MA 214: Applied Statistics

Instructor: Ashis Gangopadhyay

ANOVA
Part II

Where We've Been

We were discussing….

- how to develop a model for a continuous response variable based on one categorical covariate (factor). This is called one-way ANOVA.

Where We're Going

- We want to generalize the by considering multiple categorical covariates (factors).
- In particular, we want to discuss the models with two factors. This called two-way ANOVA.

Two-factor Analysis of Variance

- Frequently, we may need to investigate the effects of several factors on the response variables.
- In the blood pressure example, we may be interested in investigating the effects of treatment method (medication, exercise and diet) and the gender of the patients (male and female) on the blood pressure level. This is a two-factor ANOVA with six treatment combinations.

Two-factor Analysis of Variance

- Let's also consider the following example:
- **Example 2:** A study was conducted by a bakery on the effects of heights of display (bottom, middle and top shelf) and width of display (regular and wide) on the sales of its bread. In this study, 12 supermarkets are selected that are similar in terms of sales volume and clientele, and two of these supermarkets are randomly assigned to each of the six treatment combinations. At the end of the study period, the sales of the bakery's bread (in terms of the number of cases sold) was recorded.

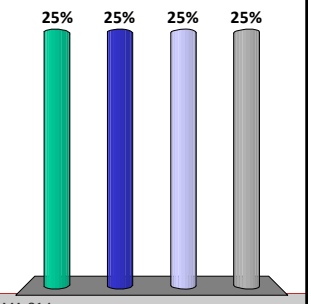
Two-factor Analysis of Variance

Factor B (Width)

Factor A (Height)	**B1 (Regular)**	**B2 (Wide)**
A1 (Bottom)	47	46
	43	40
A2 (Middle)	62	67
	68	71
A3 (Top)	41	42
	39	46

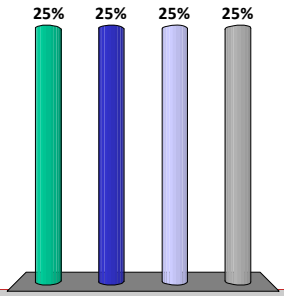
In the Bakery example, what is the number of treatments?

1. 2
2. 3
3. 5
4. ✓ 6



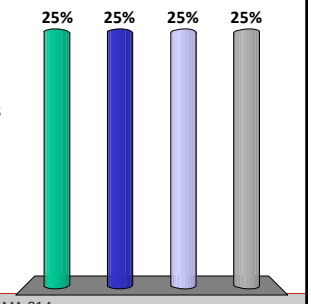
In the Bakery example, the number of replication is…

1. 1
2. ✓ 2
3. 4
4. 12



In the Bakery example, is the design balanced?

1. Yes, since the number of treatments is even.
2. Yes, since the number of replications are equal.
3. No, since the factor levels of the two factors are unequal.
4. Not meaningful, one applies for one-way ANOVA.



Two-factor ANOVA

- As we have seen in multiple regression, there are additional issues that you need to worry about when we have more than one factor (covariate) in the model. Let's consider the following two examples:
- **Example 3:** We wish to study the effect of two medications on the reduction in cholesterol level among patients suffering from chronic high cholesterol. Each drug can be used at two dosage levels (low and high), and a combination of the two drugs is being used in the study.

Two-factor ANOVA

- **Example 3 (con't):** This is a two-factor experiment with the following factors and levels:

	Factor A (Drug 1)		**Factor B (Drug 2)**	
Dosage Level:	A1 (Low)	A2 (High)	B1 (Low)	B2 (High)

Two-factor ANOVA

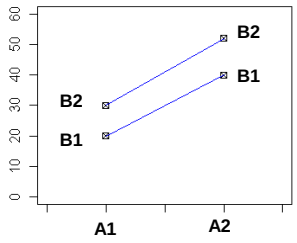
- **Example 3 (con't):** Thus, the experiment has four treatments. Ten patients are randomly selected from a group of forty patients, and are assigned to each treatment. After six months, the data on the reduction in cholesterol level was collected from each patient. The following table shows the average reduction in each treatment combination.

		Factor B	
		B1	**B2**
Factor A	**A1**	20	30
	A2	40	52

Two-factor ANOVA

Average Response Plots

(a)



Two-factor ANOVA

- **Example 4:** We have the same set-up as in example 3, but with the following average response:

	Factor B	
	B1	**B2**
A1	12	40
A2	50	10

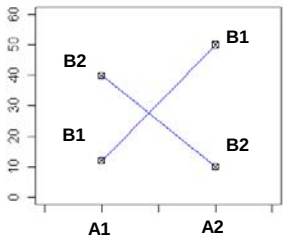
Factor A

What are the differences between these two datasets?

Two-factor ANOVA

Average Response Plots

(b)



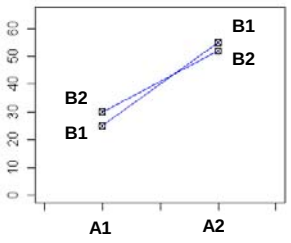
Factorial Experiment and Interaction

- A factorial experiment (or a complete factorial experiment) involves all possible combinations of factor levels of all factors. If A and B are two factors with factor levels *a* and *b* respectively, then the experiment is called an *axb* factorial experiment. The Bakery experiment is an example of a 3x2 factorial experiment.
- If a factorial experiment, if the effect one factor (say A) on the response variable depends on the factor level of the other factor(s) (say B), we say that the factors A and B interact.

Interaction

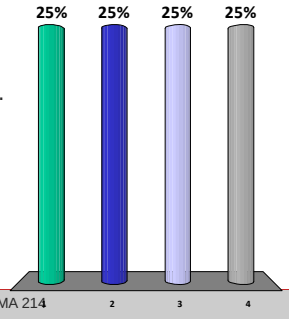
Average Response Plots

(c)



Does the average response plot in (c) indicate that the two factors interact?

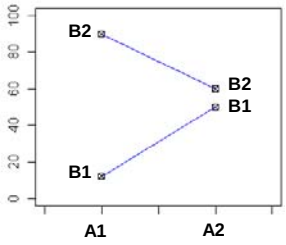
1. Yes, the lines cross.
2. Yes, since in a factorial experiment, factors always interact.
3. No, the lines in the response plot move in the same direction. ✓
4. Can't say without knowing more about the experiment.



Interaction

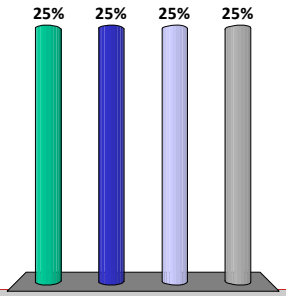
Average Response Plots

(d)



Does the average response plot in (d) indicate that the two factors interact?

1. Yes, since the direction of the lines are substantially different.
2. No, since the lines do not cross.
3. No, since the lines are not parallel.
4. Can't say without knowing more about the experiment.



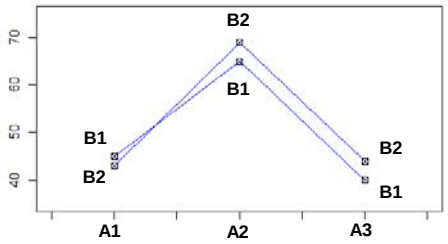
Interaction

Average Response Table for Bakery Data

	B1 (Regular)	B2 (Wide)
A1 (Bottom)	45	43
A2 (Middle)	65	69
A3 (Top)	40	44

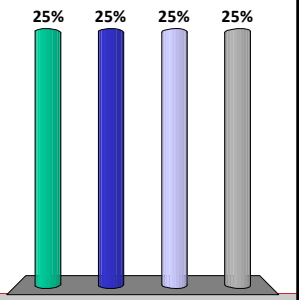
Interaction

Average Response Plot for Bakery Data

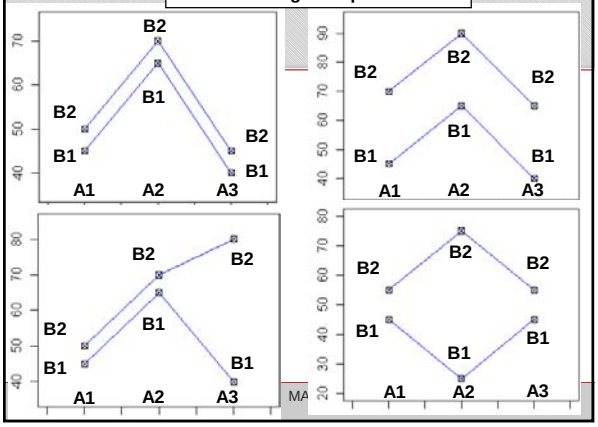


Based on the average response plot, what can we say about a possible interaction between height and width?

1. Significant interaction as the lines change direction.
2. Significant interaction as the average value at A2 is much higher than that at A1 or A3.
3. No interaction as the lines move roughly in the same direction.
4. No interaction as the average values for B1 and B2 at each levels A1, A2 and A3 are roughly the same.

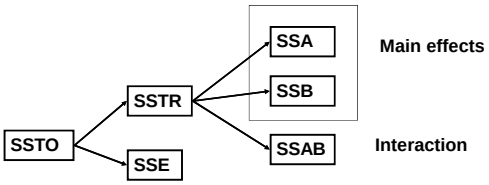


Various Average Response Plots



1.5 ANOVA for two-factor model

- **Breakdown of SS in Two-Way ANOVA**



ANOVA for two-factor model

- We consider a two factor ANOVA with factors A and B. In general we assume that there are *a* levels of Factor A, and *b* levels of factor B, and *r* replications within each treatment. Thus, the total sample size is $n = abr$.

Source	*SS*	*DF*	*MS*	*F*
Treatment	*SSTR*	*ab – 1*	*MSTR*	$F_{TR} = \frac{MSTE}{MSE}$
Error	*SSE*	*ab(r – 1)*	*MSE*	
Total	*SSTO*	*abr – 1*		

ANOVA for two-factor model

- Breakdown of *SSTR*:

Source	*SS*	*DF*	*MS*	*F*
A	*SSA*	*a – 1*	*MSA*	$F_A = \frac{MSA}{MSE}$
B	*SSB*	*b – 1*	*MSB*	$F_B = \frac{MSB}{MSE}$
AB	*SSAB*	*(a – 1)(b – 1)*	*MSAB*	$F_{AB} = \frac{MSAB}{MSE}$
Total	*SSTR*	*ab – 1*		

ANOVA for two-factor model

- *How to interpret the ANOVA?*

Step 1: First, test the following hypothesis:

H_0 : There is no treatment effect.

H_a : There is a significant treatment effect.

Decision Rule: Reject H_0 if $F_{TR} > F_\alpha$; $df = \{ab-1, ab(r-1)\}$

Note that acceptance of H_0 implies that the factors are not effective in explaining the variability of the response variable. In that case, we STOP and conclude that the model is NOT significantly significant. Otherwise, we proceed to the next step.

ANOVA for two-factor model

- *How to interpret the ANOVA?*

Step 2: Test for interaction

H_0 : There is no possible interaction between factors A and B.

H_a : There is a significant interaction between factors A and B.

Decision Rule: Reject H_0 if $F_{AB} > F_\alpha$; $df = \{(a-1)(b-1), ab(r-1)\}$

Note that rejection of H_0 implies strong interaction between factors A and B. In this case, individual inference of factor A and factor B are not meaningful. DO NOT carry out the F-test for main effects. Instead identify the best treatment by carrying out a pairwise comparison procedure on all *axb* treatments. Otherwise, if the interaction is not significant, we proceed to the next step.

ANOVA for two-factor model

- *How to interpret the ANOVA?*

Step 3: Test for main effect A

H_0 : There is no significant effect of factor A.

H_a : There is a significant effect of factor A.

Decision Rule: Reject H_0 if $F_A > F_\alpha$; $df = \{(a-1), ab(r-1)\}$

If we reject H_0, it implies that the factor levels of factor A are significant, and we can use multiple comparison procedures to identify the equivalent groups among the factor levels of factor A.

ANOVA for two-factor model

- *How to interpret the ANOVA?*

Step 4: Test for main effect B

H_0 : There is no significant effect of factor B.

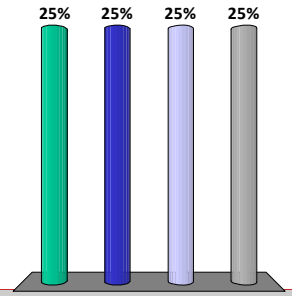
H_a : There is a significant effect of factor B.

Decision Rule: Reject H_0 if $F_B > F_\alpha; df = \{(b-1), ab(r-1)\}$

If we reject H_0, it implies that the factor levels of factor B are significant, and we can use multiple comparison procedures to identify the equivalent groups among the factor levels of factor B.

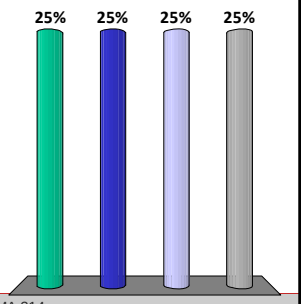
In a two factor ANOVA, the hypothesis to test first is….

1. Main effects are significant.
2. Interaction is significant.
3. ✓ Overall treatment effect is significant.
4. None of the above.



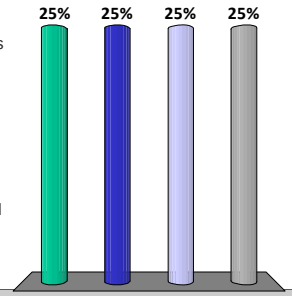
If the overall treatment effect is significant, the hypothesis to test next is…

1. The main effects are significant.
2. ✓ The interaction is significant.
3. The pairwise differences between the means of the factor levels are significant.
4. STOP and conclude that the model is statistically significant.



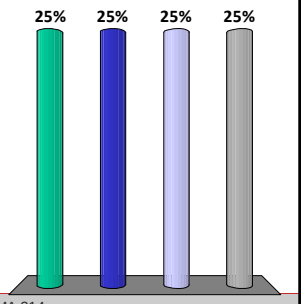
If the interaction is statistically significant, then the next step of the analysis is…

1. Test the significance of each of the two main effects.
2. Carry out pairwise comparisons between the factor levels of each factor separately.
3. ✓ Do not test the significance of the main effects, carry out pairwise comparisons between all possible treatment means instead.
4. STOP and conclude that model is not statistically sound.



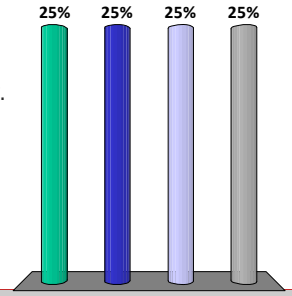
If the interaction is not significant, then the next step of the analysis is …..

1. Test for the significance of main effects.
2. Test for the significance of main effects and carry out pairwise comparisons of factor level means of all main affects.
3. ✓ Test for the significance of main effects and carry out pairwise comparisons of factor level means only for the significant main affects.
4. STOP and conclude that the model is not statistically significant.



Analysis of bakery data suggests that the best combination of height and width to maximize sales is…

1. Either bottom or top shelf & wide display
2. Either bottom or top shelf & regular display.
3. Middle shelf and wide display
4. ✓ Middle shelf, the width doesn't matter.

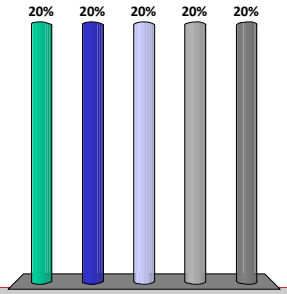


Multifactor ANOVA

- The approach to analyze two-factor ANOVA can be generalized to a multifactor settings. However, there are certain additional concerns that need to be addressed.
- Consider the problem of a three factor ANOVA. Suppose the factors are A, B and C at factor levels *a, b* and *c* respectively. This is called a *axbxc* factorial experiment.

In the three factor experiment, how many possible interaction terms are there?

1. 1
2. 2
3. 3
4. 4 ✓
5. None of the above



Multifactor ANOVA

- For the three factor experiment, the interactions are AB,AC,BC and ABC and

  SSTR=SSA+SSB+SSC+SSAB+SSAC+SSBC+SSABC
- What is the general formula for the number of interactions in a k-factor experiment?
- Thus in multifactor experiments, the number of interactions can be very large, which creates problems with analysis and interpretation.

Multifactor ANOVA

- One possible solution is to assume that higher order interactions are negligible and not include those terms in the models.
- So for example, in a three factor model, we may want to include terms up to second order interaction.
- There are other better approaches to handle the problem of large number of interaction terms in multifactor models.