**CPaT Stats Week 5: simple as possible ANOVA demo**

Recall:

|  |  |
| --- | --- |
|  |  |

**Example:** Effects of early snowmelt on alpine plant growth (Delphimiums)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Unmanipulated | Control | Treatment |  |  |
|  | 10 | 9 | 12 | N groups |  |
|  | 12 | 11 | 13 |  |  |
|  | 12 | 11 | 15 |  |  |
|  | 13 | 12 | 16 |  |  |
| Ybar-i |  |  |  | Ybar |  |
| n-i |  |  |  | N samples |  |
|  |  |  |  | **SS among** |  |

**What kinds of data do you have? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**What are your null and alternative hypotheses?**

**H0: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Ha: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**We’ll do ANOVA calculations in class, together, NOW.**

**SS-among: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**SS-Total: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**SS-within:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**F-statistic:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**ANOVA Summary of Fit**

|  |  |
| --- | --- |
| **Rsquare** |  |

**Analysis of Variance**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** | **Prob > F** |
| --- | --- | --- | --- | --- | --- |
| **Age Class** |  |  |  |  |  |
| **Error** |  |  |  |  |  |
| **C. Total** |  |  |  |  |  |

what do you think these results mean?

**Q13.** Create a vertical bar graph of mean tree height +/- 1 standard error by height class. Insert your graph below.

The table below is ‘extra’ – just for you to check your own values.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **Diff in Means** | | |
| **Level** | **Mean** | **Std Err** | **N** | **MC** | **PC** | **TC** |
| MC | 34.63 | 0.476 | 250 | 0 | 3.84 | 18.04 |
| PC | 38.47 | 0.532 | 200 | 3.84 | 0 | 14.2 |
| TC | 52.67 | 1.176 | 41 | 18.04 | 14.2 | 0 |

**Q14.** For which (if any) plots (age classes) does it appear that heights differ?

**Heights for TC are quite different, MC and PC are closer, but still different by about 10%.**

**Q15** (optional): in JMP run a Tukey’s HSD test to determine which age classes are significantly different from the others. You will want to report your F(x,x) = X.XX, p = X.XXXX (an F with x,x degrees of freedom = X.XX, p = X.XXX) and Tukey’s lowercase letters on the graph you will make in Q10 below.

**Heights for the three plots (TC, PC, and MC), which represent different age classes, are significant different (F(2,488) = 102.87, p = 0.0001.**

**Q16 (optional)**. Please interpret your results as you would in a scientific paper.

**Trees in three plots (TC, PC, and MC), representing three age classes (100, 155, and 500 years, respectively), were measured for the Thousand Year Chronosequence Study (1kcs) in the Pacific Northwest (1999-2006). A subsequent statistical analysis showed that heights in the three plots were significantly different, with older trees being taller. See Figure Q16 below.**

**Figure Q16: Tree Heights (meters) of 3 Age Classes differ significantly different (F(2,488) = 102.87, p = 0.0001).**

**Part 5: Resampling ANOVA (Not all of you will get to this part this week).**

**You will be using the same data for Part 5 as you did for Part 4, but I have formatted it to make it easier for you to do a resampling ANOVA: PSME-2-3-5-Excel**

To analyze these data using Resampling Stats you will first need to calculate a few things. First, recall how many groups we have. # groups is *a* (below).

Then, recall how the sample size is for each group; this is *n*.

In Excel, please calculate the following:

1. *Number of groups (a)*
2. *sample size for each group/treatment (n)* (report ni’s for i = 1,a if sample sizes are not all the same)
3. *overall mean* (*Ybar)*
4. *average for each group (Ybari where i=1,a)*
5. *Total Sum of Squares (SStotal)*
6. *SS Among (SSamong)*

To calculate Total Sum of Squares, use the following algorithm:

* SStotal = 0
* For i = 1, a
  + For j = 1,n
    - SStotal = SStotal + (Yij – Ybar)2

To calculate SSamong, use the following algorithm: (Ybari is average for I’th group

* SSamong = 0
* For i= 1,a
  + For j = 1,n
    - SSamong = SSamong + (Ybari - Ybar)2

**Q17. Fill in the following table with your calculations:**

What is your (actual) SSamong? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Number of groups (a)* | 3   |  |  |  | | --- | --- | --- | | **MC** | **PC** | **TC** | |
| *sample size for each group (n* | |  |  |  | | --- | --- | --- | | **250** | **200** | **41** | |
| *overall mean* (*Ybar)* | **36.4079646** |
| *average for each group (Ybari where i=1,a)* | |  |  |  | | --- | --- | --- | | **34.6316** | **38.472** | **52.66829268** | |
| *Total Sum of Squares (SStotal)* | **1844.15** |
| *SS Among (SSamong)* | **940,586,929** |
|  |  |

Using Resampling Stats, shuffle all data among treatments in Excel – recalculate SSamong and repeat and score the SSamong 1,000 times. Compare your actual SSamong to the null distribution of random SSamong values.

**Q17**. *What is your Resampling Stats p-value? Is it comparable to your ANOVA p-value from JMP? Speculate as to why or why not.*