10d: Class experiment

Here's the data we collected from our class contest:

Subject	Silence	Music	Interruption	Mean	Std. Dev
1	1	4	2	2.333	1.528
2	6	6	1	4.333	2.887
3	4	4	3	3.667	0.577
4	5	0	3	2.667	2.517
5	4	6	6	5.333	1.155
6	8	6	4	6.000	2.000
7	6	5	11	7.333	3.215
8	4	5	1	3.333	2.082
9	6	8	5	6.333	1.528
10	7	5	3	5.000	2.000
11	6	5	4	5.000	1.000
12	7	6	6	6.333	0.577
13	8	7	2	5.667	3.215
14	9	4	11	8.000	3.606
15	6	5	6	5.667	0.577
16	9	6	6	7.000	1.732
17	6	5	4	5.000	1.000
Sample size	17	17	17	N = 51	
Mean	6.000	5.118	4.588	M = 5.235	
Std. Dev	2.031	1.691	2.938	S = 2.312	

1) What assumptions would we need to make if we conducted an ANOVA on this data (to compare the means of the silence, music, and interruption groups). Are these assumptions satisfied? I had Stata conduct an ANOVA with Bonferroni post hoc tests. What conclusions can we make (if we use α=0.10)? What proportion of variance in recall is due to environment?

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	17.2941176	2	8.64705882	1.66	0.2007
Within groups	249.882353	48	5.20588235		
Total	267.176471	50	5.34352941		

Bartlett's test for equal variances: chi2(2) = 5.0728 Prob>chi2 = 0.079

Comparison of memory by environment
(Bonferroni)

Row Mean- Col Mean	1	2
2	-.882353	
	0.795	
3	-1.41176	-.529412
	0.233	1.000

2) Let's continue to assume we have independent observations. Other than environment, what factors could account for the variation in the number of words recalled? I hypothesize that students in the 8:00 section, being so early in the morning, will have worse recall than students in the 9:25 section. With this hypothesis, I can reformat the data:

Class section	Silence	Music	Interruption	Total
8:00	n = 6 mean = 4.667 std. dev = 2.338	n = 6 mean = 4.333 std. dev = 2.338	n = 6 mean = 3.167 std. dev = 1.722	N = 18 Mean = 4.056 Std. Dev = 2.127
9:25	n = 11 mean = 6.727 std. dev = 1.489	n = 11 mean = 5.545 std. dev = 1.128	n = 11 mean = 5.364 std. dev = 3.233	N = 33 Mean = 5.879 Std. Dev = 2.176
Sample size	N = 17 Mean = 6.000 Std. Dev = 2.031	N = 17 Mean = 5.118 Std. Dev = 1.691	N = 17 Mean = 4.588 Std. Dev = 2.938	N = 51 Mean = 5.235 Std. Dev = 2.312

a) We're going to conduct an AxB ANOVA on this data. First, produce a means plot and predict if we will find a significant interaction.

b) Based on the output produced below, what conclusions can we make from this study (use α=0.10)? What proportion of the variance in recall is due to environment and section? What additional analyses, if any, would you conduct on this data?

```
                    Number of obs =      51     R-squared     =  0.2179
                    Root MSE      = 2.15486     Adj R-squared =  0.1310

              Source |  Partial SS    df       MS           F     Prob > F
---------------------+------------------------------------------------------
         environment |  15.9364231     2  7.96821153        1.72     0.1913
             section |  38.7168746     1  38.7168746        8.34     0.0059
environment#section |  2.21093286     2  1.10546643        0.24     0.7891
                    |
            Residual |  208.954545    45  4.64343434
---------------------+------------------------------------------------------
               Total |  267.176471    50  5.34352941
```

3) Last year, students in a 1:00 and an evening section of MATH 301 completed the same contest. I remembered to record student majors, so I was able ask the following questions:

We did not cover this in class, but we can also run an AxBxC ANOVA to analyze the effect of 3 variables on the ability to recall words. The variables are:

A) Environment (silent vs. music vs. interruptions)
B) Class section (1:00 vs. evening)
C) Student major (IE vs. other)

IE Majors		Environment		
		Silence	Music	Interruption
Class section	1:00	n = 7 Mean = 6.4286 SD = 1.6183	n = 7 Mean = 3.4286 SD = 1.3973	n = 7 Mean = 6.0000 SD = 2.2361
	Evening	n = 3 Mean = 6.6667 SD = 0.5774	n = 3 Mean = 5.6667 SD = 1.5275	n = 3 Mean = 3.6667 SD = 3.0551

Other Majors		Environment		
		Silence	Music	Interruption
Class section	1:00	n = 4 Mean = 5.7500 SD = 3.5000	n = 4 Mean = 4.000 SD = 2.7080	n = 4 Mean = 4.2500 SD = 1.8930
	Evening	n = 3 Mean = 8.3333 SD = 1.5275	n = 3 Mean = 7.6667 SD = 1.1547	n = 3 Mean = 6.3333 SD = 0.5774

4. Sketch a graph for each table and predict whether we will find a significant interaction or not.

I then had Stata run an AxBxC ANOVA on this data:

```
                           Number of obs =        51     R-squared     =  0.4041
                           Root MSE      =   2.01869     Adj R-squared =  0.2360

                  Source |  Partial SS    df       MS           F     Prob > F
-----------------------+----------------------------------------------------
                Environ. |   28.138666     2   14.069333        3.45     0.0416
                   class |   22.6031746     1   22.6031746       5.55     0.0236
         Environ.#class |   17.8764937     2   8.93824683       2.19     0.1251
                   Major |   6.30354914     1   6.30354914       1.55     0.2210
          Environ#Major |   1.65177457     2   .825887284       0.20     0.8174
            class#Major |   21.1050473     1   21.1050473       5.18     0.0284
  Environ#class#Major |   4.42331015     2   2.21165507        0.54     0.5855
                Residual |   158.928571    39   4.07509158
-----------------------+----------------------------------------------------
                   Total |   266.705882    50   5.33411765
```

5. What conclusions can you draw? How much of the variation in memory is accounted for by environment, class section, and major?

6) Let's finish by conduct an AxS ANOVA on this data:

Subject	Silence	Music	Interruption	Mean	Std. Dev
1	1	4	2	2.333	1.528
2	6	6	1	4.333	2.887
3	4	4	3	3.667	0.577
4	5	0	3	2.667	2.517
5	4	6	6	5.333	1.155
6	8	6	4	6.000	2.000
7	6	5	11	7.333	3.215
8	4	5	1	3.333	2.082
9	6	8	5	6.333	1.528
10	7	5	3	5.000	2.000
11	6	5	4	5.000	1.000
12	7	6	6	6.333	0.577
13	8	7	2	5.667	3.215
14	9	4	11	8.000	3.606
15	6	5	6	5.667	0.577
16	9	6	6	7.000	1.732
17	6	5	4	5.000	1.000
Sample size	17	17	17	N = 51	
Mean	6.000	5.118	4.588	M = 5.235	
Std. Dev	2.031	1.691	2.938	S = 2.312	

What assumptions must be met in order to run an AxS ANOVA? Are you worried about any of these assumptions with our dataset?

Complete the AxS ANOVA summary table below. To speed things up, recall that $SS_{Total} = (N-1)s^2_{all}$. To speed things up even more, take a look at SStotal for all the previous analyses and note that the design of our analysis does not change the total amount of variation in our data. If you're really feeling lazy, look at how we would have calculated SSA on page 1. Would that calculation change for an AxS ANOVA?

Source	SS	df	MS	MSR
Environment				
Subjects				
Env x Subjects				
(ANOVA error)				
Total				