USEFUL INFO ABOUT ANOVA

R^2 in ANOVA, the proportion-of-variance-accounted-for type of effect size measure, has a very simple formula -- it's
$R^2 = SS_A / SS_{TOT}$
or with slightly different notation, SS_{BET} / SS_{TOT}, or $SS_A / (SS_A + SS_{S/A})$ since the TOTAL SS is the sum of the BET and W/IN SS (see below for example). It's also called eta-squared (η^2). Similar but better effect size measures include ω^2 (omega squared) and Cohen's f.

INDEPENDENT SAMPLES T-TEST REMINDER
The original Student's t-test says, check Levene's test to be sure it's non-significant, i.e., you've met the homogeneity of variance assumption; then since the variances are the same in the population, calculate the "pooled" variance to get a weighted average of your two sample variances, and use that s^2_p in place of s^2_1 and s^2_2 in the t denominator. If Levene's test IS significant and the variances are different in the population, i.e., you've violated the assumption, do Welch's t-test instead: Welch's t says DON'T pool the variances, just use s^2_1 and s^2_2 in the t denominator and use Welch's unintuitive formula to lower the df for the test (which SPSS does for you), making it appropriately harder to find significance, and all will be well. And the general best recommendation is actually to forget Levene and any other tests of the assumption, and just ALWAYS do Welch's t and that'll be best, because it'll take care of violations of homogeneity to the extent there are any, while not inflicting a penalty when the assumption is met.

SOME HELPFUL THINGS TO NOTE ABOUT ANOVA
The Logic of ANOVA handout is a starting point for understanding ANOVA Here I will expand upon it a little bit in ways that should make things more clear.

First, for clarity, these terms refer to the same Sources of Variance:
Numerator: Between Groups, or Factor A, or "treatment effect term"
Denominator: Within Groups, or Subjects Within Levels of A, or "error term"
So for instance you can describe the two df numbers equivalently as:
df_{num} and df_{den}
df_{BET} and $df_{W/IN}$
df_A and $df_{S/A}$
$df_{effect\ or\ trtmt}$ and df_{error}
It's better to use A and S/A instead of BETWEEN and WITHIN because soon enough there will be more than one factor (e.g., not just classical vs. popular music, but also vocal vs. instrumental) and we'll label the new one B, and another C (familiar vs. unfamiliar), and so on. We haven't used the names "effect term" and "error term" yet, and we won't call them that on this exam.

IT IS IMPORTANT that you understand and remember the **meaning of F in terms of what its numerator and denominator represent -- experimental error and individual**

differences in both, treatment effect in the numerator only. That is something conceptual that you should know about how ANOVA works.

The Logic of ANOVA summary starts out saying that **F is a ratio of two independent estimates of the SAME underlying population variance**. THAT'S IMPORTANT. If the null hypothesis IS true and there is zero treatment effect, the numerator variance (that we're now calling the numerator "Mean Square") ends up representing the SAME two things that the denominator represents: just experimental error and individual differences, with zero treatment effect. They represent the SAME things even though the numerator and denominator are calculated independently of one another, using the different formulas for MS_A and $MS_{S/A}$, respectively.

With the numerator and denominator being influenced by the SAME two things, the most likely value of F is 1. But if there IS a treatment effect, it will be a third thing in the numerator, not in the denominator, making F greater than 1. IN THAT CASE, our obtained F value is no longer really part of the F distribution per se, because **it's not a "true F ratio" -- it's a ratio of two estimates of two DIFFERENT variances**, which differ because one includes a treatment effect while the other does not.

We say we'd reject the null hypothesis when the F ratio is enough larger than 1 that we think a treatment effect really must be present in the numerator to make it so, and therefore our F isn't a "true F" in the sense just described. And by "enough larger", we mean large enough that the value of F we obtained (or one more extreme / larger than that) would occur purely by chance no more than 5% of the time under the null hypothesis of no treatment effect -- i.e., it's in the extreme 5% of the tail of the F distribution given the df for the numerator and denominator.

AN EXAMPLE OF ANOVA FOR TWO LEVELS OF FACTOR A
If there are only two levels, you could do a t-test instead of ANOVA. In that case, $F = t^2$, keeping in mind that the F numerator df is always 1 when there are just two groups (i.e., no. of groups - 1 = 1) and the denominator df are the same as the t df. Maybe it's more precise to write, $F(df_{num}, df_{den}) = F(1, df_{den}) = [t(df_{den})]^2$ -- for two groups ONLY.

Let's use the formulas from the ANOVA summary on these data:
GP 1: 4, 6, 8; M = 6; n = 3; df = 2; SS = 8; s^2 = 4
GP 2: 5, 9, 13; M = 9; n = 3; df = 2; SS = 32; s^2 = 16
TOTAL SS = $(4 - 7.5)^2$ + ... + $(13 - 7.5)^2$ = 53.5 -- the sum of squared deviations of all six scores around their overall mean of 7.5.

For comparison: The independent groups t value comes out to 1.162, which is not significant for t(4) since the .05 cutoff is 2.776. The r^2 value is $t^2 / (t^2 + df) = 1.162^2 / (1.162^2 + 4) = 0.252$, the proportion of DV variance accounted for by the IV. That 0.252 is literally the square of the r you'd get if you correlated the IV "GROUP MEMBERSHIP"

1,1,1,0,0,0, with the DV "SCORE" 4,6,8,5,9,13, which would be r = .502. Cohen's d would be M1-M2 / s_p = (6-9) / $\sqrt{10}$ = 0.949 (using the absolute value instead of leaving it negative).

Doing the same analysis in ANOVA, remember n is the number of observations in ONE group (in this case n = 3 for both groups; we'll address unequal n's later), and a is the number of levels of Factor A (in this case a = 2 -- classical vs. popular music).

$SS_{S/A}$ = SS1 + SS2 = 40
$df_{S/A}$ = df1 + df2 = 2 + 2 = 4
$MS_{S/A}$ = 40/4 = 10 -- which is exactly the same as doing the pooled variance in the t-test, if we'd done the pooling that's required by Student's t-test but not by Welch's

SS_{means} = $(6 - 7.5)^2$ + $(9 - 7.5)^2$ = 2.25 + 2.25 = 4.5 (where 7.5 is the mean of the two group means, and 6 and 9 are the two group means whose deviations from their overall mean are being squared and then summed -- hence the sum of squares of the means or SS_{means})

SS_A = n * SS_{means} = 3 * 4.5 = 13.5 -- since all groups have the same n; hopefully the only arbitrary-looking formula, for now
df_A = a - 1 = 2 - 1 = 1
MS_A = 13.5 / 1 = 13.5

$F_{(1,4)}$ = 13.5 / 10 = 1.35
But $F_{(1.4).05}$ = 7.71, so 1.35 misses being significant by quite a bit.

Notice that the square roots of those numbers are the same as the t results above:
t(4) = 1.162
$t(4)_{.05}$ = 2.776

R^2 = SS_A / SS_{TOT} = 13.5 / 53.5 = 0.252 -- exactly what we got for r^2 in the t-test above.

In a printed F table, there are clear headings for which numerator df column you're looking for and which denominator df row you're looking for, and at the intersection of the column and row you'll usually find the .05 and .01 cutoff values (and sometimes others), with the .01 cutoff value always being larger than the .05 cutoff value since a larger F is needed to be significant "at the .01 level".

How would this look as a THREE level ANOVA, with a third group that studied while listening to jazz? Here we go again:

GP 1: 4, 6, 8; M = 6; n = 3; df = 2; SS = 8; s^2 = 4
GP 2: 5, 9, 13; M = 9; n = 3; df = 2; SS = 32; s^2 = 16
GP 3: 2, 3, 4; M = 3; n = 3; df = 2; SS = 2; s^2 = 1

TOTAL SS = $(4 - 6)^2 + (6 - 6)^2 + (8 - 6)^2 + (5 - 6)^2 + (9 - 6)^2 + (13 - 6)^2 + (2 - 6)^2 + (3 - 6)^2 + (4 - 6)^2$ = 96 -- the sum of squared deviations of all NINE scores around their overall mean of 6.

$SS_{S/A}$ = SS1 + SS2 + SS3 = 8 + 32 + 2 = 42
$df_{S/A}$ = df1 + df2 + df3 = 2 + 2 +2 = 6
$MS_{S/A}$ = 42/6 = 7 -- which is exactly the same as doing the pooled variance in the t-test, if three groups were allowed in a t-test (which they're not)

SS_{means} = $(6 - 6)^2 + (9 - 6)^2 + (3 - 6)^2$ = 0 + 9 + 9 = 18 (where 6 is the mean of the THREE group means, and 6 and 9 and 3 are the three group means whose deviations from their overall mean are being squared and then summed -- hence the sum of squares of the means or SS_{means})

SS_A = n * SS_{means} = 3 * 18 = 54
df_A = a - 1 = 3 - 1 = 2
MS_A = 54 / 2 = 27

$F_{(2,6)}$ = 27 / 7 = 3.857
But $F_{(2,6).05}$ = 5.14, so 3.857 fails to reach significance.

R^2 = SS_A / SS_{TOT} = 54 / 96 = 0.5625

In the form of an ANOVA table, the three level example would be set up like this:

Source of Variance	SS	df	MS	F
A	54	2	27	3.857
S/A	42	6	7	
TOTAL	96	8		

Notice that the **A and S/A add up to the TOTAL, for both SS and df** but not for MS. And notice that by definition each MS is its SS divided by its df, and F is MS_A / $MS_{S/A}$. So there are a lot of blanks you could fill in given some of the information. For instance, told only that there were three groups of 3 subjects, and given only the two numbers below, you could fill in all the rest: the $SS_{S/A}$ would have to be 42, the df_A would be 3-1=2, the $df_{S/A}$ would be 2+2+2, the MS would be SS/df for both, and the F would be the ratio of the MS_A to the $MS_{S/A}$, just as above.

Source of Variance	SS	df	MS	F
A	54			
S/A				
TOTAL	96			