Math 3339 - 22163

Written Homework 11 (Chapters 9 and 10)

Name: James Richardson PeopleSoft ID: 1555520

Instructions:

- Homework will NOT be accepted through email or in person. Homework must be submitted through CourseWare BEFORE the deadline.
- Print out this file and complete the problems.
- Use blue or black ink or a dark pencil.
- Write your solutions in the space provided. You must show all work for full credit.
- Submit this assignment at http://www.casa.uh.edu under "Assignments" and choose **HW11**

1. *An article wanted to look at the compression strength of aluminum cans filled with strawberry drink and another sample filled with cola. Does the following data suggest that the extra carbonation of cola results in a higher average compression strength? Base your answer on a *P*-value. What assumptions are necessary for your analysis?

Beverage	Sample Size	Sample Mean	Sample SD
Strawberry drink	15	540	21
Cola	15	554	15

Ho: mean(strawberry drink) = mean(cola)
Ha: mean(strawberry drink) < mean(cola)

Two sample t test for means

t = (540 - 554) / sqrt(21^2/15 + 15^2/15) = -2.101051
P-value = pnorm(t) = 0.01781825

If our significance level is .05 then we would reject the null hypothesis, but if it was a lower significance level, such as .01, then we would fail to reject the null hypothesis.

2. *Some college students did a study of textbook pricing. They compared pries at the campus bookstore and Amazon.com for the same price. To be fair, they included the sales tax for the local store and the added shipping form Amazon. Here are the prices for a sample of 10 books.

Campus	Amazon
99.34	113.94
51.53	61.44
20.45	31.59
97.22	108.29
61.89	78.44
58.17	65.74
61.63	63.49
44.63	40.39
96.69	117.99
48.88	58.94

a. We want to determine if there is a significant difference in the price of textbooks from the campus bookstore and from Amazon.com. Which type of test would we use for this data?

Two sample t test for means

b. Determine a 95% confidence interval for the difference of the population means.

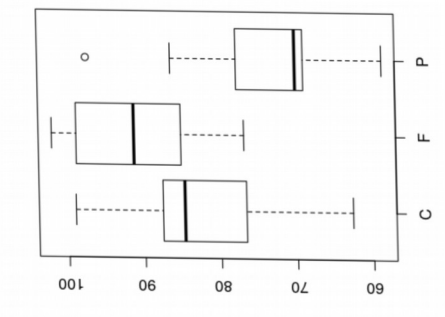
campus = c(99.34 , 51.53 , 20.45,97.22,61.89,58.17,61.63,44.63,96.69,48.88)
Amazon = c(113.94,61.44,31.59,108.29,78.44,65.74,63.49,40.39,117.99,58.94)
x1 = mean(campus) ; x2 = mean(Amazon) ; s1 = sd(campus); s2 = sd(Amazon)
t =
(x1-x2) + c(-1, 1)* t*sqrt(s1^2/n1 + s2^2/n2) = (-47.48626, 27.52226)

c. Interpret your results. Is there a substantial difference between the two ways to be textbook? Assuming that the populations remain unchanged and you have just these two sources, where would you buy?

* Problems came from Devore, Jay and Berk, Kenneth, *Modern Mathematical Statistics with Applications*, Thomson Brooks/Cole, 2007.

3. A study was conducted to examine the effect of pets in stressful situations. Fifteen subjects were randomly assigned to each of three groups to do a stressful task alone (the control group), with a good friend present, or with their dog present. The subject's mean heart rate (in beats per minutes) during the task is one measure of the effect of stress. The data has is the mean heart rates during stress with a pet (P), with a friend (F) and for the control group (C).

   The dataset name is called Stress, here is the link:
   https://www.math.uh.edu/~cathy/Math3339/data/Stress.csv
   Copy this link address.
   Import this into R studio by selecting "Import Datset" on the top right hanc
   Then input the link address in the top box.
   Select "Update," then select "Import."



   a. Make a side by side box plot of the heart rates by the three groups. To do
      boxplot(Rate~Group,data=Stress)

      Does there seem to be a difference in the heart rates of the three groups? Do any of the groups show outliers or extreme skewness?

      Yes there definitely seems to be a difference between the three groups. Group 'c' also seems to have the most outlines, or at least the farthest from the mean

   b. We want to test if there is a difference in the mean heart rates for the three groups. Give the null hypothesis of this test.

      Ho: there is no difference between heartrates no matter who is present

   c. Does the data suggest that there is a difference among the three groups? Use α = 0.05.

      The p value is very close to zero, this will cause us to reject the null hypothesis. This means that there is a difference in the means of the three groups.

```
> summary(rate.lm)

Call:
lm(formula = Rate ~ Group, data = Stress)

Residuals:
    Min      1Q  Median      3Q     Max
-19.878  -4.724  -1.221   6.179  24.055

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   82.524      2.378  34.709   <2e-16 ***
GroupF         8.801      3.362   2.617   0.0123 *
GroupP        -9.041      3.362  -2.689   0.0102 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.208 on 42 degrees of freedom
Multiple R-squared:  0.4014,    Adjusted R-squared:  0.3729
F-statistic: 14.08 on 2 and 42 DF,  p-value: 2.092e-05

> anova(rate.lm)
Analysis of Variance Table

Response: Rate
          Df Sum Sq Mean Sq F value    Pr(>F)
Group      2 2387.7 1193.84  14.079 2.092e-05 ***
Residuals 42 3561.3   84.79
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

   d. Use the Tukey's Method to determine which pair(s) differ significantly.

      > qtukey(.95, 3, 45) * sqrt(84.79/5)
      [1] 14.11451

   e. If there seems to be a difference, complete a Bonferroni pairwise test to determine which or if all the means are different from each other.

4. A sample of 200 components is selected from the output of a factory that uses three different machines to manufacture these components. Each component in the sample is inspected to determine whether or not it is defective. The machine that produced the component is also recorded. The results are in the table below. The manager wishes to determine whether or not there is a relationship between the proportion of defectives and the machine used.

	Machine		
Outcome	A	B	C
Defective	8	6	12
Non-Defective	54	62	58

a. State the null and alternative hypotheses.

Ho: The probability of a part being defective is the same for any of the three machines. Pa=Pb=Pc

Ha: The probably of defectiveness is not the same for each machine. At least one of the probabilities differs

b. Give the p-value and the decision from the test at the 5% significance level.

```
> factory=matrix(c(8,51,6,62,12,53),nrow=2, ncol = 3)
> factory
     [,1] [,2] [,3]
[1,]    8    6   12
[2,]   51   62   53
> chisq.test(factory)

        Pearson's Chi-squared test

data:  factory
X-squared = 2.6368, df = 2, p-value = 0.2676
```

c. What do you conclude from this significance test at the 5% level? State your conclusion in the context of the problem.

Fail to reject the null hypothesis which states that the probability of failure is the same for any machine

5. The U.S. National Center for Health Statistics records information on each new baby born, such as time and date of births, weight, and sex. On bit of information available from these data is the day of the week on which each baby was born. Under the proportional model, we would expect that babies should be born at the same frequency on all seven days of the week. But is this true? The following table lists the number of babies born on each day of the week in a random sample of 350 births.

Day	Number of Births
Sunday	33
Monday	41
Tuesday	63
Wednesday	63
Thursday	47
Friday	56
Saturday	47

a. Write a null and alternative hypothesis to test the claim that the proportion of births should be 1/7 for each day of the week.

Ho: p1=p2=p3=p4=p5=p6=p7=1/7
Ha: one not equal to the rest

b. Test this hypothesis giving the **test statistic** and **p-value** and **conclusion** about this test. (You can use R to get the values).

```
> days=c(33,41,63,63,47,56,47)
> chisq.test(days)

        Chi-squared test for given probabilities

data:  days
X-squared = 15.24, df = 6, p-value = 0.01847

> resid(chisq.test(days))^2
[1] 5.78 1.62 3.38 3.38 0.18 0.72 0.18
>
```

With a p-value of .018 we will reject the null hypothesis in favor of the alternative hypothesis.

For each of the following situations in problems 6 - 10, determine which of the following tests should be used. Assume necessary conditions have been met. (Answers may be used more than once or not at all.)

A. 1 sample z test B. 1 sample t test C. 2 sample t test D. Chi-squared
E. 1 proportion z test F. 2 proportion z test G. matched pairs t test

6. Consumer Reports (January 1993) stated that the mean retail cost of an AT&T model 3730 cellular phone was $600. A random sample of 10 stores in Los Angeles had a mean cost of $586.50 with standard deviation of $26.77. Does this indicate that the mean cost in Los Angeles is less than $600?

1 sample t test

7. Athabasca Fishing Lodge is located on Lake Athabasca in northern Canada. In one of their recent brochures, the lodge advertises that 75% of their guests catch northern pike over 20 pounds. Suppose that last summer 64 out of a random sample of 83 guests did in fact catch northern pikes weighing over 20 pounds. Does this indicate that the population proportion of guests who catch pikes over 20 pounds is different from 75%?

1 sample proportion z test

8. A new law has been passed giving city police greater powers in apprehending suspected criminals. For six neighborhoods, the numbers of reported crimes one year before and one year after the new law are shown. Does this indicate that the number of reported crimes have dropped?

Neighborhood	1	2	3	4	5	6
Before	18	35	44	28	22	37
After	21	23	30	19	24	29

Matched Pairs test

9. A machine has a record of producing 80% excellent, 16% good, and 4% unacceptable parts. After extensive repairs, a sample of 200 produced 157 excellent, 42 good, and 1 unacceptable part. Have the repairs changed the nature of the output of the machine?

Chi-squared test

10. A random sample of 378 hotel guests was taken one year ago, and it was found that 178 requested nonsmoking rooms. Recently, a random sample of 516 hotel guests showed that 320 requested nonsmoking rooms. Do these data indicate that the proportion of hotel guests requesting nonsmoking rooms has increased?

2 proportion z test