RSM Recap

- Effective experimentation is sequential: information gained in one experiment can help to inform future experiments.

- Screening experiments are used to identify which among a large number of factors are the ones that significantly influence the response variable.

- We follow these up with further experimentation where the goal is **response optimization**.
  - **Method of Steepest Ascent/Descent**
  - **Response Surface Designs**

The Method of Steepest Ascent/Descent

- We use the method of steepest of ascent/descent to determine *roughly* where in the x-space the optimum lies.
  - Hence, this tells us where a response surface design and a second order model would be most useful.

- The method is gradient-based and designed to identify the direction that when traversed moves you toward the optimum as quickly as possible.

- We use a $\boxed{2^{K'}}$ factorial experiment to estimate a *first-order response surface*:

  or $2^{K'-p}$

$$\boxed{\widehat{\eta} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \cdots + \widehat{\beta}_{K'} x_{K'}}$$

- The gradient of this surface is then calculated:

$$\mathbf{g} = \nabla\widehat{\eta} = \begin{bmatrix} \frac{\partial\widehat{\eta}}{\partial x_1} & \frac{\partial\widehat{\eta}}{\partial x_2} & \cdots & \frac{\partial\widehat{\eta}}{\partial x_{K'}} \end{bmatrix}^T$$

  - This gradient defines the **path of steepest ascent/ descent**

    the direction of steepest increase/decrease on the fitted surface

  - If maximizing the response is of interest, then we should ascend the surface by moving in the direction of $+\mathbf{g}$.

$$\mathbf{x}' = \mathbf{x} + \lambda\mathbf{g} \tag{1}$$

  - If minimizing the response is of interest, then we should descend the surface by moving in the direction of $-\mathbf{g}$.

$$\mathbf{x}' = \mathbf{x} - \lambda\mathbf{g} \tag{2}$$

– With a fixed step size λ we move from $\mathbf{x}$ to $\mathbf{x}'$

– We typically define the step size as

$$\lambda = \frac{\Delta x_j}{|\hat{\beta}_j|}$$

- Pick factor j, the one you know most about, or the one that is hardest to manipulate
- Δx_j is the step size of factor j in the coded units
- $\hat{\beta}_j$ is the estimated coefficient corresponding to factor j in the estimated first order response model.

- **The method of steepest ascent/descent algorithm:**

1. The first condition along the path of steepest ascent/descent is at the origin of the x-space $\boxed{\mathbf{x}_0 = [0 \;\; 0]^T}$ (i.e., the center of the $2^{K'}$ factorial design that was used to fit $\widehat{\eta}$). Data is collected and the metric of interest is calculated.

2. Then the step size λ is determined.

3. The location of the next condition is determined by formula (1) in the case of maximization and (2) in the case of minimization. Data is collected and the metric of interest is calculated.

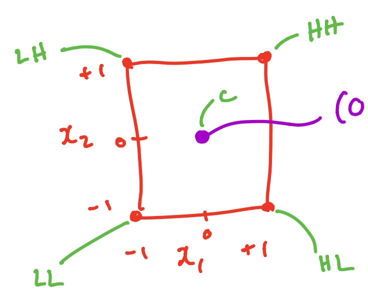
4. Repeat Step 3 until incremental improvements in the MOI cease.

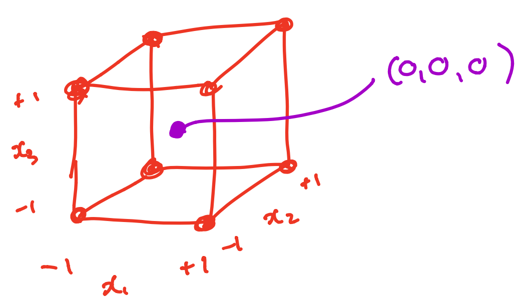
5. Return to the location of the best MOI value and *test for curvature*.
   – If the test for curvature suggests that you are not yet in the vicinity of the optimum, fit a new first order model and repeat Steps 1-4.
   – If the test for curvature suggests that you are in the vicinity of the optimum, use a response surface design to fit a full second order model and hence precisely identify the coordinates of the optimum.

Checking for Curvature

- A test for quadratic curvature is an important component of the method of steepest ascent/descent.
  – The presence of quadratic curvature signifies that you are in the vicinity of the optimum.

- Such a test is possible when a $2^{K'}$ factorial experiment is augmented with a **center point** condition
  – The center point condition is defined (in coded units) as $x_1 = x_2 = \cdots = x_{K'} = 0$
  – Located at the center of the cuboidal region defined by the $2^{K'}$ factorial conditions





- The data arising from a $2^{K'}$ factorial design is insufficient to estimate a second order linear predictor.
  - We are able to estimate the main effects and the two-factor interaction effects, but *not* the quadratic effects.

- With the addition of the center point condition, one *additional* effect may be estimated: the **pure quadratic effect**

$$\beta_{PQ} = \sum_{j=1}^{K'} \beta_{jj}$$

- A test of

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

  is a test for *overall curvature.*

- **Example:** $K' = 2$
  - When $K' = 2$ the second order linear predictor is

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

  - In a 2^2 factorial design plus a center point, we have the following five unique experimental conditions: $(x_1, x_2) \in \{(-1,-1), (+1,-1), (-1,+1), (+1,+1), (0,0)\}$ which respectively give rise to five unique variants of the linear predictor, which we define as

$$\begin{aligned}
\eta_{LL} &= \beta_0 - \beta_1 - \beta_2 + \beta_{12} + \beta_{11} + \beta_{22} \quad \beta_{PQ} \\
\eta_{HL} &= \beta_0 + \beta_1 - \beta_2 - \beta_{12} + \beta_{11} + \beta_{22} \quad \beta_{PQ} \\
\eta_{LH} &= \beta_0 - \beta_1 + \beta_2 - \beta_{12} + \beta_{11} + \beta_{22} \quad \beta_{PQ} \\
\eta_{HH} &= \beta_0 + \beta_1 + \beta_2 + \beta_{12} + \beta_{11} + \beta_{22} \quad \beta_{PQ} \\
\eta_C &= \beta_0
\end{aligned}$$

  - With only these five conditions, we cannot separately estimate β_{11} and β_{22}, but we *can* estimate $\beta_{PQ} = \beta_{11} + \beta_{22}$

  - Notice that

$$\beta_{PQ} = \frac{\eta_{LL} + \eta_{HL} + \eta_{LH} + \eta_{HH}}{4} - \eta_C$$

β_{PQ} is the difference between the average linear predictor value at the factorial conditions, minus the linear predictor value at the center point.

  - The estimate is therefore:

$$\widehat{\beta}_{PQ} = \frac{\widehat{\eta}_{LL} + \widehat{\eta}_{HL} + \widehat{\eta}_{LH} + \widehat{\eta}_{HH}}{4} - \widehat{\eta}_C.$$

* If this difference, and hence $\widehat{\beta}_{PQ}$, is small then it suggests that the response values observed in the factorial conditions are similar to those observed in the center point condition and hence that there isn't significant curvature in the response surface

* If $\widehat{\beta}_{PQ}$ is very different from zero it suggests that there is significant quadratic curvature.

- We formally test $\boxed{H_0 : \beta_{PQ} = 0}$ using t-tests (or Z-tests) in a linear (or logistic) regression model that has linear predictor:

$$\boxed{\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{PQ} x_{PQ}}$$

where

$$\boxed{x_{PQ}} = \begin{cases} 1 & \text{if } (x_1, x_2) \in \{(-1,-1), (+1,-1), (-1,+1), (+1,+1)\} \\ 0 & \text{if } (x_1, x_2) = (0, 0) \end{cases}$$

this indicates whether a response observation came from a factorial condition or the center point condition.

- If β_{PQ} is significantly different from 0 then it suggests that *both* β_{11} and β_{22} are significantly non-zero, and therefore that there is significant quadratic curvature

- For general K', we conduct a $2^{K'}$ factorial experiment with a center point and then test for curvature using a regression model with linear predictor:

$$\boxed{\eta = \beta_0 + \sum_{j=1}^{K'} \beta_j x_j + \sum_{j<l} \beta_{jl} x_j x_l + \beta_{PQ} x_{PQ}}$$

where now

$$\beta_{PQ} = \sum_{j=1}^{K'} \beta_{jj}$$

and x_{PQ} is again a binary indicator indicating whether a response value was observed in a factorial condition or the center point condition.

- No matter the value of K', the pure quadratic effect is always represented by a single term in the model
  - As such, the test for curvature is always a test of $H_0 : \beta_{PQ} = 0$ and is carried out with ordinary t-tests in a linear regression and Z-tests in a logistic regression.

- The intuitve estimate for β_{PQ} in the $K' = 2$ case also generalizes:

$$\boxed{\widehat{\beta}_{PQ} = \overline{\widehat{\eta}}_F - \widehat{\eta}_C}$$

where

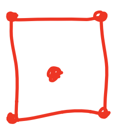
  - $\overline{\widehat{\eta}}_F$ is the average of the estimated linear predictor values in the factorial conditions
  - $\widehat{\eta}_C$ is the estimated linear predictor value at the center point

- **IMPORTANT:** This test assumes that all of the β_{jj}'s, $j = 1, 2, \ldots, K'$, have the same sign.
  - If they didn't, then it's possible that significantly large β_{jj}'s could cancel each other out, making $\beta_{PQ} = \sum_{j=1}^{K'} \beta_{jj}$ close to zero ~ we're misled into think that we're not in the vicinity of quadratic curvature, even when we are.
  - This assumption is fine as long as the experiment is not conducted near a saddle point on the response surface

→ this problem can only be identified by separately estimating all of the quadratic effects with a full response surface design.

The Netflix Example

- Here we illustrate the *method of steepest descent* using the hypothetical Netflix experiment from your final project.
- We focus on the Preview Length and Preview Size factors only.
- We begin with a 2^2 factorial experiment with a center point condition. The factor levels in coded and natural units are shown in the table below.



Condition	Preview Length	x_1	Preview Size	x_2	Average Browsing Time
1	90 seconds	-1	0.2	-1	22.16 minutes
2	120 seconds	$+1$	0.2	-1	22.20 minutes
3	90 seconds	-1	0.5	$+1$	20.22 minutes
4	120 seconds	$+1$	0.5	$+1$	21.98 minutes
5	105 seconds	0	~~0.5~~ 0.35	0	22.05 minutes

U_L U_H U_L U_H

- Prior to embarking down the path of steepest descent, a curvature test was performed to determine whether this experimental region was already in the vicinity of the optimum.
- The linear regression model with linear predictor

$$\boxed{\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{PQ} x_{PQ}}$$

was fit. The resulting output is shown below.

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.04607    0.19513 112.979  < 2e-16 ***
x1           0.44828    0.09757   4.595 4.78e-06 ***
x2          -0.53894    0.09757  -5.524 4.04e-08 ***
x1:x2        0.43118    0.09757   4.419 1.08e-05 ***
xPQ         -0.40466    0.21817  -1.855   0.0639 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We do not reject $H_0: \beta_{PQ} = 0$

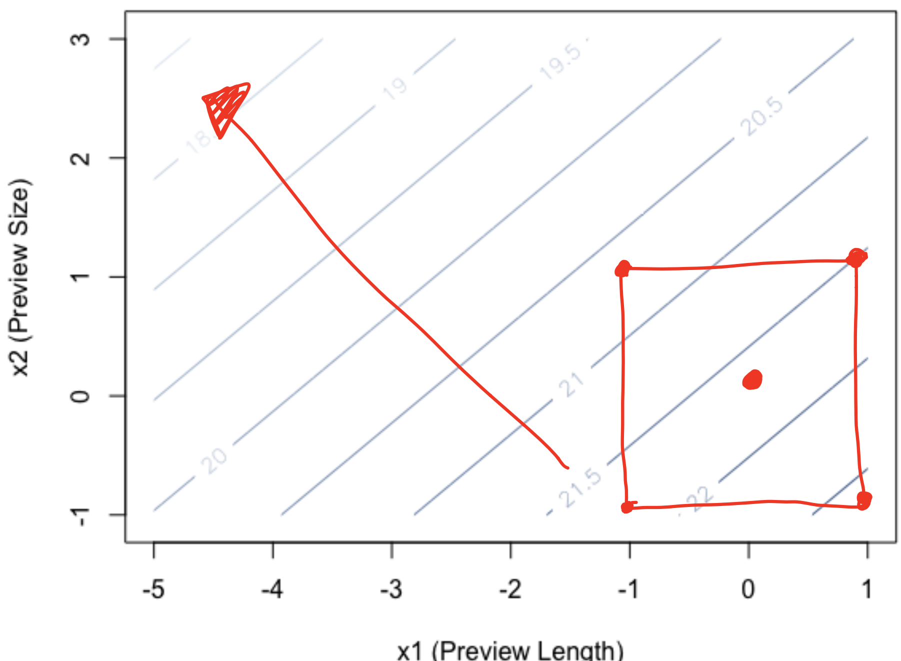
- To begin the method of steepest descent procedure, we use the aforementioned data to fit the first order regression model with linear predictor

$$\boxed{\hat{\eta} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2}$$

The model summary is shown below.

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 21.72234    0.08800 246.852  < 2e-16 ***
x1           0.44828    0.09838   4.556 5.71e-06 ***
x2          -0.53894    0.09838  -5.478 5.20e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

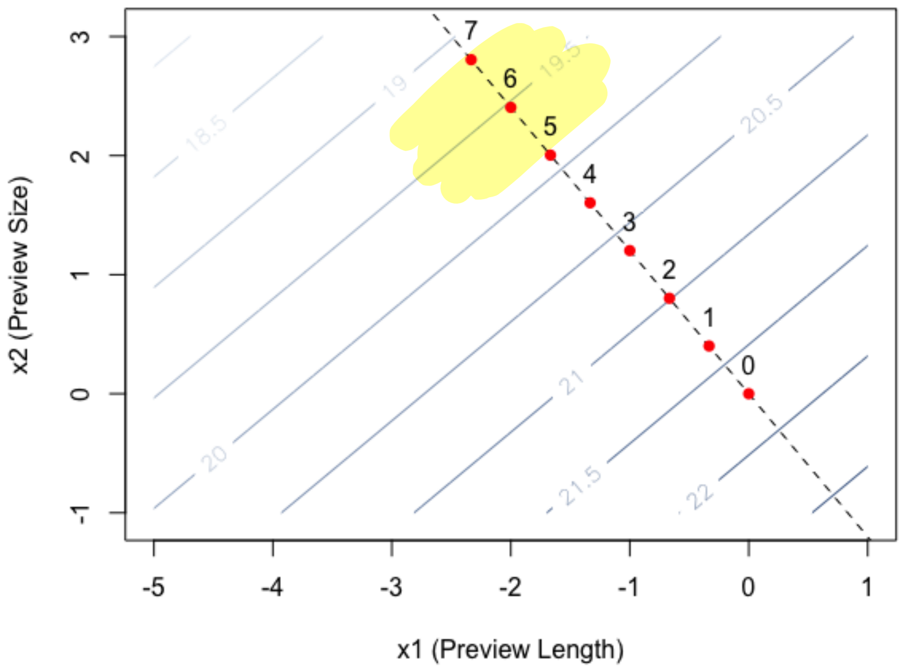
- The plot below depicts the contours of the estimated first order response surface.



- We calculate the gradient

$$\mathbf{g} = \begin{bmatrix} \overset{\frac{\partial \hat{\eta}}{\partial x_1}}{\widehat{\beta}_1} & \overset{\frac{\partial \hat{\eta}}{\partial x_2}}{\widehat{\beta}_2} \end{bmatrix}^T = [0.44828, -0.53894]$$

- This path of steepest descent is depicted by the dashed black line in the plot below. The red dots signify the experimental conditions conducted along this path, beginning from the center point $(x_1, x_2) = (0, 0)$.



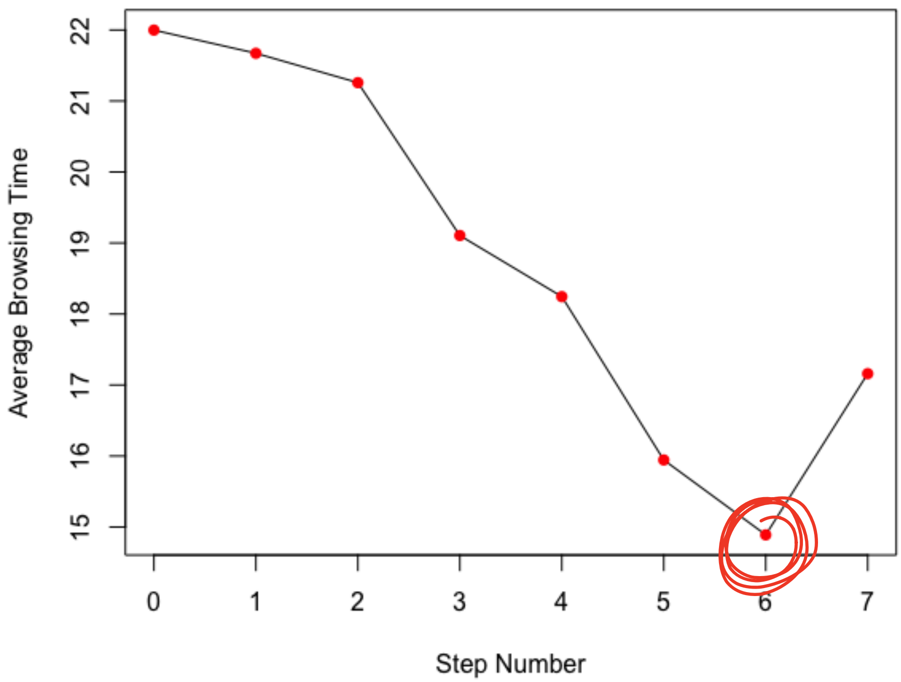
- The locations in coded and natural units for each of these conditions are provided in the table below
  - Note that a step size of

$$\lambda = \frac{1/3}{|0.44828|}$$

was used, where the value 1/3 was chosen to ensure steps of 5 seconds in Preview Lengths.

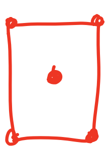
Step	Preview Length	x_1	Preview Size	x_2	Average Browsing Time
0	105 seconds	0	0.3500000	0	22.00 minutes
1	100 seconds	-1/3	0.4101118	.4007454	21.67 minutes
2	95 seconds	-2/3	0.4702236	0.8014908	21.26 minutes
3	90 seconds	-1	0.5303354	1.202236	19.11 minutes
4	85 seconds	-4/3	0.5904472	1.602982	18.24 minutes
5	80 seconds	-5/3	0.6505591	2.003727	15.94 minutes
6	75 seconds	-2	0.7106709	2.404472	14.89 minutes
7	70 seconds	-7/3	0.7707827	2.805218	17.16 minutes

- The average browsing time in each condition is reported in table above and visualized in the plot below.



- Clearly that Step 6 corresponded to the lowest observed average browsing time and so we should perform another test of curvature in this region to determine whether we've reached the vicinity of the optimimum.

- In order to do so, another 2^2 factorial experiment with a center point needs to be run. The factor levels in coded and natural units for this next experiment are shown in the table below.



Condition	Preview Length	x_1	Preview Size	x_2	Average Browsing Time
1	U_L 60 seconds	-1	0.6 U_L	-1	14.57 minutes
2	90 seconds	$+1$	0.6	-1	18.17 minutes
3	U_H 60 seconds	-1	0.8 U_H	$+1$	18.22 minutes
4	90 seconds	$+1$	0.8	$+1$	18.65 minutes
5	75 seconds	0	0.7	0	14.83 minutes

- Once again we fit a linear regression model with linear predictor

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{PQ} x_{PQ}$$

The resulting output is shown below.

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 14.83111    0.18918  78.397   <2e-16 ***
x1           1.00913    0.09459  10.669   <2e-16 ***
x2           1.03291    0.09459  10.920   <2e-16 ***
x1:x2       -0.79191    0.09459  -8.372   <2e-16 ***
xPQ          2.57337    0.21151  12.167   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Reject $H_0: \beta_{PQ} = 0$
$\therefore$ we conclude there is significant quadratic curvature

- This investigation should now be followed up by a response surface experiment so that a full second order model may be fit and the optimum identified.

Central Composite Designs

- The goal of a response surface experiment is to be able to fit a full second order response surface model.
  - This requires estimating $(K' + 1)(K' + 2)/2$ effects.

- Several such designs exist, but here we study one in particular: the **central composite design** (CCD).

- A CCD is typified by three different types of experimental conditions:

  i **two-level factorial** conditions

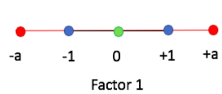
  ii a **center point** condition

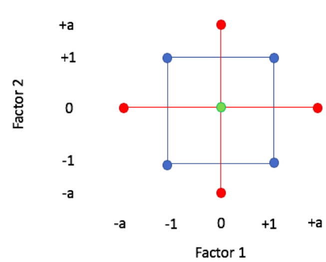
  iii **axial**, or *star*, conditions

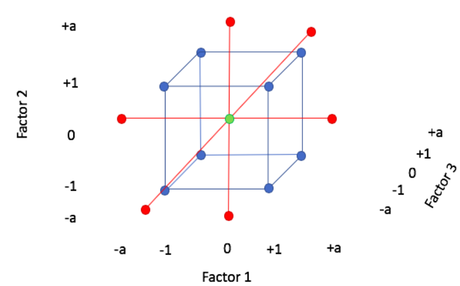
  - The factorial conditions constitute a full $2^{K'}$ factorial design.
  - The center point condition sits at $x_1 = x_2 = \cdots = x_{K'} = 0$ in the center of the factorial ones.
  - The axial conditions sit ‘outside’ of the factorial ones at $\pm a$ on each of the K' factors’ axes.

- When investigating K' factors the central composite design therefore requires $2^{K'} + 2K' + 1$ distinct experimental conditions

- These designs may be visualized geometrically as we see in the figures below, for $K' = 1, 2, 3$.







- The design matrices that give rise to these designs (for $K' = 1, 2, 3$) are shown below:

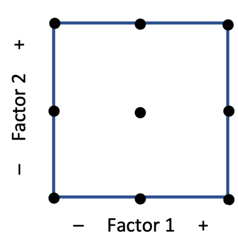
Condition	x_1
1	-1
2	$+1$
3	$-a$
4	$+a$
5	0

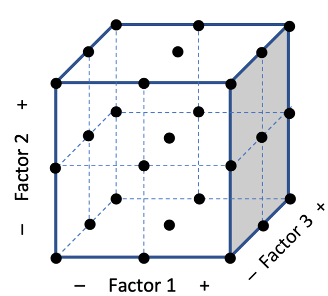
Condition	x_1	x_2
1	-1	-1
2	$+1$	-1
3	-1	$+1$
4	$+1$	$+1$
5	$-a$	0
6	$+a$	0
7	0	$-a$
8	0	$+a$
9	0	0

Condition	x_1	x_2	x_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	$+1$	$+1$	$+1$
9	$-a$	0	0
10	$+a$	0	0
11	0	$-a$	0
12	0	$+a$	0
13	0	0	$-a$
14	0	0	$+a$
15	0	0	0

- **Choosing a**:
  - The value of a is determined by the experimenter, and may be chosen to balance both practical and statistical concerns.
  - The experimenter must be mindful of the constraints imposed by the region of operability and whether the natural-unit counterpart to a is something inconvenient/infeasible.
  - Barring practice constraints, two common choices for a are $a = 1$ and $a = \sqrt{K'}$.

- **a = 1**:
  - The CCD reduces to a $\boxed{3^{K'}}$ design
  - It is referred to as *face-centered central composite design*
  - A benefit is that it requires just 3 (not 5) levels for every factor
  - Another benefit is that it is a cuboidal design and so it inherets all of the usual conveniences associated with orthogonal cuboidal designs





- **a** $=\sqrt{\boldsymbol{K'}}$:
  - In this design the axial conditions at an equal distance from the center point as the factorial conditions
  - Such a design is referred to as *spherical* since it places all axial and factorial conditions on a sphere of radius $\sqrt{K'}$
  - The benefit of such equal spacing is that it ensures that the estimate of the response surface at each condition is equally precise. — Designs with this property are called rotatable.
- No matter the choice of a, the CCD facilitates estimation of the full second order response surface model, and hence identification of the optimum.

The Lyft Example

- We illustrate the design and analysis of a central composite experiment in the context of a common ride-sharing problem.
- Suppose that Lyft is interesed in designing a promotional offer that maximizes ride-bookings during an experimental period.
- Previous screening experiments evaluated the influence of discount amount, discount duration, ride type, time-of-day, and the method of dissemination. It was found that the most important factors were discount amount (x_1) and discount duration (x_2).
- A previous steepest ascent exercise also suggested that the optimal discount duration is somewhere in the vicinity of 4.5 days and the optimal discount amount is somewhere in the vicinity of 50%.
- To find optimal values of these factors a follow-up two-factor central composite design was run in order to fit a second-order response surface model.
- The experimental conditions (in both coded and natural units) are shown in the table below:

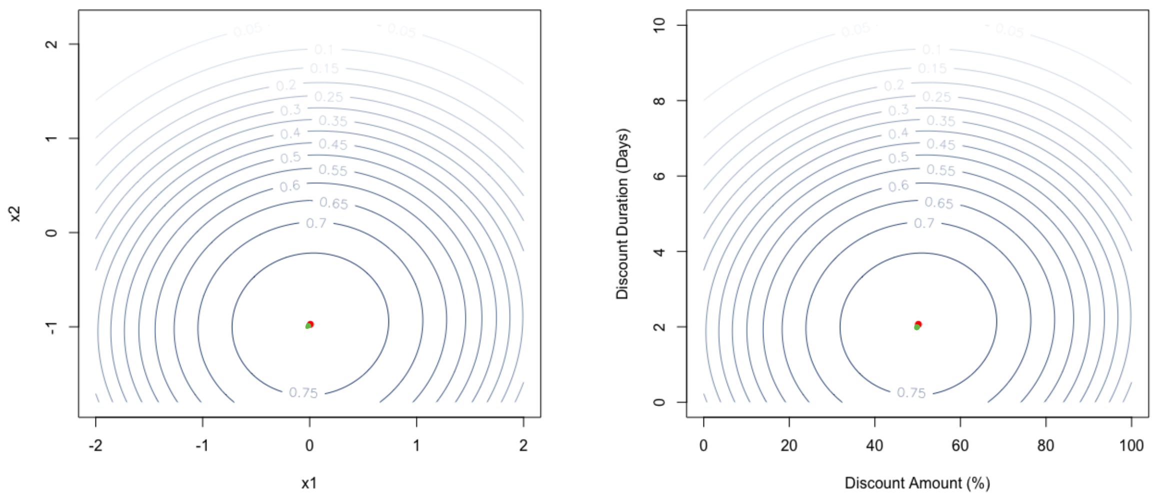
Condition	Discount Amount	x_1	Discount Duration	x_2	Booking Rate
1	25%	-1	2 days	-1	0.71
2	75%	$+1$	2 days	-1	0.32
3	25%	-1	7 days	$+1$	0.71
4	75%	$+1$	7 days	$+1$	0.35
5	85%	$+1.4$	4.5 days	0	0.53
6	15%	-1.4	4.5 days	0	0.50
7	50%	0	8 days	$+1.4$	0.26
8	50%	0	1 day	-1.4	0.78
9	50%	0	4.5 days	0	0.72

- **NOTE:** that the experimenters had intended to perform axial conditions at $a = \sqrt{2}$ but the corresponding discount amounts and discount durations were (14.64466%, 85.35534%) and (0.9644661 days, 8.035534 days). In the interest of defining experimental conditions with practically convenient levels they opted for $a = 1.4$ yielding the discount amounts and durations shown in the table above.
- $n = 500$ users were then randomized into each of these $m = 9$ conditions and for each user, whether they booked a ride in the experimentation period was recorded.
  - The booking rates in each condition are also shown in the table above.

- The output from the fitted second order logistic regression model is shown below.

```
Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.94284    0.09952   9.474  < 2e-16 ***
x1           0.03881    0.03307   1.174    0.241
x2          -0.80684    0.03568 -22.612  < 2e-16 ***
x1:x2        0.03392    0.04846   0.700    0.484
I(x1^2)     -0.44207    0.05788  -7.637 2.22e-14 ***
I(x2^2)     -0.41448    0.05931  -6.989 2.77e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Contour plots of the fitted response surface are shown in figure below.



Coded units   Natural Units

- The stationary point for this second order model is located (in coded units) at $x_1 = 0.006565206$, $x_2 = -0.973047233$
  - In the natural units this corresponds to a discount rate of $\boxed{50.16\%}$ that lasts for $\boxed{2.07}$ days.
  - The predicted booking rate at this point is 0.7918, with a 95% prediction interval given by (0.7691, 0.8144)
- A slightly less optimal but more practically feasible promotion would be a 50% discount lasting 2 days
  - This achieves a booking rate of 0.7917 with a 95% prediction interval of (0.7693, 0.8141)

$\vec{x}_s = -\frac{1}{2} B^{-1} b$

$B = \begin{bmatrix} \hat{\beta}_{11} & 0.5\hat{\beta}_{12} \\ 0.5\hat{\beta}_{12} & \hat{\beta}_{22} \end{bmatrix}$ $\vec{b} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}$

What to do if we have ≥1 categorical factor in addition to our numeric factor(s)

RSM with Categorical Factors

- Everything that has been discussed thus far with respect to central composite designs and response surface optimization has assumed that the factors under experimentation are quantitive (i.e., the factors have numeric levels)

- In the presence of one or more categorical factors we need to take additional care.

- When categorical factors are present, we can think of there being different response surfaces that relate the response to the quantitative factors at each of the factorial combinations of the categorical factors' levels.

- Thus, the general strategy is to enumerate all factorial combinations of the categorical factors' levels and employ the methods of response surface methodology independently within each.
  - Perform the method of steepest ascent/descent independently on each surface ✓
  - Perform CCDs independently on each surface ✓
  - Independently fit second order models for each surface ✓
  - Independently identify the stationary point on each surface ✓

- Among all of the candidate surfaces, the one with the most optimal optimum is the 'winner'
  - The factor levels (numeric and categorical) that gave rise to it should be defined as the optimal operating conditions.

Ex: 2 numeric factors: x_1, x_2

2 categorical factors: x_3, x_4

3 levels (L, M, H)

2 levels (L, H)