The purpose of this little document is to allow you to understand where the test statistic for the ANOVA originates. Here are the data from the example in your project:

group		test scores							
1) constant sound	7	4	6	8	6	6	2	9	
2) random sound	5	5	3	4	4	7	2	2	
3) no sound	2	4	7	1	2	1	5	5	

Hypotheses are as follows:

$$H_0: \mu_{\text{constant sound}} = \mu_{\text{random sound}} = \mu_{\text{no sound}}$$

H_1: at least one of the means is significantly different from the others

By the way, your hypotheses for any ANOVA will **always** look like this: the null is each average equal, and the research is "at least one of the means is significantly different from the others".

We need to find two statistics: the variance **between** samples, which is called the "sum of squares of the treatment", or SS(treatment), and the variance **within** samples, which is called the "sum of squares error", or SS(error). SS(treatment) basically lets us know if there is a large deviation from sample to sample, as would be seen if our treatment was actually doing something.   SS(error) is really a control; it measures whether or not our treatment is actually doing something, by combining variances.

To construct our test statistic, we need to find the individual values of $SS(\text{treatment})$ and $SS(\text{error})$. Start by finding the mean of each of the samples:

$$\bar{x}_{\text{constant sound}} = 6 \qquad \bar{x}_{\text{random sound}} = 4 \qquad \bar{x}_{\text{no sound}} = 3.375$$

Next, take all sample values together and find the average of ALL of them (your text calls this $\bar{\bar{x}}$, or "x double bar"...other books call it the "grand mean"):

$$\bar{\bar{x}} = \frac{\text{sum of all data, regardless of group}}{\text{total data points across groups}} = \frac{107}{24}$$

To create SS (treatment), we evaluate the following:

$$SS(\text{treatment}) = \sum n\left(\bar{x} - \bar{\bar{x}}\right)^2 = 8\left(6 - \frac{107}{24}\right)^2 + 8\left(4 - \frac{107}{24}\right)^2 + 8\left(3.375 - \frac{107}{24}\right)^2 = \frac{361}{12}$$

(please note that **n** does not have to be the same for each sample...in our case, it just happens to be)

To find SS(error), first find the variances of each sample (remember, the sample variance is the *square* of the sample standard deviation):

$$s^2_{\text{constant sound}} = \frac{34}{7} \qquad s^2_{\text{random sound}} = \frac{20}{7} \qquad s^2_{\text{no sound}} = \frac{271}{56}$$

Then calculate SS(error) as follows:

$$SS(\text{error}) = \sum (n-1)s^2 = (8-1)\frac{34}{7} + (8-1)\frac{20}{7} + (8-1)\frac{271}{56}$$

$$= \frac{703}{8}$$

Please note that, as freaky as these computations look, they're simply based on the definition of variance, which, as you'll recall from way back in the day, is

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Now, in order to finally create our test statistic, we turn the each sum of squares into a mean (or average) square, as follows:

Let N = total number of values in all samples combined = 24

Let k = number of different samples = 3

$$MS(\text{treatment}) = \frac{SS(\text{treatment})}{k-1} = \frac{\frac{361}{12}}{2} = \frac{361}{24}$$

$$MS(\text{error}) = \frac{SS(\text{error})}{N-k} = \frac{\frac{703}{8}}{21} = \frac{703}{168}$$

And, at long last, our test statistic (called F, for the statistician R. A. Fisher):

$$F = \frac{MS(\text{Treatment})}{MS(\text{Error})} = \frac{\dfrac{361}{24}}{\dfrac{703}{168}} = \frac{133}{37}$$

What you just created, with this F statistic, is a ratio of two variances. The rationale behind it is fairly straightforward: If the averages among the different samples are the same, then the numerator and denominator should have the same value (within random allowable variation). In other words, if there does not seem to be any effect of our experimental procedure (as H_0 claims there isn't), the treatment and error terms should be approximately equal, and the F – statistic is approximately equal to 1.

However, if one (or more) of our treatments **does** appear to have an effect, the numerator's ratio (which measures variation among samples) should grow larger out of proportion to the denominator (which measures common variation among all of our tested populations). Then, the value of the F – statistic should be larger than 1.

As you now notice, our F – statistic has a value of about 3.59. It's larger than 1...but how much larger than 1 does it need to be to mark a **significant** difference in the means of the populations? Well, we need a reference sheet, and you've got it in your textbook. Since we're testing variances, we can't use the distributions we're accustomed to (z, t, or χ^2)...we need to use the F – distribution.

In the appendix of your textbook, locate the table(s) for the F distribution. Notice that there are a number of them. This is because the value of the F critical value depends on both the number of degrees of freedom in the numerator <u>and</u> the number of degrees of freedom in the denominator[1]. To find these, use the following:

<u>**Numerator DF $= k - 1$**</u>

<u>**Denominator DF $= N - k$**</u>

(notice that these are the same values we had to use to turn SS variances into MS variances[2])

So, for our little example, our numerator has 2 degrees of freedom, and our denominator has 21. Let's test at the 5% level of significance...like the χ^2 distribution, this test is always right – tailed. Using the $\alpha = 0.05$ chart with 2 DF in the numerator, and DF in the denominator, I arrive at a test statistic of $F = 3.4668$:

---

[1] This is a total pain in the rear. Technology makes it easier. More on that in your project, though...
[2] Thus, the total DF is : $(N - k) + (k - 1) = N - 1$

Fail to reject Ho

Reject H0

↑   ↑
critical value test statistic
F = 3.4668   F = 3.59
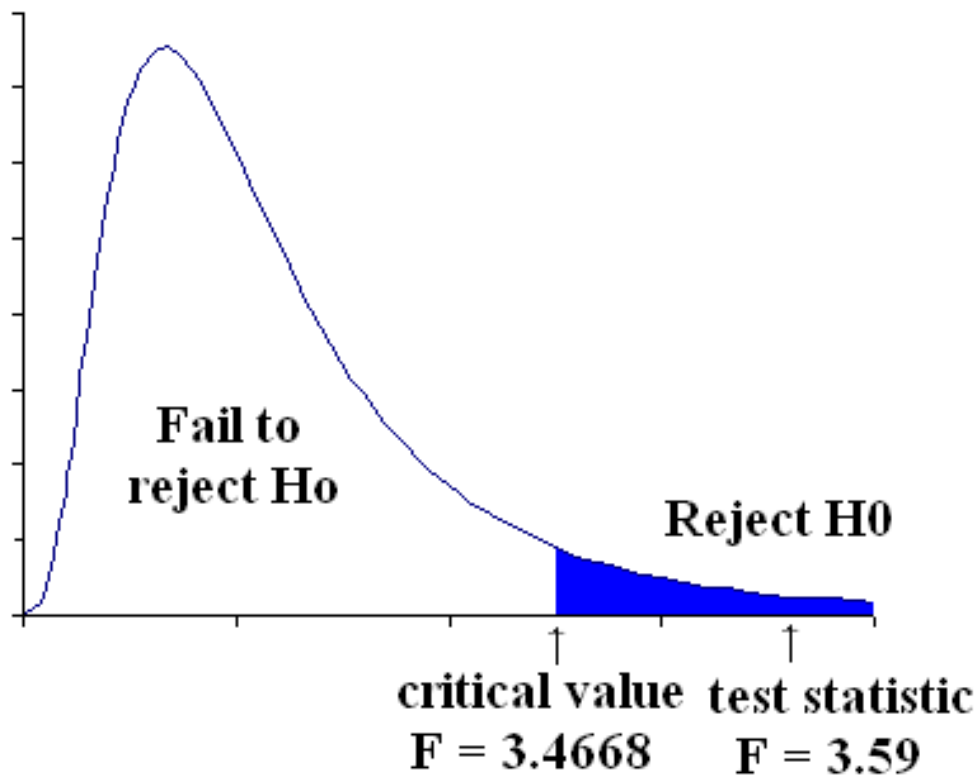
   And so, at the end of all of this, we can see that, since our test statistic falls in the rejection region, we reject H_0 and suggest that the data indicates that at least one of the population means is significantly different that the others[3].

---

[3] Of course, ANOVA doesn't tell us *which* mean is different, nor does it guarantee that our treatment was what caused the difference. For that, you need other statistical methods outside the scope and time allotted for this course. In other words, if you reject H_0, you open up a HUGE can of worms that has to be dealt with...of course, that's where the fun begins! Alas, for us, though, it ends there.