A Very Brief Introduction to Two-Level Designs

- Factorial experiments are the most informative means of exploring several design factors.

- But this may require a larger number of experimental conditions than is practically feasible.

- As a compromise we might consider **two-level factorial experiments**

- Such an experiment is typically used for **factor screening**

- Factoring screening is predicated on the **Pareto Principle**:

  Only a few factors are important

- We will discuss two types of two-level factorial experiments:
  - $\mathbf{2^K}$ **factorial designs**

    K design factors

  - $\mathbf{2^{K-p}}$ **fractional factorial designs**

    K-p design factors

Designing 2^K Factorial Experiments

- 2^K factorial experiments involve K design factors, each at two levels

- These experiments are typically used for factor screening
  - **Primary Goal:** Determine which among the K factors significantly influence the response variable
  - **Secondary Goal:** Determine which combination of levels is optimal

- The design of the experiment involves:
  1. Choose the MOI and response variables
  2. Choose the design factors

     Choose K that you wanna learn about
  3. Choose the levels of the design factors

     ↳ with the goal of factor screening, we want to give influential factors as fair an opportunity as possible to show themselves as being influential.

     ↳ pick levels that are different
  4. Define experimental conditions
  5. Assign n experiment units to each condition

     ↳ balance is not necessarily, it's just notationally convenient

- In two-level experiments we regard the two levels of a factor as *low* and *high* values of that factor

- We represent each factor by a binary variable:

$$x = \begin{cases} -1 & \text{if the factor is at its “low” level} \\ +1 & \text{if the factor is at its “high” level} \end{cases}$$

- With the factor levels coded in this way, each experimental condition can be identified by a unique combination of plus and minus ones

- The experimental design can be completely summarized by the **design matrix**
  - 2^K rows and K columns of plus and minus ones
  - The ± 1 entries are organized such that each row corresponds to a unique condition and the columns correspond to each of the factors
  - The design matrix provides a prescription for running the 2^K factorial experiment
- 2^1 **Example:**
- 2^2 **Example:**
- 2^3 **Example:**

$$\begin{bmatrix} -1 & -1 & -1 & \vdots \\ +1 & -1 & -1 & \\ -1 & +1 & -1 & \\ +1 & +1 & -1 & \\ \vdots & \vdots & +1 & \vdots \\ & & +1 & \\ & & +1 & \\ & & +1 & \end{bmatrix}$$

- 2^K experiments may also be visualized geometrically as K-dimensional hypercubes
  - Vertices correspond to the unique configurations of the K factors' levels, and hence experimental conditions

orthogonality?

- **Examples:**

Intuitive Analysis of 2^K Experiments

- Primary goal of a 2^K factorial experiment is factor screening
  - Interest lies primarily in estimation of main and interaction effects

- The **main effect** of a factor is defined as the expected change produced by changing that factor from its low to its high level

- The **interaction effect** between two factors quantifies the difference between the main effect of one factor at the two levels of the other

conditioned

- **Toy Example:** Factors A and B are investigated in a 2^2 factorial experiment with $n = 3$

Condition	Factor A	Factor B	Response (y)	Average Response ($\overline{y}$)
1	-1	-1	{1,1,2}	4/3
2	$+1$	-1	{3,4,5}	12/3
3	-1	$+1$	{2,1,3}	6/3
4	$+1$	$+1$	{1,2,5}	8/3

- Intuitive estimate of the main effect of A:

$$\widehat{ME}_A = \overline{y}_{A^+} - \overline{y}_{A^-} = \frac{\overline{y}_{A^+\cap B^-} + \overline{y}_{A^+\cap B^+}}{2} - \frac{\overline{y}_{A^-\cap B^-} + \overline{y}_{A^-\cap B^+}}{2}$$

avg. response when A is high

$$= \frac{12/3 + 8/3}{2} - \frac{4/3 + 6/3}{2}$$

- Intuitive estimate of the main effect of B:

$$\widehat{ME}_B = \overline{y}_{B^+} - \overline{y}_{B^-} = \frac{\overline{y}_{A^-\cap B^+} + \overline{y}_{A^+\cap B^+}}{2} - \frac{\overline{y}_{A^-\cap B^-} + \overline{y}_{A^+\cap B^-}}{2} = -\frac{1}{3}$$

expect response goes down by 1/3 when B goes from low to high.

- To evaluate whether factors A and B interact, we should compare the main effect of A when B is at its high level to the main effect of A when B is at its low level

$$\widehat{ME}_{A|B^+} = \overline{y}_{A^+\cap B^+} - \overline{y}_{A^-\cap B^+} = 2/3$$

$$\widehat{ME}_{A|B^-} = \overline{y}_{A^+\cap B^-} - \overline{y}_{A^-\cap B^-} = 8/3$$

since $2/3 \neq 8/3$, we know there's A:B interaction

- The interaction effect is defined as the average difference between the conditional main effects:

$$\begin{aligned}
\widehat{IE}_{AB} &= \frac{\widehat{ME}_{A|B^+}}{2} - \frac{\widehat{ME}_{A|B^-}}{2} \\
\star &= \frac{\widehat{ME}_{B|A^+}}{2} - \frac{\widehat{ME}_{B|A^-}}{2} \\
&= \frac{\overline{y}_{A^+\cap B^+} + \overline{y}_{A^-\cap B^-}}{2} - \frac{\overline{y}_{A^+\cap B^-} + \overline{y}_{A^-\cap B^+}}{2}
\end{aligned}$$

"A:B interaction is the same as B:A interaction"

- If a third factor C were involved, we may define the three-way ABC interaction as

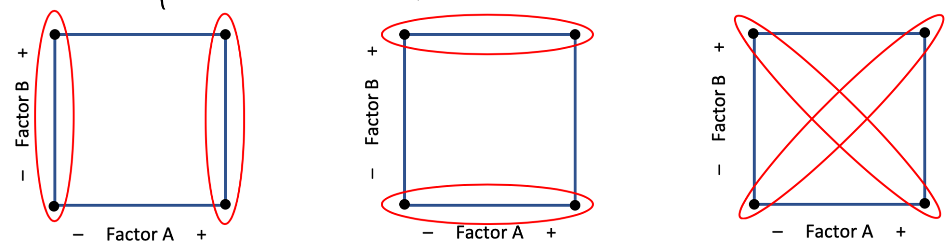
$$\widehat{IE}_{ABC} = \frac{\widehat{IE}_{AB|C^+}}{2} - \frac{\widehat{IE}_{AB|C^-}}{2}$$

$$\star = \frac{\widehat{IE}_{AC|B^+}}{2} - \frac{\widehat{IE}_{AC|B^-}}{2}$$

$$\star = \frac{\widehat{IE}_{BC|A^+}}{2} - \frac{\widehat{IE}_{BC|A^-}}{2}$$

- So what actually happened here?

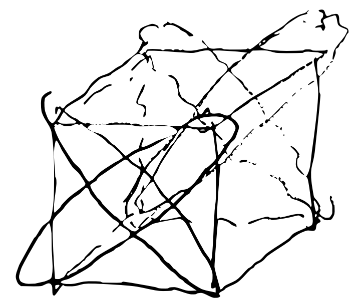
ME. of A. ME of B. IE. of A:B



ME of A: avg. response in the rightmost corners

− avg. . . . − − . leftmost . . − −

IE of A:B:C



- These intuitive comparisons are still relevant when the response variable is binary

geometric averages of "odds" and take ratios

$$\widehat{ME}_A = \sqrt{\frac{\overline{y}_{A+\cap B-}}{1-\overline{y}_{A+\cap B-}} \times \frac{\overline{y}_{A+\cap B+}}{1-\overline{y}_{A+\cap B+}}} \div \sqrt{\frac{\overline{y}_{A-\cap B-}}{1-\overline{y}_{A-\cap B-}} \times \frac{\overline{y}_{A-\cap B+}}{1-\overline{y}_{A-\cap B+}}}$$

$$\widehat{ME}_B = \sqrt{\frac{\overline{y}_{A-\cap B+}}{1-\overline{y}_{A-\cap B+}} \times \frac{\overline{y}_{A+\cap B+}}{1-\overline{y}_{A+\cap B+}}} \div \sqrt{\frac{\overline{y}_{A+\cap B-}}{1-\overline{y}_{A+\cap B-}} \times \frac{\overline{y}_{A-\cap B-}}{1-\overline{y}_{A-\cap B-}}}$$

$$\widehat{IE}_{AB} = \sqrt{\frac{\overline{y}_{A+\cap B+}}{1-\overline{y}_{A+\cap B+}} \times \frac{\overline{y}_{A-\cap B-}}{1-\overline{y}_{A-\cap B-}}} \div \sqrt{\frac{\overline{y}_{A+\cap B-}}{1-\overline{y}_{A+\cap B-}} \times \frac{\overline{y}_{A-\cap B+}}{1-\overline{y}_{A-\cap B+}}}$$

Regression Analysis of 2^K Experiments

The Model

- Fitted regression models provide an estimate of the **response surface**

  $Y = f(x_1, \ldots, x_K)$ to be approximated

- Each of the K factors is represented by the binary variables

$$x_j = \begin{cases} -1 & \text{if factor } j \text{ is at its "low" level} \\ +1 & \text{if factor } j \text{ is at its "high" level} \end{cases}$$

  for $j = 1, 2, \ldots, K$

- Since each factor is represented by a single term, the linear predictor contains:
  - An intercept β_0
  - K main effect terms correspond to $x_1, \ldots, x_K$
  - $\binom{K}{2}$ two-factor interaction terms correspond to $x_1x_2, \ldots, x_ix_j, \ldots$
  - $\binom{K}{3}$ three-factor interaction terms correspond to $x_1x_2x_3$, $x_1x_2x_4, \ldots$
  - $\vdots$
  - $\binom{K}{K} = 1$ K-factor interaction term correspond to $x_1 \times x_2 \times \cdots \times x_K$

  Totally $\sum_{j=0}^{K} \binom{K}{j} = 2^K$ terms in the model

- 2^1 **Example:**

  $\beta_0 + \beta_1 x_1$

- 2^2 **Example:**

  $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$

- 2^3 **Example:**

Estimation

- Estimation of the β's is carried out by
  - Ordinary least squares (in the case of linear regression)
  - Maximum likelihood (in the case of logistic regression)

- In both cases there is a one-to-one connection between the β estimates and the expressions for the main and interaction effects
  - Continuous response:
$$\widehat{\text{Effect}} = 2\widehat{\beta}$$
  - Binary response:
$$\widehat{\text{Effect}} = e^{2\widehat{\beta}}$$
    where β is the regression coefficient correspdong to the effect of interest

- Recall the **Toy Example**:
  - The linear predictor for that experiment is
$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$$

$3 \times 2^2 = 12$ in example

  - The linear regression model is
$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \varepsilon_i \qquad i = 1, 2, \ldots, \underbrace{n2^K}_{\text{total \# of units}}$$
    which can be written in matrix-vector notation as
$$\mathbf{Y} = X\boldsymbol{\beta} + \varepsilon$$
    where

12×1 $\quad 2\times 4$

$$\vec{Y} = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 2 \\ 1 \\ 3 \\ 1 \\ 2 \\ 5 \end{bmatrix} \quad X = \begin{array}{c} \begin{array}{cccc} & x_1 & x_2 & x_1x_2 \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array} \quad \vec{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \end{bmatrix} \quad \vec{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{12} \end{bmatrix}$$

* columns of X are all orthogonal

diagonal since columns of X are orthogonal

– The least squares estimate of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$$

$$X^T X = \begin{bmatrix} n\cdot 2^k & & & 0 \\ & n\cdot 2^k & & \\ & & n\cdot 2^k & \\ 0 & & & n\cdot 2^k \end{bmatrix} = n2^k I_{2^k}$$

$$(X^T X)^{-1} = \begin{bmatrix} 1/n\cdot 2^k & & 0 \\ & \ddots & \\ 0 & & \frac{1}{n\cdot 2^k} \end{bmatrix} = \frac{1}{n2^k} I_{2^k}$$

$$X^T \vec{Y} = \begin{bmatrix} \vec{1}^T \vec{Y} \\ \vec{x}_1^T \vec{Y} \\ \vec{x}_2^T \vec{Y} \\ \vec{x}_{12}^T \vec{Y} \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum_{i:A^+} Y_i - \sum_{i:A^-} Y_i \\ \sum_{i:B^+} Y_i - \sum_{i:B^-} Y_i \\ \sum_{i:A^+ \cap B^+} Y_i + \sum_{i:A^- \cap B^-} Y_i - \sum_{i:A^+ \cap B^-} Y_i - \sum_{i:A^- \cap B^+} Y_i \end{bmatrix}$$

$$\vec{\beta} = (X^T X)^{-1} X^T \vec{Y} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_{12} \end{bmatrix}$$

– Notice that

$$2\widehat{\boldsymbol{\beta}} = \begin{bmatrix} 5 \\ 10/6 \\ -1/3 \\ -1 \end{bmatrix} = \begin{bmatrix} 2\overline{y} \\ \widehat{ME}_A \\ \widehat{ME}_B \\ \widehat{IE}_{AB} \end{bmatrix}$$

- In general:
  - $\mathbf{Y}$ is an $N \times 1$ vector of response observations
  - ε is an $N \times 1$ random vector of error terms
  - $\boldsymbol{\beta}$ is a $2^K \times 1$ vector of regression coefficients
  - X is the $N \times 2^K$ **model matrix** containing plus and minus ones

· each column represents a different effect (i.e. term in the linear predictor)

* Columns of X are orthogonal only when the experiment is balanced, i.e. equal # of units in each condition.

★ Covariance of any $\hat{\beta}$'s is zero.

  - Due to the orthogonality of the model matrix, any effect (whether main or interaction) is estimated as
$$\widehat{\text{Effect}} = 2\widehat{\beta} = \frac{\mathbf{x}^T\mathbf{Y}}{n2^{K-1}}$$
where $\mathbf{x}$ is the column of X corresponding to the effect of interest, and β is the corresponding regression coefficient

Hypothesis Testing

- The significance of main and interaction effects is determined by testing hypotheses that set the relevant β's equal to 0

- But now, because each effect is represented by just a single term, the hypotheses of interest involve just a single β

- In the **Toy Example**, if we wanted to determine the significance of factor A we simply test

$$H_0 : \beta_1 = 0$$

or if we want to determine whether the A:B interaction is significant, we test

$$H_0 : \beta_{12} = 0$$

- Hypotheses like these are tested with ordinary significance tests for individual regression coefficients
  - t-tests in the case of linear regression
  - Z-tests in the case of logistic regression

- But if for some reason we still want to test hypotheses about several β's simultaneously, we can compare full and reduced models with the usual
  - Partial F-tests in the case of linear regression
  - Likelihood ratio tests in the case of logistic regression

Credit Card Example

- To illustrate a complete analysis of a 2^K factorial experiment, we consider an example from Montgomery (2019) in which an experiment was performed to test new ideas to improve the conversion rate of credit card offers. For this example, the response is binary – indicating whether an individual signed up for a credit card as a result of the offer – and so an analysis based on logistic regression is performed.

- A 2^4 factorial experiment was carried out to investigate four factors and their influence on credit card sign ups. The four factors and each of their levels are summarized in the table below.

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 $2^4 = 16$ unique combinations of these factor levels produced 16 experimental conditions, each of which was assigned $n = 7500$ units. Practically speaking, 16 credit card offers were devised (one corresponding to each condition) and each was mailed to 7500 customers. The design matrix and a summary of the conversion rates are provided in the table below

Condition	Factor 1	Factor 2	Factor 3	Factor 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%

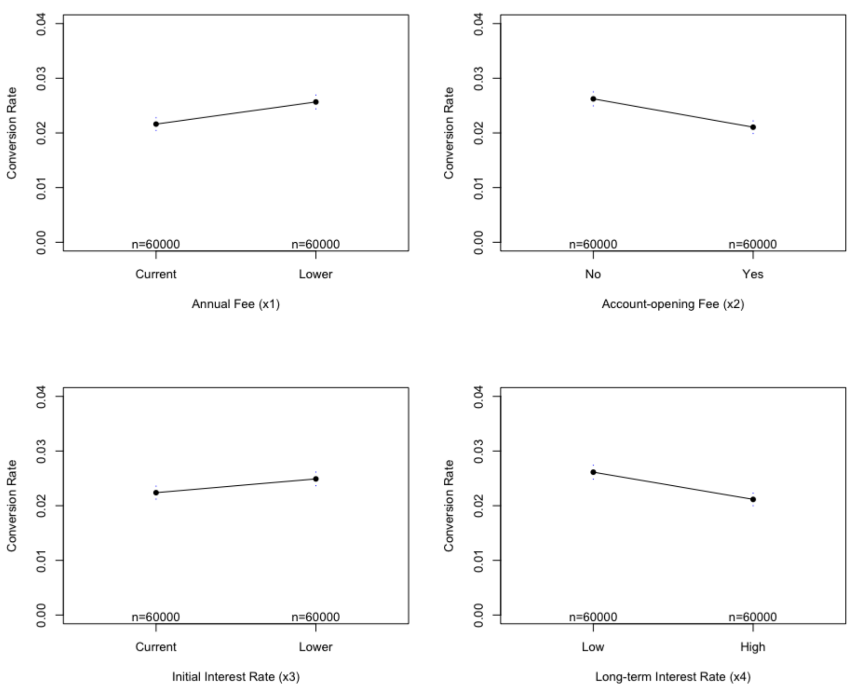
- Using this data we fit a logistic regression model with the following linear predictor

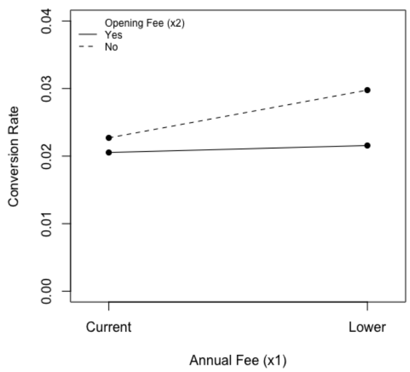
$$\begin{aligned}\beta_0 \;&+\; \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \\ &+\; \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 \\ &+\; \beta_{123} x_1 x_2 x_3 + \beta_{124} x_1 x_2 x_4 + \beta_{134} x_1 x_3 x_4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

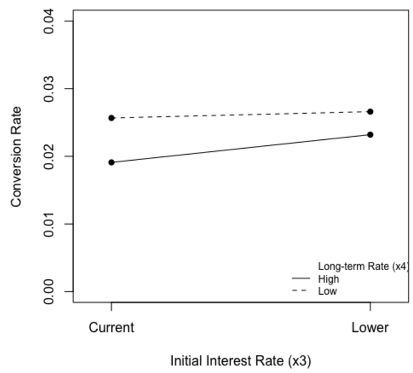
- The regression output associated with this model is:

```
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
```

- We now know which main and interaction effects are significant
  - Let’s use main and interaction effect plots to help us interpret these effects.







Optional Exercises:

- Calculations: 9
- Proofs: 13
- R Analysis: 11, 12(a), 21(a), 26(a)