Blocking

Matthew W. Wheeler

Last time there were two things that we needed to go over more. The big one is :

Additional Covariates

Why are additional covariates so important?

If you are randomizing out extra variability, does it matter?

Types of Additional Covariates:

Blocks: These are things we can include in our experimental design. That is we assign experimental units to Blocks.

Observed: Covariates that we can not assign to the experimental unit, but come along for the ride and may impact the result. For example, you can't assign IQ or height.

Though they are essentially analyzed the same we are more interested in blocks, because they can affect how we randomly assign units.

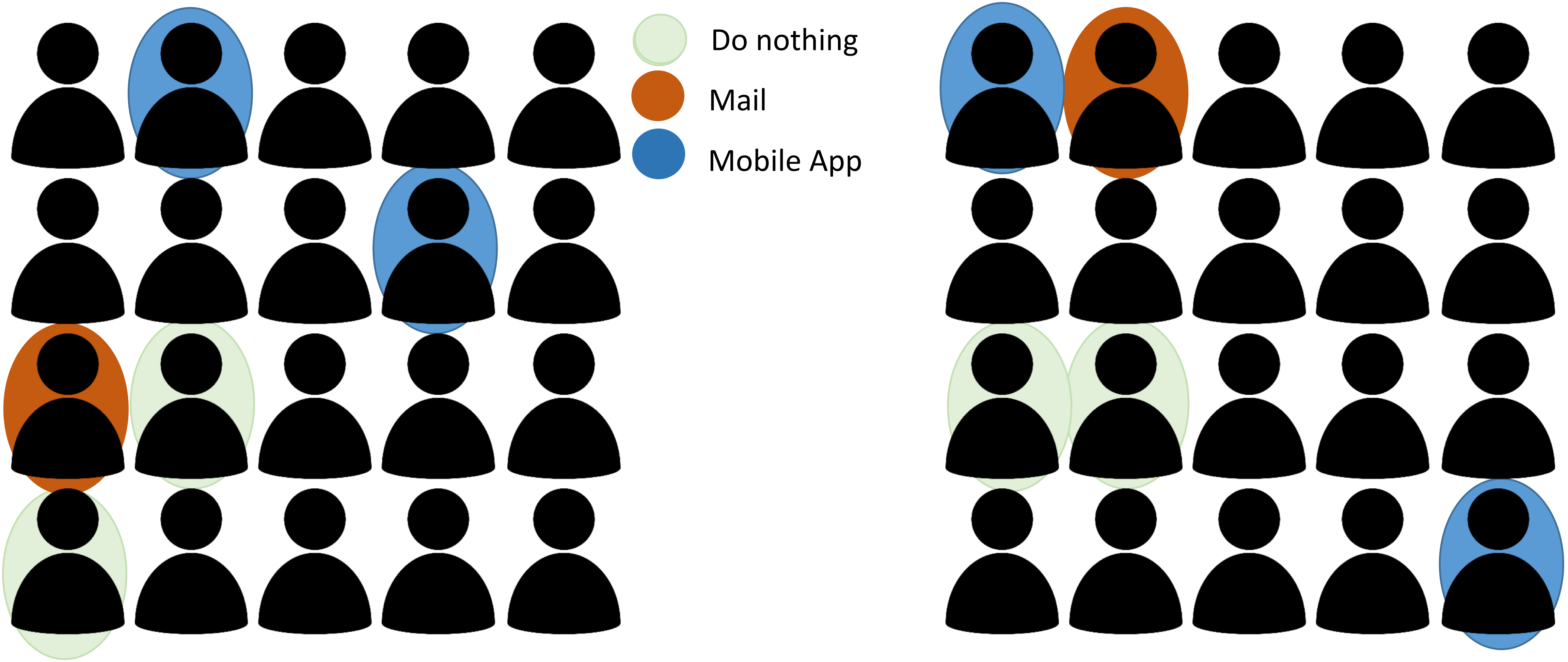
Let’s think about the direct marketing campaign example from last class. What if I was the Kroger Company and I had two different Brands?

For example, Kroger has both Kroger Stores and Harris Teeter Stores.

It is very possible that Harris Teeter Shoppers are different from Kroger shoppers (for example I always spend more in Teeter because it is more expensive), but if I am essentially using the exact same marketing methods, I need to account for people who go to Harris Teeter and people who go to Kroger.

Kroger

Harris Teeter
Neighborhood Food & Pharmacy



That is, I have two **BLOCKS** one is Kroger the other is Harris Teeter.

I randomly select customers from these blocks. These blocks are an extra variable I have no control over as an experimenter, but I can account for in my model. I can not make these two franchises the same, but I can randomly sample customers from each franchise.

NOTE: This is not the same as extra ‘observed’ variables or extra factors in my experiment. WHY?

Example 1

To save energy, an engineer investigated the effects of different types of insulation for the same home design. We are interested in looking into differences in cooling costs for the months of June, July and August.

Insulation Type

	1	2	3	4
June	$74.44	$68.75	$71.34	$65.47
July	$89.96	$73.47	$83.62	$72.33
August	$82.00	$71.23	$79.88	$70.87

Factor: Insulation type, with four different types of insulation.

Block : Month of observation.

Experimental Unit: House

Measurement: Amount spent on cooling house.

Tests of Interest: Differences between all insulation types.

Additionally 2 and 4 are similar structurally, are they different than 1 and 3?

Experiment Wise Error Rate: α=0.05

```
/* Insulation example 1 for
   Class 3 Notes*/
data insulation;
       input month insulation cost @@;
       cards;
1 1 74.44 2   1 89.96       3 1    82.00
1 2 68.75 2 2 73.47 3 2 71.23
1 3 71.34 2 3 83.62 3 3 79.88
1 4 65.47 2 4 72.33 3 4 70.87
;
/*
 Estimate the individual effects of the
 given experimental design
*/
proc glm data=insulation;
       class month insulation; *we have one block and one factor;
       model cost = month insulation; *cost is a linear model of the month and the
                                                              type of insulation;
       lsmeans insulation/cl adjust=TUKEY; *multiple effects adjustment TUKEY;
       contrast 'Insulation 2 and 4 -vs- 1 and 3' insulation 0.5 -0.5 0.5 -0.5;
       estimate 'Insulation 2 and 4 -vs- 1 and 3' insulation 0.5 -0.5 0.5 -0.5;
run;
quit;
```

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	514.2602667	102.8520533	15.41	0.0023
Error	6	40.0515333	6.6752556		
Corrected Total	11	554.3118000			

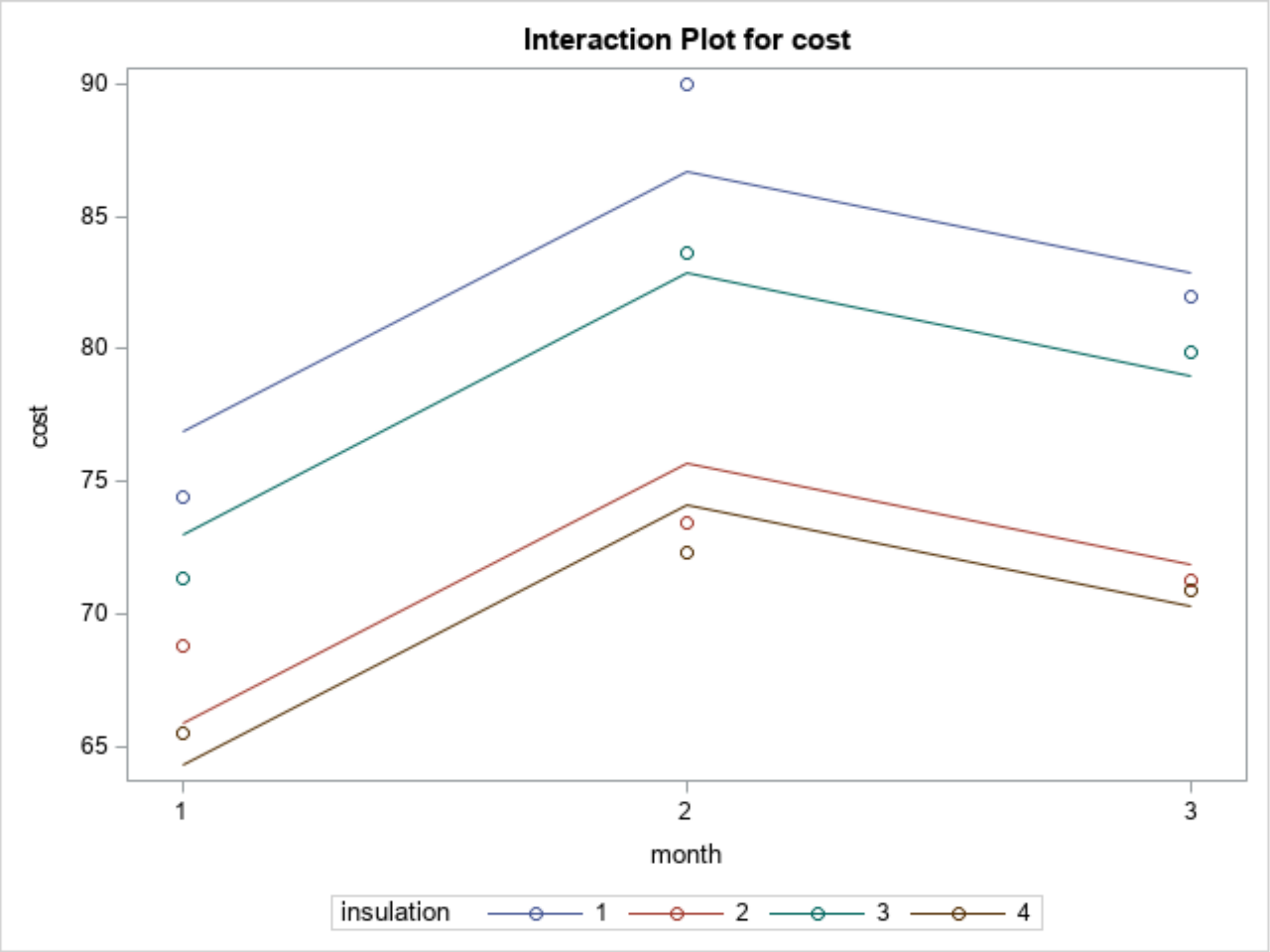
R-Square	Coeff Var	Root MSE	cost Mean
0.927745	3.432056	2.583652	75.28000

Type I Sums of Squares

Source	DF	Type I SS	Mean Square	F Value	Pr > F
month	2	196.9154000	98.4577000	14.75	0.0048
insulation	3	317.3448667	105.7816222	15.85	0.0029

Something is going on here

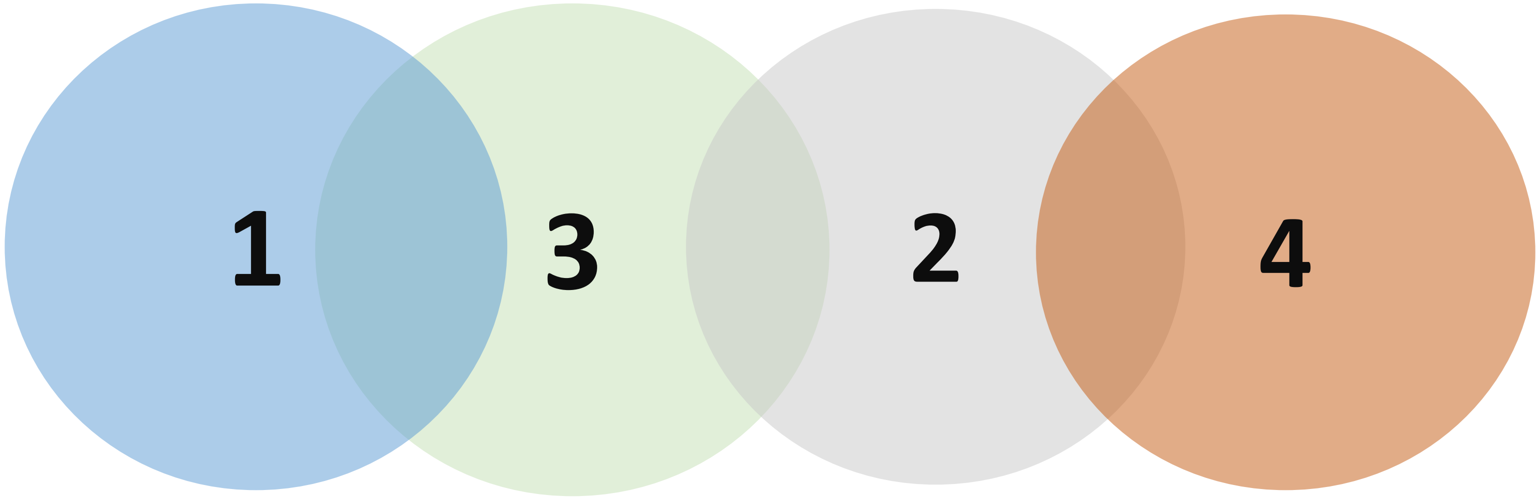
Does the month interact with the type of insulation?



Least Squares Means for effect insulation Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: cost				
i/j	1	2	3	4
1		0.0080	0.3473	0.0040
2	0.0080		0.0550	0.8714
3	0.3473	0.0550		0.0236
4	0.0040	0.8714	0.0236	

Least Squares Means for Effect insulation				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	10.983333	3.680717	18.285950
1	3	3.853333	-3.449283	11.155950
1	4	12.576667	5.274050	19.879283
2	3	-7.130000	-14.432617	0.172617
2	4	1.593333	-5.709283	8.895950
3	4	8.723333	1.420717	16.025950

So from above we have this weirdness:



There are no groups that are completely separate. What do I do?

One Possible Solution, but what about the Type 1 Error rate?

```
contrast 'Insulation 2 and 4 -vs- 1 and 3' insulation 0.5 -0.5 0.5 -0.5;
estimate 'Insulation 2 and 4 -vs- 1 and 3' insulation 0.5 -0.5 0.5 -0.5;
```

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Insulation 2 and 4 -vs- 1 and 3	1	291.2645333	291.2645333	43.63	0.0006

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Insulation 2 and 4 -vs- 1 and 3	9.85333333	1.49167194	6.61	0.0006

I can use a Bonferroni adjustment, but technically I have 7 comparisons. Six original, plus the other one, so to control for the type I error my $\alpha^* = 0.05/7 = 0.007$.

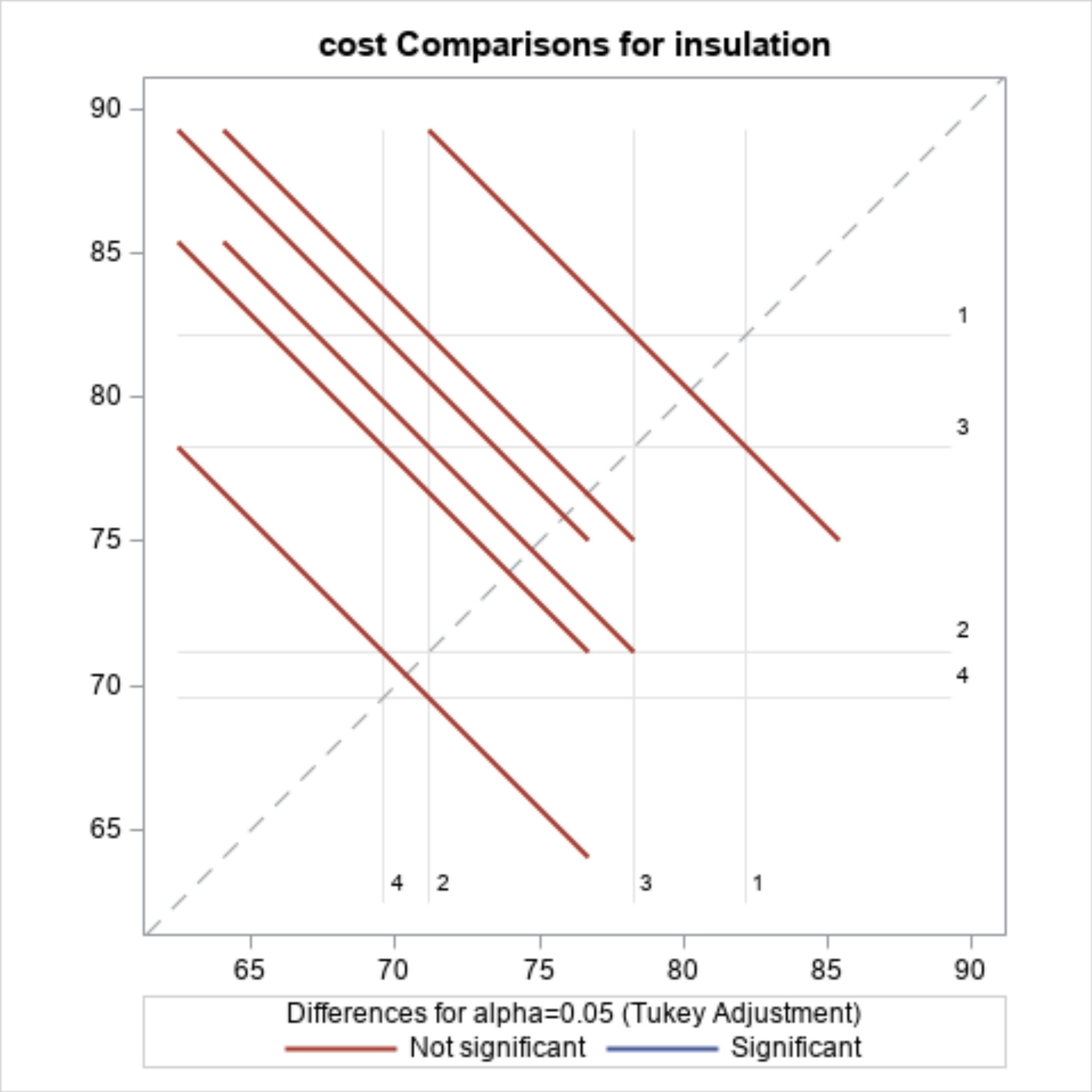
I can barely conclude that 2 and 4 are different from 1 and 3 at a simultaneous error of 0.05 that the two types of insulation are different.

I also didn't test if the blocks had significant difference, because I really didn't care. If I did, my overall Type I error rate would not necessarily be controlled.

What if I didn't use the blocking?

```
/*
 Estimate the individual effects of the
 given experimental design
*/
proc glm data=insulation;
       class  insulation; *we have one block and one factor;
       model cost =  insulation; *cost is a linear model of the month and the
                                                        type of insulation;
       lsmeans insulation/cl adjust=TUKEY; *multiple effects adjustment TUKEY;
       contrast 'Insulation 2 and 4 -vs- 1 and 3' insulation 0.5 -0.5 0.5 -0.5;
       estimate 'Insulation 2 and 4 -vs- 1 and 3' insulation 0.5 -0.5 0.5 -0.5;
run;
quit;
```

Month is not in the model.



cost Comparisons for insulation
90
85
80
75
70
65
65
70
75
80
85
90
1
3
2
4
4
2
3
1
Differences for alpha=0.05 (Tukey Adjustment)
Not significant
Significant

Barely have enough power just for this contrast (if I didn't do multiple comparisons).

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Insulation 2 and 4 -vs- 1 and 3	1	291.2645333	291.2645333	9.83	0.0139

So what have we learned?

Blocks are extra sources of variability that we account for in our sampling plan.

By removing that variability, we can get better estimates of what we are studying.

They are things we are not interested in studying, but they an change the results if we don't account for them.

What we have been dealing with are “complete” designs.

That is everyone factor/block gets the same number of observations.

<u>This is unrealistic</u>. In a real study, we will not be able to always get the same number of experimental units for each treatment or within each block.

This makes all of the pretty book formulas blow up. One reason I have not and will not show you them is that SAS doesn’t really care.

But you will notice some results will change (e.g. Type I and Type III sums of squares, mean estimates vs. lsmean estimates in proc glm)

What Breaks Down

Tukey's test (as well as most other tests other than Bonferonni break down). You either

a) Use Bonferroni

b) Search for another method that will work (most likely there isn't one)

Type I and Type III sums of squares mean different things and may give diverging answers.

It is usually preferable to do Type III sums of squares. This is the variability that is accounted for by the Factor/Block given all of the other Factor/Blocks are in the model

Example 2

A company is interested in determining the optimal amount of flower to use in making it's cookies "chewier". In this study, three flower amounts are considered and chocolate chip cookies are made in two kitchens by two different chefs (each chef visits both kitchens).

Factor: Amount of flour in the cookie recipe.

Blocks : Chef and Kitchen

Experimental Unit: Cookie Batch

Measurement: Cookie Chewiness.

Tests of Interest: Differences between all flower types.

Also is there are difference between Chefs?

Experiment Wise Error Rate: α=0.05

```
/* Cookie Chewiness for example 2
   class 3*/
data chew;
      input chef kitchen flour chew @@;
      cards;
      1 1 1 1.620 1 1 1 1.342 1 2 1 2.669 1 2 1 2.687
      1 2 1 2.155 1 2 1 4.000 1 1 2 3.228 1 1 2 5.762
      1 2 2 6.219 1 2 2 8.207 1 1 3 6.615 1 1 3 8.245
      1 1 3 8.077 1 2 3 11.37 2 1 1 2.282 2 2 1 4.233
      2 2 1 4.664 2 2 1 3.002 2 2 1 4.506 2 2 1 6.385
      2 2 1 3.696 2 1 2 5.080 2 1 2 4.741 2 1 2 4.522
      2 2 2 4.647 2 2 2 4.999 2 2 2 5.939 2 1 3 8.240
      2 1 3 6.330 2 1 3 9.453 2 1 3 7.727 2 2 3 7.809
      2 2 3 8.942
;
```

```
proc glm data = chew;
      *Three class variables;
      class chef kitchen flour;
      *Only main effects for now;
      model chew = flour chef kitchen;
      lsmeans flour/cl adjust=bon;
      *above line adjusts using a Bonferroni adjustment;
      run;
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
flour	2	138.8967359	69.4483680	49.05	<.0001
chef	1	0.6148981	0.6148981	0.43	0.5153
kitchen	1	18.1398510	18.1398510	12.81	0.0013

Notice the differences now that it is not a complete block design.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
flour	2	155.2477358	77.6238679	54.83	<.0001
chef	1	0.0980783	0.0980783	0.07	0.7943
kitchen	1	18.1398510	18.1398510	12.81	0.0013

What are Type I and Type III SS

Each batch of flower is different and I have controlled at the α = 0.05

Least Squares Means for Effect flour				
i	j	Difference Between Means	Simultaneous 95% Confidence Limits for LSMean(i)-LSMean(j)	
1	2	-2.440255	-3.755811	-1.124699
1	3	-5.712716	-7.102940	-4.322493
2	3	-3.272461	-4.647213	-1.897708

Least “chewy” flour amount: Type 1

Chewiest flour amount : Type 3

Are the Chef’s different : ?

Question: Is the Type III SSE enough to answer the last question?

So let’s recap:

Blocks are thing we can assign experimental units, this effects how we design our experiment.

Sometimes we randomly sample a group of blocks, and then randomly assign treatments within the blocks.

These sampling plans are outside of the scope of the class, but you need to know they exist.

A/B testing recap (so far..)

1) The first thing is to define your measurements.

What are you looking to achieve?

This will define every question after this.

A/B testing recap (so far..)

2) Are there any blocks that might have extra variability, which when accounted for will give you more information about the one factor of interest?

For example, you might have multiple populations (Kroger Shoppers/Harris Teeter Shoppers).

My favorite is the Chevy Nova, rumor has it is that it didn’t go over so well in Spain. If you know any Spanish you will understand why.

A/B testing recap (so far..)

3) Plan your treatment comparisons. Focus on what you are really interested in; you need to plan them ahead of time. This prevents "data snooping" and you coming up with some answer that seems right but isn't when it goes to market.

It is better to come back with nothing than to suggest that management make a decision, spend money on the change, and not get any results after implementation. You do it once, it is bound to happen, too much you might be unemployed.

A/B testing recap (so far..)

4) We have only hinted at it, but the next two classes will be about power. You need to have a large enough sample to be able to see if a difference exists. You can't have too big of a sample because it costs money and time.

Most of your time should be going into finding the optimal sample size.

5) **Implementation**: Assign your experimental units to Blocks/Treatments etc.

This is the part of the experimental design that ensures you get correct results in the analysis.