Introduction to Design of Experiments and Optimal Design

Overview of Design of Experiments (DOE) and Optimality

Importance of DOE in Statistical Modeling

Design of Experiments (DOE) is a structured approach for planning experiments to efficiently explore the relationship between factors and responses. By strategically choosing experimental runs, DOE enables precise estimation of model parameters and reliable predictions. Optimal design theory extends DOE by selecting runs that maximize information or minimize variance according to a chosen criterion. This approach is crucial when experiments are costly or time-consuming, as it seeks to extract the most information from limited runs.

Optimal Design Criteria – Focus on I-Optimality

Optimal designs are characterized by criteria like **D-optimality, A-optimality, and I-optimality**, each targeting a different aspect of estimation precision. I-optimality (also known as V-optimality) is especially relevant when the goal is accurate prediction across the design space. An I-optimal design minimizes the average prediction variance over the experimental region, which means it provides more uniform precision for predicted responses at all points in the design space.

Criterion	Optimization Goal	Focus
D-Optimality	Maximizes det(X X)	Overall parameter estimate precision
A-Optimality	Minimizes trace((X X)^(-1))	Average variance of parameter estimates
I-Optimality	Minimizes integrated variance of predictions	Prediction accuracy across factor space

Setting the Stage: DOE in R and the Tidyverse

R Environment for Experimental Design

R provides a rich ecosystem for DOE, including packages for classical designs and algorithmic optimal design. Packages like AlgDesign and skpr offer functions to generate D-, I-, and other optimal designs, evaluate their statistical properties, and visualize design efficiency.

Leveraging the Tidyverse for Data and Design Management

The tidyverse set of packages integrates well with DOE tasks by simplifying data manipulation and visualization. In the context of optimal design, tidyverse tools can be used to:

- Define candidate sets of experimental runs
- Filter or modify candidate points based on practical constraints
- Examine and visualize designs

Example of creating a candidate set using tidyverse functions:

```r
library(tidyverse)
# Define candidate set for 2 factors (x1, x2 each in {-1,0,1})

candidate_set <- expand.grid(x1 = c(-1, 0, 1), x2 = c(-1, 0, 1))
print(candidate_set)
```

This tibble of candidates can then be fed into optimal design algorithms in R. The tidyverse thus helps maintain clarity and efficiency in managing DOE data, which becomes increasingly important as the number of factors and candidate runs grows.