**Mat 363 Fall 2011 Exam 4 Take Home Part**

**Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Note: to get full credit, you do need to do the 5 steps of hypothesis testing in problems 2-6. The conclusions should be stated in terms of the problem. A conclusion that says “we reject the null hypothesis” will not get any credit.

1. (10 pts) Let X1, X2,…,X10 be a random sample from a Poisson distribution with parameter λ. Previous observations suggest that λ=.5. Changes in the conditions of the processes were implemented to decrease the mean. We would like to see if there is enough evidence the mean of the Poisson distribution has decreased.
   1. Write the hypothesis of this test.
   2. Let Y= X1+X2+…+X10. What is the distribution of Y?
   3. If we set rejection region C={Y≤2}, calculate the significance level of the test.
   4. Draw the power function of this test.

### Just in case you are interested in doing this in R:

### dpois(x, lambda) # computes the density function

ppois(x, lambda) # computes the cumulative distribution function

qpois(p, lambda) # computes the quantile function (inverse of cumulative)

rpois(n, lambda) # gives n random samples

for the Poisson distribution with parameter lambda. The cumulative can help you draw the power function:

x<-seq(0,lambda,by=0.1)

plot(x,F(x))

(you have to enter the right value for lambda and the right function for F). Talk to me if you have doubts.

1. (10 pts) A recent survey found the following prices per gallon of regular unleaded gasoline at Albany, NY: 3.46, 3.57, 3.50, 3.49, 3.79, 3.56, 3.50, 3.57, 3.59, 3.48, 3.65, 3.59, 3.42, 3.6, 3.57, 4.53. Is there sufficient evidence to claim that the average unleaded gas price in Albany is greater than 3.50? Do the 5 steps of hypothesis testing.

R (or Excel) can help you with the running of the test, but you have to interpret what you got and do the 5 steps of a test of hypothesis: to enter data in a vector called x:

x<-scan() # hit enter only once!!!! Now start typing each number (no commas) and hit enter once after each number. When you are done, hit enter again. Alternatively, you can enter the values in an excel worksheet and copy the column (or row). After x<- scan() , paste the values you just copied (Ctl-v).

t.test(y,mu= ) # enter the value for mu in the Null hypothesis.

Include a graphical display of the data:

boxplot(x)

1. A company that manufactures light bulbs claims that a particular type of light bulb will last 850 hours on average with standard deviation of 50. A consumer protection group thinks that the manufacturer has overestimated the lifespan of their light bulbs by about 40 hours. How many light bulbs does the consumer protection group have to test in order to prove their point with reasonable confidence?
2. Set up the hypothesis for this test and explain what are the assumptions of your test.
3. Calculate the minimum sample size so that the power of the test at the level suggested by the consumer protection group is .90. To help you, we could use the statistical package R. The power function depends on the significance level, the normalized difference of the means (µ1- µ0)/s), and the sample size. You can find the exact formula in the book. I found that R has a procedure that computes the different elements involved in the power function. Here is the code in R that helps you find the value of n. Since I am giving you the code, I want to set up the equation from where the value of n will be solved.

**library(pwr)**

**pwr.t.test(d=(**µ1- µ0)/s**,power=0.9,sig.level=0.05,type="one.sample", alternative= )**

Notes: (i) you have to enter the actual value for d. (ii) If your alternative hypothesis was one sided, enter **alternative="less" (or “greater”),** and if it was two-sided enter **alternative="two.sided".**

1. Next, suppose we have a sample of size 10, how much power do we have, keeping all of the other numbers the same? We can use the same command inputting the sample size instead of the power:

**pwr.t.test(d==(**µ1- µ0)/s**,n=10,sig.level=0.05,type="one.sample", alternative= )**

1. (10 pts) The gas prices in problem 2 were sampled in Albany, NY. Compare those prices with the following sample from Buffalo: 3.38, 3.3, 3.61, 3.77, 3.72, 3.24, 3.74, 3.57, 3.90, 3.44, 3.4, 3.39, 3.41, 3.21, 3.19 . Is there evidence that the average price of gas is higher in Buffalo? Conduct the 5 steps of test of hypothesis. See note below for the t.test in R for 2 samples. Explain which test you decided to use and why. Include a graphical display of your data. ( You can display two groups x and y together with the command: boxplot(x,y)).
2. (10 pts) With the data in problems 2 and 7, run a test of hypothesis to see if the variance in the gas prices is the same in the two cities. Clearly state the rejection region.
3. (10 pts) Thirty magazines were ranked by educational level of their readers. Three magazines were randomly selected from the first, second, and third ten magazines. Six advertisements were randomly selected from each of the nine selected magazines. The magazines were:

Group 1 Highest educational level: 1. Scientific American 2. Fortune 3. The New Yorker.

Group 2 Medium educational level: 4. Sports IIlustrated 5. Newsweek 6. People

Group 3 Lowest educational level : 7. National Enquirer 8. Grit 9 True Confessions

For each advertisement, the data in file Magazine.xlsx were observed.

**Variable Names:**

WDS = number of words in advertisement copy

SEN = number of sentences in advertising copy

3SYL = number of 3+ syllable words in advertising copy

MAG = magazine (1 through 9 as above)

GRP = educational level (as above)

Questions arise as to whether significant differences exist in the characteristics of advertising copy among the magazines or the groups of magazines. Also relevant to readability are the number of words per sentence and the proportion of three syllable words in the copy. (a) Test whether there are significant differences in the number of words in advertisement copy among the 3 educational level groups. Make a graphical display of the variables of interest and follow the 5 steps of hypothesis testing. (b) Test whether there are significant differences in the number of words in advertisement copy among the 9 magazines. Make a graphical display of the variables of interest and follow the 5 steps of hypothesis testing.

**Commands for t.test on R:**

# independent 2-group t-test  
t.test(y1,y2) # where y1 and y2 are numeric

# paired t-test  
t.test(y1,y2,paired=TRUE) # where y1 & y2 are numeric

# one samle t-test  
t.test(y,mu=3) # Ho: mu=3

**Notes about the power function:**

|  |  |
| --- | --- |
| pwr.t.test {pwr} | R Documentation |

Power calculations for t-tests of means (one sample, two samples and paired samples)

**Description:** Computes power of tests or determine parameters to obtain target power (similar to as power.t.test).

**Usage:** pwr.t.test(n = NULL, d = NULL, sig.level = 0.05, power = NULL,

type = c("two.sample", "one.sample", "paired"), alternative =c("two.sided","less","greater"))

|  |  |
| --- | --- |
| **Arguments**  n | Number of observations (per sample) |
| d | Effect size |
| sig.level | Significance level (Type I error probability) |
| power | Power of test (1 minus Type II error probability) |
| type | Type of t test : one- two- or paired-samples |
| alternative | a character string specifying the alternative hypothesis, must be one of "two.sided" (default), "greater" or "less" |