# Lab 7: More Hypothesis Testing: on Goodness of Fit (Chapter 4) and on Means of Two Populations (Chapter 5)

## Objectives

* Learn how to do hypothesis tests on the form of the distribution of data by using the goodness of fit test statistic
* Expand on hypothesis tests of means from one to two populations of data

## Goodness of Fit

Up to now, most of our hypothesis testing has been on statistics of data; namely, the variance or mean. We can also do hypothesis testing on the form of the distribution of data. Most often, we deal with data that is normally distributed. However, if we are not, **we can use the chi-squared goodness of fit test statistic to form a hypothesis test on the form of the distribution.** First, the data is arranged in a histogram with k number of bins. We then calculate the test statistic:

Where O is the observed frequency of the ith bin, and E is the expected frequency of the ith bin based on the form of distribution that we are testing. Example 4-19 on page 221 of the textbook goes into more detail on how to use the goodness of fit test statistic. **Note that you can use JMP to both create and analyze fits of histograms, as we did in the beginning of the semester**. But in the absence of JMP, the goodness of fit test can be used to calculate by hand.

## Hypothesis Testing on Means of Two Populations

In chapter 4, we worked with hypotheses on statistics relating to one population. Now, we will consider tests on two or more populations of data. Instead of comparing a mean or variance to a certain number, we will be comparing the means or variances to each other. **So, before our null hypothesis was ; now, our null hypothesis deals with the differences between the means:**

**The tests work similarly to those in chapter 4; when we are testing the difference between means of two populations and the variance is known, we use the Z test. If the variance is unknown, we use the T test**. However, to take into account the two populations, we calculate our Z and T test statistics differently, according to the table below.

|  |  |
| --- | --- |
| **Test** | **Test Statistic** |
| Difference in means of two populations, variance known (5-2) |  |
| Difference in means of two populations, variance unknown (5-3) | where |

## Lab 7 Exercises

You can use JMP or Excel for any calculations, as long as you are able to explain your answers and show your work.

1. Using the class weight data uploaded on Blackboard, decide whether the distribution is normal or lognormal using:
   1. probability plots (section 3.6 of the textbook)
   2. chi -squared goodness of fit test (section 4.10)

For the following data, = 56.7, = 59.2, S1 = S2 = 3.1, n1 = n2 = 16:

1. State the appropriate hypothesis test to see if the two populations’ means are the same.
2. Which test statistic are you going to use and why?
3. Find the P-value for this test.
4. Find the 95% confidence interval bounds on the sample mean difference.
5. What is the practical meaning of this interval?