Boston University
Department of Mathematics and Statistics

MA 214

Applied Statistics

Lab Session 3: Two-Sample Inference

In this lab we will use the JMP software to review the ways to calculate and interpret descriptive statistics of data, and to explore the tools of statistical inference based on two samples.

Preparation

We will be using the same data set as last week, so if necessary, please again download the JMP data set called Televisions.jmp from the Blackboard site under the Course Documents link. Recall that in this file, for each of the forty largest countries in the world (according to 1990 population figures), data are given for the country's life expectancy at birth, number of people per television set, and number of people per physician. SOURCE: *The World Almanac and Book of Facts 1993* (1993), New York: Pharos Books.

Variable Descriptions:

Columns

1	Country
2	Life expectancy
3	People per television
4	People per physician
5	Female life expectancy
6	Male life expectancy

Missing values are denoted with *.

Questions

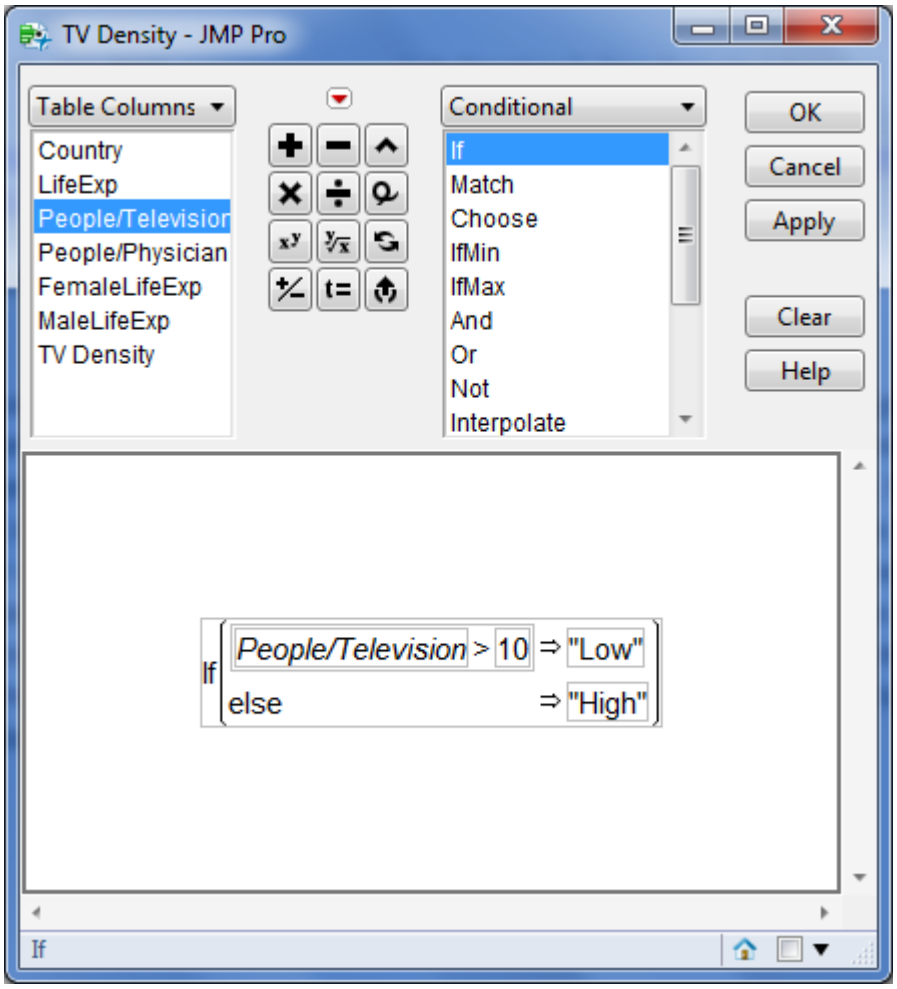
1) Experts from the University of Queensland, Australia, write: “TV viewing time may have adverse health consequences that rival those of lack of physical activity, obesity and smoking; every single hour of TV viewed may shorten life by as much as 22 minutes.” SOURCE:

http://www.telegraph.co.uk/health/healthnews/8702101/Every-hour-of-TV-watching-shortens-life-by-22-minutes.html

Let us investigate the claim in this recent news article by taking a closer look at our data set.

(a) After opening the Televisions.jmp file in JMP, go to the empty column to the right of MaleLifeExp and right-click and select the "New Column…" option. Call the new column "TV Density" and set the "Data Type" to "Character" and the "Modeling Type" to "Nominal", then press the OK button. We will use this column to classify each country in our lists as one that has either a "Low" or "High" amount of televisions, using the information contained in the "People/Television" column. To do this, right-click on the new "TV Density" column and select the "Formula…" option, and then click on the word "Conditional" in the list of functions on the right, and then on the word "If" once the menu on the right changes to the list of "Conditional" functions.

Now, in the box at the bottom of the window, set up the following conditional statement (**Note:** you can make the "People/Television" column name appear just by clicking on it in the list of Table Columns on the left) and then press the OK button:



If a country has more than 10 people for each television, we will say that it has a “Low” TV density; otherwise, we say it has a “High” TV density.

(a) Now that we have classified each country in terms of its TV density, let’s take a fresh look at the distribution of life expectancies by making separate box plots for each type of TV density. To do this, go to the “Analyze -> Fit Y by X” option, and then place “LifeExp” in the “Y, Response” role and “TV Density” in the “X, Factor” role. After pressing the OK button, you will see two side-by-side dot plots for each type of TV density. Go to the red triangle, and from the Display Options menu, select Box Plots to superimpose box plots onto these dot plots. Comment on the difference between these two box plots, and the apparent effect of televisions on human life expectancy.

The median for two groups, "lower" and "higher" life expectancy, is about 61 and 72, respectively. The IQR for the lower life expectancy group is about 51 to 70 and for the higher life expectancy group it is about 64 to 79.

Based on the two box plots, we can see the clear difference between the two groups, low and high TV Density. The low TV density group tends to have a significantly lower life expectancy, with a median life expectancy of only 61 years. On the other hand, the high TV density group tends to have a much higher life expectancy, with a median life expectancy of 72.

However, it's important to note that correlation does not imply causation. While these boxplots show a relationship between TV density and life expectancy, they do not prove that the presence of more televisions directly causes an increase in life expectancy. There could be other factors at play that are influencing both TV density and life expectancy.

For example, TV density might be a proxy for wealth or development level of a region. Areas with more TVs per person might be wealthier or more developed, which could lead to better healthcare and thus higher life expectancy. Similarly, areas with fewer TVs per person might be less developed or poorer, with less access to healthcare, leading to lower life expectancy.

To establish a causal relationship, we would need to control for these other factors, which might require more complex statistical analysis or experimental design. It's also possible that the relationship is not linear – for instance, there might be diminishing returns to life expectancy at very high levels of TV density.

(b) We wish to carry out an appropriate statistical inference procedure to find out if there is any statistical difference between the mean life expectancies of the “Low” and “High” TV density countries. Write down the null and alternative hypothesis.

The null hypothesis (H0) would be that there is no difference in the mean life expectancies between countries with low TV density and high TV density. Mathematically, this can be represented as: H0: μ1 = μ2, where μ1 is the mean life expectancy in the low TV density group, and μ2 is the mean life expectancy in the high TV density group.

The alternative hypothesis (Ha) would be that there is a difference in the mean life expectancies between countries with low TV density and high TV density. Mathematically, this can be represented as:
Ha: μ1 ≠ μ1

(c) Now, let's conduct an appropriate test to decide about the null and alternative hypothesis. There are two options for this:

i. From the red triangle on the box plot window, choose "Means/ANOVA/Pooled t" option. Write down the appropriate p-value of the test. What is your conclusion?

In the "Pooled t Test" section, the "Prob > |t|" value is the p-value for the two-tailed t-test. The notation "< 0.0001*" indicates that the p-value is less than 0.0001, which is much smaller than the common significance level of 0.05.

This small p-value suggests that it is very unlikely we would observe such a large difference in mean life expectancies between the low and high TV density groups if the null hypothesis were true (i.e., if there were no real difference in mean life expectancies between the groups).

Therefore, we reject the null hypothesis and conclude that there is a statistically significant difference in mean life expectancies between the low and high TV density groups. This aligns with the boxplot observations you described earlier.

The "Difference" value of 12.3295 in the "Pooled t Test" section is the difference in means between the high and low TV density groups, which also aligns with the boxplot observations.

The "Analysis of Variance" section also provides a p-value (under "Prob > F"), which is less than 0.0001, indicating that the difference in means is statistically significant according to the ANOVA test as well.

ii. From the red triangle on the box plot window, choose “t-test” option. Write down the appropriate p-value of the test. What is your conclusion?

The results from the t-test with the assumption of unequal variances (also known as Welch's t-test) also show a p-value of less than 0.0001 (as indicated by "Prob > |t|: < 0.0001*"). This is much smaller than the common significance level of 0.05.

This small p-value suggests that it is very unlikely we would observe such a large difference in mean life expectancies between the low and high TV density groups if the null hypothesis were true (i.e., if there were no real difference in mean life expectancies between the groups).

Therefore, we reject the null hypothesis and conclude that there is a statistically significant difference in mean life expectancies between the low and high TV density groups. This conclusion is consistent with the conclusion from the pooled t-test.

The "Difference" value of 12.3295 is the difference in means between the high and low TV density groups, which also aligns with the boxplot observations.

iii. What is the difference between these two t-tests and what do you need to know to decide which p-value is more appropriate?

The two t-tests we've conducted are:

1. **Pooled t-test**: This test assumes that the variances of the two populations (in this case, the life expectancies in low and high TV density countries) are equal. This is also known as the assumption of homogeneity of variances.

2. **Welch's t-test**: This test does not assume equal variances. It is more robust and can be used even when the two populations have unequal variances and/or unequal sample sizes.

The choice between these two tests depends on whether the assumption of equal variances is reasonable for the data. If we have reason to believe

that the variances are equal (or nearly so), the pooled t-test may be appropriate. If not, or if we're unsure, it's generally safer to use Welch's t-test, as it is more robust to violations of the assumption of equal variances.

To decide which test is more appropriate, you would typically perform a test for equality of variances, such as Levene's test or the F-test. If this test shows a significant result ($p < 0.05$), it suggests that the variances are not equal, and you should use Welch's t-test. If the test is not significant ($p >= 0.05$), it suggests that the variances are equal, and you can use the pooled t-test.

However, in practice, many statisticians prefer to use Welch's t-test by default, because it has good performance under a variety of conditions and because the assumption of equal variances is often not met in real-world data.

(d) Let us see whether it is reasonable to assume that both of these populations have the same variance. Perform a test of unequal variances in JMP by going to the red triangle on your box plot window and selecting the “Unequal Variances” option. By default, JMP will perform many tests to see whether or not the variances are equal. The test of interest to us which has been discussed in lecture can be found at the last line under the “Tests that the Variances are Equal” section that appears. Specifically, it is the “F Test 2-sided” test under this section. What is the p-value of this test, and what is your conclusion about the population variances based on it? Clearly state the hypothesis and your conclusion.

The F-test for equality of variances (the "F Test 2-sided" row in the table) has a p-value of 0.1853. This is greater than the common significance level of 0.05.

Therefore, we fail to reject the null hypothesis and conclude that there is not a statistically significant difference in variances between the life expectancies in low and high TV density countries. In other words, based on this test, it is reasonable to assume that the variances of the two populations are equal.

However, it's important to note that failing to reject the null hypothesis is not the same as proving the null hypothesis. While this test suggests that the variances are equal, it does not definitively prove that they are. It simply means that, based on the data we have, we do not have enough evidence to conclude that the variances are different.

Also, note that all the other tests for equality of variances (O’Brien, Brown-Forsythe, Levene, Bartlett) also have p-values greater than 0.05, which

supports the conclusion from the F-test.

Given these results, it would be reasonable to use the pooled t-test for comparing the means of the two groups, as this test assumes equal variances.

(e) Based on your answer in part (d), state which of the tests in part c is more appropriate? In this context, for your final conclusion, did the choice of the test make any difference?

Based on the results from the F-test for equality of variances, it appears that the assumption of equal variances is reasonable for your data. Therefore, the pooled t-test, which assumes equal variances, would be an appropriate test to use.

However, it's important to note that both the pooled t-test and Welch's t-test (which does not assume equal variances) gave very similar results in this case. Both tests resulted in a p-value of less than 0.0001, leading to the rejection of the null hypothesis and the conclusion that there is a statistically significant difference in mean life expectancies between the low and high TV density groups.

So, in this particular case, the choice of test did not make a difference to the final conclusion. Both tests led to the same conclusion about the difference in mean life expectancies between the two groups.

In other situations, especially when the sample sizes and/or variances are very different between the two groups, the choice of test could potentially lead to different conclusions.

(f) The CEO of a large television manufacturing company reads the news article referenced in part **(a)** and becomes enraged after seeing a subsequent decline in TV sales. She hires her own independent researchers (you) to determine the validity of the findings of the experts at the University of Queensland. After showing her your JMP output and conclusions from parts **(c)** and **(d)**, she decides that the experts at the University of Queensland were completely wrong in their conclusions, and that actually television sets have a clear positive influence on the life expectancy of people around the world. How can you explain the different conclusions reached by you, the CEO, and the experts at the University of Queensland?

The discrepancy in the conclusions drawn by the University of Queensland experts, the CEO, and our analysis can be attributed to the difference in the nature of the

studies and the interpretation of the results.

1. **University of Queensland Study**: The statement from the University of Queensland experts seems to be based on a study that looked at the impact of TV viewing time on individual health and life expectancy. This is a different question than the one addressed in our analysis. They might have conducted a longitudinal study tracking individual health outcomes and TV viewing habits over time. Their conclusion suggests a negative impact of prolonged TV viewing on health, which could be due to sedentary behavior associated with excessive TV watching.

2. **CEO's Interpretation**: The CEO's conclusion that "television sets have a clear positive influence on the life expectancy of people around the world" is an overinterpretation of our analysis. Our analysis found a correlation between high TV density and higher life expectancy at the country level, but this does not imply causation. It's possible that TV density is a proxy for other factors like wealth, development, or access to healthcare, which are the real drivers of the observed difference in life expectancy. It's also important to note that TV density is not the same as TV viewing time, which was the focus of the University of Queensland study.

3. **Our Analysis**: Our analysis compared life expectancy in countries with low and high TV density. We found a statistically significant difference in life expectancy between these two groups. However, this does not mean that TVs themselves are causing this difference. Correlation does not imply causation, and there could be many confounding factors at play.

Therefore, the different conclusions are due to different research questions and different interpretations of the data, which in turn tells us that it is important to carefully consider the nature of the data and the question being asked when interpreting statistical results.

2) Last week we computed confidence intervals separately for the true average life expectancy for males and females, and discussed the possibility that on average, females tend to live longer lives than males. We will now further study the difference between life expectancies of males and females around the world by applying the theory of two-sample tests.

(a) Based on our data set, what type of two-sample test for this question would be appropriate? In other words, do we have independent samples, or paired data, and why?

Given that the dataset includes separate life expectancy values for males and females for each of the forty largest countries, we have paired data. The life

expectancy of males and females are paired by each country.

Therefore, a paired samples test would be the most appropriate statistical test to use if we want to compare the mean life expectancy of males versus females.

The reason we use a paired test in this situation is because the life expectancy of males and females in the same country are likely not independent. They are both influenced by the same country-specific factors, such as healthcare quality, diet, lifestyle, and socioeconomic conditions. By using a paired test, we can control for these country-specific factors and focus on the difference in life expectancy between males and females.

(b) Assuming that the data are paired, we will simplify the problem and apply a one-sample t test to the difference between female life expectancy and male life expectancy. Create a new column in the JMP data table called “Difference”, and set the “Data Type” to “Numeric” and the “Modeling Type” to “Continuous”. Then, right-click on the new column and select the “Formula...” option. This time, the formula that you should enter in the bottom box is the difference between FemaleLifeExp and MaleLifeExp; i.e., *FemaleLifeExp – MaleLifeExp*.

After making this column, take a look at its distribution and summary statistics by using the Analyze -> Distribution command. Perform a hypothesis test to determine whether the true average difference between female and male life expectancies is equal to 0 or greater than 0.

Based on the results from the one-sample t-test, the p-value is approximately 0.04209. This is less than the common significance level of 0.05.

Therefore, we reject the null hypothesis and conclude that there is a statistically significant difference in life expectancy between males and females, with females having a longer life expectancy on average.

The mean difference in life expectancy between females and males in your sample is 5.075 years, which is significantly greater than 0. This suggests that, on average, females in the countries in our sample have a life expectancy that is about 5 years longer than males.

However, since this is a sample of the 40 largest countries by population, so the results might not generalize to all countries. Also, while this analysis suggests a difference in life expectancy between males and females, it does not identify the reasons for this difference. Other factors not included in this analysis may also influence life expectancy.

(c) What is a 95% confidence interval for the true average difference between female and male life expectancies? How does this interval reflect the conclusion you reached above in part **(b)**?

The 95% confidence interval for the true average difference between female and male life expectancies, as provided in the summary statistics, is approximately (4.300, 5.850).

This interval means that we are 95% confident that the true average difference in life expectancy between females and males (with females living longer) is between 4.300 years and 5.850 years.

This confidence interval supports the conclusion reached in part (b) because the entire interval is greater than 0. If the null hypothesis were true (i.e., there was no difference in life expectancy between females and males), we would expect the confidence interval to include 0. However, since the entire confidence interval is above 0, this suggests that females have a longer life expectancy than males on average.