Assignment 9

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.

Answer 12.1

I am presently working towards loosing weight. This is a very good situation where I can use DOE to see what factors (exercise and diet with their variations) will help or have the largest impact on my targets. It can also help in identifying important interactions between the factors.

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.

Answer 12.2

Experimental Design

With the current selection we have 10 factors and 16 runs.Factorial designs are useful as screening experiments because they require relatively few runs to estimate main and interaction effects.

Randomization

While we have no control over how the data will be collected, by selecting a random sample from the data, we are incorporating randomization into the model.

Blocking

Blocking is not used in this design. This is a screening experiment and we are interested in the effects of all the factors.

```r
library(FrF2)
```

```r
set.seed(42)
des<-FrF2(nfactors = 10,nruns = 16,
          factor.names = c('Yard', 'Parking','Rooms','Basement', 'Price',
                           'Location','School Nearby', 'Supermarket Nearby',
                           'Luxary Home','Train Station Nearby'))
```

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
b. Geometric
c. Poisson
d. Exponential
e. Weibull

Answer 13.1

A probability distribution is a statistical function that describes all the possible values and likelihoods that a random variable can take within a given range. This range will be bounded between the minimum and maximum possible values, but precisely where the possible value is likely to be plotted on the probability distribution depends on a number of factors. These factors include the distribution's mean (average), standard deviation, skewness, and kurtosis.

1. Binomial Applying for a n Data Scientist jobs with observation being selected or rejected. Thus, the number of times a person gets selected might follow a binomial distribution
2. Geometric
   If I randomly ask people which university they went to, until I meet someone who went to Georgia Tech. Thus, in this example p is the probability of finding someone who went to Georgia Tech after asking X number of people. In this case, we can say that X follows a geometric distribution.
3. Poisson
   The number of car accidents per hour in a day might follow a Poisson Distribution
4. Exponential
   In the above example the time between two accidents might follow the exponential distribution.
5. Weibull Modeling the time to failure of an engine by a car manufacturer might follow Weibull distribution.

Question 13.2

In this problem you, can simulate a simplified airport security system at a busy airport. Passengers arrive according to a Poisson distribution with Lambda 1=5 per minute (i.e., mean interarrival rate mu1 = 0.2 minutes) to the ID/boarding check queue, where there are several servers who each have exponential service time with mean rate mu2 = 0.75 minutes. [Hint: model them as one block that has more than one resource.]

After that, the passengers are assigned to the shortest of the several personal-check queues, where they go through the personal scanner (time is uniformly distributed between 0.5 minutes and 1 minute).

Use the Arena software (PC users) or Python with SimPy (PC or Mac users) to build a simulation of the system, and then vary the number of ID/boarding-pass checkers and personal-check queues to determine how many are needed to keep average wait times below 15 minutes. [If you're using SimPy, or if you have access to a non- student version of Arena, you can use Lambda 1 = 50 to simulate a busier airport

Answer 13.2

I started this assignment using the student version of Arena but quickly hit the road block with the limitation of the student version. I then started this exercise in the Virtual Lab of Gatech. Below is the approach I followed with various results summarized.MOst of this approach was discussed in the office hour

Basic Model: 5 Scanners and 1 ID Check resource

1. Created the two processes required ID check and Scanners. ID Check is an Exponential distribution and Security lines have a uniform distribution as per the assignment.
2. Scanner sorting decision box is used to manage the queue.The expression used to manage the the traffic is as follows:

a) Scanner 1: Minimum of all the scanner queues

b) Scanner 2 to 6: Queue greater than previous Scanner and less than the rest of the scanners.

The model is tested with 5 replications/iterations for a 24 hour cycle for this simulation.The results are depicted in hours.

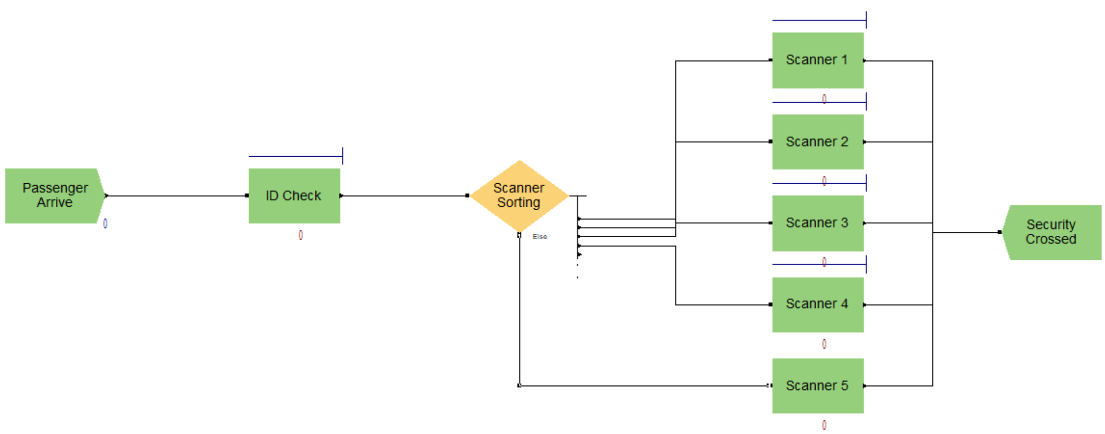

Figure 1: Base Model (Image from newer student version of Arena).

Model 1: 5 Scanner and ID Check (1 resource)

This model was able to bring down the waiting time to much lower than 15 mins but there are resources which are not used as can be seen in the table above with Scanner 4 and 5 zero average waiting time.

Time

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	0.00169648	0.00	0.00118066	0.00269609	0.00	0.04815562
Scanner 1.Queue	0.00113673	0.00	0.00074639	0.00149554	0.00	0.04460377
Scanner 2.Queue	0.01935788	0.01	0.01702631	0.02649539	0.00700480	0.04620413
Scanner 3.Queue	0.01210888	0.00	0.01052612	0.01511450	0.00019510	0.04295537

Other

Number Waiting	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	0.1033	0.05	0.06877372	0.1668	0.00	9.0000
Scanner 1.Queue	0.06550278	0.02	0.04210863	0.08553215	0.00	1.0000
Scanner 2.Queue	0.01897277	0.02	0.00780373	0.04921313	0.00	6.0000
Scanner 3.Queue	0.02544774	0.02	0.01432754	0.04222727	0.00	1.0000
Scanner 4.Queue	0.00	0.00	0.00	0.00	0.00	0.00
Scanner 5.Queue	0.00	0.00	0.00	0.00	0.00	0.00

Figure 2: Table 1.

Model 2: 3 Scanner and ID Check (1 resource)

Time

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	2.9545	0.17	2.7298	3.0620	0.00	6.3412
Scanner 1.Queue	2.8880	0.16	2.6814	2.9988	0.00	6.3214
Scanner 2.Queue	2.9598	0.15	2.7450	3.0381	0.02101023	6.3253
Scanner 3.Queue	2.9894	0.16	2.7642	3.0908	0.04888463	6.3266

Other

Number Waiting	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	180.27	14.87	162.02	193.74	0.00	382.00
Scanner 1.Queue	45.8058	2.97	42.5128	48.3350	0.00	99.00
Scanner 2.Queue	45.4659	2.98	42.1468	48.0057	0.00	99.00
Scanner 3.Queue	45.1529	2.96	41.8702	47.6566	0.00	99.00

Figure 3: Table 2.

Based on the previous model we saw that we were not using all the resources and thus for this model we used only 3 scanners. The results above show that the average wait time now is much larger using 3 scanners and 1 resource at ID check. We can will further improvise to find the best model.

Model 3: 4 Scanner and ID Check (1 resource)

With this model we can see that we were able to bring down the average waiting time from the previous model but not to the extent we want to. The first three model helped to see how many scanners we can work with. In the coming model I worked to optimize the resources in the ID check in the subsequent models.

Time

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	1.2780	1.11	0.1362	2.4467	0.00	11.4875
Scanner 1.Queue	0.4123	0.45	0.02764666	0.8744	0.00	11.6209
Scanner 2.Queue	2.3422	1.97	0.2253	4.3902	0.00019510	11.6278
Scanner 3.Queue	2.8791	1.97	0.4639	4.6307	0.00946939	11.6433
Scanner 4.Queue	3.1693	1.83	0.8663	4.8769	0.00977743	11.6548

Other

Number Waiting	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	129.62	99.67	17.2554	223.76	0.00	699.00
Scanner 1.Queue	18.3133	14.27	3.3101	34.5830	0.00	99.00
Scanner 2.Queue	18.0778	14.17	3.1948	34.2327	0.00	99.00
Scanner 3.Queue	17.8607	14.13	3.0824	34.0196	0.00	99.00
Scanner 4.Queue	17.5782	14.03	2.8962	33.6407	0.00	99.00

Figure 4: Table 3.

Model 4: 4 Scanner and ID Check (6 resource)

Time

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	0.00001061	0.00	0.00000365	0.00001881	0.00	0.00748741
Scanner 1.Queue	0.00000952	0.00	0.00000034	0.00002514	0.00	0.00745713

Other

Number Waiting	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	0.00063789	0.00	0.00022371	0.00116871	0.00	2.0000
Scanner 1.Queue	0.00057718	0.00	0.00002021	0.00156202	0.00	1.0000
Scanner 2.Queue	0.00	0.00	0.00	0.00	0.00	0.00
Scanner 3.Queue	0.00	0.00	0.00	0.00	0.00	0.00
Scanner 4.Queue	0.00	0.00	0.00	0.00	0.00	0.00

Figure 5: Table 4.

With 6 resources at the ID check the model performed well but again Scanner 2,3,4 are unused.

Model 5: 4 Scanner and ID Check (4 resource)

The final model gave the best results in terms of using all the resources and also bringing the average below 15 minutes.There are many ways to optimize this simulation and adding more constraints also takes time run those simulations.

Time

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	0.00030703	0.00	0.00018393	0.00045405	0.00	0.02221562
Scanner 1.Queue	0.00023539	0.00	0.00013838	0.00030671	0.00	0.02151725
Scanner 2.Queue	0.00785156	0.00	0.00607242	0.00896525	0.00183569	0.01789663
Scanner 3.Queue	0.00818076	0.01	0.00	0.01205672	0.00	0.01564276
Scanner 4.Queue	0.00165636	0.00	0.00	0.00828182	0.00	0.00828182

Other

Number Waiting	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
ID Security Check.Queue	0.01867115	0.01	0.01061447	0.02726173	0.00	4.0000
Scanner 1.Queue	0.01418001	0.01	0.00795696	0.01808317	0.00	1.0000
Scanner 2.Queue	0.00367917	0.00	0.00126509	0.00709749	0.00	1.0000
Scanner 3.Queue	0.00126984	0.00	0.00	0.00215053	0.00	1.0000
Scanner 4.Queue	0.00006902	0.00	0.00	0.00034508	0.00	1.0000

Figure 6: Table 5.