Lecture: Overview of Experimental Design

Dr. Goutam Chakraborty

Outline

- Experiments (or, tests)
  - Why do we need experiments?
  - Is experimentation relevant in the world of big data?
  - Examples of experiments that are done by businesses (off-line and online) or for academic research
  - What are some key issues in doing these experiments correctly?

2

Basic Issues in Experiments

- Why do we need experiments (tests)?
  - To understand and quantify cause-and-effect relationship between variables
- Why do we care about *cause-and-effect* relationships?
- Why do we need it in the world of big data (in business applications)?
- Why do we need it in the academic world (why not just use surveys or secondary data)?
- What are some of the problems in understanding cause-an-effect relationships particularly *in social sciences*?

3

Cause and Effect in Business Applications

- A credit card marketer mailed two different offers of balance transfer checks to two groups of customers from its housefile (current customer base).
- One group was offered 1.99% APR on transferred balances (from other credit cards), and the other group was offered 3.99% APR on transferred balances (from other credit cards).
- The response rate from the first group was 4.5%. For the second group, response rate was 3.0%
- For simplicity, assume that
  - Response rate is the **only issue** that the marketer is interested in. (Clearly there could be others, such as amount of balance transferred or how long customer carries the balance.)
  - The two response rates are **statistically significantly** different.
- Could we conclude that the low APR **causes** the response rate to go up?

4

Cause and Effect in Academic Research

- What's the problem with establishing "cause-and-effect" in survey data?
  - Contemporaneous covariation
- If we just SEM (Structural Equation Modeling) then we will be good – right?
  - Not really!
- Look at articles in journals such as *JCR*, *JCP* and so on

5

The Role of Tests in Marketing (Applied)

- Experiment and Testing in Marketing is nothing new!
- Historically speaking, the ability to test (almost anything such as prospect list, offer copy or, offer premium) is one of the biggest advantages of direct (database) marketers over mass marketers
- With the advent of the Internet and web, we now have the ability to design and run a test, get the data, analyze the data, and make decisions in a *very short time*
- However, businesses often fail to take full advantage of the power of tests because they use rules such as "*test one thing at a time*" or "*test a few packages*" or, do *A/B testing*

6

What Is an Experiment (or Test)?

- The manipulation (or change) of one (or more) variables by the manager so that its effect on one (or more) other variables can be **unambiguously** determined.
- **Independent variables** (or *factors*) – these are the ones being changed by the manager. The values that a factor takes are called the *levels* of the factor.
  - Sometimes factors are *not manipulated, but observed/measured* such as days of the week we run a promotion on our web site or, customer's lifetime value.
- **Dependent variable** - the variable that the manager hopes will change because of the change in the independent variable.

7

Steps in Conducting a Test

- Decide what you want to change and what you expect this change to achieve.
  - independent variables (factors), levels of the factors, dependent variable
- Compute sample size. This depends on the following:
  - desired confidence level in test results.
  - level of sampling error you are willing to tolerate in test results.
- Structure and perform the test (i.e., administer properly).
  - use randomization.
  - compare test with a control group (typically what you have been doing historically).
- Analyze the data from the test (calculate statistics and conduct appropriate statistical tests of significance).
- Make your decision

8

Determining Sample Size

- The idea is that we want to know **before** we collect data (or before we run tests) what our sample size (n) should be to guarantee a particular *level of confidence and a specified sampling error*.
- Statistical issues that affect sample size:
  - Desired level of confidence
  - Acceptable error limit
  - Power of test
- Pragmatic issues that affect sample size:
  - Cost (as well as time/effort) in getting the desired sample
  - External considerations

9

Administering the Test Correctly

- **Key concept: Randomization**
  - Subjects (test units) must be assigned **at random** to the different levels of independent variable (or factor).
  - Let's revisit the credit card marketer who is using offers of either 1.99% or 3.99% APR. The marketer uses the following rule:
    - If the zip code of a customer's address is less than or equal to 49999, then the customer gets the 1.99% offer.
    - If the zip code is greater than 49999, then the customer gets the 3.99% offer.
  - What do you think about this procedure?

10

More on Randomization

- **Key concept: Randomization**
  - Many factors (other than the one manipulated) could influence the dependent variable in any test.
  - The goal for the experimenter is to **control (hold equal)** as many of these **extraneous** variables across the groups as possible.
  - In most cases with large sample sizes, randomization will produce groups that will be *roughly equal* on all measured/unmeasured characteristics.

11

Lecture: Single Factor ANOVA

Outline

- Basics of single factor design
- Model, Mechanics and Assumptions for such design
- Analysis issues

13

Single Factor, Multiple Levels Experiment

- Dependent variable, factor and factor levels are selected by the analyst
- Name of the design : Completely Randomized
  - Subjects are chosen randomly from population of interest (for generalizability)
  - Subjects (or, test units) are assigned randomly to each level of the factor (or, *treatment condition*) for averaging out extraneous variation
- Typically, dependent variable is measured using interval scale (continuous)
  - Binary response may be analyzed using the same technique (for large data)
- Analysis name: ANOVA (Analysis of Variance)

14

An Example of Single Factor Experiment

- A charity organization wants to test whether the *type of package* in which an appeal for $ is sent has any effect on *donation* generated from such offers.
- The company currently uses a *large envelope* and wants to test *three new* packages: a medium envelope, a small envelope, and a self-mailer
  - Content of the appeal will be the same
- The company chooses 2,000 names **randomly** from its housefile.
- Each member is **randomly** assigned to receive an identical appeal in one of the following four types of mail packages:
  - large envelope (control group), medium envelope, small envelope, or self-mailer

The ANOVA Model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

Y_{ij} the j^{th} value of the response variable for the i^{th} factor level.

μ the overall population mean of the response

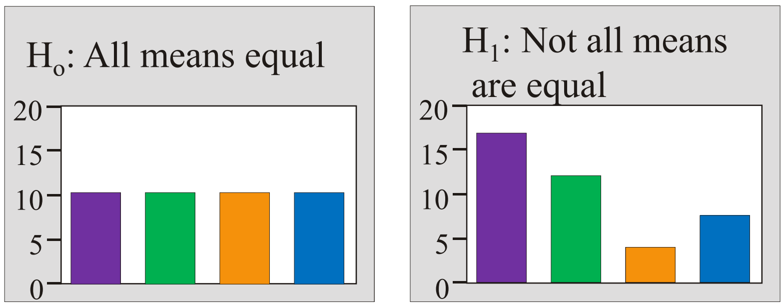
α_i the difference between the population mean of the i^{th} factor level and the overall mean, μ. This is referred to as the effect of level i.

ε_{ij} the error term, or residual. All variability other than that caused by the factor is included here.

Note that it's a linear model!

16

The ANOVA Hypothesis (F-test): Look at Means of Response for All Subjects in Each Factor Level

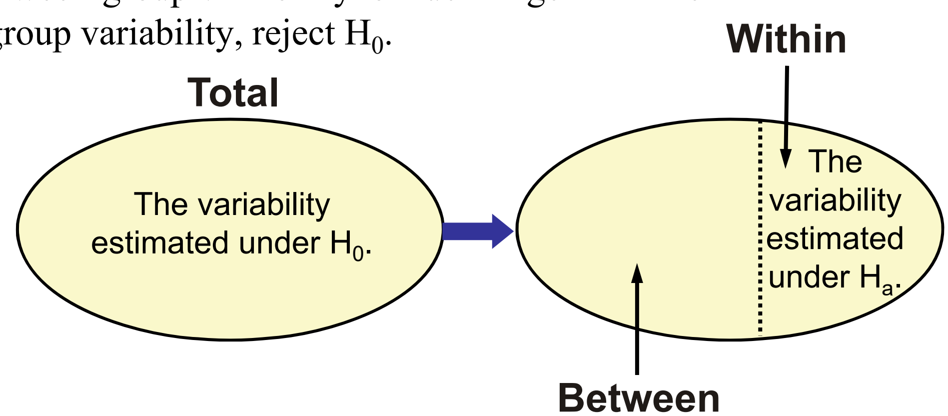


H_0: $\mu_{\text{Self-mailer}} = \mu_{\text{Small}} = \mu_{\text{Large}} = \mu_{\text{Medium}}$

17

Partitioning Variability in ANOVA

If the between-group variability is much larger than the within-group variability, reject H_0.



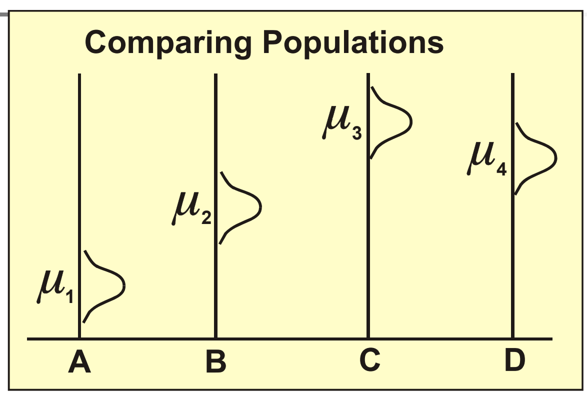
18

Mechanics of ANOVA

Total variation	the sum of the squared differences between each observed value and the overall mean, $\Sigma\Sigma(Y_{ij}-\mu)^2$. This is the overall variability in the response variable and is equal to the between-group variation plus the within-group variation. It is the corrected total sum of squares.
Between-group variation	the sum of the squared differences between the mean for each group and the overall mean, $\Sigma n_i(\alpha_i)^2$. This is the variability explained by the independent variable. It is the model sum of squares.
Within-group variation	the sum of the squared differences between each observed value and the mean for its group, $\Sigma\Sigma(Y_{ij}-(\mu+\alpha_i))^2$. This is the variability not explained by the independent variable. It is the error sum of squares.

19

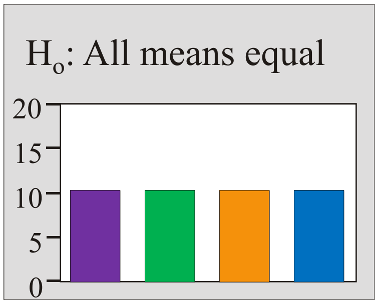
Assumptions of ANOVA

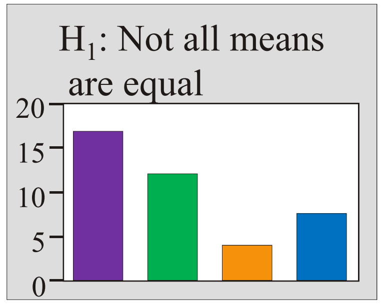


- independent observations
- normally distributed residuals
- equal variances for each group

20

Follow-up Tests after (F-test)





H_0: $\mu_{\text{Self-mailer}} = \mu_{\text{Small}} = \mu_{\text{Large}} = \mu_{\text{Medium}}$

What can we conclude after we reject the above H_o?
Can we say that $\mu_{\text{Self-mailer}} \neq \mu_{\text{Small}}$? Or, $\mu_{\text{Large}} \neq \mu_{\text{Medium}}$?

21

Different Follow-Up Tests

- Depends on researcher's goals and preferences
- If the researcher wants to compare all tests and control groups with each other, then choices are:
  - Tukey's HSD (honest significant difference), Bonferroni, Scheffe…
- If the researcher wants to compare just the control group against each of the test group, then use Dunnett's test

22

Demo Single Factor ANOVA

Data set: Package4

Demo Procedure

- Open file: Package4
- Variables: ID, Package (4 levels), Donation (net in $ after subtracting mailing and processing costs)
- Note: overall mean of donation across all 2,000 observations = $4.46
- Analyze > Fit Y by X > Select Donation as Y, Package as X factor > OK
  - Click red Triangle next to Oneway Analysis .. > Select Means and Std Dev
  - Click red Triangle next to Oneway Analysis .. > Select Means/ANOVA
- Interpret results
- Do follow up tests (red triangle): Dunnett and Tukey's HSD
- Test Assumptions (red triangle): Unequal variances

24

Lecture: Factorial Design and ANOVA

Dr. Goutam Chakraborty

Outline

- More on Experiments and ANOVA
  - What if I want to change multiple variables such as:
    - the image on our landing page (control image vs. test image)
    - the discount for ordering today (control is no discount, test is 5% off)
  - One possibility: we change one variable at a time and then make our decisions sequentially
  - Another possibility : we change them *simultaneously* but do it in a statistically correct way instead of a haphazard fashion
    - On the web, A/B testing often does this in an *ad hoc* way!
  - What if the dependent variable is not continuous but binary (1/0) such as if a person responded to our offer or not?

26

Two-Way ANOVA and Factorial Design

- **Factorial designs**: two or more **factors** (with any number of levels) varied simultaneously.
- Combinations of levels of different factors are called *treatments*.
- Each subject must be chosen **randomly** form population of interest and assigned **randomly** to any one of the treatments.
- An example: A credit card marketer wants to test two factors, **APR** and **Fee**, in a "direct mail offer of balance transfer checks."
  - **APR** has two levels: Control (4.99% the one that the marketer has been using in the past) and Test (2.99%, the one marketer wants to try out)
  - **Fee** has two levels: Control (no transaction fee), Test (transaction fee of 1% of transferred amount)

27

Testing Multiple Factors (Example)

- Note that the number of treatments in this example is 2×2, or 4.
- The marketer chose a **random sample** of 1,000 from his house file.
- Each sample member **assigned at random** to receive any 1 of these 4 treatments (combinations of APR and Transaction Fee).
- The marketer tracked the response from each offer (coded as 1=yes, 0=no) as well as how much balance was transferred from this experiment.
- Data file: `Balance Transfer`.
- How should we analyze the data and how should we explain the results from this experiment?

28

Plan of Analysis (Continuous Dependent Variable)

- Calculate average balance transfer for:
  - Each level of factor 1 (APR)
  - Each level of factor 2 (Fee)
  - Each treatment combinations
- Are these average balance transfer amounts different from each other?
- Test Global Null hypothesis in ANOVA followed by:
  - Test for one factor (main effect test), Test for another factor (main effect test) and Test for interaction effects between the two factors
    - Will be handled automatically in ANOVA when you specify model correctly
  - Follow-up tests (as appropriate)
    - To be selected by the analyst

29

Plan of Analysis (Binary Dependent Variable)

- Build a LR model for predicting Response (Binary):
- Test Global Null hypothesis in LR followed by:
  - Test for one factor (main effect test), Test for another factor (main effect test) and Test for interaction effects between the two factors
    - Will be handled automatically in LR when you specify model correctly
  - Use odds-ratios and/or LR model coefficients to explain results

30

Plan of Analysis (Binary Dependent Variable but Have Lots of Data)

- Change the binary variable (such as Response) from nominal to continuous
- Calculate average response rate for :
  - Each level of factor 1 (APR)
  - Each level of factor 2 (Fee)
  - Each treatment combinations
- Are these average response rates different from each other?
- Test Global Null hypothesis in ANOVA followed by:
  - Test for one factor (main effect test), Test for another factor (main effect test) and Test for interaction effects between the two factors
    - Will be handled automatically in ANOVA when you specify model correctly
  - Follow-up tests (as appropriate)
    - To be selected by the analyst

31

Additional Comments

- Test a few important factors (instead of one) and set each factor at a small number of levels.
- Use *factorial design* or its variants to test for interactions.
- Whenever possible, use *equal sample size* for each treatment condition.
- Use **true randomization**. Do not compromise this for the sake of convenience.
- Consider using a **back test** before making a final decision about changing your **control** offer.
- Learn about following topics (will be covered in later courses):
  - Use of **blocking/matching** to reduce unwanted variation whenever possible.
  - **Analysis of covariance** (ANCOVA) and **Multivariate analysis of variance** (MANOVA)

Demo: Analyzing Data from Factorial Design

Data Set: Balance Transfer

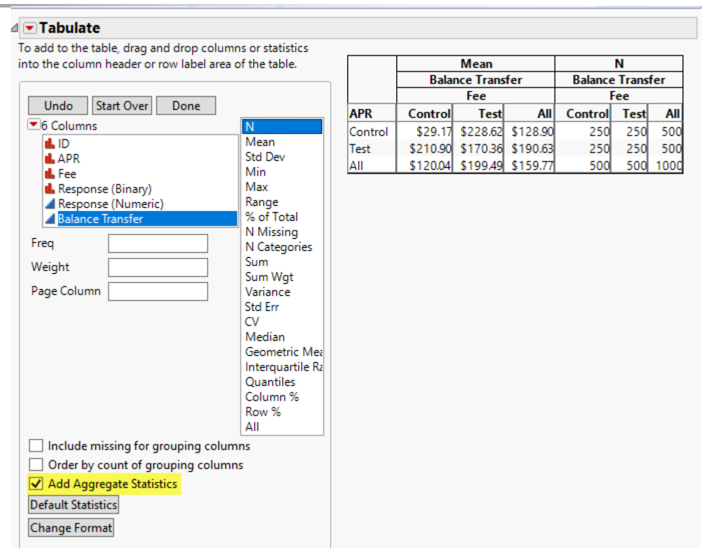
Outline

- Balance Transfer Data (2 factors each at 2 levels), 2 potential target variables (how many people responded and how much did they transfer)
- Demo of ANOVA with factorial design using 2 factors
  - With interval dependent variable (amount transferred)
  - With binary dependent variable (whether responded or not)
- Discus issues in analyzing data from higher order factorials (3 or more factors)

34

Demo Procedure (Target or Dependent Variable: Balance Transfer)

- JMP > Analyze> Tabulate > Control-Click and select Balance Transfer and Mean > Click Statistics and select Mean > drag to Resulting Cells > Click APR and drag it to row box > Click Fee and drag it to column header right-above balance transfer) > Click N and drag next to Mean > Click Add aggregate statistics box.



35

Table Re-created by Using Numbers from the Tables in the Previous Slide

Fee

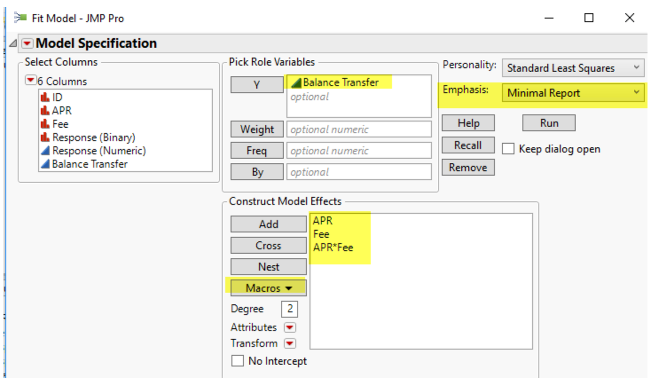
APR		Control (no fee)	Test (1% of transfer)	*Average*
	Control (4.99%)	$29.17	$228.62	*$128.90*
	Test (2.99%)	$210.90	$170.36	*$190.63*
	Average	*$120.04*	*$199.49*	$159.77

2-way ANOVA provides three tests: Main effect of APR, Main effect of Fee, and Interaction effect of APR*Fee.

36

Demo Procedure (Contd.)

- JMP > Analyze > Fit Model > Move Balance Transfer to Y > Control-Click both APR and Fee, then click macros drop-down arrow and select Full Factorial > Change Emphasis to Minimal Report > Run



37

JMP Output

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	6130770	2043590	9.6199
Error	996	211584888	212435	Prob > F
C. Total	999	217715658		<.0001*

Parameter Estimates

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
APR	1	1	952895.2	4.4856	0.0344*
Fee	1	1	1578234.5	7.4293	0.0065*
APR*Fee	1	1	3599640.0	16.9447	<.0001*

38

JMP Output: Test of Main Effect of APR

		Fee		
		Control (no fee)	Test (1% of transfer)	*Average*
APR	Control (4.99%)	?	?	*$128.90*
	Test (2.99%)	?	?	*$190.63*
	Average	*?*	*?*	

The sample average balance transfers are $128.90 and $190.63 for control and test APR. The main effect tests for APR are testing if the average balance transfer amounts are different between the control and test APR in the population.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
APR	1	1	952895.2	4.4856	0.0344*
Fee	1	1	1578234.5	7.4293	0.0065*
APR*Fee	1	1	3599640.0	16.9447	<.0001*

39

JMP Output: Test of Main Effect of Fee

		Fee		
		Control (no fee)	Test (1% of transfer)	*Average*
APR	Control (4.99%)	?	?	*?*
	Test (2.99%)	?	?	*?*
	Average	*$120.04*	*$199.49*	

The sample average balance transfers are $128.90 and $190.63 for control and test Fee. The main effect tests for Fee are testing if the average balance transfer amounts are different between the control and test Fee in the population.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
APR	1	1	952895.2	4.4856	0.0344*
Fee	1	1	1578234.5	7.4293	0.0065*
APR*Fee	1	1	3599640.0	16.9447	<.0001*

40

Question to Ponder About:

- If the manager ran two sequential tests (first with APR and then with Fee) and made sequential decisions based on each test, then what would he have concluded? (assume: data pattern would have remained the same)
- First test with APR : Conclusion will be that Test APR is better with higher ($190.63) average balance transfer than control APR ($128.90)
- Second test with Fee: Conclusion will be that Test Fee is better with higher ($199.49) average balance transfer than control Fee ($120.04)
- This seems to suggest that "Test APR and Test Fee" will be best combination!

41

JMP Output: Test of Interaction Effect of APR*Fee

		Fee: Control (no fee)	Fee: Test (1% of transfer)	*Average*
APR	Control (4.99%)	$29.17	$228.62	?
APR	Test (2.99%)	$210.90	$170.36	?
	Average	?	?	

The interaction effects is testing the patterns of the means inside the table. Specifically, it is testing if the effect of APR is the same whether Fee is set at control or, at test.

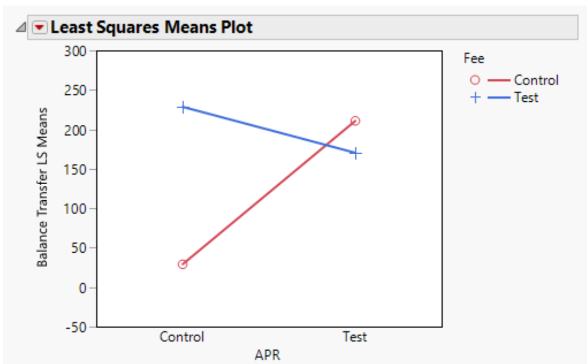
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
APR	1	1	952895.2	4.4856	0.0344*
Fee	1	1	1578234.5	7.4293	0.0065*
APR*Fee	1	1	3599640.0	16.9447	<.0001*

42

JMP Plot and Means

JMP > red triangle in Effect Details for APR*Fee > LSMeans Plot > Select create an interaction plot > select Fee as terms for overlay > red triangle Least squares means plot > Deselect show confidence limits (optional)



APR*Fee

Least Squares Means Table

Level	Least Sq Mean	Std Error
Control,Control	29.17200	29.150275
Control,Test	228.62000	29.150275
Test,Control	210.90400	29.150275
Test,Test	170.36400	29.150275

43

What about Follow-up Tests?

- Should we do follow-up tests for APR or Fee?
  - When there are *only two levels of a factor*, the significant main effect means that the two levels are different.
    - So, no need to do a follow-up test
- But what about the combinations (there are four levels) of APR*Fee?
    - JMP will do both Dunnett's and Tukey's HSD test on these means.
    - From red triangle of APR*Fee in Effects Details, select LSMeans Dunnett > Select Control,Control as level > OK
    - From red triangle of APR*Fee in Effects Details, select LSMeans Tukey's HSD

44

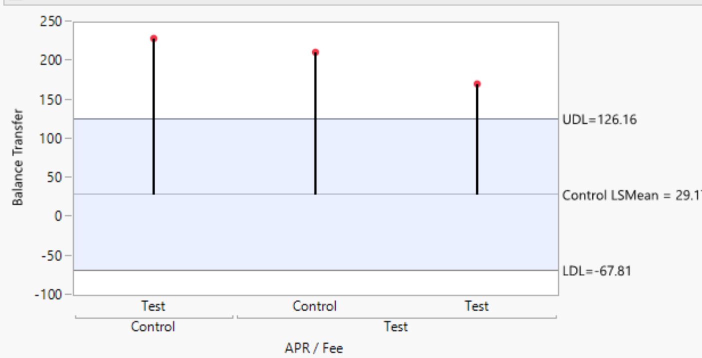
JMP: Follow up Tests (Dunnett)

LSMeans Differences Dunnett

α= 0.050 Q= 2.35257 Control=Control,Control Adjustment = Dunnett

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Control,Test	Control,Control	199.4480	41.22471	102.4641	296.4319	<.0001*
Test,Control	Control,Control	181.7320	41.22471	84.7481	278.7159	<.0001*
Test,Test	Control,Control	141.1920	41.22471	44.2081	238.1759	0.0019*

Control Differences



α = 0.05, Control = Control,Control

45

JMP: Follow up Tests (Tukey's HSD)

LSMeans Differences Tukey HSD

α= 0.050 Q= 2.57343

LSMean[i]	Mean[i]-Mean[j] / Std Err Dif / Lower CL Dif / Upper CL Dif	LSMean[j] Control,Control	Control,Test	Test,Control	Test,Test
Control,Control	Mean[i]-Mean[j]	0	-199.45	-181.73	-141.19
	Std Err Dif	0	41.2247	41.2247	41.2247
	Lower CL Dif	0	-305.54	-287.82	-247.28
	Upper CL Dif	0	-93.359	-75.643	-35.103
Control,Test	Mean[i]-Mean[j]	199.448	0	17.716	58.256
	Std Err Dif	41.2247	0	41.2247	41.2247
	Lower CL Dif	93.3589	0	-88.373	-47.833
	Upper CL Dif	305.537	0	123.805	164.345
Test,Control	Mean[i]-Mean[j]	181.732	-17.716	0	40.54
	Std Err Dif	41.2247	41.2247	0	41.2247
	Lower CL Dif	75.6429	-123.81	0	-65.549
	Upper CL Dif	287.821	88.3731	0	146.629
Test,Test	Mean[i]-Mean[j]	141.192	-58.256	-40.54	0
	Std Err Dif	41.2247	41.2247	41.2247	0
	Lower CL Dif	35.1029	-164.35	-146.63	0
	Upper CL Dif	247.281	47.8331	65.5491	0

Level			Least Sq Mean
Control,Test	A		228.62000
Test,Control	A		210.90400
Test,Test	A		170.36400
Control,Control		B	29.17200

Levels not connected by same letter are significantly different.

46

How to Analyze Response Rate as a Dependent Variable

- Theoretically correct way is to use LR as a model
- JMP > Analyze > Fit Model > Move Response (Binary) to Y > Click and change Target Level to 1 > Control-Click both APR and Fee, then click macros drop-down arrow and select Full Factorial > Run

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	18.10933	3	36.21867	<.0001*
Full	319.96044			
Reduced	338.06977			

RSquare (U)	0.0536
AICc	647.961
BIC	667.552
Observations (or Sum Wgts)	1000

Fit Details

Parameter Estimates

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
APR	1	1	11.6312003	0.0006*
Fee	1	1	14.6371313	0.0001*
APR*Fee	1	1	24.7452952	<.0001*

47

How to Analyze Response Rate as a Dependent Variable (Contd.)

- With large sample size, you may use Response (Numeric)
- JMP > Analyze > Fit Model > Move Response (Numeric) to Y > > Control-Click both APR and Fee, then click macros drop-down arrow and select Full Factorial > Change Emphasis to Minimal Report > Run > then do LSMeans plot, Tukey's HSD etc.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	2.676000	0.892000	9.6476
Error	996	92.088000	0.092458	Prob > F
C. Total	999	94.764000		<.0001*

Parameter Estimates

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
APR	1	1	0.4000000	4.3263	0.0378*
Fee	1	1	0.6760000	7.3114	0.0070*
APR*Fee	1	1	1.6000000	17.3052	<.0001*

APR*Fee

Least Squares Means Table

Level	Least Sq Mean	Std Error
Control,Control	0.02000000	0.01923100
Control,Test	0.15200000	0.01923100
Test,Control	0.14000000	0.01923100
Test,Test	0.11200000	0.01923100

Least Squares Means Plot

48

Higher Order Factorial Designs

- The procedure and method covered in this session can easily be extended to the following:
  - Two-factor designs where each factor has *three or more levels*
    - Use **follow-up tests** (as we did in the single factor ANOVA) to explore whether the means of the levels of a factor are different from each other.
- *Multi-factor* (3 or more factors each with 2 or more levels) design issues
  - Start with testing the **highest order** interaction (for three-factor it is the three-way interaction). If that interaction is significant, then tell your story based on that interaction.
  - It may be *misleading to interpret the lower order interactions or main effects* when the higher order interaction is significant.
  - The sample size requirements go up very quickly as we consider multi-factor designs. One way, to get around this issue is by using special designs such as "*fractional factorials*."

49