Homework 5

In this homework we cover the basics behind testing, including type I and type II errors, rejection regions and p-values. We also look at the differences between one and two sided tests. The last question gives JMP practice.

(1) Suppose you are willing to put 1 out of 200 innocent people in prison. You do the test H_0 : person is innocent against H_A : person is guilty. What is the type I error in this case?

Can you also calculate the type II error?

(2) What is wrong with these statements?

- (i) A researcher tests the following hypothesis $H_0 : \bar{x} = 23$.
- (ii) A company wants to test that the mean mileage of their cars is greater than 40 miles per gallon. They state their null hypothesis as $H_0 : \mu > 40$.
- (iii) The z-statistics is equal to -1.6. Because -1.6 is less than 5%, the null hypothesis is rejected at the 5% level.

(3) Select the appropriate null and alternative hypothesis in each of the following cases:

- (i) The Batt recently changed the format of their opinion page, you want to see what the students think of this change. You take a random sample of students and select those who regularly read the Batt. They are asked to indicate their opinions on the changes using a 5 point scale. -2 if the new format is much worse than the old, -1 if the new format is worse than the old, 0 if the old format is the same as the old, 1 if the new format is better than the old and 2 if the new format is much better than the old.
- (ii) The average square footage of one bedroom apartments in a new student housing development is advertised to be 880 square feet. A student group thinks that the apartments are smaller than advertised. They hire an engineer to measure a sample of apartments to test their suspicion.

(4) A test statistic of the null $H_0 : \mu = \mu_0$ gives the z-transform $z = 2$.

- (a) What is the p-value if the alternative is $H_A : \mu > \mu_0$.
- (b) What is the p-value if the alternative is $H_A : \mu < \mu_0$.
- (c) What is the p-value if the alternative is $H_A : \mu \neq \mu_0$.

(5) Suppose that I want to test the hypothesis H_0 : that $\mu = 6$ against the alternative $H_A : \mu \neq 6$. The population standard deviation is assumed to be $\sqrt{3}$ and the sample size 40. I do the test at the 5% level and I am unable to reject the null.

- (i) Make a plot of the distribution of the sample mean under the null hypothesis (giving the mean and standard deviation under the null). Indicate on the plot where the sample mean must lie if we cannot not reject the null at the 5% level (the interval in which $\bar{X}$ must lie). Hint: The non-rejection region.

(ii) Is this statement correct: 'I could have made a type II error'.
TRUE or FALSE? Give a reason for your answer.

(6) There has been a lot of speculation that the average height of A&M students has increased over the past 20 years. It is known that in 1992 the average height of an A&M student was 165 cm.

Some students are wondering whether there has been a change in student heights. They cannot question all A&M students, but they want to randomly sample some students and draw conclusions about the mean based on the sample.

(a) They would like the 90% confidence interval to have length 5cm (the margin of error to be equal to 2.5). Using a guess of 30cm for the standard deviation, how many students should they question?

(b) Suppose that the mean height of an A&M student is now μ_0 and the standard deviation is known to be 30cm. 100 A&M students are random drawn and the sample mean is evaluated.

(i) What is the mean and standard error of the sample mean?

(ii) Make a sketch of the distribution of the sample mean, stating all the assumptions you make.

(iii) Suppose the population mean is $\mu_0 = 165$ (it has not changed since 1992). Calculate the probability that the sample mean is greater than 172 given that the population mean is 165.

(iv) Using the data above (sample mean is $\bar{X} = 172$, which is calculated from a sample of 100 students and the population standard deviation is 30) test the hypothesis the mean height of students has increased since 1992. State precisely your null and alternative and do the test at the 5% level.

(v) Using the data above (sample mean is $\bar{X} = 172$, which is calculated from a sample of 100 students and the population standard deviation is 30) test the hypothesis the mean height of students has decreased since 1992. State precisely your null and alternative and do the test at the 5% level.

(vi) Using the data above (sample mean is $\bar{X} = 172$, which is calculated from a sample of 100 students and the population standard deviation is 30) test the hypothesis the mean height of students has changed since 1992. State precisely your null and alternative and do the test at the 5% level.

(7) In the above question we assumed the standard deviation was known, even if it were estimated from the data there is little difference between using the t-distribution or the normal distribution since the sample size is large (remember the only reason we use a t-distribution instead of a normal is to correct for the fact that for small samples the estimated (sample) standard deviation tends to underestimate the amount of variability in the data).

Let us now see whether female heights have changed since 1992. Let us suppose in 1992 it is known that the mean height of females was 163cm. We want to see whether female heights have changed since 1992.

(i) State the null and alternative hypothesis of interest.

(ii) A random sample of 30 heights is taken:

156.14	148.03	151.25	153.48	140.33	156.72	166.32	167.73	179.30	174.13
174.84	159.90	170.22	177.16	166.48	190.56	171.34	149.48	191.35	149.79
157.92	122.74	208.53	152.34	179.85	133.59	161.93	176.21	201.60	167.79

The sample mean and sample standard deviation of the above data is $\bar{x} = 165.2$ and $s = 18.93$ (remember since we have estimated the standard deviation we need to use a t-distribution with 29dfs).

Do the test at the 5% level.

(8) Return to the cow weight data analysed in HW4. It is known that the mean weight of 8 week old calves until recently was 137 pounds. But in recent years, there has been speculation that changes in farming practices (using corn feed instead of grass) has led to an **increase** in mean weight of 8 week old calves. The standard deviation for the weight of a calf is known to be 17 pounds. Use the option `Analyse -> Distribution -> Then place Wt8 into Y, column -> Then press okay -> You should now see the histogram, right click on the red triangle next to Wt 8 -> select Test Mean (placing the mean under the null and standard deviation 17 in the correct places)` to test the hypothesis that the mean weight has increased. State precisely your null and alternative.

(i) What is the result of the test at the 5% level?

(ii) What is the result of the test at the 1% level?

(iii) Based on your answer above, would you say that there is some <u>or</u> overwhelming evidence that the mean weight has increased?