.3, 9.6, 4.1, 9.9, 7.6, 9, 7.2, 8.5

   (a) Create a QQ plot and histogram of the weights. Do you think it is reasonable to assume that the population distribution is normal? Explain your answer

   4 points: 2 for QQnorm plot; 2 for assumption not well met*Answer: No, Data is strongly left skewed in histogram. Sample size of 20 is a bit small to be relying on CLT - probably ok?. Also the QQ plot shows prominent curvature, and more curvature than I see in most runs ofqqnorm(rnorm(20)).*

(b) Regardless of your answer to (a), use R to perform the bootstrap with 3000 resamplings to create a 90% CI for μ. (Show your R code and its output - you can copy and paste the code from Tuesday's (10/23) Notes.)

6 points: 2 for n=3000, 2 for 90% level; 2 points for final interval *Answer: I got an interval of (6.938446 8.693850) with seed=1.*

(c) Now construct a 90% t CI for μ by hand and compare it to that which you found via bootstrap. Which would you tell a scientist to use?

2 points for interval constructed by hand; -1 if only used t.test() *Answer:* $\bar{X} \pm t_{.05,19} \frac{s}{\sqrt{n}} = 8.06 \pm 1.729 \frac{1.960773}{\sqrt{20}} = (7.3019, 8.8180)$. *This interval is a bit higher and a bit narrower. I would tell them the wider bootstrap better reflects the uncertainty due to sample size and evidence in sample that population may not be normal. Students could also argue t test assumptions are roughly met so it is find to use. Both intervals are an approximation.*

(d) Suppose last years pumpkins had a mean weight of 8.2 lbs. Perform a bootstrap test at $\alpha = 0.1$ to see if there is evidence that the mean weight has changed. Make sure to specify your null and alternative hypothesis, report p values, and draw your conclusions in context. How does this conclusion compare to what was found in (b)?

6 points; 2 for hypotheses; 2 points for p value; 1 for interpretation; 1 for comparison*Answer:* $H_o : \mu = 8.2$; $H_a : \mu \neq 8.2$. *I got a p value of* $pval = 2 * (1031)/3000) = 0.687$ *which is no evidence against the null. Evidence suggests mean weight has not changed. 8.2 falls in the bootstrap interval, so we are not surprised that we failed to reject the null.*

(e) Would a two-sided t test reject the null specified in (d)? Compare your answer to what you found in (c). Find an approximate p value for the t test by hand (check value you get in R).

6 points; 2 for no; 2 points comparison; 2 for approximate p value*No. Since the two-sided confidence interval in c. contained 8.2, a two-sided t test will fail to reject* $H_o : \mu = 8.2$; $H_a : \mu \neq 8.2$. $t_{obs} = \frac{\bar{x}-\mu_o}{s/\sqrt{n}} = \frac{8.06-8.2}{1.96/\sqrt{20}} = -0.31$ *and* $2 * P(T_{19} < -0.31) = 2 * 0.3799 = 0.759$; *exact pvalue in R:* 0.753. *From table on df=19 line, students could only get pvalue* > 0.5

(f) If there is strong evidence that the median weight of his pumpkins from the jumbo patch is different from 15, then he feels like he will need to give specific directions to his staff on how to sort them. Let M be the population median. Use the sign test to test: $H_o : M = 15$ vs $H_A : M \neq 15$ at $\alpha = 0.05$. Compute the p value and make a conclusion in the context of the problem.

He sampled 8 pumpkins from his jumbo patch and found weights of:

$$12.6, 12.9, 14.8, 14.3, 19.1, 10.2, 11.4, 9.3$$

What advice would you give the farmer in light of the statistics - what is a limitation to our test?

6 points; 2 for definition and observed B (ok if used number of negatives); 2 points for p value; 2 for conclusion at 5% level *Observed statistic is b=1 since only one pumpkin weight is larger than 15.* $B \sim Bin(8, .5)$. *Since the alternative is two-sided, we need to compute:* $2*min(P(B \leq 1), P(B \geq 1)) = 2*P(B \leq 1) = 2*(P(B = 0)+P(B = 1)) = 2(0.004+0.031) = .07$.

We have moderate evidence against the null, but insufficient evidence at 5% level to reject null and suggest median is different from 15. This is a very small sample size, so we have little power to reject the null. I would recommend to him that he look at more pumpkins (as he doesn't want to upset customers).

3. An electric car company's sales manager is interested in the salaries of people who are on the wait list for their most affordable model and wonder if it is different from the median salary at their company of $135,000. They send a survey to a random sample of 2000 of the people on the waitlist and 85 of the respondents report having salaries above $135,000 and 103 have salaries below. The rest of the surveys were not returned.

   (a) Carry out and interpret a sign test of the hypothesis: $H_o : M = \$135,000$ vs $H_A : M \neq \$135,000$ where M is the median salary of the people on the waitlist.

   Answer: Let B=number of positive (values above median). $B \sim Bin(103{+}85 = 188, .5)$ *and observed test statistic* $b = 85$. *Because 2 sided test, we need* $p-value = 2*min(P(B \geq 85), P(B \leq 85))$. $P(B \geq 85) = 0.9171694$ *and* $P(B \leq 85) = 0.1074609$. *So* $p - value = 2 * P(B \leq 85) = 2 * 0.1074609 = 0.2149218$. *which is very weak evidence against null. No evidence median salary of customers on wait list is different from median salary at company.*

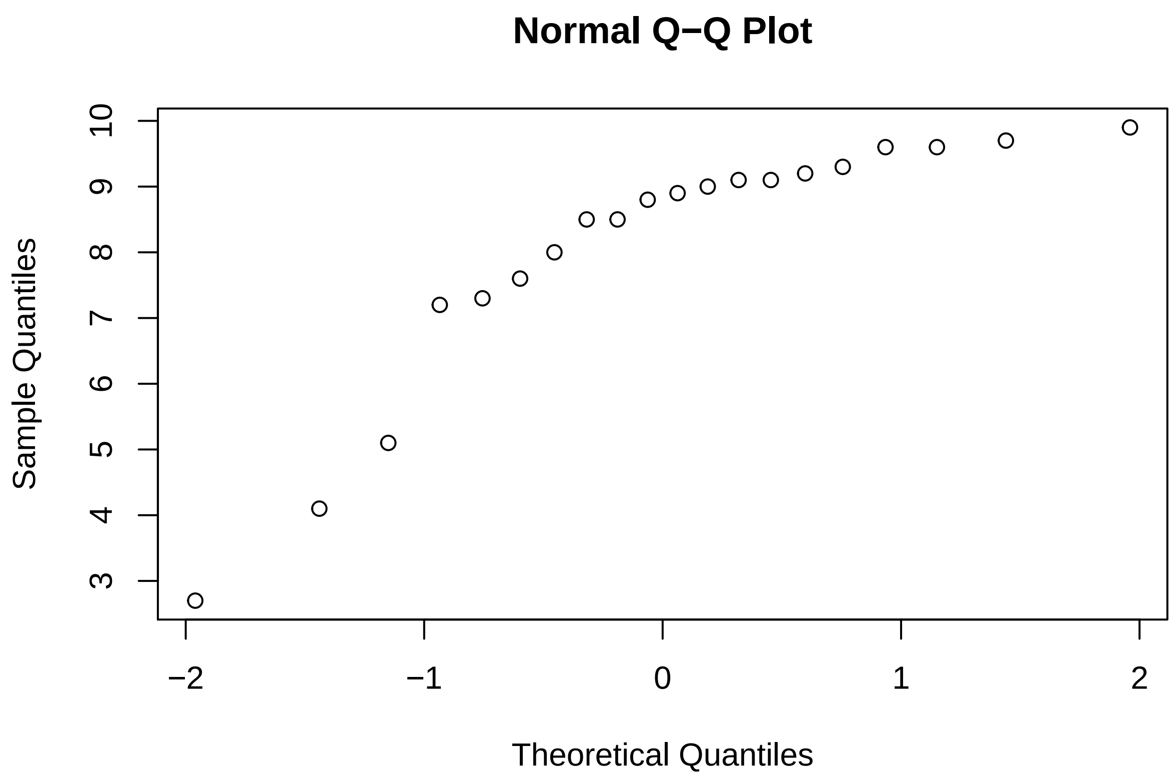
   (b) Carry out and interpret a proportion test of the hypothesis: $H_o : p = 0.5$ vs $H_A : p \neq 0.5$, where p is the proportion of people on the waitlist with salaries above the median salary at the company.

   5 points; 2 points for z value, 1 for p value; 2 for interpretation *Answer: Using* $\hat{p} = 85/188 = 0.452$ *point estimator of true proportion* $z = \frac{.452-.50}{\sqrt{\frac{.5*.5}{188}}} = -1.316286$ *and* $p - value = 2 * P(Z < -1.316286) = 2 * 0.09403903 = 0.1880781$, *This test similarly gives us weak/insufficient evidence against the null at typical significance levels (5% or 10%) so we fail to reject the null. Insufficient evidence that proportion of people on waitlist with salaries above median salary is different from 50%.*

   (c) Why should the car company's sales manager be cautious about putting too much weight on the results of this survey (hint: think about the data collection)?

   Answer: These values are based on 188 respondants from a survey of 2000 people. The people who voluntarily report their income may be systematially different from the whole population of people on the wait list. Perhap people who have low salaries are more likely to respond because they are proud of making such a large investment with their limited funds - or perhaps they would be less likely to respond because they are concerned they won't be prioritized because of their low salary. If they requested salary information before anyone could get on the wait list, then the sales manager would have a census and could actually calculate the true population median (or mean).

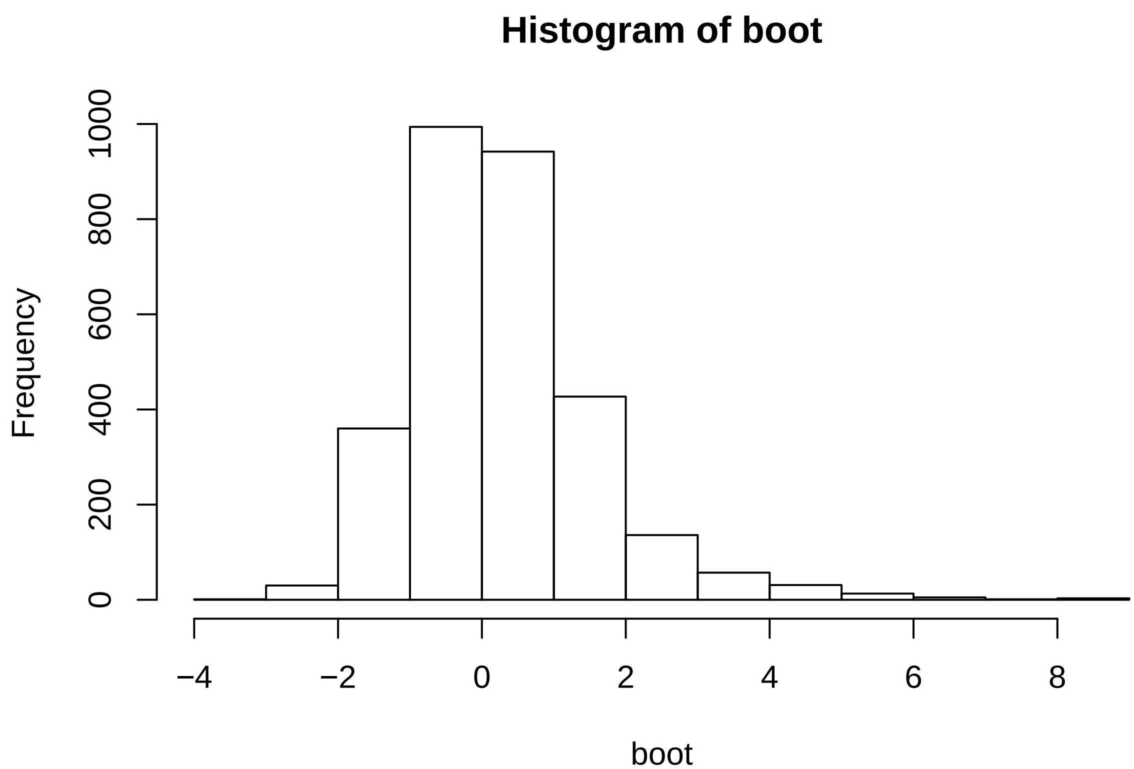
```r
#Question 2
q2data<-c(9.6, 8.8, 5.1, 9.7, 9.1, 8.9, 8, 9.2, 2.7, 9.1, 8.5, 7.3, 9.3, 9.6, 4.1, 9.9, 7.6, 9, 7.2, 8.5
qqnorm(q2data)
```




```
hist(q2data)
```


```
bootstrap = function(x, n.boot) {
  n = length(x)
  x.bar <- mean(x)
```

```r
  t.hat <- numeric(n.boot) # create vector of length n.boot zeros
  for(i in 1:n.boot) {
    x.star <- sample(x, size=n, replace=TRUE)
    x.bar.star <- mean(x.star)
    s.star <- sd(x.star)
    t.hat[i] <- (x.bar.star - x.bar) / (s.star / sqrt(n))
    }
  return(t.hat)
}
set.seed(1)
boot = bootstrap(q2data, 3000)
hist(boot)
```




```r
#2b
# Find quantiles for a 90% confidence interval.
t.lower <- quantile(boot, probs=.05) # This is our t_{1 - alpha.2}.
t.upper <- quantile(boot, probs=.95) # This is our t_{alpha/2}
quantile(boot, probs=c(0.025, 0.05, 0.1, .90, .95, .975))
```

```
##      2.5%        5%       10%       90%       95%     97.5%
## -1.720962 -1.445687 -1.144637  1.772069  2.558042  3.402754
```

```r
# Make the interval.
n = length(q2data)
x.bar = mean(q2data)
s = sd(q2data)
ci.low  = x.bar - t.upper * s / sqrt(n) # This is our lower interval endpoint.
ci.high = x.bar - t.lower * s / sqrt(n) # This is our upper interval endpoint.
interval = c(ci.low, ci.high)
names(interval) = NULL # Strip the now reversed quantile names.
```

```r
print(interval)
```

```
## [1] 6.938446 8.693850
```

```r
#2c
t.test(q2data, conf.level = .90)
```

```
##
##  One Sample t-test
##
## data:  q2data
## t = 18.383, df = 19, p-value = 1.466e-13
## alternative hypothesis: true mean is not equal to 0
## 90 percent confidence interval:
##  7.301875 8.818125
## sample estimates:
## mean of x
##      8.06
```

```r
#2d Hypothesis Test
mu.0=8.2
(t.obs=(x.bar-mu.0)/(s/sqrt(n)))  #-0.3193123
```

```
## [1] -0.3193123
```

```r
summary(boot > t.obs)   #1969 Above, 1031 below
```

```
##    Mode   FALSE    TRUE
## logical    1031    1969
```

```r
summary(boot == t.obs)    # none exactly equal
```

```
##    Mode   FALSE
## logical    3000
```

```r
(m.upper<-sum(boot>t.obs))  #1969
```

```
## [1] 1969
```

```r
(m.lower<-sum(boot<t.obs))  #1031
```

```
## [1] 1031
```

```r
(pval=2*(1031)/3000)  #2 -sided p value (double the smaller of the two extremes)  0.687
```

```
## [1] 0.6873333
```

```r
(pval=2*min(m.upper, m.lower)/3000)
```

```
## [1] 0.6873333
```

```r
#2e 2-sided t test
t.test(q2data, conf.level=0.90, mu=8.2)
```

```
##
##  One Sample t-test
##
## data:  q2data
## t = -0.31931, df = 19, p-value = 0.753
## alternative hypothesis: true mean is not equal to 8.2
## 90 percent confidence interval:
```

```
##  7.301875 8.818125
## sample estimates:
## mean of x
##      8.06
```

```r
#Question 6
area.above<-sum(dbinom(x=85:188, size=188, prob=.5))
area.below<-sum(dbinom(x=0:85, size=188, prob=.5))
(pval=2*min(area.above, area.below))
```

```
## [1] 0.2149218
```

```r
binom.test(x=85, n=188,alternative="two.sided")
```

```
##
##  Exact binomial test
##
## data:  85 and 188
## number of successes = 85, number of trials = 188, p-value = 0.2149
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.3795852 0.5262037
## sample estimates:
## probability of success
##              0.4521277
```

```r
#part b
(85/188-.5)/(sqrt(.5*.5/188))
```

```
## [1] -1.312785
```

```r
1-pnorm(1.312785)
```

```
## [1] 0.0946277
```