Study & Analysis of Household WiFi Speed

Huafeng Zhang

1. Introduction

Since the internet is an important source of information, people need high-speed WiFi to study, work and keep in touch with friends and colleagues. People realize that WiFi speed is related to whether there is an obstacle between the router and their devices, how far their devices are from their router, and the number of users. In this experiment, I will conduct a balanced three-factor factorial study to test the main effects of these three variables, as well as the potential effects of interactions among these three variables on WiFi speed. My questions of interest include the following:

- Do obstacles affect WiFi speed?
- How does the distance between WiFi router and WiFi users affect the WiFi speed?
- How does the number of WiFi users affect WiFi speed?
- Is there any interaction among these three factors?

2. Study Design Protocol

In this experiment, there are three fixed factors: **obstacle** (2 levels: with or without), **distance** between WiFi router and WiFi users (three levels: 10, 30 and 50 feet), and **number of WiFi users** (four levels: 1, 2, 3, or 4 users, all using Apple devices). The response variable is WiFi download speed for the chosen user measured in Mbps using the *testmy.net* website. There are five independent replications at each of the 24 ($2 \times 3 \times 4$) treatment combinations. So the design size is 120 ($2 \times 3 \times 4 \times 5$).

With an assistant, I collected data in the early morning, to minimize noise from nearby networks. In order to maintain the independence of the 120 experimental units, we used Google Incognito tabs to prevent the webpage from referencing stored WiFi speed data from our previous run. Furthermore, no other activities that require WiFi were allowed during the experiment.

- The obstacle was simulated by putting a computer monitor and a board to simulate a wall containing wiring between the WiFi router and WiFi users.
- The distance between router and users was measured indoors in a straight line free of obstacles.
- I had each group of 1 to 4 users download the same file (12 MB data) at the same time.
- The order of the 120 experimental runs is completely randomized.

After collecting the data, I first used diagnostic plots to check the assumptions for a three-factor analysis of variance (ANOVA), which is used for testing the effects of these three variables on WiFi speed. Before drawing conclusions about significant differences, I ran a power analysis to find the power for the given design size (120) and the 24 mean estimates.

3. Methods

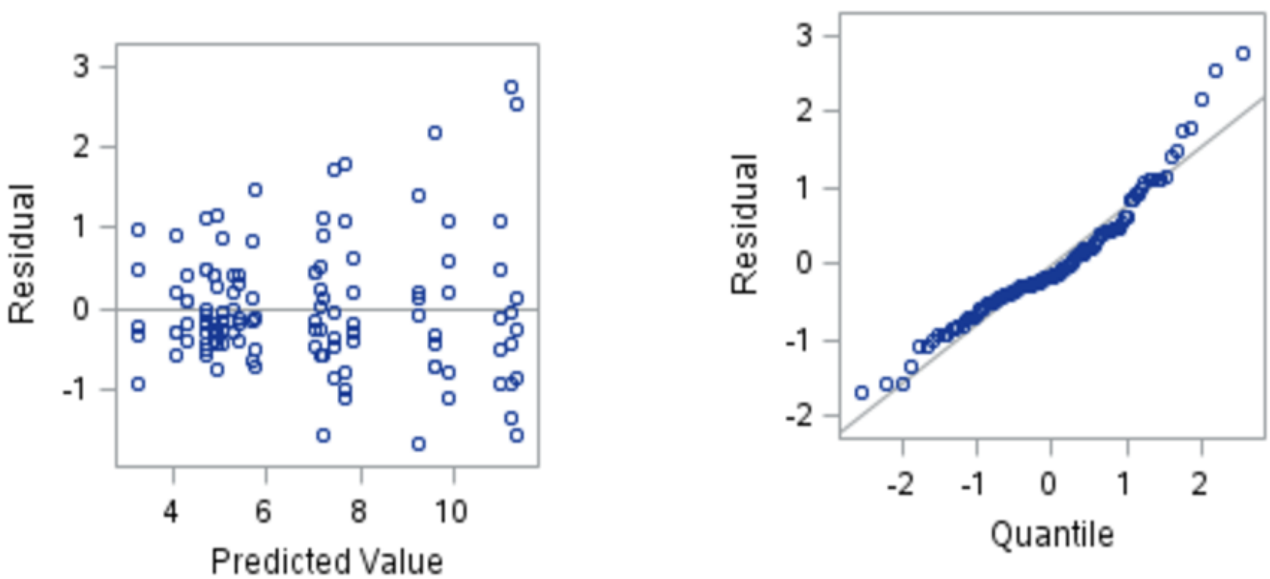
In this design, all three factors are fixed, and the model for the three-factor completely randomized design is $y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \xi_{ijkl}$, where:

- y_{ijkl} is the WiFi speed for l^{th} observation from the $(i, j, k)^{th}$ treatment.
- μ is the baseline mean of WiFi speed.
- τ_i, β_j, γ_k are the three main effects for the factors obstacle, distance, and number of users, respectively.
- $\tau\beta_{ij}$, $\tau\gamma_{ik}$, $\beta\gamma_{jk}$ are the two-factor interaction effects for the interaction *obstacle * distance,obstacle * number of users*, and *distance * number of users*.
- $\tau\beta\gamma_{ijk}$ are the three factor interaction effects for the *obstacle * distance * number of users* interaction.
- ξ_{ijkl} is the random error of the l^{th} observation from the $(i, j, k)^{th}$ treatment.

The assumptions of using this generalized linear model for the three-factor factorial design include the following:

- The 120 experimental units are independent from each other.
- The residuals have constant variance.
- The residuals are normally distributed, i.e. $\xi_{ijkl} \sim$ IID N(0, σ^2).

Using the model above yields the plots in Figure 1. The Residuals vs. Fitted Values plot (Figure 1(a)) is used to assess homogeneity of variance and absence of outlier assumptions, while the Normal Q-Q plot (Figure 1(b)) is used to assess normality of the residuals. An increasing pattern of the variability of the residuals appears in Figure 1(a), which indicates that there are severe violations of the assumption of homogeneity of variance. Figure 1(b) suggests that the distribution of the residuals is right skewed. An initial inspection of the diagnostic plots (Figure 1) from this model suggests the need for a transformation of the response **Speed**.



(a) Residual vs. Fitted Values

(b) Normal Quantitle-Quantile (Normal Q-Q)

Figure 1: Diagnostics Plots

Therefore,a Box-Cox procedure was used to find the optimal transformed scale λ of the response. Figure 2 suggests that the optimal λ is 0.4, which maximizes the percentage of variance explained by the chosen model.

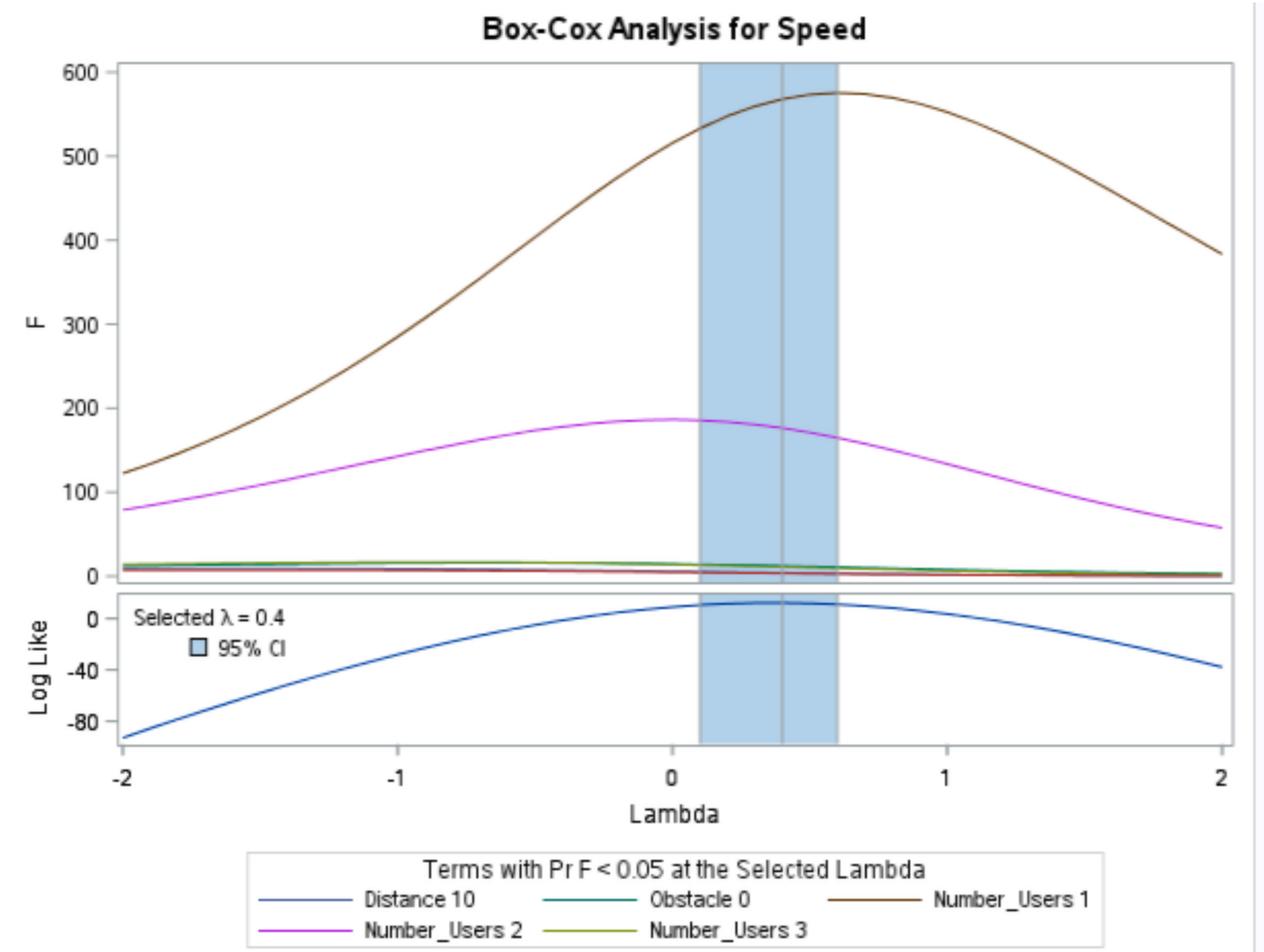


Figure 2: Box-Cox Transformation Using the TRANSREG Procedure

However, as we can see from Figure 3, the values of log(Speed) and $Speed^{0.4}$ are close to each other within the range of my WiFi speed (4-15 Mbps), and R^2 for the model using log(Speed), 0.9144, is close to R^2 for the model using $Speed^{0.4}$ (0.9149). Since it is easier to interpret the results when using log(Speed), I transformed Speed to log(Speed).

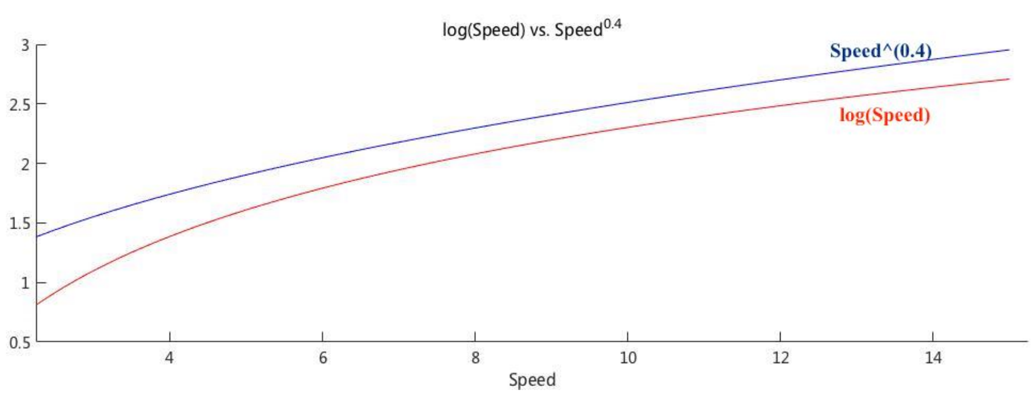


Figure 3: log(Speed) vs. $Speed^0.4$

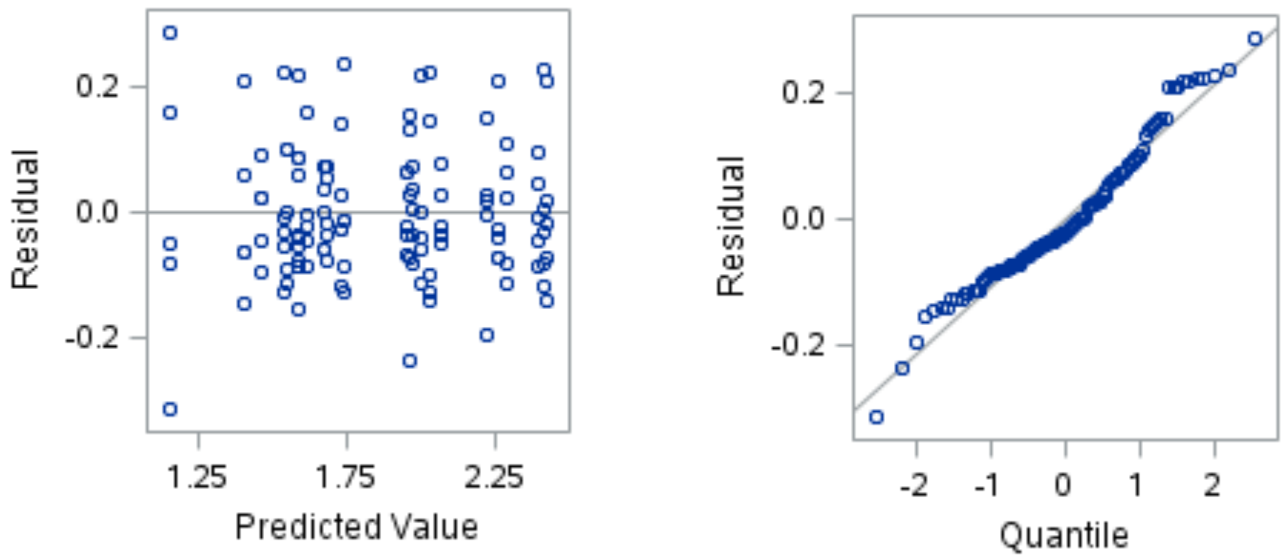
After transforming the response variable WiFi speed logarithmically, the new model for this analysis is $y_{ijkl} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + \xi_{ijkl}$, where:

- y_{ijkl} is the log WiFi speed for l^{th} observation from the $(i, j, k)^{th}$ treatment.
- μ is the baseline mean of log WiFi speed (log(Speed).
- τ_i, β_j, γ_k are the main effects for the factors obstacle, distance, and number of users on the mean of log(Speed), respectively.
- $\tau\beta_{ij}$, $\tau\gamma_{ik}$, $\beta\gamma_{jk}$ are the two-factor interaction effects for the interaction *obstacle * distance,obstacle * number of users*, and *distance * number of users*.
- $\tau\beta\gamma_{ijk}$ are the three factor interaction effects on the mean log(Speed) for the *obstacle * distance * number of users* interaction.
- ξ_{ijkl} is the random error of the l^{th} observation from the $(i, j, k)^{th}$ treatment, assumed that $\xi_{ijkl} \sim$ IID N(0, σ^2).

The hypotheses involving this study include the following:

- H_0: $\tau_1 = \tau_2 = 0$, H_a: $\exists\ \tau_i \neq 0$.
- H_0: $\beta_1 = \beta_2 = \beta_3 = 0$, H_a: $\exists\ \beta_j \neq 0$.
- H_0: $\gamma_1 = \gamma_2 = \gamma_3 = 0$, H_a: $\exists\ \gamma_k \neq 0$.
- H_0: $(\tau\beta)_{11} = (\tau\beta)_{12} = ... = (\tau\beta)_{23} = 0$, H_a: $\exists\ (\tau\beta)_{ij} \neq 0$.
- H_0: $(\tau\gamma)_{11} = (\tau\gamma)_{12} = ... = (\tau\gamma)_{24} = 0$, H_a: $\exists\ (\tau\gamma)_{ik} \neq 0$.
- H_0: $(\beta\gamma)_{11} = (\beta\gamma)_{12} = ... = (\beta\gamma)_{34} = 0$, H_a: $\exists\ (\beta\gamma)_{jk} \neq 0$.
- H_0: $(\tau\beta\gamma)_{111} = (\tau\beta\gamma)_{12} = ... = (\tau\beta\gamma)_{234} = 0$, H_a: $\exists\ (\tau\beta\gamma)_{ijk} \neq 0$.

Diagnostics plots were created for this new model (Figure 4). The Residuals vs. Fitted Value plot did not suggest severe violations of the constant variance assumption (though there is slightly more variability for the smallest fitted value). There are no clear outliers. Most of the residuals fall on a straight line in the Normal Q-Q plot, which suggests that there is no serve deviations from normality in the residual distribution. The model also assumes independence of the observations, which is met because we used Google Incognito tabs which don't store previous data, so that early runs could not affect later ones. So it is reasonable to assume that the experimental units in this study don't affect each other. All together, the results above suggest no serious violations of the model assumptions for using an ANOVA F-test.



(a) Residual vs. Fitted Value

(b) Normal Quantitle-Quantile (Normal Q-Q)

Figure 4: Diagnostics Plots (after using logarithm transformation)

4. Results

After checking the assumptions, an ANOVA F-test was conducted using the new model. Figure 5 below suggests the following results:

- There is strong evidence that the obstacles have effects on the mean of log(Speed). Thus, we reject the null hypothesis that H_0: τ_i = 0 (F(1, 96) = 21.32, p-value $<$ 0.0001, α=0.05).
- There is moderate evidence that the distance between WiFi router and WiFi users has effects on the mean of log(Speed). Thus, we reject the null hypothesis that H_0: β_j = 0 (F(2, 96) = 5.08, p-value = 0.008, α=0.05).
- There is strong evidence against the null hypothesis that there is no effect of the number of WiFi users on the mean of log(Speed). Hence, we reject the null hypothesis that H_0: γ_k = 0 (F(3,96) = 307.93, p-value $<$ 0.0001, α = 0.05).
- There is strong evidence that there is a Distance*Obstacle interaction effect on the mean of log(Speed). So we reject the null hypothesis that H_0: $(\tau\beta)_{ij}$ (F(2,96) = 19.43, p-value $<$ 0.0001, α = 0.05).

- There is no evidence that there is a Distance*Number_User interaction effect on the mean of log(Speed). So we fail to reject the null hypothesis that H_0: $(\beta\gamma)_{jk}$ (F(6,96) = 1.69, p-value = 0.131, α = 0.05).
- There is moderate evidence that there is a Obstacle*Number_User interaction effect on the mean of log(Speed). So we reject the null hypothesis that H_0: $(\tau\gamma)_{ik}$ (F(3,96) = 4.02, p-value = 0.01, α = 0.05).
- There is no evidence that there is a three-way interaction Distance*Obstacle*Number_User interaction effect on the mean of log(Speed). So we fail to reject the null hypothesis that H_0: $(\tau\beta\gamma)_{ijk}$(F(6,96) = 1.52, p-value = 0.1807, α = 0.05).

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Distance	2	0.14384173	0.07192086	5.08	0.0080
Obstacle	1	0.30169275	0.30169275	21.32	<.0001
Distance*Obstacle	2	0.54986030	0.27493015	19.43	<.0001
Number_Users	3	13.07116117	4.35705372	307.93	<.0001
Distance*Number_User	6	0.14369391	0.02394898	1.69	0.1310
Obstacle*Number_User	3	0.17061515	0.05687172	4.02	0.0097
Distan*Obstac*Number	6	0.12881242	0.02146874	1.52	0.1807

Figure 5: ANOVA Table

The table of 60 parameter estimates involved in this new model were stated in the Appendix. The R^2 = 0 of this model is 0.9144, which suggests that the model explains 91.44% of the variance of the response log(Speed). A post-hoc power analysis was then conducted using the square root of MSE (0.118952), the actual total design size (120), and the 24 mean estimates. Figure 6 below includes actual power for each term in the model.

Computed Power			
Index	Source	Test DF	Power
1	Distance	2	0.477
2	Obstacle	1	0.929
3	Distance*Obstacle	2	0.997
4	Number_Users	3	>.999
5	Distance*Number_Users	6	0.293
6	Obstacle*Number_Users	3	0.484
7	Distance*Obstacle*Number_Users	6	0.352

Figure 6: Post Hoc Power Analysis

The results suggest that this design has great powers in finding effects of Obstacle (power = 0.929), Distance*Obstacle (power = 0.997), and Number_Users (power $>$ 0.999) on the mean of log(Speed), if these effects actually exist. It has moderate power in detecting an effect of Distance (power = 0.477) and Obstacle*Number_Users (power = 0.484) on the mean of log(Speed). Although the power for Distance*Number_Users (power = 0.293), and for Distance*Obstacle*Number_Users (power = 0.352) are relatively lower than the powers for other terms in the model, these two powers are not extremely low. So it is appropriate to use this study design for detecting the effects of the factors of interest on the mean of log(Speed).

5. Scope of Inference

The Apple devices and WiFi router are not randomly selected, so inferences of the study results cannot extend beyond this sample. All three factors are fixed, so difference of the effects of these factors are limited within the range of factors in this study design. However, we did randomly assign the treatments to four different Apple devices, so we can state that the differences in the mean of log(Speed) were due to this random assignment of the three fixed factors obstacle, distance, and number of WiFi users.

Appendix

The Table of Parameter estimates

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	1.147929899	B	0.05319677	21.58	<.0001
Distance 10	0.434692071	B	0.07523159	5.78	<.0001
Distance 30	0.308617125	B	0.07523159	4.10	<.0001
Distance 50	0.000000000	B	.	.	.
Obstacle 0	0.440804467	B	0.07523159	5.86	<.0001
Obstacle 1	0.000000000	B	.	.	.
Distance*Obstacle 10 0	-0.487309810	B	0.10639353	-4.58	<.0001
Distance*Obstacle 10 1	0.000000000	B	.	.	.
Distance*Obstacle 30 0	-0.283691952	B	0.10639353	-2.67	0.0090
Distance*Obstacle 30 1	0.000000000	B	.	.	.
Distance*Obstacle 50 0	0.000000000	B	.	.	.
Distance*Obstacle 50 1	0.000000000	B	.	.	.
Number_Users 1	1.109954095	B	0.07523159	14.75	<.0001
Number_Users 2	0.813236136	B	0.07523159	10.81	<.0001
Number_Users 3	0.250233709	B	0.07523159	3.33	0.0012
Number_Users 4	0.000000000	B	.	.	.
Distance*Number_User 10 1	-0.269588340	B	0.10639353	-2.53	0.0129
Distance*Number_User 10 2	-0.366509684	B	0.10639353	-3.44	0.0008
Distance*Number_User 10 3	-0.148030189	B	0.10639353	-1.39	0.1673
Distance*Number_User 10 4	0.000000000	B	.	.	.
Distance*Number_User 30 1	-0.277534701	B	0.10639353	-2.61	0.0105
Distance*Number_User 30 2	-0.316719216	B	0.10639353	-2.98	0.0037
Distance*Number_User 30 3	-0.159064871	B	0.10639353	-1.50	0.1382
Distance*Number_User 30 4	0.000000000	B	.	.	.
Distance*Number_User 50 1	0.000000000	B	.	.	.
Distance*Number_User 50 2	0.000000000	B	.	.	.
Distance*Number_User 50 3	0.000000000	B	.	.	.
Distance*Number_User 50 4	0.000000000	B	.	.	.
Obstacle*Number_User 0 1	-0.286849282	B	0.10639353	-2.70	0.0083
Obstacle*Number_User 0 2	-0.399218888	B	0.10639353	-3.75	0.0003

Obstacle*Number_User 0 3	-0.109249236	B	0.10639353	-1.03	0.3071
Obstacle*Number_User 0 4	0.000000000	B	.	.	.
Obstacle*Number_User 1 1	0.000000000	B	.	.	.
Obstacle*Number_User 1 2	0.000000000	B	.	.	.
Obstacle*Number_User 1 3	0.000000000	B	.	.	.
Obstacle*Number_User 1 4	0.000000000	B	.	.	.
Distan*Obstac*Number 10 0 1	0.132286292	B	0.15046318	0.88	0.3815
Distan*Obstac*Number 10 0 2	0.383450293	B	0.15046318	2.55	0.0124
Distan*Obstac*Number 10 0 3	0.137288472	B	0.15046318	0.91	0.3638
Distan*Obstac*Number 10 0 4	0.000000000	B	.	.	.
Distan*Obstac*Number 10 1 1	0.000000000	B	.	.	.
Distan*Obstac*Number 10 1 2	0.000000000	B	.	.	.
Distan*Obstac*Number 10 1 3	0.000000000	B	.	.	.
Distan*Obstac*Number 10 1 4	0.000000000	B	.	.	.
Distan*Obstac*Number 30 0 1	0.238415541	B	0.15046318	1.58	0.1164
Distan*Obstac*Number 30 0 2	0.352282795	B	0.15046318	2.34	0.0213
Distan*Obstac*Number 30 0 3	0.139963012	B	0.15046318	0.93	0.3546
Distan*Obstac*Number 30 0 4	0.000000000	B	.	.	.
Distan*Obstac*Number 30 1 1	0.000000000	B	.	.	.
Distan*Obstac*Number 30 1 2	0.000000000	B	.	.	.
Distan*Obstac*Number 30 1 3	0.000000000	B	.	.	.
Distan*Obstac*Number 30 1 4	0.000000000	B	.	.	.
Distan*Obstac*Number 50 0 1	0.000000000	B	.	.	.
Distan*Obstac*Number 50 0 2	0.000000000	B	.	.	.
Distan*Obstac*Number 50 0 3	0.000000000	B	.	.	.
Distan*Obstac*Number 50 0 4	0.000000000	B	.	.	.
Distan*Obstac*Number 50 1 1	0.000000000	B	.	.	.
Distan*Obstac*Number 50 1 2	0.000000000	B	.	.	.
Distan*Obstac*Number 50 1 3	0.000000000	B	.	.	.
Distan*Obstac*Number 50 1 4	0.000000000	B	.	.	.

Note: The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable.