Chapter 6

Hypothesis Testing

Course Website: `www.math.mun.ca/~sneddon/st2500`

These handouts are modifications of lab notes prepared by Lauren Granter.

During the next two lab sessions, we will study some examples to see how Minitab can be used for hypothesis testing. We will look at 1-sample tests about the population mean and proportion, and 2-sample tests to compare means.

6.1 Test for Population Mean

Hypothesis testing on the population mean (μ) is broken down into two situations: large sample ($n \geq 30$) and small sample ($n < 30$).

6.1.1 Test for Population Mean: Large Sample

EX: The U.S. Commerce Dept. is interested in the average house price of all new houses sold in the U.S. They believe that this average price is now above \$195,000. They selected a random sample of 345 homes, and found their average was \$201,400. Does the data support their claim? Use $\alpha = 0.01$. Assume the standard deviation of the price of all new homes sold is \$38,000.

Since the sample size $n = 345$ is large, we use the test procedure that uses the following test statistic:

$$z = \frac{\bar{x} - \mu_o}{\sigma/\sqrt{n}}$$

The question claims that the average (mean) is above \$195,000:

$$H_o : \mu = 195000, \quad H_a : \mu > 195000$$

We do this in Minitab as follows, since the alternative is the mean **greater than** 195000:

1. Select **Stat–Basic Statistics–1 Sample Z**
2. Select **Summarized Data**, and enter 345 for **Sample size** and 201400 for **mean**.
3. Enter 38000 in the **Standard deviation** box.
4. Enter 195000 for **Test Mean**.
5. Select **Options** and change **Alternative** to **greater than**.
6. Select **OK**.

The output is shown below:

```
One-Sample Z

Test of mu = 195000 vs > 195000
The assumed standard deviation = 38000

                          95%
                        Lower
  N    Mean  SE Mean    Bound     Z   p # p-value
345  201400     2046   198035  3.13   0.001
```

From the output: $z = 3.13$, p-value $= P(z > 3.13) = 0.001$.
Rejection region ($\alpha = 0.01$): Reject H_o if $3.13 > z_{0.01}$.
From the normal distribution table, $z_o = 2.33$. Since $3.13 > 2.33$, we reject H_o. Alternatively, we can say p-value $< \alpha = 0.01$, so reject H_o.
Therefore it seems very likely that the average house price for new homes is greater than \$195000.

6.1.2 Test for Population Mean: Small Sample

EX: The data file **pallet.mtw** on the course website contains a sample of the weights of wooden pallets of 2 types of shingles ("Boston" and "Vermont").

1. Does it appear that the mean weight of the Boston shingles is less than 3150 pounds? Test at $\alpha = 0.05$.

2. Does it appear that the mean weight of the Vermont shingles differs from 3700 pounds? Test at $\alpha = 0.05$.

3. Evaluate whether the assumption needed to conduct the tests in (1) and (2) has been seriously violated.

We'll go through (1), and you can work on (2) and (3) on your own.

1. For the test about μ with small sample size, we do the following in Minitab:

   The first step is to download the data from the webpage onto your Desktop, then select **Open Worksheet** in Minitab to get the data into Minitab.

   In this case, we have a small sample size, so we need the test procedure that uses the following test statistic:

   $$t = \frac{\bar{x} - \mu_o}{s/\sqrt{n}}$$

   In this case:

   $$H_o : \mu = 3150, \quad H_a : \mu < 3150$$

   We do the following in Minitab:

   (a) Select **Stat–Basic Statistics–1 Sample t**
   (b) Select the Boston column for **Samples in columns**.
   (c) Set **Test mean** as 3150.
   (d) Select **Options** and change **Alternative** to **less than**
   (e) Select **OK**.

   Enter your answer from Minitab in the space below.

2. Does it appear that the mean weight of the Vermont shingles differs from 3700 pounds? Test at $\alpha = 0.05$.

   In the space below, write down the hypotheses, test statistic and conclusions.

3. Evaluate whether the assumption needed to conduct the tests in (1) and (2) has been seriously violated.

   Enter your answer below:

6.2 Test for Population Proportion

EX: A study of 828 travellers showed that 567 of them purchased plane tickets on an airline website in the past 12 months. The major airlines believe that more than 65% of all travellers purchase their tickets on airline websites. Do the data support this? Test at $\alpha = 0.10$.

In a test about a proportion, we are only studying one test statistic:

$$z = \frac{\hat{p} - p_o}{\sqrt{p_o(1 - p_o)/n}}$$

where $\hat{p}$ is the sample proportion.

In this example, our sample size is $n = 828$ and $x = 567$, which is the number of people with the characteristic we are interested in. Our sample proportion is $\hat{p} = 567/828 = 0.685$. Our hypotheses are

$$H_o : p = 0.65 \quad H_a : p > 0.65$$

We conduct our test for p in Minitab as follows:

1. Select **Stat-Basic Statistics-1 Proportion**
2. Select **Summarized data** and enter 828 after **Number of trials** and 567 after **Number of events**
3. Select **Options** and **greater than** for **Alternative**, choose **Use test and interval based on normal distribution** and **greater than** for **Alternative**. Click **OK**.
4. Click **OK**.

The output is as follows:

```
Test and CI for One Proportion

Test of p = 0.65 vs p > = 0.65

                                       95%
                                     Upper
Sample     X    N  Sample p          Bound  Z-Value  P-Value
1        567  828  0.684783       0.658225     2.10    0.018
```

NOTE: The term **Sample p** is our $\hat{p}$.
Report your conclusions in the space below.

6.3 Comparing Means: Two Populations

Next we discuss how to use Minitab to compare means in two populations.

6.3.1 Independent Populations: Large Sample Sizes

If we have large samples ($n_1 \geq 30$ AND $n_2 \geq 30$) we use the test statistic

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\sigma_1^2/n_1) + (\sigma_2^2/n_2)}}$$

Unfortunately, Minitab won't do this test procedure for us. So we'd have to use our calculators to do it on our own.

6.3.2 Independent Populations: Small Sample Sizes

EX: Return to example on weights of shingles. Do the data suggest that Vermont shingles weight more than Boston shingles? Use $\alpha = 0.1$ to decide.

Since we have small sample sizes, we use the test statistic

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

If we call the weight of Boston shingles as population 1, and the Vermont shingles population 2, write down the appropriate hypotheses:

Now we do the following in Minitab:

1. Select **Stat–Basic Statistics–2 Sample t**
2. Select**Samples in different columns**, putting Boston in **First**, Vermont in **Second**.
3. Select **Assume Equal Variances**.
4. Select **Options** and change **Alternative** to **less than**. Leave **Test difference** as 0.
5. Select **OK**.

State your test statistic and conclusion below, based on the Minitab output.

Does it appear that the normality assumption is satisfied?

6.3.3 Paired Data

EX: This dataset contains the prices of 15 books at Miami University and at Amazon.com. The data is in the file **textbook.mtw** at the course website.

Is there evidence of a difference between the average price of texts at the university bookstore and Amazon.com?

This is **paired data**: we have data in 2 groups, but the 2 groups are related (same textbooks, but being sold at 2 different locations). So the paired t-test is appropriate.

I'll leave you to figure most of the details of this one, but I'll tell you the starting point: **Stat – Basic Statistics – Paired t**. Enter your details below: