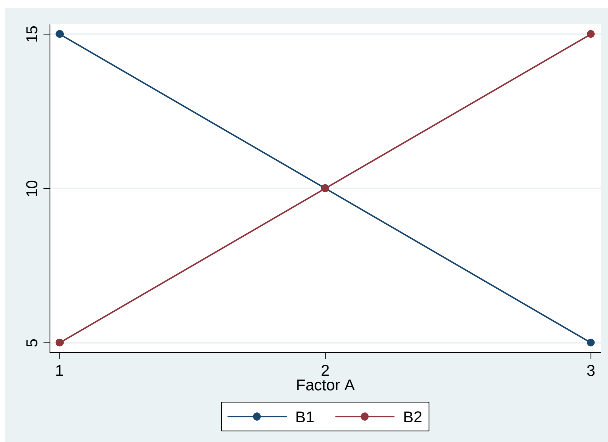
STATISTICS 110/201 PRACTICE FINAL EXAM on ANOVA

1. Consider a two-factor experiment in which one factor is “Restaurants” and five restaurants are used in the study. Give a set of circumstances under which the restaurant factor would be considered fixed, and a set of circumstances under which it would be considered random.

*The factor would be considered fixed if those 5 restaurants were of interest. It would be considered random if the 5 restaurants were randomly chosen to represent a much larger collection of restaurants. For example, a company might own hundreds of restaurants, but want to try a new menu or a new method of greeting customers, etc., at a small set of them to see what works.*

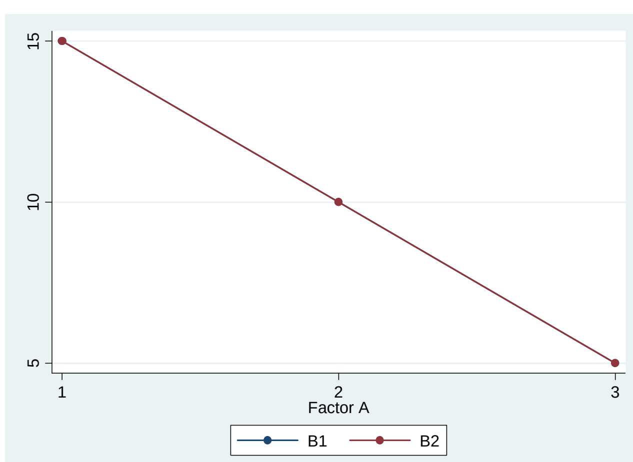
2. Sketch a picture of possible cell means for the following scenarios:
a. Factor A has 3 levels, Factor B has 2 levels. There is an AB interaction, but no A or B main effects.

*Here’s one possibility:*




b. Factor A has 3 levels, Factor B has 2 levels. There is an effect for Factor A, but no interaction and no Factor B effect.

*Here is one possibility (the lines are right on top of each other for B1 and B2):*



3. A study was done to see if meditation would reduce blood pressure in patients with high blood pressure. There were 100 people available for the study. Half of the patients were randomly chosen to learn meditation and told to practice it for half an hour a day. The other half was told not to alter their regular daily routine. Blood pressure measurements were taken at the beginning of the study, after 5 weeks, and after 10 weeks.

a. Specify the factors in this study, and for each one, state whether it is fixed or random, and the number of levels.

*Factor A = Meditation or not, 2 levels, fixed*
*Factor B = Time of measurement, 3 levels, fixed*
*Factor C = Patient, 50 levels, random*

*(You could specify them in a different order.)*

b. Specify whether the factors are crossed or nested with the other factors.

*Patient is nested under meditation group, but crossed with time period. Meditation group and time period are crossed with each other.*

c. Why was it important to include a group that was told to continue their regular routine, rather than just have everyone learn meditation?

*There may be conditions in the environment that cause more (or less) stress in everyone during the period of the study. For example, if it was during a holiday period, or when there was particularly bad news in the world, or the weather was particularly good, everyone would be affected. So it is important to have a "control group" to compare to the meditation group.*

4. Comment briefly on the following statements:
a. In one factor ANOVA where the factor is fixed, a highly significant F* ($p < .001$) indicates that the k population means, $\mu_1$, , $\mu_k$ are all different.

*The statement is not true. The test only tells us that at least one of the means differs from the others.*

b. In one factor ANOVA, we need to use multiple comparisons (like the Tukey procedure) because it is impossible to compare k means all at once.

*Not quite right. We use multiple comparisons to control the overall chance of making an erroneous conclusion, over all of the tests or confidence intervals examined.*

5. A student analyzed data for a one-way analysis of variance situation for which there were 3 levels of the factor, and 21 people measured at each level. Unfortunately, after running the analysis, the student lost the computer output. She said "All I remember is that one of the mean squares was 100 and the other one was 500, but I can't remember which was which. Oh, and I remember that the p-value for the test was about .01."

a. Based on this information, can you construct the analysis of variance table? (I've provided headings to remind you of the table structure.) If so, fill it in. If not, explain why not. If you think you can partially fill it in, do that.

| Source | SS | df | MS | F* | p-value |
|---|---|---|---|---|---|
| *Model* | *1000* | *3 – 1 = 2* | *500* | *5* | *.01* |
| *Error* | *6000* | *3(21 – 1) = 60* | *100* | | |
| *Total* | *7000* | *62* | | | |

*Note that the larger MS must be for the model and not for error, because an F* less than one would not lead to a p-value of .01 in any case.*

b. In the statement of the question, it wasn't specified whether the factor was fixed or random. Write the null and alternative hypotheses being tested in each of the two cases. Make sure you define any symbols you use.

*If the factor is fixed, the null hypothesis can be written as* $H_0$: $\mu_1 = \mu_2 = \mu_3$ *where* = $\mu_i$ *is the population mean for the response variable for factor level i.*

*If the factor is random, the null hypothesis can be written as* $H_0$: $\sigma_\mu^2 = 0$*, i.e. the variance of the population of possible means is 0. This is equivalent to saying that all of the possible factor levels have the same mean for the response variable.*