HW9

2023-10-20

##Question 12.1 Describe a situation or problem from your job, everyday life, current events, etc., for which a design of experiments approach would be appropriate.

A design of experiments approach would be useful for optimizing the asset allocation in an investment portfolio. With the proliferation of AI-driven investment strategies and robo-advisors, there are countless combinations of different providers to choose from. Companies like Betterment, Wealthfront, SoFi, and Vanguard all offer algorithmic investing services with varying methods and fee structures.

A design of experiment could help determine the ideal allocation between these AI platforms to maximize returns for a given risk tolerance. We could vary the percentage allocated to each provider systematically while tracking performance over time. This would reveal any interactions between the different AI methodologies and enable discovering an optimal asset mix. Key experimental factors would be the relative amounts allocated to passive vs. active quant strategies, fundamental factor-based vs. alternative data models, and low-fee vs. high-fee offerings. A well-designed experiment would efficiently map the response surface for portfolio return as a function of the allocations across these dimensions.

##Question 12.2 To determine the value of 10 different yes/no features to the market value of a house (large yard, solar roof, etc.), a real estate agent plans to survey 50 potential buyers, showing a fictitious house with different combinations of features. To reduce the survey size, the agent wants to show just 16 fictitious houses. Use R's FrF2 function (in the FrF2 package) to find a fractional factorial design for this experiment: what set of features should each of the 16 fictitious houses have? Note: the output of FrF2 is "1" (include) or "-1" (don't include) for each feature.

```
options(repos = c(CRAN = "https://cran.rstudio.com/"))
install.packages("FrF2")
```

```
##
## The downloaded binary packages are in
##  /var/folders/rk/s8z97k_16n34c36n84qb_zdm0000gn/T//RtmpkB5Qmf/downloaded_packages
```

```
library(FrF2)
```

```
## Loading required package: DoE.base
```

```
## Loading required package: grid
```

```
## Loading required package: conf.design
```

```
## Registered S3 method overwritten by 'DoE.base':
##   method           from
##   factorize.factor conf.design
```

```
##
## Attaching package: 'DoE.base'
```

```
## The following objects are masked from 'package:stats':
##
##     aov, lm
```

```
## The following object is masked from 'package:graphics':
##
```

```
##     plot.design

## The following object is masked from 'package:base':
##
##     lengths
```

```
# Define 10 factors
factors <- c("Yard", "Solar", "Pool", "Basement", "Garage", "Deck", "Fireplace", "Hardwood", "Granite",

# Generate 16 run design
design <- FrF2(nfactors = 10, nruns = 16, repeat.only = TRUE)

# Print design with factor names
print(design)
```

```
##     A  B  C  D  E  F  G  H  J  K
## 1  -1 -1  1  1  1 -1 -1 -1 -1  1
## 2  -1 -1 -1  1  1  1  1 -1  1 -1
## 3   1  1  1  1  1  1  1  1  1  1
## 4   1 -1  1  1 -1  1 -1  1 -1 -1
## 5  -1  1  1  1 -1 -1  1 -1  1 -1
## 6  -1  1  1 -1 -1 -1  1  1 -1  1
## 7  -1  1 -1 -1 -1  1 -1  1  1 -1
## 8   1 -1 -1 -1 -1 -1  1 -1 -1 -1
## 9  -1  1 -1  1 -1  1 -1 -1 -1  1
## 10  1  1 -1 -1  1 -1 -1 -1  1  1
## 11 -1 -1  1 -1  1 -1 -1  1  1 -1
## 12 -1 -1 -1 -1  1  1  1  1 -1  1
## 13  1  1  1 -1  1  1  1 -1 -1 -1
## 14  1  1 -1  1  1 -1 -1  1 -1 -1
## 15  1 -1 -1  1 -1 -1  1  1  1  1
## 16  1 -1  1 -1 -1  1 -1 -1  1  1
## class=design, type= FrF2
```

```
print(factors)
```

```
##  [1] "Yard"      "Solar"     "Pool"      "Basement"  "Garage"    "Deck"
##  [7] "Fireplace" "Hardwood"  "Granite"   "SmartHome"
```

This outputs a 16 run fractional factorial design with the combinations of 10 features that should be shown in each fictitious house. The output from FrF2 is +/- 1 indicating whether to include or exclude each factor in each run. For example, the first house would have a large yard, solar roof, pool, basement, etc. while excluding a garage, deck, fireplace, hardwood floors, and smart home technology. This design allows estimating the main effects of each feature efficiently with only 16 survey responses.

##Question 13.1 For each of the following distributions, give an example of data that you would expect to follow this distribution (besides the examples already discussed in class).

a. Binomial - Whether a loan applicant passes a credit check (yes/no outcome)

b. Geometric - The number of 0.25% interest rate hikes by the Fed before unemployment starts to increase

c. Poisson - The number of login attempts to a website before a valid username/password combination is entered

d. Exponential - The time between loan applications being received after an interest rate cut advertisement

e. Weibull - The length of time a stock price remains at an all-time high before experiencing a downward correction

##Question 13.2 (see python script for code and discussion)