

# A/B Testing and Beyond

## Designed Experiments for Data Scientists



# Week 6

Wednesday October 11<sup>th</sup>, 2017



# Outline

- Analyzing Factorial Experiments
  - Continuous Responses
  - Binary Responses
- Two-Level Factorial Experiments
  - $2^k$  Factorial Experiments
  - $2^{k-p}$  Fractional Factorial Experiments



# RECAP



# Recap

- Primer on linear regression
- Experiments with Multiple Conditions
  - Comparing means
  - Comparing proportions
  - The multiple comparison problem
- Experiments with Multiple Factors
  - Factorial vs. One-factor-at-a-time
  - Designing a factorial experiment



# Multivariate Experiments

## Designing a Factorial Experiment

The design is conceptually simple:

- Pick your design factors
- Pick their levels
- Your experimental conditions are all of the different combinations of these factors' levels

If you have  $k$  factors with  $m_1, m_2, \dots, m_k$  levels, respectively, the corresponding factorial experiment will have

$$M = m_1 m_2 \cdots m_k$$

experimental conditions



# Multivariate Experiments

## Designing a Factorial Experiment

Once units have been assigned to each condition, the response variable is measured on all of them

Using the collected data we

- (1) Identify which factors are influential, and
- (2) Identify which combination of factors is optimal

To do (1) we will apply regression techniques

To do (2) we will use two sample  $t$ -,  $Z$ - or  $\chi^2$ -tests



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

We discuss these concepts in the context of the following example:

Suppose, again, Instagram is experimenting with ads to understand their influence on user engagement.

Again we assume the response variable ( $Y$ ) is session duration (measured in minutes)

But now we assume we have two design factors





# Multivariate Experiments

Analyzing a Factorial Experiment – Continuous  $Y$

Factor 1: Ad Frequency

- None (coded as 0)
- 7:1 (coded as 1)
- 4:1 (coded as 2)
- 1:1 (coded as 3)

Factor 2: Ad Type

- Photo (coded as 1)
- Video (coded as 2)



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

### Factor 1: Ad Frequency

- None (coded as 0)
- 7:1 (coded as 1)
- 4:1 (coded as 2)
- 1:1 (coded as 3)

### Factor 2: Ad Type

- Photo (coded as 1)
- Video (coded as 2)

This leads to  $4 \times 2 = 8$   
unique conditions

Assume we randomize  
 $n=1000$  units to each  
and measure  $Y$



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

Frequency: None  
Type: Photo

Frequency: None  
Type: Video

Frequency: 7:1  
Type: Photo

Frequency: 7:1  
Type: Video

Frequency: 4:1  
Type: Photo

Frequency: 4:1  
Type: Video

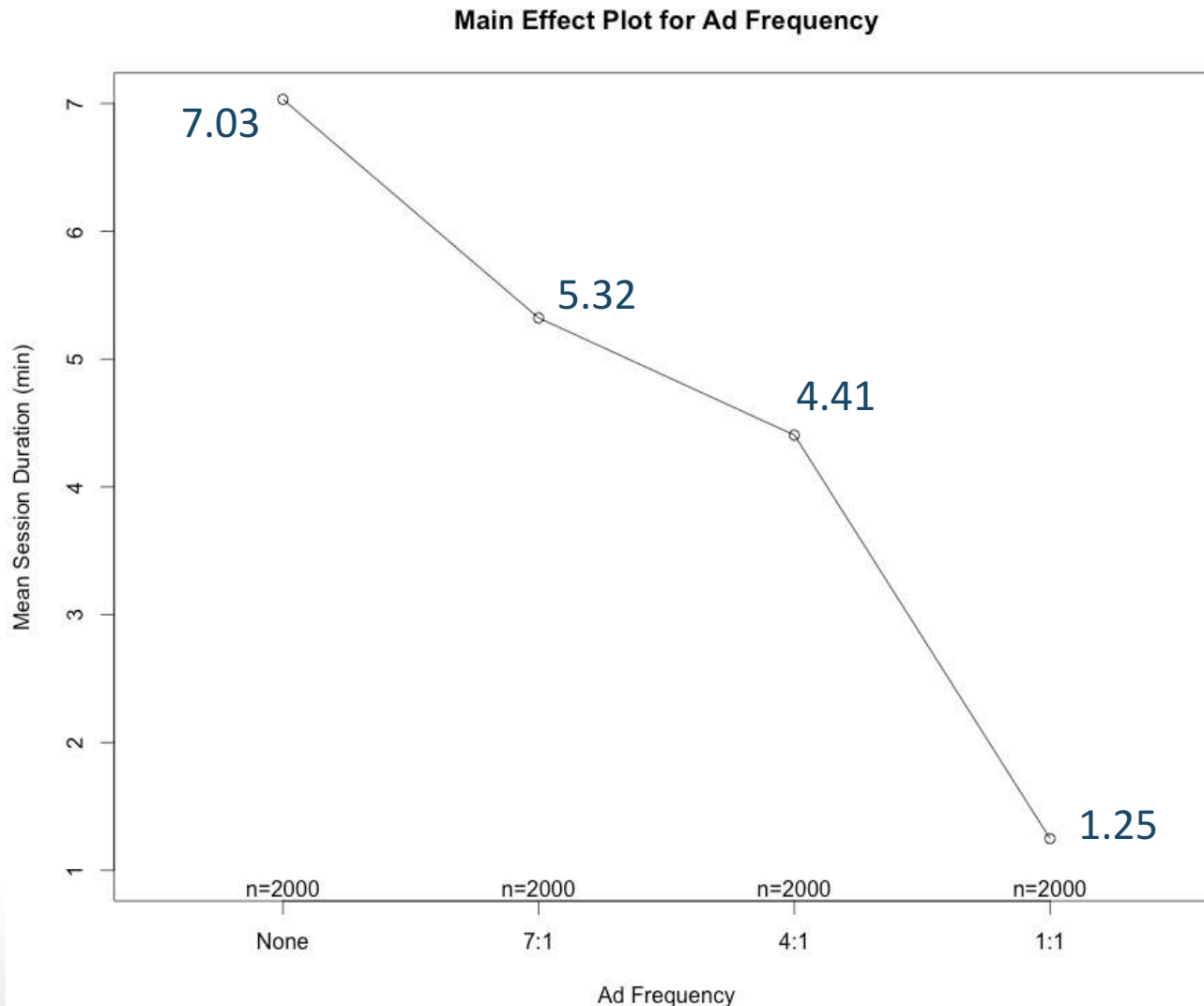
Frequency: 1:1  
Type: Photo

Frequency: 1:1  
Type: Video



# Multivariate Experiments

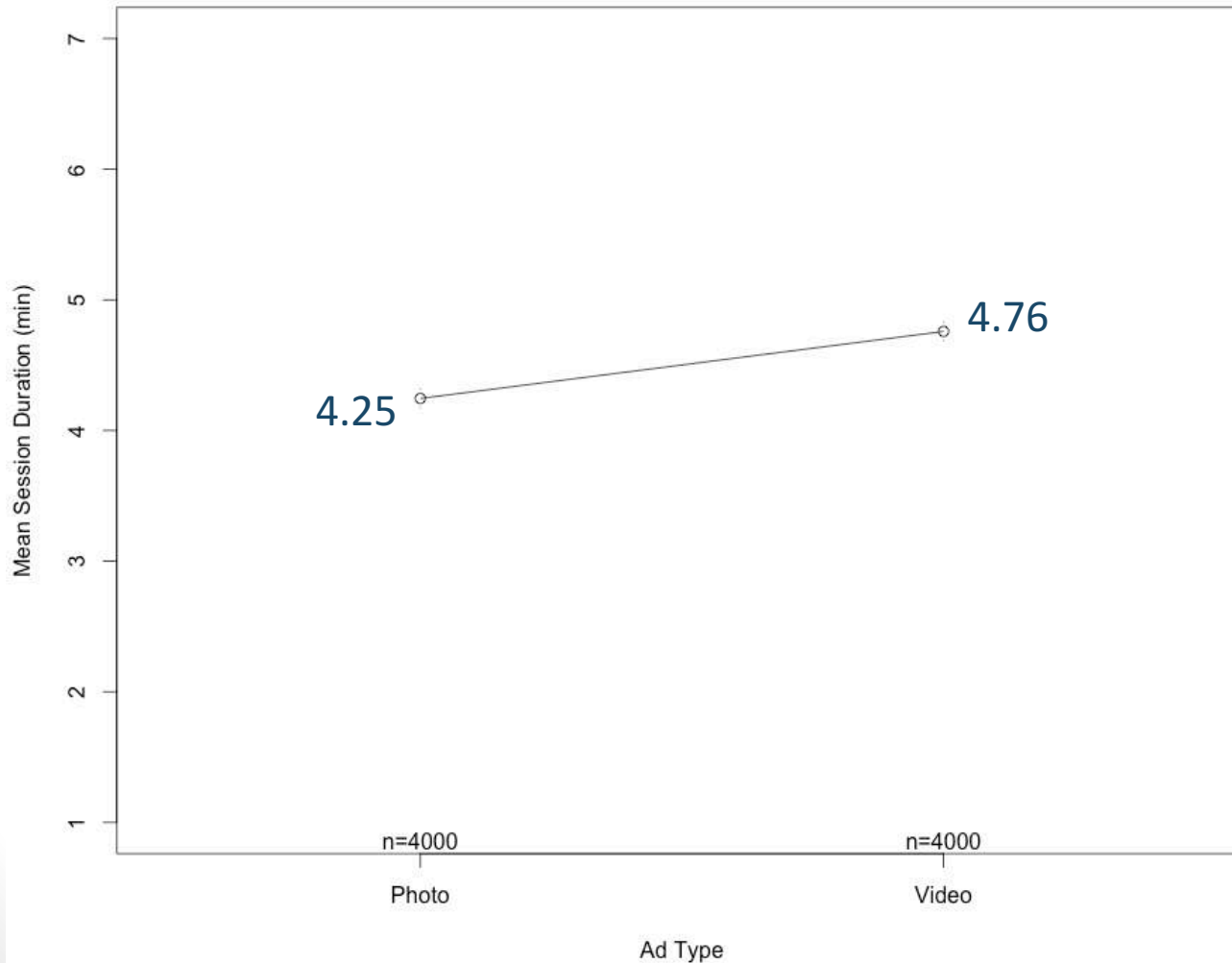
## Analyzing a Factorial Experiment – Continuous $Y$



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

Main Effect Plot for Ad Type



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

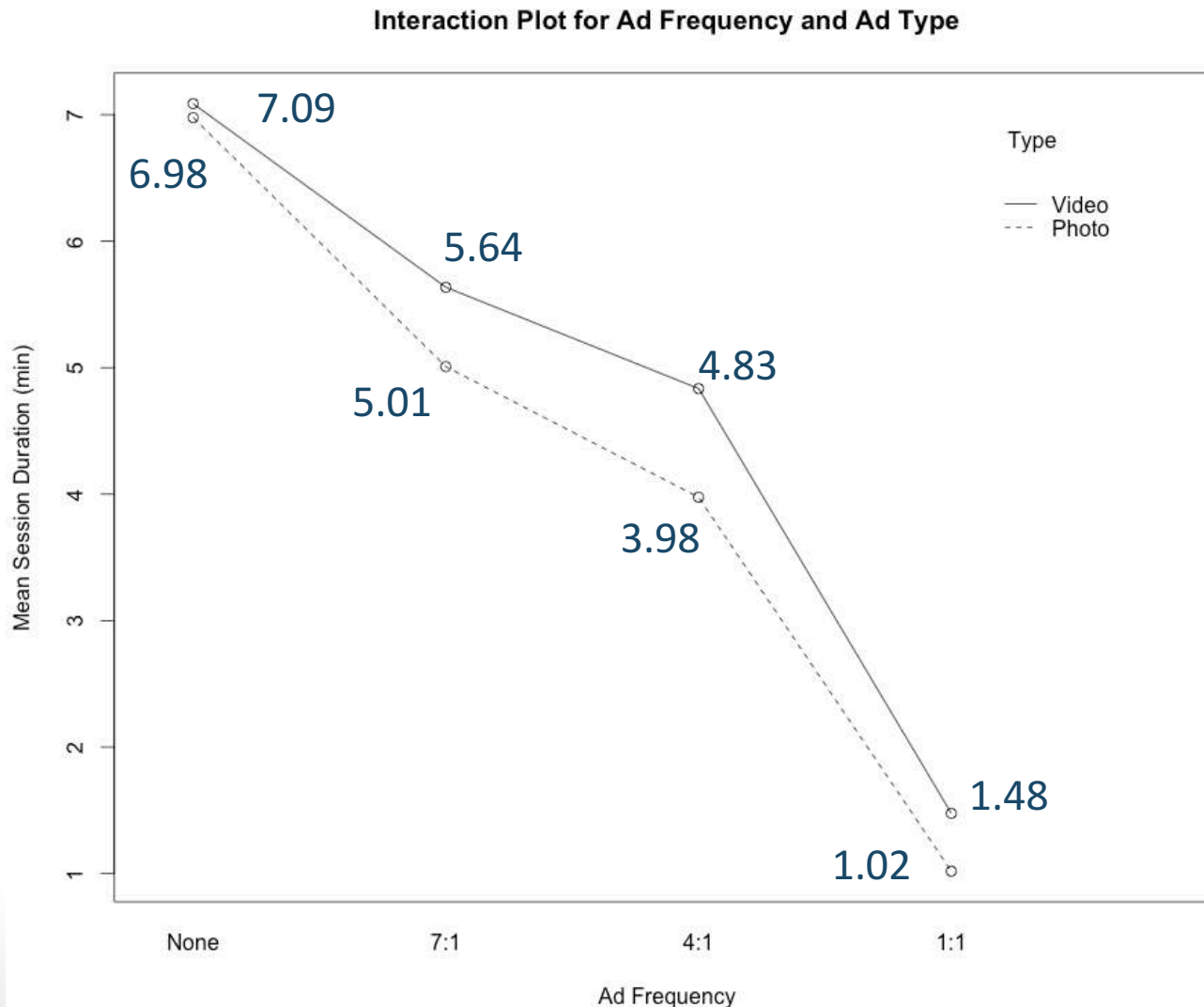
The main effect plots tell us:

- Session duration decreases as ad frequency increases
- Session duration is slightly longer for video ads vs. photo ads
- The influence of ad frequency is larger than the influence of ad type



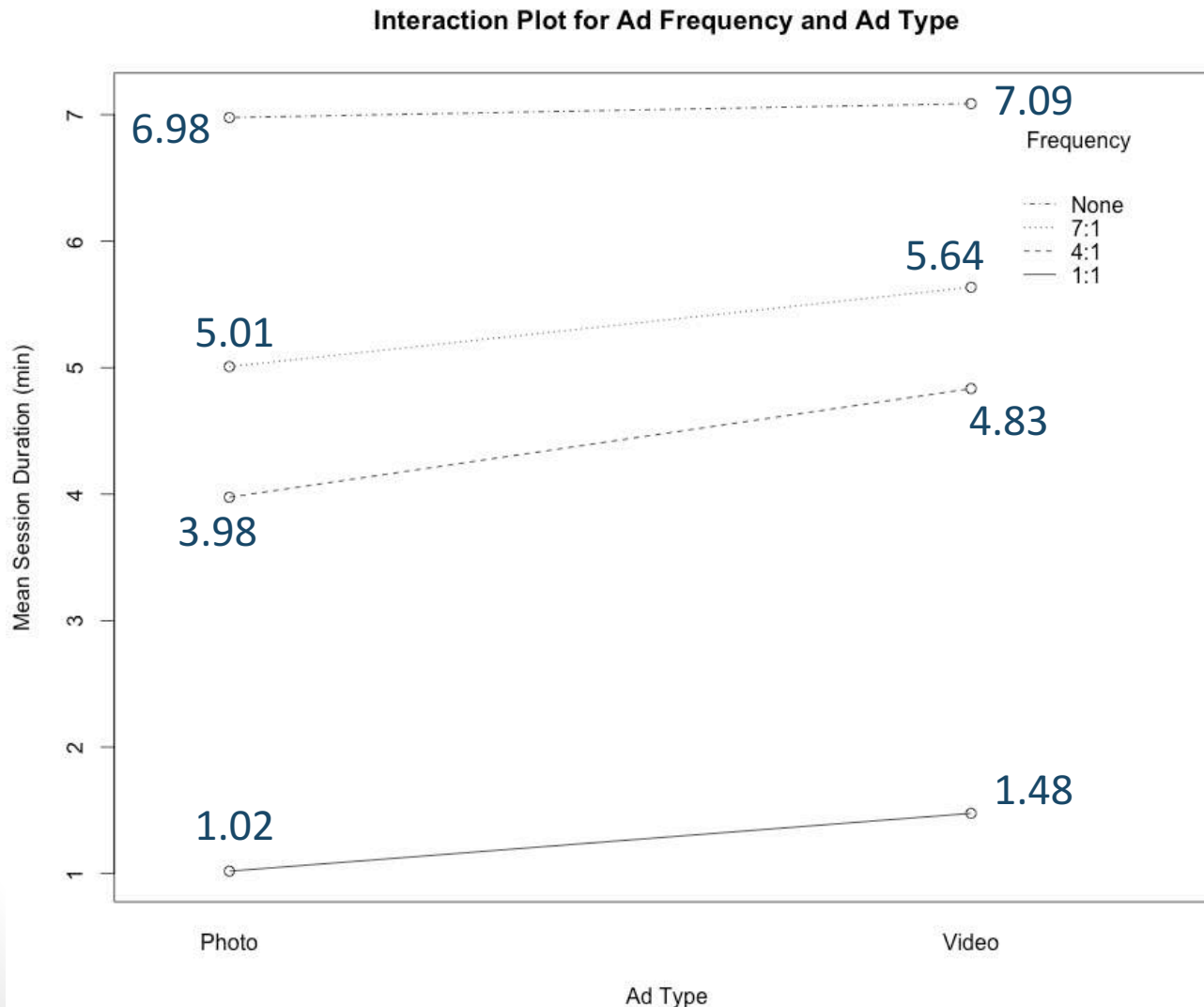
# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$





# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

The **interaction effect** plots tell us:

- The effect of ad frequency is not quite the same for both ad types
- The effect of ad type is not quite the same for all ad frequencies
- Thus an interaction is present

To formally decide whether the main and interaction effects are significant, we use **linear regression**



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

Linear regression models used for this purpose should contain

- Indicator variables for each factor; the number of indicators for a particular factor is equal to the number of levels of that factor, minus 1.
  - This allows us to evaluate main effects
- $k$ -way products of the indicator variables for the  $k$  different factors.
  - This allows us to evaluate interaction effects



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

The linear regression model appropriate for the Instagram factorial example is

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} \\ + \beta_5 x_{i1} x_{i4} + \beta_6 x_{i2} x_{i4} + \beta_7 x_{i3} x_{i4} + \epsilon_i$$

where

- $x_{i1} = 1$  if unit  $i$  is in the 7:1 condition
- $x_{i2} = 1$  if unit  $i$  is in the 4:1 condition
- $x_{i3} = 1$  if unit  $i$  is in the 1:1 condition
- $x_{i4} = 1$  if unit  $i$  is in the video condition



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

Main effects become irrelevant in the context of interaction, and so it is common practice to first decide whether the interaction effect is significant

Note that  $\beta_5 = \beta_6 = \beta_7 = 0$  removes the interaction terms from the model and so a test of

$$H_0: \beta_5 = \beta_6 = \beta_7 = 0 \text{ vs. } H_A: \beta_j \neq 0$$

formally tests whether the interaction effect is significant for at least one of  $j = 5, 6, 7$



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

If the interaction effect is significant (i.e., we do not reject  $H_0$ ) we must be careful to only draw conclusions regarding the effect of one factor in the context of the levels of the other factor

However, if the interaction effect is not significant (i.e., we do not reject  $\beta_5 = \beta_6 = \beta_7 = 0$ ) we may use the **reduced main effects model**:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \epsilon_i$$

which can be used to evaluate the significance of main effects



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

The expected response, based on this model, in each of the “photo” conditions is shown below.

Frequency	Expected Response
None	$E[Y_i   x_{i1} = x_{i2} = x_{i3} = 0, x_{i4} = 0] = \beta_0$
7:1	$E[Y_i   x_{i1} = 1, x_{i4} = 0] = \beta_0 + \beta_1$
4:1	$E[Y_i   x_{i2} = 1, x_{i4} = 0] = \beta_0 + \beta_2$
1:1	$E[Y_i   x_{i3} = 1, x_{i4} = 0] = \beta_0 + \beta_3$



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

The expected response, based on this model, in each of the “video” conditions is shown below.

Frequency	Expected Response
None	$E[Y_i   x_{i1} = x_{i2} = x_{i3} = 0, x_{i4} = 1] = \beta_0 + \beta_4$
7:1	$E[Y_i   x_{i1} = 1, x_{i4} = 1] = \beta_0 + \beta_1 + \beta_4$
4:1	$E[Y_i   x_{i2} = 1, x_{i4} = 1] = \beta_0 + \beta_2 + \beta_4$
1:1	$E[Y_i   x_{i3} = 1, x_{i4} = 1] = \beta_0 + \beta_3 + \beta_4$



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

- Notice that the expectations in each row are identical if  $\beta_1 = \beta_2 = \beta_3 = 0$
- Thus, ad frequency does not significantly influence the response if  $\beta_1 = \beta_2 = \beta_3 = 0$
- We formally test whether the main effect of ad frequency is significant by testing

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0 \text{ vs. } H_A: \beta_j \neq 0$$

for at least one of  $j = 1, 2, 3$





# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

- Notice that the expected response for photo vs. video ads becomes the same if  $\beta_4 = 0$
- Thus, ad type does not significantly influence the response if  $\beta_4 = 0$
- We formally test whether the main effect of ad type is significant by testing

$$H_0: \beta_4 = 0 \text{ vs. } H_A: \beta_4 \neq 0$$



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

- All of these hypothesis tests correspond to simultaneously setting a subset of the  $\beta$ 's equal to zero
- Thus, each of these tests generates a **reduced model** with fewer terms than the corresponding full model
- In each case we compare the full and reduced models to decide if they seem significantly different – rejecting  $H_0$  if they do
- This is done formally with a **partial  $F$ -test**



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

- The partial  $F$ -test compares the mean squared errors between the full and reduced models (similar to the  $F$ -test for overall significance in a linear regression)
- The test statistics and p-values associated with this test are provided in standard linear regression ANOVA output like `anova()` in R.



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

```
lm(formula = Time ~ Frequency * Type)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.7276	-0.5474	-0.0020	0.5499	4.4332

Coefficients:

	Estimate	Std.Error	t value	Pr(> t )
(Intercept)	6.97785	0.02824	247.104	< 2e-16 ***
Frequency7:1	-1.96929	0.03994	-49.312	< 2e-16 ***
Frequency4:1	-3.00204	0.03994	-75.173	< 2e-16 ***
Frequency1:1	-5.95856	0.03994	-149.206	< 2e-16 ***
TypeVideo	0.10993	0.03994	2.753	0.00592 **
Frequency7:1:TypeVideo	0.51768	0.05648	9.166	< 2e-16 ***
Frequency4:1:TypeVideo	0.74924	0.05648	13.266	< 2e-16 ***
Frequency1:1:TypeVideo	0.34731	0.05648	6.150	8.14e-10 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.893 on 7992 degrees of freedom

Multiple R-squared: 0.8497, Adjusted R-squared: 0.8496

F-statistic: 6455 on 7 and 7992 DF, p-value: < 2.2e-16



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

Analysis of Variance Table

Response: Time

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Frequency	3	35353	11784.3	14778.187	< 2.2e-16 ***
Type	1	527	527.3	661.318	< 2.2e-16 ***
Frequency:Type	3	149	49.8	62.398	< 2.2e-16 ***
Residuals	7992	6373	0.8		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '1'



# Multivariate Experiments

## Analyzing a Factorial Experiment – Continuous $Y$

The p-values in the ANOVA table are sufficiently small so we conclude:

- Ad frequency has a significant main effect
- Ad type has a significant main effect
- The interaction between these factors is also significant

This means that both factors should be considered when trying to optimize session duration.

To determine which condition is optimal we can use a series of pairwise t-tests



# LOGISTIC REGRESSION – A PRIMER



# Logistic Regression

- Linear regression is an effective method of modeling the relationship between a single response variable ( $Y$ ), and one or more explanatory variables ( $x_1, x_2, \dots, x_p$ )
- However, ordinary linear regression is appropriate for continuous response variable – particularly when  $Y \sim N(\mu, \sigma^2)$
- When a response variable is binary, i.e.  $Y \sim \text{BIN}(1, \pi)$ , we use logistic regression instead





# Logistic Regression

- In the context of regression models we call

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

the linear predictor

- In ordinary linear regression we relate the expected response to this linear predictor as follows:

$$E[Y|x_1, x_2, \dots, x_p] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

- But in logistic regression, the expected value of  $Y$  is  $\pi \in [0,1]$ , and it's unrealistic to constrain the linear predictor to the  $[0,1]$  interval



# Logistic Regression

- So instead of equating the linear predictor to  $E[Y]$ , in logistic regression we relate them through a **link function** that maps  $[0,1]$  to the real numbers
- The link function chosen for logistic regression is the **logit function**

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

- Inverting this gives

$$E[Y|x_1, x_2, \dots, x_p] = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p}}$$



# Logistic Regression

- $\beta_0$  is interpreted as the **log-odds** that  $Y = 1$  when  $x_1 = x_2 = \dots = x_p = 0$
- Thus  $e^{\beta_0}$  is interpreted as the **odds** that  $Y = 1$  (i.e., the odds that a unit performs the action of interest) when  $x_1 = x_2 = \dots = x_p = 0$
- It can also be shown that  $\beta_j$ , for  $j = 1, 2, \dots, p$ , is a **log-odds ratio**
- Specifically,  $e^{\beta_j}$  is interpreted as the **odds ratio** comparing the odds that  $Y = 1$  when  $x_j = x$  vs. when  $x_j = x + 1$



# Logistic Regression

- Parameter estimation in a logistic regression model is typically carried out use **maximum likelihood estimation**
- As a consequence tests for individual regression coefficients, such as

$$H_0: \beta_j = 0 \text{ vs. } H_A: \beta_j \neq 0$$

are carried out using **Z-tests**

- To test whether several  $\beta$ 's are simultaneously zero, and hence to compare the **full** model to a **reduced** one we use **likelihood ratio tests**



# Logistic Regression

- The test statistic for the likelihood ratio test is referred to as the **deviance** and is defined as

$$\begin{aligned}\Lambda &= 2[\log \text{likelihood}(\text{full model}) \\ &\quad - \log \text{likelihood}(\text{reduced model})] \\ &= 2\log \left( \frac{\text{likelihood}(\text{full model})}{\text{likelihood}(\text{reduced model})} \right)\end{aligned}$$

- If  $H_0$  is true (i.e., there is no difference between the two models) this approximately follows a  $\chi^2$  distribution with  $l$  degrees of freedom
- $l$  is the number of restrictions imposed by  $H_0$



# Logistic Regression

- In the previous example the design of the experiment dictated the following linear predictor

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ + \beta_5 x_1 x_4 + \beta_6 x_2 x_4 + \beta_7 x_3 x_4$$

- And a test of

$$H_0: \beta_5 = \beta_6 = \beta_7 = 0$$

was used to formally evaluate whether the interaction effect is significant



# Logistic Regression

- If the design of the experiment was identical, except the response variable was binary, the logistic regression model would be

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} \\ + \beta_4 x_{i4} + \beta_5 x_{i1} x_{i4} + \beta_6 x_{i2} x_{i4} + \beta_7 x_{i3} x_{i4}$$

and

$$H_0: \beta_5 = \beta_6 = \beta_7 = 0$$

would be tested using a likelihood ratio test whose null distribution is the  $\chi^2$  distribution with  $l = 3$  degrees of freedom, since  $H_0$  involves 3  $\beta$ 's



# Logistic Regression

- Any easy way to remember the degrees of freedom in a likelihood ratio test is

$$l = p - q$$

where  $p$  is the number of regression coefficients in the full model and  $q$  is the number of regression coefficients in the reduced model

- We will see an application of logistic regression to analysis of factorial experiments in the context of a two-level factorial experiment





# TWO-LEVEL FACTORIAL EXPERIMENTS



# TWO-LEVEL FACTORIAL EXP'S

- Factorial experiments are an **informative** and **efficient** means of exploring several factors
- **Advantage:** every possible combination of factor levels is tested
  - We do not risk missing an optimal combination
- **Disadvantage:** every possible combination of factor levels is tested
  - Such experiments can become very large
- **Compromise:** two-level factorial experiments



# TWO-LEVEL FACTORIAL EXP'S

## Two-Level Factorial Experiments

- When investigating  $k$  factors, two-level factorial experiments are the smallest possible factorial experiments
- Such experiments are typically used for factor screening
- Pareto Principle: only *a vital few* factors are important relative to the *trivial many*
- The purpose of a screening experiment is to identify this small number of influential factors



# TWO-LEVEL FACTORIAL EXP'S

## Two-Level Factorial Experiments

Here we discuss two particular types of two-level factorial experiments for investigating  $k$  factors:

- $2^k$  factorial experiments
  - These investigate each of the unique  $2^k$  conditions
- $2^{k-p}$  fractional factorial experiments
  - These investigate just a *fraction* of the unique  $2^k$  conditions



# $2^k$ FACTORIAL EXPERIMENTS

## Designing $2^k$ Factorial Experiments

**Step 1:** Choose  $k$  factors that are expected to influence the response in some way

**Step 2:** Choose two levels for each factor to experiment with

- It's important to choose levels that provide the largest opportunity for an influential factor to be noticed
- Levels should be chosen that are quite different from one another; even a very influential factor may not appear to be influential if the factor levels are too similar.



# $2^k$ FACTORIAL EXPERIMENTS

## Designing $2^k$ Factorial Experiments

**Step 3:** The experimental conditions are defined to be the unique combination of these factors' levels

- There will be  $2^k$  of them

**Step 4:** Assign experimental units to each of the  $2^k$  conditions

- For ease of notation, we assume that the experiment is balanced and  $n$  units are assigned to each condition
- The sample size  $n$  can be determined by power analyses based on two-sample tests that account for the multiple comparison problem



# $2^k$ FACTORIAL EXPERIMENTS

## Analyzing $2^k$ Factorial Experiments

Analysis in this context is performed using regression:

- Linear regression (when  $Y$  is continuous)
- Logistic regression (when  $Y$  is binary)

In either case the linear predictor of the model has

- $\binom{k}{1} = k$  main effect terms
- $\binom{k}{2}$  two-factor interaction terms
- $\vdots$
- $\binom{k}{k} = 1$   $k$ -factor interaction term



# $2^k$ FACTORIAL EXPERIMENTS

## Analyzing $2^k$ Factorial Experiments

The linear predictor associated with a  $2^3$  factorial experiment is

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3$$

For two-level designs we code the levels of the  $x$ 's using  $\pm 1$ 's where -1 corresponds to the *low* level of the factor and +1 corresponds to the *high* level

Note that when a factor is categorical, the labels *low* and *high* are arbitrary





# $2^k$ FACTORIAL EXPERIMENTS

## Analyzing $2^k$ Factorial Experiments

But when a factor is numeric, the *low* and *high* levels correspond to the small and large values of the factor

- In this case the translation from **natural units** to **coded units** can be done using the following formula

$$x_C = \frac{x_N - (x_H + x_L)/2}{(x_H - x_L)/2}$$

where  $x_H$  and  $x_L$  correspond to the high and low values of the factor as recorded in the natural units and  $x_N$  is any value of the factor in the natural units



# $2^k$ FACTORIAL EXPERIMENTS

## Analyzing $2^k$ Factorial Experiments

- Using the  $\pm 1$  coding, each experimental condition can be identified by a unique combination of plus and minus ones
- The design of the experiment can be displayed in what is known as a **design matrix**
- Consider the design matrix for a  $2^3$  factorial experiment
  - Rows correspond to unique conditions
  - Columns indicate, for a given condition, which level each factor should be set at



# $2^k$ FACTORIAL EXPERIMENTS

## Analyzing $2^k$ Factorial Experiments

Condition	Factor 1	Factor 2	Factor 3
1	-1	-1	-1
2	+1	-1	-1
3	-1	+1	-1
4	+1	+1	-1
5	-1	-1	+1
6	+1	-1	+1
7	-1	+1	+1
8	+1	+1	+1

- Design matrices provide a prescription for running  $2^k$  factorial experiments



# $2^k$ FACTORIAL EXPERIMENTS

## Analyzing $2^k$ Factorial Experiments

- Using the data collected from such a study we fit a linear or logistic regression model
- Apart from this difference, the models are similar in that
  - they are based on exactly the same linear predictor
  - we can evaluate the significance of main and interaction effects by performing tests concerning individual or multiple  $\beta$ 's (although the specific tests that are used differ in the two settings)



# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

Montgomery (2017) presents an example in which a credit card company ran a  $2^4$  factorial experiment to test new ideas of improving conversion rates on credit card offers.

Factor	Low (-)	High (+)
Annual fee ( $x_1$ )	Current	Lower
Account-opening fee ( $x_2$ )	No	Yes
Initial interest rate ( $x_3$ )	Current	Lower
Long-term interest rate ( $x_4$ )	Low	High

The  $2^4=16$  combinations of these levels defined 16 credit card offers which were each sent to  $n = 7500$  people



Condition	$x_1$	$x_2$	$x_3$	$x_4$	Sign-ups	Conversion Rate
1	-1	-1	-1	-1	184	2.45%
2	+1	-1	-1	-1	252	3.36%
3	-1	+1	-1	-1	162	2.16%
4	+1	+1	-1	-1	172	2.29%
5	-1	-1	+1	-1	187	2.49%
6	+1	-1	+1	-1	254	3.39%
7	-1	+1	+1	-1	174	2.32%
8	+1	+1	+1	-1	183	2.44%
9	-1	-1	-1	+1	138	1.84%
10	+1	-1	-1	+1	168	2.24%
11	-1	+1	-1	+1	127	1.69%
12	+1	+1	-1	+1	140	1.87%
13	-1	-1	+1	+1	172	2.29%
14	+1	-1	+1	+1	219	2.92%
15	-1	+1	+1	+1	153	2.04%
16	+1	+1	+1	+1	152	2.03%



# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

Using the data we fit a logistic regression model with the following linear predictor:

$$\begin{aligned} &\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 \\ &+ \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 \\ &+ \beta_{34} x_3 x_4 + \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 \\ &+ \beta_{134} x_1 x_3 x_4 + \beta_{234} x_2 x_3 x_4 + \beta_{1234} x_1 x_2 x_3 x_4 \end{aligned}$$

We do so using the `glm()` function in R with a logit link function. The model summary is shown on the next slide.



## Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-3.739697	0.019342	-193.347	< 2e-16	***
x1	0.080845	0.019342	4.180	2.92e-05	***
x2	-0.106211	0.019342	-5.491	3.99e-08	***
x3	0.058248	0.019342	3.011	0.00260	**
x4	-0.108086	0.019342	-5.588	2.29e-08	***
x1:x2	-0.055164	0.019342	-2.852	0.00434	**
x1:x3	-0.004794	0.019342	-0.248	0.80426	
x2:x3	-0.006967	0.019342	-0.360	0.71868	
x1:x4	-0.013178	0.019342	-0.681	0.49566	
x2:x4	0.010625	0.019342	0.549	0.58280	
x3:x4	0.038079	0.019342	1.969	0.04899	*
x1:x2:x3	-0.009646	0.019342	-0.499	0.61799	
x1:x2:x4	0.010629	0.019342	0.550	0.58265	
x1:x3:x4	-0.002543	0.019342	-0.131	0.89539	
x2:x3:x4	-0.020946	0.019342	-1.083	0.27885	
x1:x2:x3:x4	-0.009496	0.019342	-0.491	0.62347	

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '1'

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26854 on 119999 degrees of freedom

Residual deviance: 26741 on 119984 degrees of freedom

AIC: 26773

Number of Fisher Scoring iterations: 6





# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

- All main effects are significant
- Two two-factor interactions are significant
- All three-factor interactions are insignificant
- All four-factor interactions are insignificant

This suggests that a reduced model with

$$\begin{aligned}\beta_{13} = \beta_{14} = \beta_{23} = \beta_{24} = \beta_{123} = \beta_{124} = \beta_{134} \\ = \beta_{234} = \beta_{1234} = 0\end{aligned}$$

may be appropriate. The summary of this model is shown on the next slide



## Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.73961	0.01934	-193.316	< 2e-16 ***
x1	0.08214	0.01920	4.279	1.88e-05 ***
x2	-0.10834	0.01920	-5.644	1.66e-08 ***
x3	0.05886	0.01916	3.072	0.00212 **
x4	-0.11068	0.01916	-5.777	7.61e-09 ***
x1:x2	-0.05706	0.01920	-2.972	0.00296 **
x3:x4	0.04051	0.01916	2.115	0.03447 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 26854 on 119999 degrees of freedom

Residual deviance: 26744 on 119993 degrees of freedom

AIC: 26758

Number of Fisher Scoring iterations: 6



# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

- All main effects are significant
- Two two-factor interactions are significant
- All three-factor interactions are insignificant
- All four-factor interactions are insignificant

This suggests that a reduced model with

$$\begin{aligned}\beta_{13} = \beta_{14} = \beta_{23} = \beta_{24} = \beta_{123} = \beta_{124} = \beta_{134} \\ = \beta_{234} = \beta_{1234} = 0\end{aligned}$$

may be appropriate. The summary of this model is shown on the next slide



# $2^k$ FACTORIAL EXPERIMENTS

Example: Conversion on Credit Card Offers

The formal test of

$$H_0: \beta_{13} = \beta_{14} = \beta_{23} = \beta_{24} = \beta_{123} = \beta_{124} \\ = \beta_{134} = \beta_{234} = \beta_{1234} = 0$$

and hence a comparison of the reduced model to the full model involves a likelihood ratio test using

- $\log \text{likelihood}(\text{full}) = -13370.45$
- $\log \text{likelihood}(\text{reduced}) = -13371.91$

Which give a deviance statistic of  $\Lambda = 2.9244$  and p-value of

$$P(T \geq 2.9244) = 0.9672$$

where  $T \sim \chi^2_{(9)}$



# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

- Thus we do not reject  $H_0$  which implies that the reduced model is adequate
- Therefore, only the main effects and two of the two-factor interactions are significant
- The next two slides provide main effect and interaction effect plots for each of the significant effects



# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

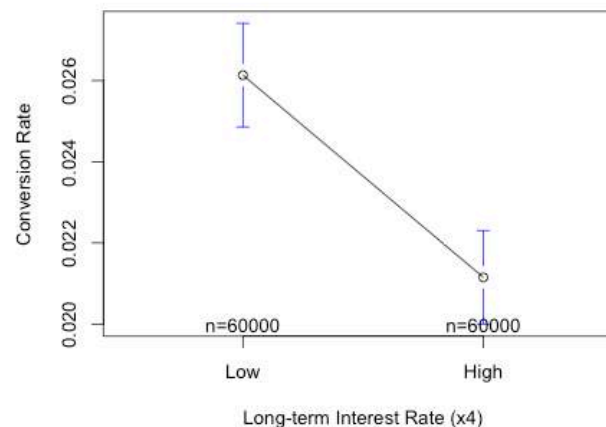
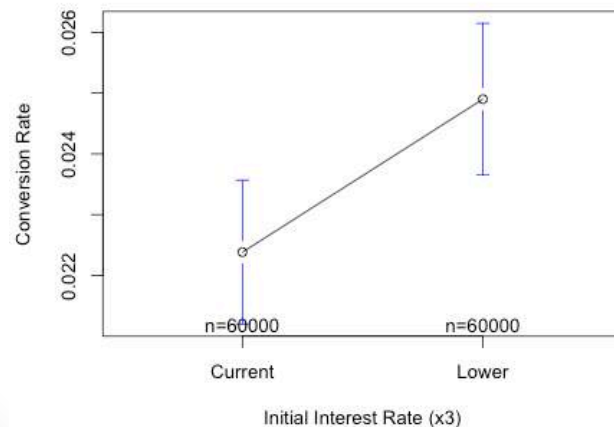
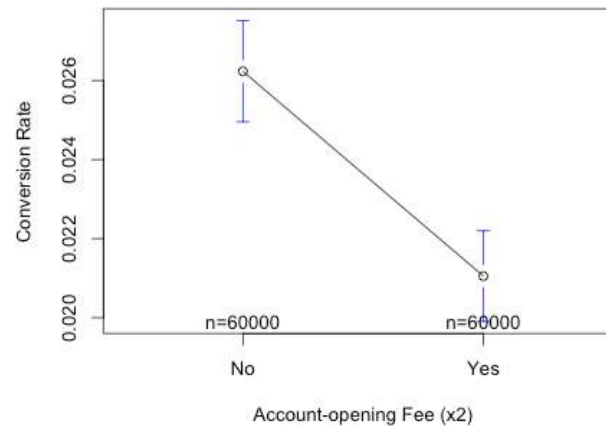
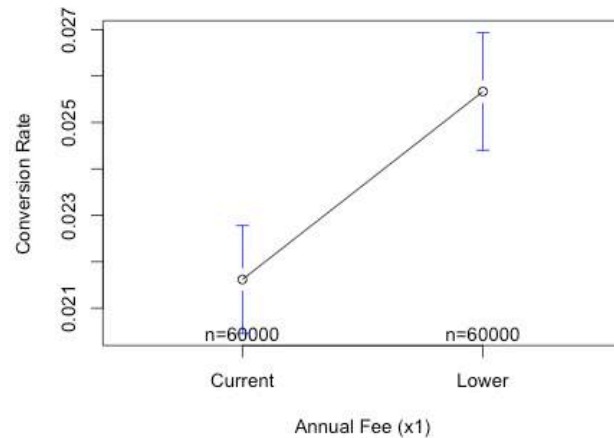
- The main effect plots suggest that:
  - lower annual fees, no account-opening fee, lower initial interest rates and lower long-term interest rates are all associated with increased conversion rates
- The interaction effect plots suggest that:
  - lower annual fees are associated with high conversion rates when there is no account-opening fee; when there is an account opening fee, the annual fee is not as influential
  - as long as the long-term interest rate is low, the initial interest rate doesn't really matter



# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

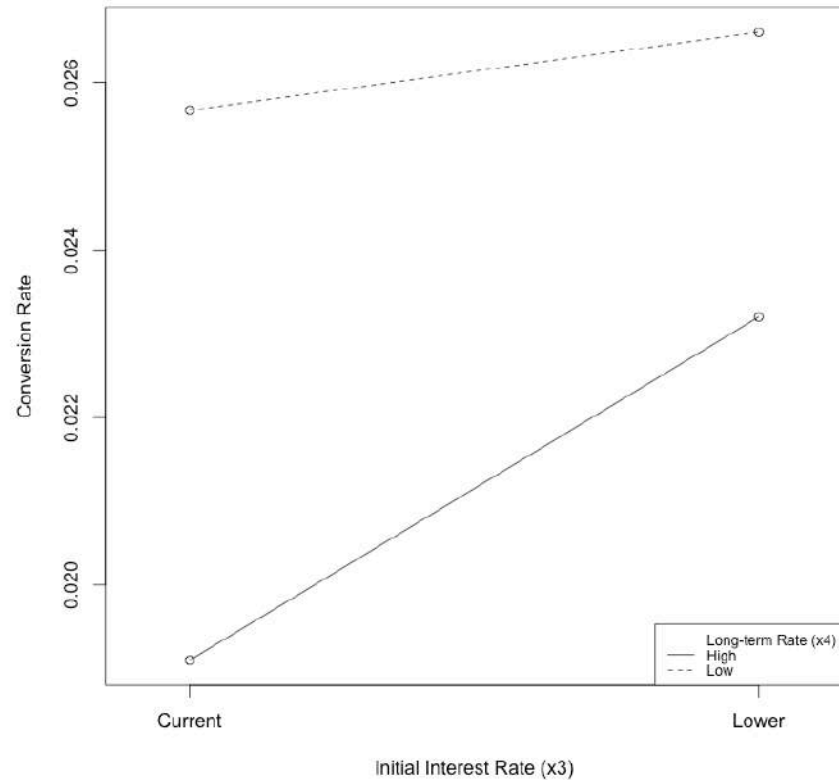
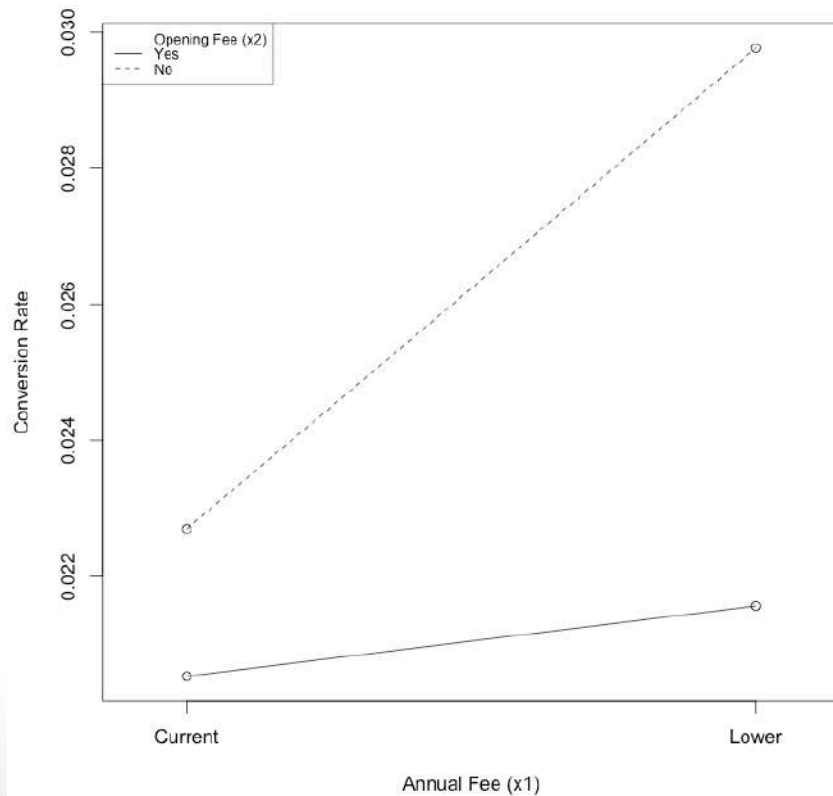
Main Effect Plots



# $2^k$ FACTORIAL EXPERIMENTS

## Example: Conversion on Credit Card Offers

Interaction Plots





# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- $2^k$  factorial experiments are a useful special case of a general factorial experiment
  - They minimize the number of levels being investigated, and hence reduces the overall number of experimental conditions
- BUT they still investigate **all possible** combinations of the factor levels – which can be a lot!
  - With  $k = 8$  factors the  $2^k$  factorial experiment has 256 conditions



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- Alternatively we could use a  $2^{k-p}$  fractional factorial experiment which also investigates  $k$  factors, but with just a fraction of the conditions
- Rather than performing  $2^k$  conditions, we perform  $2^{k-p}$  specially selected conditions which still allow us to estimate main effects and potentially important interaction effects
- **With these experiments we can investigate a relatively large number of factors with a relatively small number of conditions**
- However, we sacrifice the ability to separately estimate *all* main and interaction effects



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

**Motivation:** the linear predictor for a  $2^k$  factorial experiment consists of

- $\binom{k}{1} = k$  main effect terms
- $\binom{k}{2}$  two-factor interaction terms
- $\vdots$
- $\binom{k}{k} = 1$   $k$ -factor interaction term

That's a total of  $\sum_{i=1}^k \binom{k}{i} = 2^k - 1$  terms



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- Of these  $2^k - 1$  terms, only  $k + \binom{k}{2}$  of them are main effects and two factor interactions – the remaining correspond to higher order interaction terms
- If  $k = 8$ , there are 8 main effects, 28 two-factor interactions and 219 higher order interactions, many of which are likely to be insignificant



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- Principle of effect sparsity: in the presence of several factors, variation in the response is likely to be driven by a **small number of main effects and low-order interactions**
- Thus, it is typically a waste of resources to estimate these higher order interaction terms
- It is a better use of these resources is to estimate the main effects and low-order interactions of a larger number of factors
- So how do we do this?



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

First let's discuss  $p$ :

- Investigating  $k$  factors in a full factorial experiments takes  $2^k$  conditions
- If we'd like to investigate  $k$  factors in **half as many** conditions, we use a  $2^{k-1}$  experiment
- If we'd like to investigate  $k$  factors with just a **quarter of** the conditions, we use a  $2^{k-2}$  experiment
- In general, if we'd like to investigate  $k$  factors in  **$(1/2)^p$  as many conditions**, we use a  $2^{k-p}$  experiment



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- If a full factorial approach requires  $2^k$  conditions and we only want  $2^{k-p}$ , we need to choose **which**  $2^{k-p}$  conditions to experiment with
- For instance, if  $k = 5$  and  $p = 2$ , then the goal is to investigate 5 factors in  $2^3 = 8$  conditions (where normally 32 conditions would be required with the full factorial approach).
- The question is, among these 32 conditions, which 8 do we choose to for the  $2^{5-2}$  fractional design?



Condition	A	B	C	D	E
1	-1	-1	-1	-1	-1
2	+1	-1	-1	-1	-1
3	-1	+1	-1	-1	-1
4	+1	+1	-1	-1	-1
5	-1	-1	+1	-1	-1
6	+1	-1	+1	-1	-1
7	-1	+1	+1	-1	-1
8	+1	+1	+1	-1	-1
9	-1	-1	-1	+1	-1
10	+1	-1	-1	+1	-1
11	-1	+1	-1	+1	-1
12	+1	+1	-1	+1	-1
13	-1	-1	+1	+1	-1
14	+1	-1	+1	+1	-1
15	-1	+1	+1	+1	-1
16	+1	+1	+1	+1	-1





Condition	A	B	C	D	E
17	-1	-1	-1	-1	+1
18	+1	-1	-1	-1	+1
19	-1	+1	-1	-1	+1
20	+1	+1	-1	-1	+1
21	-1	-1	+1	-1	+1
22	+1	-1	+1	-1	+1
23	-1	+1	+1	-1	+1
24	+1	+1	+1	-1	+1
25	-1	-1	-1	+1	+1
26	+1	-1	-1	+1	+1
27	-1	+1	-1	+1	+1
28	+1	+1	-1	+1	+1
29	-1	-1	+1	+1	+1
30	+1	-1	+1	+1	+1
31	-1	+1	+1	+1	+1
32	+1	+1	+1	+1	+1



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- To answer this, we consider an extended version of the design matrix associated with a full  $2^3$  factorial experiment

Condition	A	B	C	AB	AC	BC	ABC
1	-1	-1	-1	+1	+1	+1	-1
2	+1	-1	-1	-1	-1	+1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- When it comes to fitting a regression model, each column in this matrix is used to estimate the corresponding effect a particular effect
- For instance:
  - the AB column is used to estimate  $\beta_{AB}$ , the interaction effect between factors A and B
  - the ABC column is used to estimate  $\beta_{ABC}$ , the interaction effect between factors A, B and C
- Now recall the effect sparsity principle: if an interaction is likely to be negligible, why not use its column to dictate the levels of an extra factor?



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

For example:

- Let's use the  $\pm 1$ 's in the ABC column as a prescription for when to run D at its low and high levels
- Let's use the  $\pm 1$ 's in the BC column as a prescription for when to run E at its low and high levels



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

Condition	A	B	C	AB	AC	E=BC	D=ABC
1	-1	-1	-1	+1	+1	+1	-1
2	+1	-1	-1	-1	-1	+1	+1
3	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	-1	+1	-1	-1	-1
5	-1	-1	+1	+1	-1	-1	+1
6	+1	-1	+1	-1	+1	-1	-1
7	-1	+1	+1	-1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1

- Here we say that D and ABC are **aliased** and E and BC are **aliased**
- 'D=ABC' and 'E=BC' are called the **design generators**



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- When two terms are aliased, their effects become **confounded**
- For instance the D=ABC column estimates the ABC interaction **and** the main effect of D
- So the coefficient  $\beta_{ABC}$  quantifies the joint effects of the ABC interaction and the main effect of factor D
- Thus, **we cannot separately estimate** the main effect of D from the ABC interaction effect



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- Thus confounding results from aliasing a new main effect with an existing interaction
- As such, it is important to think carefully about **which** interaction to choose as an alias
- It is best to avoid aliasing a new factor with an interaction that is likely to be significant (since separately estimating significant effects is desirable)
- So high order interaction terms (that are unlikely to be significant) are good choices for aliases



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- This notion is quantified by the **resolution** of the fractional factorial design
- A design is of resolution  $R$  if main effects are aliased with interaction effects involving at least  $R - 1$  factors
- In the design we've been discussing main effects are aliased with two- and three-factor interactions
- Thus it is a resolution III experiment denoted by

$$2_{III}^{5-2}$$





# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- In general, higher resolution designs are to be preferred over lower resolution designs.
- For instance, resolution IV and V designs are to be preferred over a resolution III designs
- In these cases main effects will not be confounded with two-factor interactions
- Since two-factor interactions are typically important, it is best if their effects are not confounded with main effects



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Designing $2^{k-p}$ Fractional Factorial Experiments

- The resolution of a fractional factorial experiment is determined by two things:
  1. The degree of fractionation desired (i.e., the size of  $p$  relative to  $k$ )
  2. The design generators chosen for aliasing
- The degree of fractionation is typically determined by resource constraints – how many conditions can you manage?
- Given the degree of fractionation ( $p$ ) we typically choose design generators to maximize resolution



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Analyzing $2^{k-p}$ Fractional Factorial Experiments

- The analysis of these fractional factorial experiments is based on regression models
  - Linear regression (if  $Y$  is continuous)
  - Logistic regression (if  $Y$  is binary)
- In fact, the analysis is not very different from what we saw in the credit card example
  - We perform individual and simultaneous hypothesis tests to compare full and reduced models
  - This allows us to evaluate the significance of various main and interaction effects



# $2^{k-p}$ FRACTIONAL FACTORIAL

## Analyzing $2^{k-p}$ Fractional Factorial Experiments

### The wrinkle:

- Here the effects estimated in these models are confounded with other effects
- So we can't be 100% certain that a given effect is due to say a main effect, or perhaps the interaction it is aliased with
- But, if the resolution is high, we hope that important effects are aliased with high-order interactions (that are likely negligible)
- This provides confidence that significant effects are not due to the high-order interactions



# Take Home Exercises

Using R or Python, formally do the pairwise comparisons to find the optimal condition in each of the examples presented here. Be sure to account for the multiple comparison problem.



See you next week!

