Activity #6: Repeated Measures ANOVA & Experimental Design

In this activity, we're going to learn how to conduct a repeated measures ANOVA (sometimes called an AxS ANOVA).

ANOVA : AxS ANOVA :: Independent samples t-test : Dependent samples t-test

1. In a one-way ANOVA, the total sums of squares among observations is partitioned into two components:

   ______________________________ (a measure of the variance due to the treatment effect), and

   ______________________________ (a measure of the unexplained, random error variance).

   Sums of squares represent: ____

   These sums of squares are divided by their degrees of freedom to get ________________________________

   which represent ____.

   We then calculate the ratio of these means squares and compare that ratio to the ______ distribution.

   Under the null hypothesis, we expect this ratio to equal ______.

   If the null hypothesis is false, we expect the mean square ratio to be ____________________________.

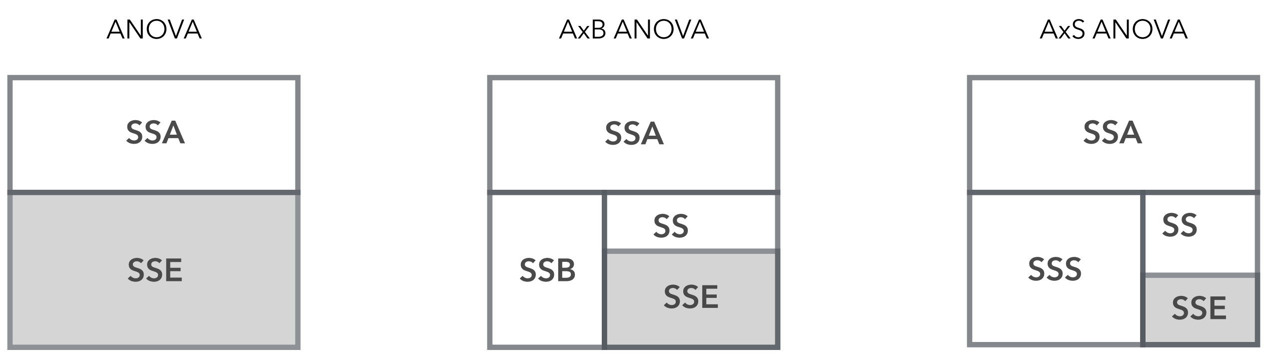
   The power of a statistical hypothesis test refers to: ____

   If we want to increase the power of our ANOVA, we could:

   _______________ our sample size, _______________ our α-level, or _______________ the size of MS_E.

   If MS_E is small, our mean square ratio (F-statistic) will be ______________ and we have a better chance of rejecting H_0.

2. The following figures attempt to explain how the variance is partitioned under one-way, AxB, and AxS ANOVA:

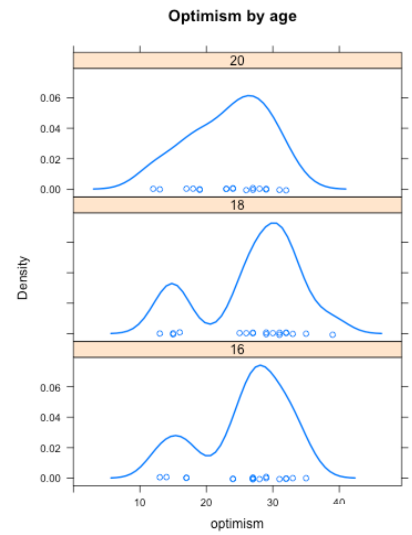


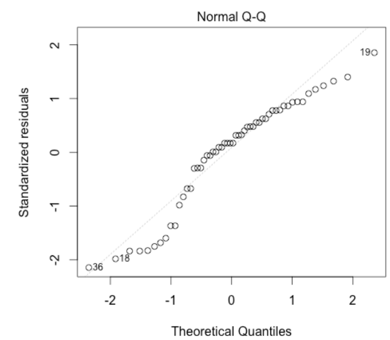
   Explain how the AxB and AxS ANOVA may increase the power of our statistical test.

Scenario: How optimistic are you? Are you more or less optimistic than you were at age 16?

50 subjects were given an optimism test at age 16. These same subjects were given the same test at ages 20 and 24. The data from the 18 subjects who completed all 3 surveys is displayed below:

subject	age 16	age 18	age 20
1	35	39	32
2	32	35	31
3	33	32	28
4	32	32	29
5	31	33	26
6	29	30	29
7	29	31	27
8	27	29	27
9	27	31	24
10	28	27	24
11	27	27	23
12	27	26	23
13	24	29	19
14	24	25	19
15	17	16	18
16	17	15	17
17	14	15	12
18	13	13	13
n =	18	18	18
mean =	25.89	26.94	23.39
sd =	6.579	7.487	5.922





Bartlett test of homogeneity of variances
Bartlett's K-squared = 0.9167, df = 2, p-value = 0.6323

3. Using the information provided above, evaluate the assumptions necessary to conduct a one-way ANOVA.

4. For now, let's pretend as though the data come from 3 independent groups (a random sample of 18 subjects at each age). If that were true, we could conduct a one-way ANOVA and find:

Source	**SS**	**df**	**MS**	**MSR (F)**
age	120	2	60	1.34
Error	2285	51	44.8	p-value= 0.271
Total	2405	53	MS_{total}	$\eta^2 = 0.05$

What conclusions can we draw?

Interpret the eta-squared value.

5. The data from this study did <u>not</u> come from 3 independent groups; it came from the same 18 subjects measured 3 different times. This kind of study is called a *repeated measures* design (in which treatments are *nested within* subjects). Why we might want to conduct this *repeated measures design* (or AxS ANOVA) over a one-way ANOVA?

Advantages of AxS ANOVA: ____

Disadvantages of AxS ANOVA: ____

6. Let's take another look at our data. In the last column, I've calculated mean optimism levels for each subject.

subject	age 16	age 18	age 20	means
1	35	39	32	35.33
2	32	35	31	32.67
3	33	32	28	31.00
4	32	32	29	31.00
5	31	33	26	30.00
6	29	30	29	29.33
7	29	31	27	29.00
8	27	29	27	27.67
9	27	31	24	27.33
10	28	27	24	26.33
11	27	27	23	25.67
12	27	26	23	25.33
13	24	29	19	24.00
14	24	25	19	22.67
15	17	16	18	17.00
16	17	15	17	16.33
17	14	15	12	13.67
18	13	13	13	13.00
n =	18	18	18	54
mean =	25.89	26.94	23.39	25.407
sd =	6.579	7.487	5.922	6.736

What are some reasons why two subjects in this study differ in optimism? Some of the variation in optimism could be due to:

- The effect of age
- Pre-existing individual differences (subject effect)
- The interaction between subjects and age (or variation unexplained by the subject and the age effects)/

The age effect is what we're interested in estimating. In our one-way ANOVA, all other variation (within the age groups) was treated as random unexplained error.

In this AxS design, we can partition this error variance further (by explaining some of it).

We know that some of the variation in optimism within each age group is due to pre-existing individual differences. For example, subject #3 tends to have higher optimism than subject #18. Since these *subject-effects* can be estimated by our experimental design, we can explain these effects and remove them from the unexplained error variance.

The rest of the variation in optimism could be described as the unknown interaction between a subject and his or her age. Some people will become more optimistic over time, some will become less optimistic, and others will have other *paths*.

Before we get to the formulas used to calculate these mean squares, let's take some time to think about what we're doing. We are saying we can explain all of the variation within each age group through subject and subject-interaction effects. Because these sources of variation are explained, the remaining error variance shrinks. If MSE (the denominator of our ratio) is smaller, then our MSR will be larger, which will give us a greater probability of rejecting the null hypothesis (increase power).

Notice that, in this study, we measured a subject's optimism only once per age group. If we had measured each subject multiple times at each age, those measurements would vary, and we'd be introducing some unexplained error variance.

7. Sketch a diagram explaining how one-way and AxS ANOVA partition the total sum of squares.

8. The following table summaries how to calculate a summary table for an AxS design:

Source	SS	df	MS	MSR (F)
Treatment	$\sum pn\left(\bar{X}_{A_i} - M\right)^2$	$a-1$	SS_A / df_A	
Subjects	$\sum pa\left(\bar{X}_{S_i} - M\right)^2$	$n-1$	SS_S / df_S	
Treatment x Subjects	$\sum\sum p\left(\bar{X}_{S_iA_j} - \bar{X}_{S_i} - \bar{X}_{A_i} + M\right)^2$ $= SS_T - SS_A - SS_S$ $= SS_{\text{one-way ANOVA error}} - SS_S - SS_{WS}$	$(a-1)(n-1)$	SS_{AxS} / df_{AxS}	
*Within-subjects	$\sum\sum\sum\left(X_{ijk} - \bar{X}_{S_iA_j}\right)^2$	$an(p-1)$	SS_{WS} / df_{WS}	
**One-way ANOVA error	$\sum\sum p\left(X_{ijk} - \bar{X}_{A_i}\right)^2$ $= SS_S + SS_{AxS} + SS_{WS}$	$an-a$	SS_E / df_E	
Total	$\sum\sum\sum\left(X_{ijk} - M\right)^2$ $= SS_A + SS_S + SS_{AxS} + SS_{WS}$ $= SS_A + SS_{\text{one-way ANOVA error}}$ $= \left(n_{ijk} - 1\right)s^2_{total}$	$an-1$	MS_{total}	η^2

Notes: * = When p=1 (each subject is measured only once per treatment), we will have no within-subject variation.
** = We do not need to calculate this row, but it may simplify the other calculations

Notation: a = number of groups
n = number of subjects within each treatment (column)
p = number of measurements per subject per treatment

9. Verify the following AxS ANOVA summary table. Notice the SSA, SSE, and SStotal values are the same as what they were in the one-way ANOVA.

Source	**SS**	**df**	**MS**	**MSR (F)**
Treatment	120	2	60	
Subjects	2182	17	128.3	
Treatment x Subjects	103	34	3.03	
One-way ANOVA error	2182 + 103 = 2285	17 + 34 = 51	44.8	
Total	120 + 2182 + 103 = 2405	2+17+34 = 53	MS_{total}	η^2

$$SS_A = 18[(25.9-25.4)^2 + (26.9-25.4)^2 + (23.4-25.4)^2] = 18(0.25+2.25+4) = 120$$ *(same as ANOVA)*

$$SS_E = \sum\sum(X-\overline{X}_{A_j})^2 = \sum(n-1)s_a^2 = (18-1)[(6.58^2+7.49^2+5.92^2)] = 2285$$ *(same as one-way ANOVA)*

$$SS_T = SS_A + SS_E = (35-25.4)^2 + (39-25.4)^2 + \ldots + (13-25.4)^2 = 2405$$ *(same as one-way ANOVA)*

$$SS_S = 3[(35.3-25.4)^2 + (32.7-25.4)^2 + \ldots + (13.7-25.4)^2 + (13.0-25.4)^2] = 2(726.43) = 2182$$

$$SS_{AS} = SS_E - SS_S = 2285 - 2182 = 103$$

10. The expected values of all the mean square values (assuming the null hypothesis is true) are displayed below. Explain what each mean square represents. If we're interested in the treatment effect (effect of age of optimism), which mean square ratio should we calculate?

$MS_A =$

$$E[MS_A] = \sigma^2 + \alpha\eta + \alpha$$

$MS_S =$

$$E[MS_S] = \sigma^2 + \eta$$

$MS_{AxS} =$

$$E[MS_{AxS}] = \sigma^2 + \alpha\eta$$

$MS_E =$

$$E[MS_E] = \sigma^2$$

11. Calculate the appropriate MSR, estimate the p-value, and draw an appropriate conclusion. Explain why this conclusion does or does not match the conclusion we made from our original one-way ANOVA.

12. Recall that in our one-way ANOVA, eta-squared = 0.05 (meaning 95% of the variation in optimism was unexplained). Calculate and interpret eta-squared for this AxS ANOVA.

η^2 = ______________________________

What proportion of variation in optimism is due to: pre-existing subject characteristics = ______________________

treatment effects (age) = ______________________________

13. Suppose we calculate another MSR using MSS and MSA. Do we expect this MSR to be large or small? What would a large value of MSR represent?

14. From our answers to the two previous question, we can quantify the advantage of the AxS ANOVA (over a one-way ANOVA). As long as our subjects differ from one another (and stay somewhat consistent to themselves) across all treatments, we will observe a significant subject-effect.

A larger subject effect means that our unexplained (error) variance will be smaller. This, in turn, means the denominator of our mean square ratio will be smaller. This causes our MSR to be larger, which makes it more likely that we'll reject the null hypothesis.

The value of $\frac{MS_S}{MS_{AxS}}$ indicates how worthwhile it was for us to conduct an AxS ANOVA instead of a one-way ANOVA.

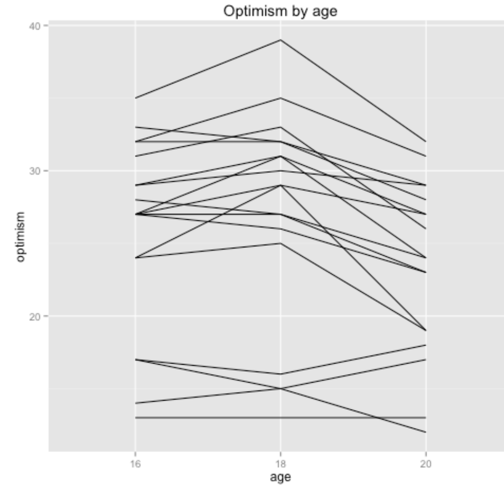
If this mean square ratio is significant, then we've eliminated a significant chunk of the unexplained variance. If it is <u>not</u> significant, then we did not reduce the error variance (and may have wasted time/resources).

We can also measure the advantage of using AxS ANOVA by noting:

$$MS_{E\text{ (one-way)}} = \frac{SS_S + SS_{AxS}}{df_S + df_{AxS}}$$

$MS_{AxS} = MS_{E\text{ (one-way)}}\left(1 - r_{pooled}\right)$, where r is the correlation of subject scores across groups.

15. To the right, I've provided a profile plot for all 18 subjects in this study. What can we discern from this plot?



16. Below, I've written-out the formal model for the AxS ANOVA. Remember our model identifies potential sources of variation. Interpret each piece of the model.

$$\text{One-way ANOVA: } Y_{ijk} = \mu + \alpha_j + e_{ijk}$$

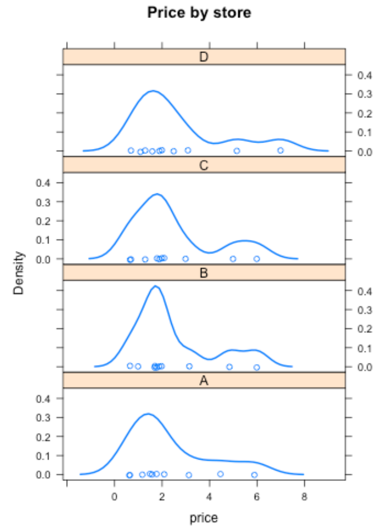
$$\text{AxB ANOVA: } Y_{ijk} = \mu + \alpha_j + \beta_k + \alpha_j\beta_k + e_{ijk}$$

$$\text{AxS ANOVA: } Y_{ijk} = \mu + \alpha_j + \eta_i + \alpha_j\eta_i + e_{ijk}$$

17. Optional discussions, if time permits:
    (1) sphericity,
    (2) fixed vs. random effects,
    (3) missing data and imputation
    (4) proportionality assumption for AxB ANOVA

Scenario: Do some stores charge higher prices than others? The prices of 10 items were found at 4 stores.

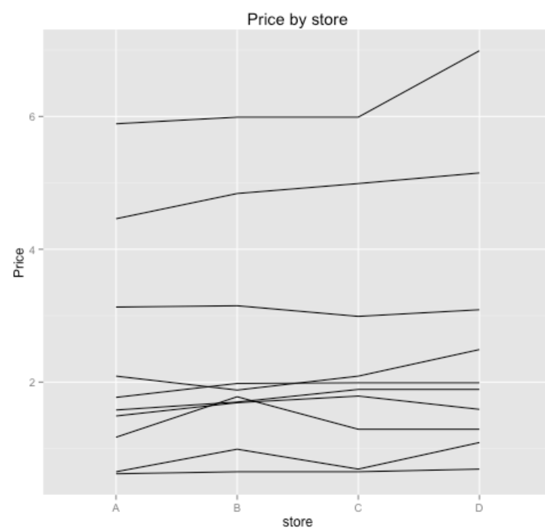
Item	Store A	Store B	Store C	Store D	means
aspirin	4.46	4.84	4.99	5.15	4.860
bread	1.58	1.7	1.89	1.89	1.765
cereal	3.13	3.15	2.99	3.09	3.090
eggs	0.65	0.99	0.69	1.09	0.855
ground beef	2.09	1.88	2.09	2.49	2.138
detergent	5.89	5.99	5.99	6.99	6.215
lettuce	1.17	1.78	1.29	1.29	1.383
milk	1.49	1.69	1.79	1.59	1.640
potatoes	1.77	1.98	1.99	1.99	1.933
tomato soup	0.62	0.65	0.65	0.69	0.653
n =	10	10	10	10	40
mean =	2.285	2.465	2.436	2.626	2.453
sd =	1.178	1.707	1.765	1.988	1.732



Bartlett test of homogeneity of variances
Bartlett's K-squared = 0.2701, df = 3, p-value = 0.9655

The summary table and profile plot are displayed below:

Source	**SS**	**df**	**MS**	**MSR (F)**
Store	0.5859	3	0.19529	4.344 p = 0.0127
Items	115.2	9	12.8	
Store x Items	1.2137	27	0.04495	
One-way ANOVA error	116.4137	36		
Total	116.9996	39	MS_{total}	η^2



Source: http://ww2.coastal.edu/kingw/statistics/R-tutorials/repeated.html

18. What conclusions can you make from this study?.

19. Calculate an effect size and determine if it was worthwhile to run an AxS ANOVA on this data.

20. Describe any follow-up analyses you would want to conduct on this data.

Scenario: Suppose we're interested in studying the effects of various textbooks on the achievement of statistics students. If we sampled 3 classrooms and randomly assigned them to use one of the textbooks, our data would look like this:

Textbook A	**Textbook B**	**Textbook C**
Classroom #1 n = 16 students Mean, SD	Classroom #4 n = 10 students Mean, SD	Classroom #7 n = 5 students Mean, SD
Classroom #2 n = 20 students Mean, SD	Classroom #5 n = 15 students Mean, SD	Classroom #8 n = 10 students Mean, SD
Classroom #3 n = 30 students Mean, SD	Classroom #6 n = 14 students Mean, SD	Classroom #9 n = 30 students Mean, SD

On it's surface, this might appear to be a simple ANOVA (or possibly an AxB ANOVA). Notice that we only have one factor (textbook type) and that the observations within each treatment are grouped.

That is why this is called a Groups Within Treatments Design.

21. Would you be willing to assume that all 66 observations within the Textbook A group are independent? Why?

22. How many independent observations do we have in this study: 150 (students) or 9 (classrooms)?

Scenario: In Iowa, the local government may be organized in a variety of ways: mayor and council with all officers elected at large; mayor elected at large with council members elected by district; council members elected by district and mayor elected from within the council; variations of these organizational types with and without a hired city manager; etc.

A political scientist wondered if one form of organization resulted in more "contented" citizens than another. He identified four types of government (A1, A2, A3, A4) to be studied. He then developed an interview-based scale by which a citizen could express his/her satisfaction with the local government's responsiveness to the needs and concerns of citizens.

Five communities were identified that are organized under each type of government. Within each community, 50 heads of households were drawn at random by use of telephone directories, tax roles, water/sewer billings, and other sources. Each resident who could be located was interviewed to obtain a measure of overall satisfaction with the local government. The data are summarized below:

Govt. Type A1	**Govt. Type A2**	**Govt. Type A3**	**Govt. Type A4**
$n_{11} = 43$	$n_{12} = 42$	$n_{13} = 39$	$n_{14} = 49$
$\overline{X}_{11} = 15.2$	$\overline{X}_{11} = 19.1$	$\overline{X}_{13} = 18.4$	$\overline{X}_{14} = 20.4$
$s_{11}^2 = 50.3$	$s_{11}^2 = 42.5$	$s_{13}^2 = 52.6$	$s_{14}^2 = 51.0$
$n_{21} = 41$	$n_{22} = 48$	$n_{23} = 41$	$n_{24} = 45$
$\overline{X}_{21} = 16.4$	$\overline{X}_{22} = 16.3$	$\overline{X}_{23} = 19.2$	$\overline{X}_{24} = 17.6$
$s_{21}^2 = 42.8$	$s_{22}^2 = 48.7$	$s_{23}^2 = 48.2$	$s_{24}^2 = 43.4$
$n_{31} = 46$	$n_{32} = 40$	$n_{33} = 47$	$n_{34} = 47$
$\overline{X}_{31} = 18.1$	$\overline{X}_{32} = 14.7$	$\overline{X}_{33} = 20.4$	$\overline{X}_{34} = 15.8$
$s_{31}^2 = 49.6$	$s_{32}^2 = 50.1$	$s_{33}^2 = 40.4$	$s_{34}^2 = 40.5$
$n_{41} = 40$	$n_{42} = 50$	$n_{43} = 40$	$n_{44} = 38$
$\overline{X}_{41} = 15.0$	$\overline{X}_{42} = 15.2$	$\overline{X}_{43} = 19.7$	$\overline{X}_{44} = 16.6$
$s_{41}^2 = 38.3$	$s_{42}^2 = 35.3$	$s_{43}^2 = 38.2$	$s_{44}^2 = 44.2$
$n_{51} = 45$	$n_{52} = 45$	$n_{53} = 33$	$n_{54} = 41$
$\overline{X}_{51} = 15.3$	$\overline{X}_{52} = 14.2$	$\overline{X}_{53} = 18.3$	$\overline{X}_{54} = 18.1$
$s_{51}^2 = 45.9$	$s_{52}^2 = 48.4$	$s_{53}^2 = 42.7$	$s_{54}^2 = 56.9$

23. Notice how stratified random sampling was used. First, a sample of 5 communities within each government type was taken. Then, within each community, a sample of 50 individuals was taken. We have to ask ourselves the following question:

    Do we have reason to believe the individuals (minor units) within each community (major unit) are similar to one another? In other words, do we believe there are dependencies within each community?

    If the answer to this question is yes, we have to run the study with 20 observations (20 communities or major units)
    If the answer to this question is no, we can run this study with 860 observations (860 individuals)

    Why would we want to run this study with 860 observations instead of 20 observations?

24. We will begin by examining the minor units (860 individuals). If we determine that significant dependencies exist within each major unit (community), we will have to run our analysis with only 20 observations. If we determine that there are no significant dependencies within each community, we'll run our analysis with 860 observations. Here are the calculations used in our **minor unit analysis**.

Note: A = Treatment Groups; K = Major Units (groups within each treatment)

Minor Unit Analysis				
Source	Sums of Squares	Degrees of freedom	Mean Square	Mean Square Ratio
Treatment (Government) (A)	$\sum n_{A_j}\left(\overline{X}_{A_j} - M\right)^2$	$a-1$	$\frac{SS_A}{df_A}$	
Between Groups Within Treatments (GwA)	$\sum\sum n_{K_j}\left(\overline{X}_{K_j} - \overline{X}_{A_j}\right)^2$	$K-a$	$\frac{SS_{GwA}}{df_{GwA}}$	$\frac{MS_{GwA}}{MS_W}$
Within Groups (W)	$\sum\sum\sum\left(X_{ikj} - \overline{X}_{K_j}\right)^2$ or $\sum(n_{K_j}-1)s^2$	$N-K$	$\frac{SS_W}{df_W}$	

Here are the calculations based on our data. Note that weighted means are used in all calculations (weighted by the number of observations within each community).

Minor Unit Analysis				
Source	Sums of Squares	Degrees of freedom	Mean Square	Mean Square Ratio
Treatment (Government) (A)	1618.09	3	539.36	
Between Groups Within Treatments (GwA)	1655.79	16	103.49	2.28 (sig.)
Within Groups (W)	38187.9	840	45.64	

25. What do the various mean squares represent? What does our mean square ratio represent? Given that our mean square ratio is significant, what do we conclude?

A = distance between treatments

GwA = distance between the groups within each treatment (if there are no dependencies, we'd expect this to be small)

W = distance within each community (random error not due to community or government types)

MSR = how much of an impact community membership has on the dependent variable

We conclude that there are dependencies within each community, therefore we cannot treat all 860 observations as independent observations. We'll have to treat our communities as 20 random observations.

26. If we did not find significance, we could treat all 860 observations as random observations. To do this, we would just have to calculate SSerror = SSw + SSGwA and run a regular ANOVA.

Since we found a significant "groups within treatments" effect, we have to run a major units analysis. To do this, we need to recalculate the treatment means using unweighted means (ignore the sample size within each community). We then complete the following summary table:

Major Unit Analysis (following a significant GwA effect)				
Source	Sums of Squares	Degrees of freedom	Mean Square	Mean Square Ratio
Treatment (Government) (A)	$\sum k_j \left(\overline{X}^*_{A_j} - M^*\right)^2$	$a-1$	$\frac{SS_A}{df_A}$	$\frac{MS_A}{MS_{GwA}}$
Groups Within Treatments (GwA)	$\sum\sum\left(\overline{X}_{K_j} - \overline{X}^*_{A_j}\right)^2$	$K-a$	$\frac{SS_{GwA}}{df_{GwA}}$	

Here are the calculations based on our data. State the conclusion based on this analysis:

Minor Unit Analysis				
Source	Sums of Squares	Degrees of freedom	Mean Square	Mean Square Ratio
Treatment (Government) (A)	36.9	3	12.3	5.28 (sig)
Between Groups Within Treatments (GwA)	37.34	16	2.33	