

Chapter 24: Comparing Means

Let's follow the same path through this as we did when we talked about comparing two proportions.

The Sampling Distribution (Theory)

Making Two Things Into One

Do you remember how we turned two proportions into one?

We subtracted! We'll do that here, too. Thus, $\bar{x}_1 - \bar{x}_2$ will be the statistic that estimates the parameter $\mu_1 - \mu_2$. We need to figure out some things about the distribution of that statistic...

The Center

We laid the groundwork for this a long time ago! If we subtract two random variables, how do we find the mean of the new resulting random variable?

Just subtract the original means!

Equation 1 - The Mean of the Sampling Distribution

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

The Spread

...and what do we do for spread?

That's right—we switch to variance, and we add the individual variances. The variances of the individual variables are $\sigma_{\bar{x}_1}^2 = \frac{\sigma_1^2}{n_1}$ and $\sigma_{\bar{x}_2}^2 = \frac{\sigma_2^2}{n_2}$.

Equation 2 - The Variance of the Sampling Distribution

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

...but since we don't know σ (since it is a parameter), we must make a substitution. What did we do in the previous chapter? Let's do that again. Thus: change σ into s , and take the square root.

Equation 3 - The Standard Error (Two Sample)

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The Shape

In the previous chapter, replacing the population standard deviation with the sample standard deviation forced us to switch distributions—from normal to Student's t . I wonder what happens when you switch two population standard deviations... wouldn't it be cool if the resulting statistic had a Student's t distribution, just like before?

Alas—for reasons that probably don't really interest you (they are complicated)—it doesn't. For years, people made a very weak assumption and forged ahead with a different distribution. In

the 1940's someone finally found a way to use a Student's t distribution—but this requires a really nasty calculation of degrees of freedom!

Degrees of Freedom

...but now that we have nice little calculators that can do this calculation with no effort on our part, we'll use it!

Here's the formula for those degrees of freedom, in case you were wondering:

Equation 4 - The Satterthwaite Approximation

$$df \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(\frac{s_1^2}{n_1} \right)}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2} \right)}{n_2 - 1}}$$

You will occasionally find references to a conservative degrees of freedom—the smaller sample size minus one. There really isn't any reason to do that anymore (unless your teacher demands it, of course). Perhaps the most important reason to NOT use that approach is that your calculator doesn't use it!

Confidence Intervals

Formula

The formula for the confidence interval follows the same pattern that all intervals have followed (estimate plus or minus critical value times measure of variation).

Equation 5 - Two Sample Confidence Interval for Means

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

We'll use the Satterthwaite Approximation for the degrees of freedom—and no, you don't have to show the work! Just state the df you are using (i.e., copy the value from your calculator).

Requirements

This procedure requires that both samples were obtained randomly from their respective populations. As usual, this will either be given, or assumed true.

Be careful here! There are some problems where the two groups aren't really two samples...rather, a single sample was randomly split into two groups. If this is the case, then you CANNOT mention "two random samples" on the AP Exam and receive full credit. Thus, your first condition should either read as *both samples were obtained randomly* or that *the two groups were formed through random assignment*.

Also, it is required that both samples are small relative to their respective population sizes (less than 10%). As usual, this ought not to be a problem, and I'll rarely mention it as an actual issue.

Finally, it is required that both population variables have normal distributions. This will rarely be known, so we'll technically fail this requirement. Remember that the t procedures are robust, though—when the requirements aren't met, we might still be able to continue. In particular, we

had a set of guidelines of what we needed to see (or assume about) the sample distribution in order to continue. Because the process of combining random variables produces a shape that is more symmetric than the original shapes, we don't need to see quite as much in the two samples here. Thus, use the old guidelines, but consider the combined sample size when deciding whether or not it is safe to continue with the procedure.

Okay, I'll remind you: if the combined sample size is less than 15, then the samples need to be pretty normal (no obvious skew); for combined sample sizes between 15 and 40, neither sample can have *strong* skew; and for combined sample sizes over 40, pretty much anything goes.

Examples

[1.] The wind speed in a certain city was measured over several days in May. The results (miles per hour) are shown below.

Table 1 - Wind Speeds in May

7.4	8.0	12.6	11.5	14.3	14.9	8.6	13.8	20.1	8.6	6.9
12.0	18.4	11.5	9.7	9.7	16.6	9.7	12.0	16.6	14.9	
12.0	14.9	5.7	7.4	9.7	9.2	10.9	13.2	11.5	8.0	

The wind speed in the same city was then measured over several days in June. The results are shown below.

Table 2 - Wind Speeds in June

8.6	9.7	16.1	9.2	8.6	14.3	9.7	6.9	13.8	11.5
8.0	13.8	11.5	14.9	20.7	9.2	11.5	10.3	6.3	1.7
8.0	8.0	10.3	11.5	14.9	8.0	4.6	6.3	10.9	9.2

Estimate the difference in mean wind speed between May and June with 95% confidence.

This calls for a two sample t interval for difference in means. This interval requires that each sample was obtained randomly, and that each population of wind speeds is normally distributed.

I'll have to assume that the samples are random samples from their populations. I don't know anything about the population distributions...but with a combined sample size of 61, we should be able to continue anyway. I'll go ahead and graph the data just to make sure there isn't anything weird going on...

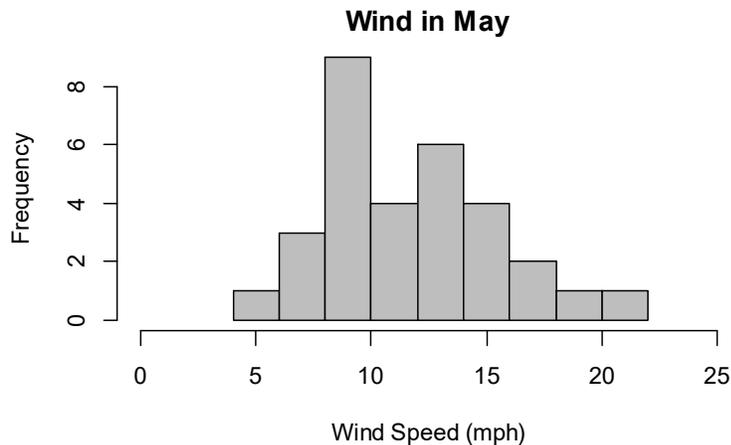


Figure 1 - Histogram of May Wind Speeds

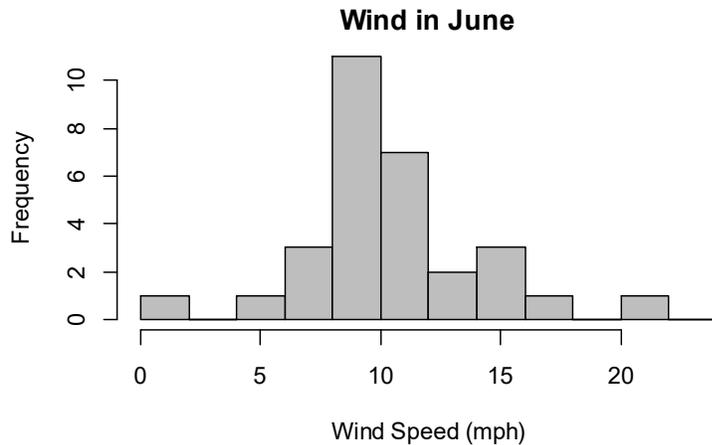


Figure 2 - Histogram of June Wind Speeds

Everything looks fine. I'll continue.

95% confidence and 58.435 degrees of freedom, $t^* = 2.0014$. The interval is

$$11.623 - 10.267 \pm 2.0014 \sqrt{\frac{3.531^2}{31} + \frac{3.769^2}{30}} = (-0.5172, 3.229).$$

I am 95% confident that the actual difference in mean wind speeds between May and June are between -0.5172 mph (June is windier) and 3.229 mph (May is windier).

[2.] The strength of an earthquake is measured by the amount of movement (acceleration) caused by the quake. Here are the measurements (*measured in g's—multiples of Earth's surface gravity*) for a certain earthquake in California:

Table 3 - Accelerations from Earthquake #1

0.264	0.263	0.230	0.147	0.286	0.157	0.237	0.133
0.055	0.097	0.129	0.192	0.147	0.154	0.060	0.057

Here are the measurements for another quake in California:

Table 4 - Accelerations from Earthquake #2

0.123	0.133	0.073	0.097	0.096	0.230	0.082	0.110	0.022
0.110	0.094	0.040	0.050	0.022	0.070	0.080	0.033	0.017

Estimate the difference in peak acceleration between these two earthquakes with 90% confidence.

This calls for a two sample t interval for difference in means. This interval requires that each sample was obtained randomly, and that each population of acceleration measurements is distributed normally. I'll have to assume that these are random samples of measurements. I don't know anything about the distributions of the populations...with a combined sample size of 34, I can continue as long as neither sample shows strong skew. Here are the graphs of the samples:

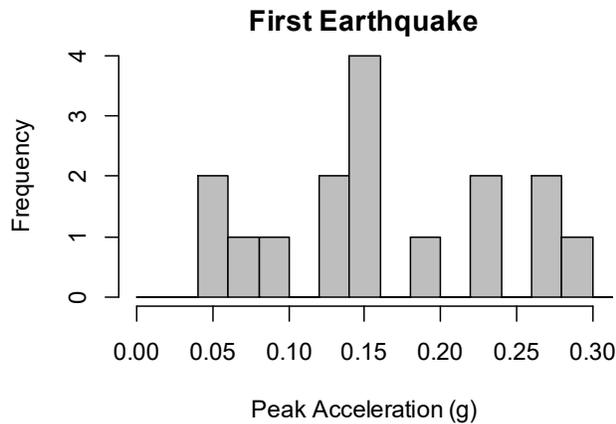


Figure 3 - Histogram for First Earthquake

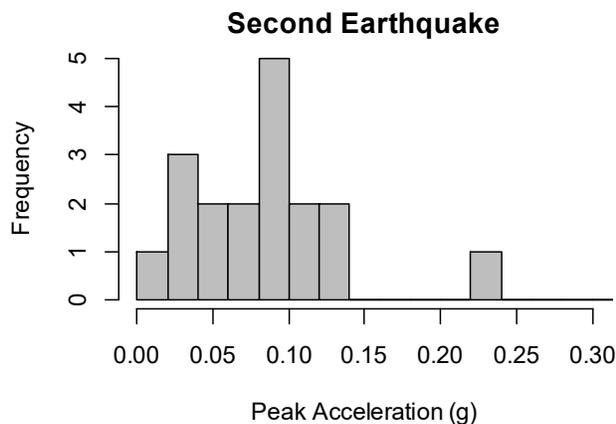


Figure 4 - Histogram for Second Earthquake

Neither appears to be too skew—I think it is safe to continue.

With 25.941 degrees of freedom, $t^* = 1.7058$. The interval is

$$0.163 - 0.0823 \pm 1.7058 \sqrt{\frac{0.076^2}{16} + \frac{0.052^2}{18}} = (0.0422, 0.1191).$$

I am 90% confident that the actual difference in mean peak acceleration for these two earthquakes is between 0.0422g and 0.1191g.

Tests of Significance

Hypotheses

Since the null hypothesis is a statement of no difference, we'll always test that our new parameter ($\mu_1 - \mu_2$) equals zero.

Equation 6 - Hypotheses for a Two Sample Test for Means

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

Requirements

Happily, the requirements for the two sample test are the same as the requirements for the two sample interval!

Some of you may be wondering about pooling...the answer is **no**. In the case of proportions, there is a good argument for pooling. That argument doesn't really work in the case of means. Thus—**do not pool** when using the two sample procedures for means! There are a lot of textbooks (and people!) who still pool—but there is really very little reason to pool now that we have calculators that will do everything for us.

The Test Statistic and Everything Else

The test statistic follows the same pattern that every other one has followed!

Equation 7 - Test Statistic for a Two Sample Test for Means

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

We'll continue to use the Satterthwaite degrees of freedom. The p -value is computed in the same way that it has always been computed, and the conclusion to the test is also the same.

Joy!

Examples

[3.] Bigger chickens mean bigger profits for farmers—thus, a diet that increases the mass of a chicken is desirable. The effect of two different diets was measured for two groups of newly hatched chicks. The following are the body masses (grams) of two groups of chicks three weeks after hatching.

Table 5 - Chick Masses for Example 3

diet 1	331	167	175	74	265	251	192	233	309	150
diet 2	256	305	147	341	373	220	178	290	272	321

Do the data provide evidence of a difference in the masses of chicks that are fed one of these two diets?

I'll let μ_1 represent the population mean mass of chicks fed diet number 1, and μ_2 represent the population mean mass of chicks fed diet number two.

$H_0 : \mu_1 - \mu_2 = 0$ (there is no difference in the mean mass for chicks fed these diets)

$H_a : \mu_1 - \mu_2 \neq 0$ (there is a difference in the mean mass for chicks fed these diets)

This calls for a two sample t test for difference of means. This requires that the feed type was randomly assigned to a feeding group, and that both populations of chick masses are distributed normally. I'll have to assume that the feed type was randomly assigned. I don't know anything about the population distributions—with a combined sample size of 20, I can continue if the samples show no strong skew. Here are the graphs:

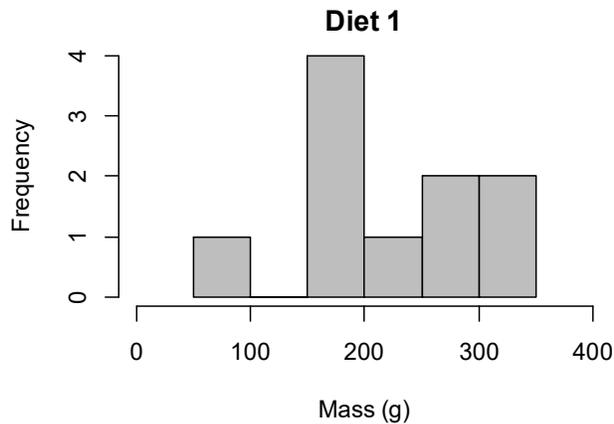


Figure 5 - Chick Masses with Diet #1

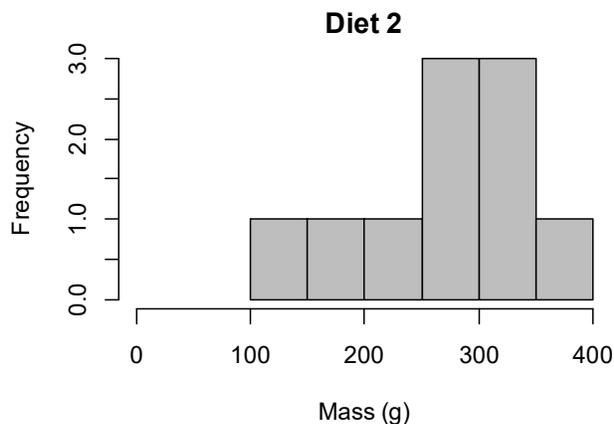


Figure 6 - Chick Masses with Diet #2

There is some skew, but I don't think that it is enough to make me stop. I'll use $\alpha = 0.05$.

$$t = \frac{214.7 - 270.3}{\sqrt{\frac{78.138^2}{10} + \frac{71.623^2}{10}}} = -1.6588; \text{ with } 17.865 \text{ degrees of freedom,}$$

$$2P(t < -1.6588) = 0.1146.$$

If there is no difference in the actual mean mass for chicks fed these two diets, then I can expect to find a difference in sample means of at least 55.6g in about 11.46% of samples.

Since $p > \alpha$, this occurs often enough to attribute to chance at the 5% level; this is not significant, and I fail to reject the null hypothesis.

There is no evidence of a difference in the actual mean mass of chicks fed these two diets (neither one appears to make bigger chickens).

[4.] The effect of Vitamin C on tooth growth was measured by giving guinea pigs either orange juice (which contains Vitamin C) or ascorbic acid (pure Vitamin C). Here are the tooth lengths (mm) of the guinea pigs that were given Orange Juice:

Table 6 - Tooth Length after using Orange Juice

15.2	21.5	17.6	9.7	14.5	10.0	8.2	9.4	16.5	9.7
23.3	23.6	26.4	20.0	25.2	25.8	21.2	14.5	27.3	25.5
22.4	24.5	24.8	30.9	26.4	27.3	29.4	23.0	26.4	19.7

Here are the tooth lengths of the guinea pigs that were given ascorbic acid:

Table 7 - Tooth Length after using Ascorbic Acid

4.2	11.5	7.3	5.8	6.4	10.0	11.2	11.2	5.2	7.0
16.5	15.2	17.3	22.5	17.3	13.6	14.5	18.8	15.5	23.6
18.5	33.9	25.5	26.4	32.5	26.7	21.5	23.3	29.5	16.5

Is there evidence that ascorbic acid produces greater tooth growth than orange juice?

I'll let μ_O represent the population mean tooth growth when using orange juice, and μ_A represent the population mean tooth growth when using ascorbic acid.

$H_0 : \mu_A - \mu_O = 0$ (orange juice and ascorbic acid result in equal mean tooth growth)

$H_a : \mu_A - \mu_O > 0$ (ascorbic acid results in a larger mean tooth growth)

This calls for a two sample t test for difference of means. This requires that the treatments were assigned randomly, and that both populations of tooth growth are distributed normally. I'll assume that the treatments were assigned randomly. I don't know anything about the populations...with a combined sample size of 60, I should be able to continue almost without regard to how the samples look. Naturally, I'll look anyway.



Figure 7 - Histogram for Orange Juice Group

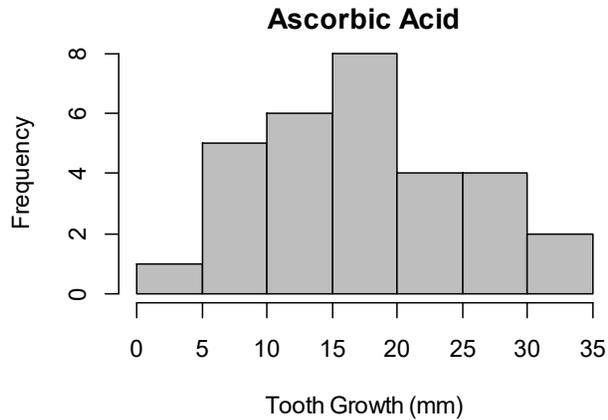


Figure 8 - Histogram for Ascorbic Acid Group

No problems here. I'll continue.

I'll let $\alpha = 0.05$.

$$t = \frac{16.963 - 20.663}{\sqrt{\frac{8.266^2}{30} + \frac{6.606^2}{30}}} = -1.9153; \text{ with } 55.309 \text{ degrees of freedom, } P(t > -1.9153) = 0.9697.$$

If there is no difference in the population mean tooth growth when using orange juice or ascorbic acid, then I can expect to find samples where the mean tooth growth for ascorbic acid is greater than that from when using orange juice, or up to 3.7mm shorter, in about 96.97% of samples.

Since $p > \alpha$, this occurs often enough to attribute to chance at the 5% level. It is not significant; I fail to reject the null hypothesis.

There is no evidence that the mean tooth growth when using ascorbic acid is greater than that when using orange juice.

(...which should have been obvious when you looked at the data!)