Homework Assignment for Lesson 4

Please submit ONE document to the Dropbox for Homework 4. This can be a Word document (preferred) or a pdf file. **NOTE: For Question 1 and 2, which require use of statistical software, at least one analysis must be done in SAS. Both software programs will perform the analysis, and so the choice is up to you which one you apply to which question.**

**1)** Consider the following data, the result of a 2-factor factorial experiment with 5 replications. Treatment combinations were assigned at random to the 20 experimental units.

Experimental units.

A B resp

1 1 12.9

1 1 11.3

1 1 11.7

1 1 12.1

1 1 12.3

1 2 13.7

1 2 12.8

1 2 13.6

1 2 13.1

1 2 13.5

2 1 14.2

2 1 14.5

2 1 13.9

2 1 13.6

2 1 14.4

2 2 13.5

2 2 13.1

2 2 13.3

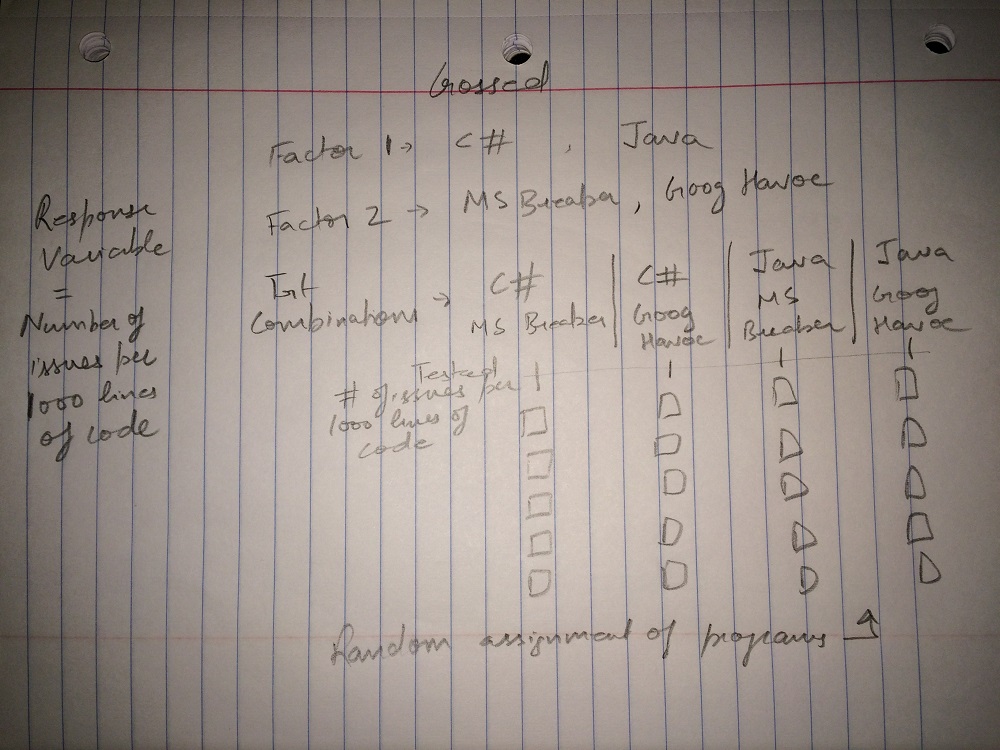
2 2 13.1

2 2 13.4

Make up and describe an experiment for which these data were obtained. This can be an experiment reflecting your area of study, or be a hobby, sports, anything. You can modify the response variable by adding or subtracting a constant amount to each value, so you can adjust the response variable values to be realistic for your experiment. (Hint: Do an exploratory data analysis (EDA) and run the ANOVA before you invent the experiment. It is important to think ahead on what the results will mean).

The IT department of the company wants to purchase a code analyzer for their two primary programming languages (C# and Java). The company decides to evaluate the two product leaders who have products that work with both C# and Java languages. In order to perform the evaluation a study is setup where 20 existing software programs are taken – 10 each of C# and Java (represented by A – 1 denotes C# and 2 denotes Java) and randomly tested with the two shortlisted products (denoted by B – 1 denotes product MS Breaker and 2 denotes Goog Havoc). The response variable is the number of issues discovered by the product per 1000 lines of code.

1. (10 pts) Provide a Study Diagram for your experiment



b) (10 pts) State the null and alternative hypotheses

The null hypothesis:

Main Effect of Factor A (Programming language):

**H0**:μ1.=μ2.   
**HA**: not all μi. are equal

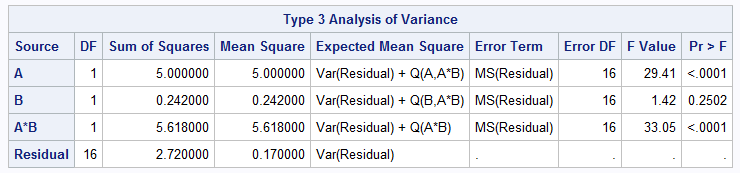
Main Effect of Factor B (Code analysis product):

**H0**:μ.1=μ.2   
**HA**: not all μ.j are equal

A × B Interaction (performance of code analysis product for different programming languages):

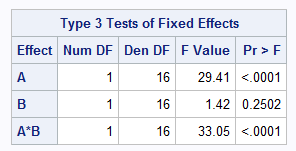
**H0**: there is no interaction (the code analysis works equally effectively for both languages)  
**HA**: an interaction exists (the code analysis effectiveness varies with programming language)

c) (10 pts) Run an ANOVA and include any output you consider important (e.g., diagnostics).



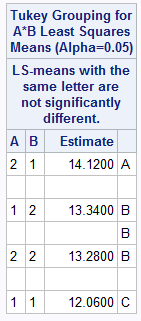
Indicates that the interaction term is significant. Since the interaction term is significant we do NOT interpret the individual factors (programming language or code analysis product).

Effectively we conclude: HA: an interaction exists (the code analysis effectiveness varies with programming language)

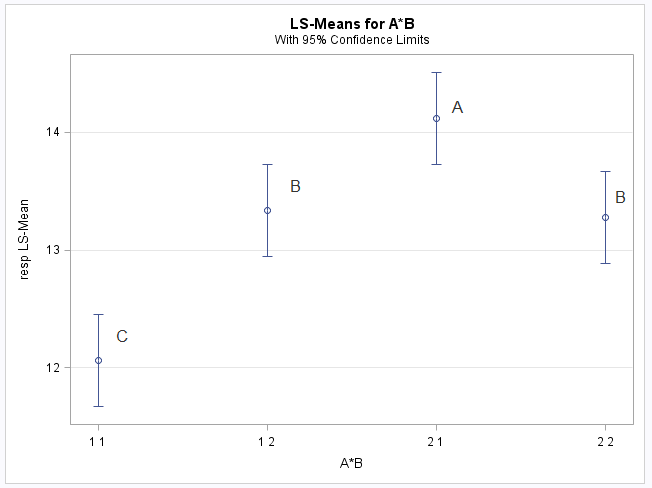


d) (10 pts) Do a mean comparison (where appropriate based on the ANOVA) using the Tukey adjustment**.**

So now that we have looked at the ANOVA output and see the significant interaction term, we know that we want to generate the LSmeans for the interaction effect (i.e., the treatment combinations) for mean comparisons and plotting our figure



e) (10 pts) Produce a final graph (or graphs, if appropriate, depending on the results) to show the outcome of your experiment. The graph(s) can be either bar charts or means plots, but needs to include error bars (either +/- 1 standard error, or 95% CI limits) and Tukey mean comparison results. Include a figure caption that explains the symbols used. Provide a *brief* interpretation of the results.



The interpretation of the result is very interesting. We find that product MS Breaker (j=1) effectiveness varies significantly with the choice of programming language. It performs significantly better for Java (i = 2) as compared to C# (i = 1). On the other hand Goog Havoc (j = 2) works equally effectively for both the languages.

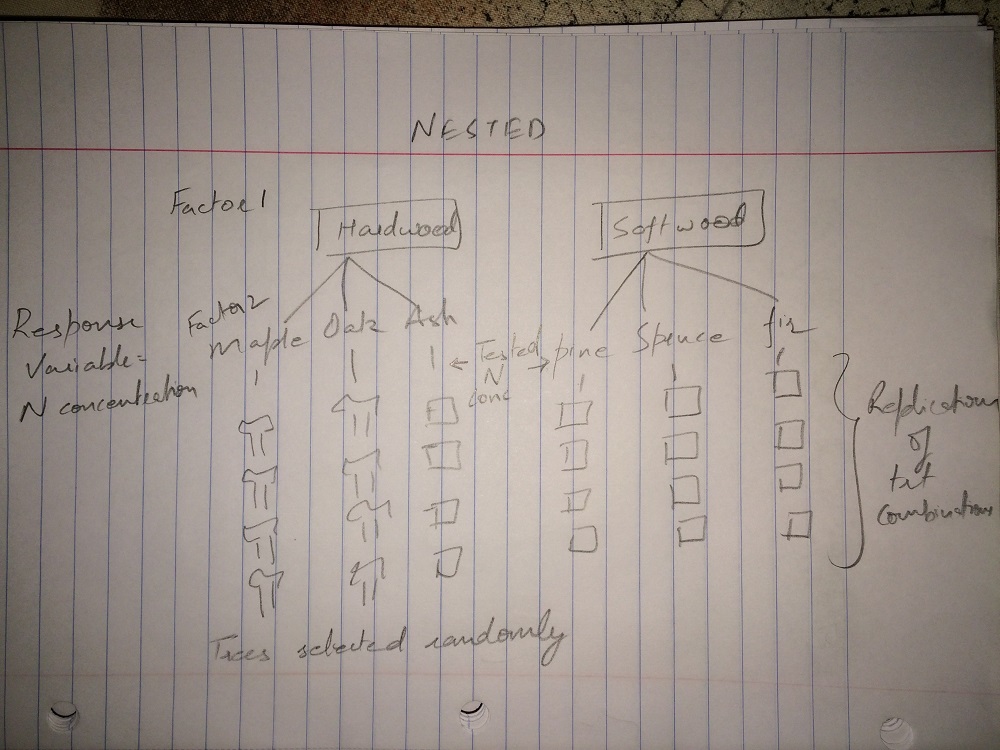
This is also illustrated in the bar chart below:

Figure Caption: Mean error rate per 1000 lines of code for the two programming languages for each of the two products. Bar height indicates the mean and error bars are +/- 1 standard error. Means sharing the same letter do not differ significantly at the 95% confidence level based on the Tukey mean comparison method.

**2)** Consider a hypothetical study in which hardwood chips are believed to differ from softwood chips in nitrogen concentration. They were specifically interested in 6 tree species currently used in biomass fuel production (3 species of hardwoods, 3 species of softwoods). Four samples of each species were tested for N concentration.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Type | Species | Nconc | | softwood | pine | 12 | | softwood | pine | 13 | | softwood | pine | 11 | | softwood | pine | 12 | | softwood | spruce | 15 | | softwood | spruce | 19 | | softwood | spruce | 17 | | softwood | spruce | 17 | | softwood | fir | 10 | | softwood | fir | 12 | | softwood | fir | 11 | | softwood | fir | 17 | | hardwood | maple | 18 | | hardwood | maple | 20 | | hardwood | maple | 21 | | hardwood | maple | 16 | | hardwood | oak | 20 | | hardwood | oak | 14 | | hardwood | oak | 17 | | hardwood | oak | 15 | | hardwood | ash | 19 | | hardwood | ash | 22 | | hardwood | ash | 21 | | hardwood | ash | 21 | |  |  |

a) (10 pts) Provide a study diagram for this experiment.



b) (10pts) Write out the statistical model for the ANOVA (You can use words instead of subscripted Greek symbols.)

For factor type, we have 2 choices:

H0:μsoftwood=μhardwood vs.  HA: Not all equal

or

H0:αsoftwood=αhardwood=0 vs HA: Not all αi=0.

So for the nested factor (Species, nested within Type) we have the Null Hypothesis:

H0= all βj(i)=0H0= all βj(i)=0 vs. HA= not all βj(i)=0

c) (20 pts) Run the ANOVA, extract relevant output to submit, and conduct mean comparisons (Tukey method) where factors are significant.

Factor coding (-1, 0, +1)

Factor Information

Factor Type Levels Values

Type Fixed 2 hardwood, softwood

Species(Type) Fixed 6 ash(hardwood), maple(hardwood), oak(hardwood), fir(softwood),

pine(softwood), spruce(softwood)

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Type 1 140.17 140.167 31.74 0.000

Species(Type) 4 96.83 24.208 5.48 0.005

Error 18 79.50 4.417

Total 23 316.50

**Comparisons for Nconc**

**Tukey Pairwise Comparisons: Response = Nconc, Term = Type**

Grouping Information Using the Tukey Method and 95% Confidence

Type N Mean Grouping

hardwood 12 18.6667 A

softwood 12 13.8333 B

Means that do not share a letter are significantly different.

**Tukey Simultaneous 95% CIs**

**Tukey Pairwise Comparisons: Response = Nconc, Term = Species(Type)**

Grouping Information Using the Tukey Method and 95% Confidence

Species(Type) N Mean Grouping

ash(hardwood) 4 20.75 A

maple(hardwood) 4 18.75 A

spruce(softwood) 4 17.00 A B

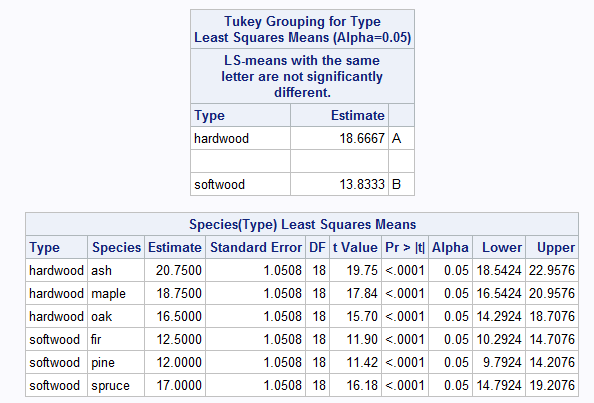
oak(hardwood) 4 16.50 A B C

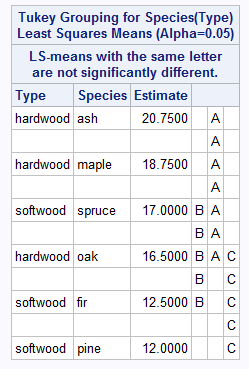
fir(softwood) 4 12.50 B C

pine(softwood) 4 12.00 C

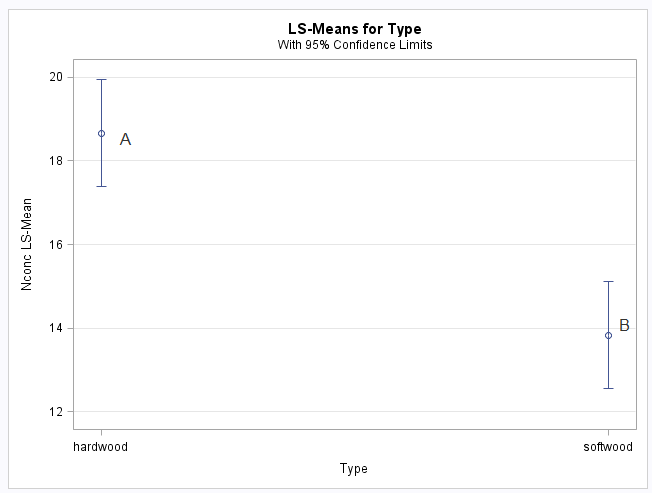
Means that do not share a letter are significantly different.

Outputs from SAS





d) (10 pts) Graph significant results in bar chart or means plot form showing the means, error bars, and results of the mean comparisons. Describe the graph in a Figure Caption.



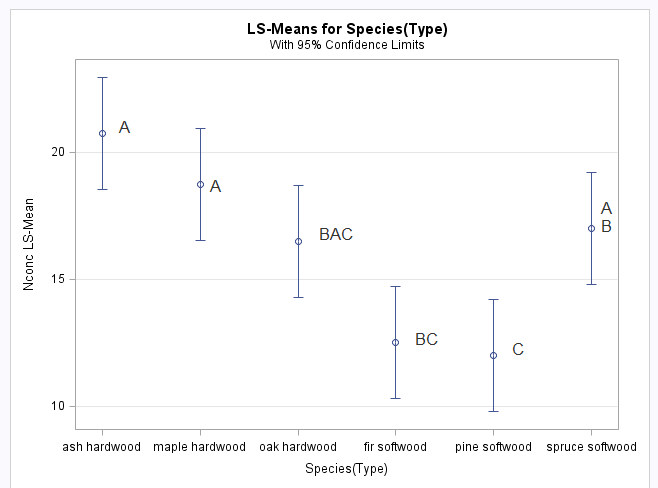


Figure Caption: Mean nitrogen concentration of trees (3 species of hardwoods, 3 species of softwoods). Bar height indicates the mean and error bars are +/- 1 standard error. Means sharing the same letter do not differ significantly at the 95% confidence level based on the Tukey mean comparison method.