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**

- In these investigations, response optimization requires investigating and characterizing response surfaces of the form
$$\mathrm{E}[Y] = f(x_1, x_2, \ldots, x_{K'})$$
(for a continuous response) and
$$\log\left(\frac{\mathrm{E}[Y]}{1-\mathrm{E}[Y]}\right) = f(x_1, x_2, \ldots, x_{K'})$$
(for a binary response).

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 $(K'+1)(K'+2)/2$ coefficients.

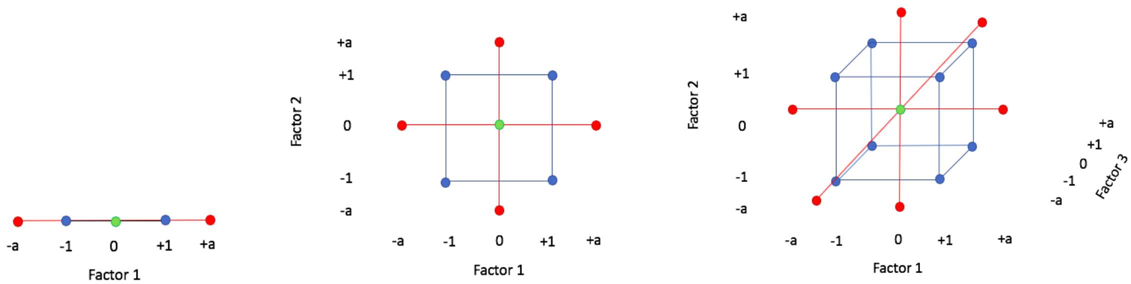
- Several such designs exist, but here we study one in particular: the **central composite design** (CCD).

- 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. Note that a is defined in coded units.

- 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.

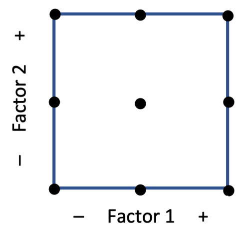
Condition	x_1
1	-1
2	$+1$
3	$-a$
4	$+a$
5	0

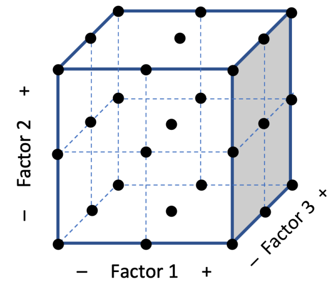
Condition	x_1	x_2
1	-1	-1
2	$+1$	-1
3	-1	$+1$
4	$+1$	$+1$
5	$-a$	0
6	$+a$	0
7	0	$-a$
8	0	$+a$
9	0	0

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 practical constraints, two common choices for a are $a = 1$ and $a = \sqrt{K'}$.

- $\mathbf{a = 1}$:
  - The CCD reduces to a $3^{K'}$ factorial 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 inherits some of the usual conveniences associated with orthogonal cuboidal designs





- **a =$\sqrt{K'}$:**
  - In this design the axial conditions are 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.

- No matter the choice of $a > 0$, the CCD facilitates estimation of the full second order response surface model, and hence identification of the optimum.

The Lyft Example (CCD)

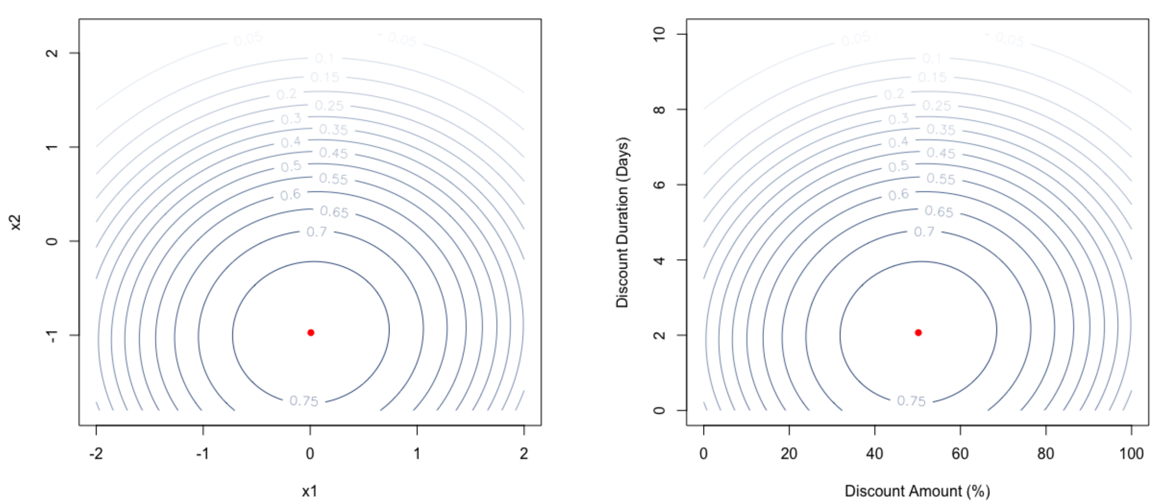
- We illustrate the design and analysis of a central composite experiment in the context of a common ride-sharing problem.
- Suppose that Lyft is interesed in designing a promotional offer that maximizes ride-bookings during an experimental period.
- Previous screening experiments evaluated the influence 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:
- **NOTE:** that the experimenters had intended to perform axial conditions with $a = \sqrt{2}$, 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 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

- 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 the figure below.



- 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 50.16% that lasts for 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)

The Method of Steepest Ascent/Descent

- In the previous example we saw the value of performing a response surface design, fitting a full second order model, and hence identifying the location of the optimum.
  - However, the utility of the second order model depends on whether the response surface design was run in the vicinity of quadratic curvature (i.e., in the vicinity of the optimium).
  - As such, it's wise to perform a response surface design *only after* we've determined *roughly* where in the x-space the optimum lies.

- This is the purpose of 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 $2^{K'}$ factorial experiment to estimate a *first-order response surface*:

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

- 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**
  - 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

just one factor

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

- pick one factor that you know most about, or one that is hardest to manipulate.
- Δx_j is the "isolated" step size in factor j (in coded units)
- $\widehat{\beta}_j$ is from the estimated first-order 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 $\mathbf{x}_0 = [0 \;\; 0]^T$ (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.

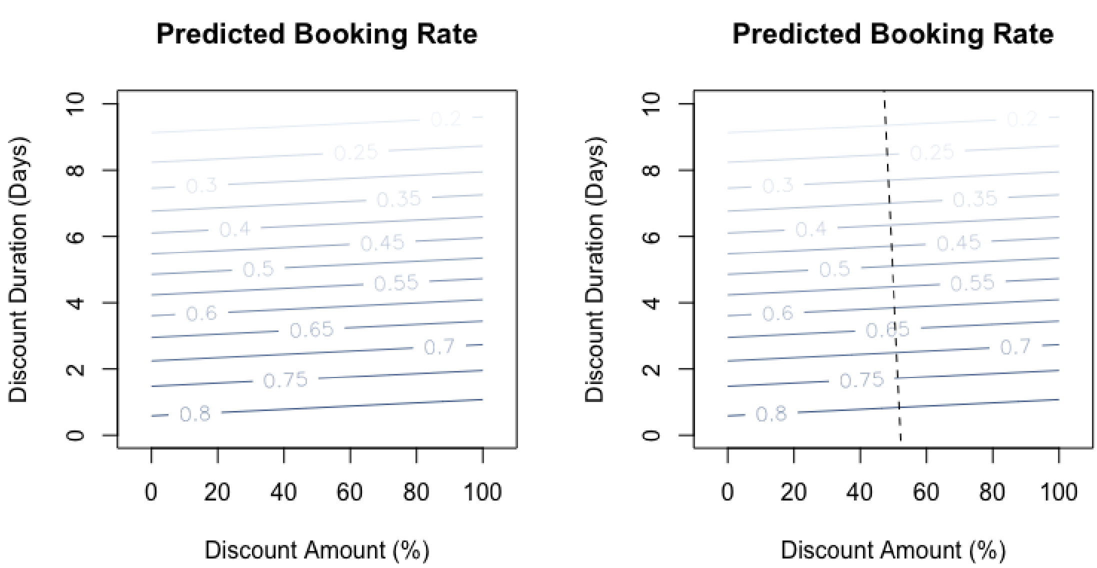
  4. Repeat Step 3 until incremental improvements in the MOI cease.

  5. Return to the location of the best MOI value and assess *curvature*. This may be done informally with a plot or formally with a hypothesis test as described in Section 8.3.2 of the Course Notes.
     - If evidence suggests you *are not* the vicinity of the optimum, fit a new first order model and repeat Steps 1-4.
     - If evidence suggests you *are* 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.

The Lyft Example (MOSA)

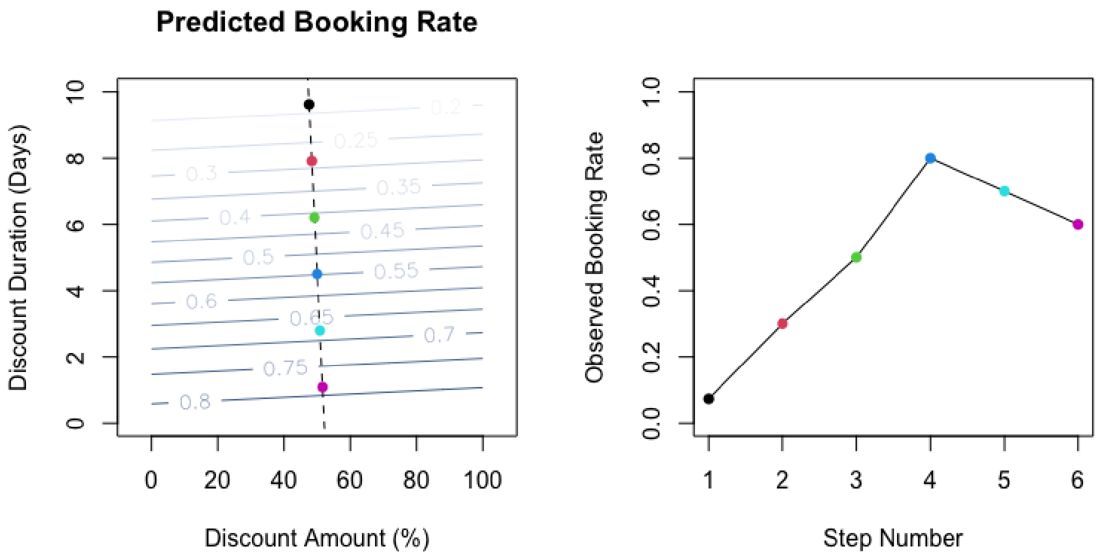
- In the Lyft example seen previously, the CCD was centered at 50% discount lasting 4.5 days. It was a acknowledged that the method of steepest ascent was used to inform this decision. Here will illustrate this.

- Suppose that data from a 2^2 factorial design is used to fit the first order surface visualized below. The left plot visualizes the fitted response surface, and the right plot visualizes the surface together with the gradient $\mathbf{g}$.



- This gradient defines the path of steepest ascent/descent, which if we move along in the *positive* direction, will direct us toward the optimum.

- The first step along this path is the black dot (at a discount of roughly 45% for 9.5 days) visualized in the left plot below. The observed booking rate is 0.07 (visualized in the right plot below).



- After a fixed step size λ along this path, we arrive at the red dot (at a discount of roughly 47% for 8 days) visualized in the left plot above. The observed booking rate is 0.3 (visualized in the right plot above).

- As we can see, this process continues for four more steps, until we are sure we've identified some quadratic curvature.

- We see that along the path of steepest ascent, the observed booking rate was highest at Step 4 (blue dot), in the neighbourhood of a 50% discount for 4.5 days. This motivates moving forward with a CCD centered at this location, which is precisely what we did.

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.