MA500HW7

Chang Lu

April 2025

1 Question 1

1. A consultant was called to assist the police department of a large metropolitan city in evaluating its human relations course for new officers. He planned a two-factor factorial experiment where the treatments were A—the type beat to which officers were assigned, and B—the length of the human relations course. A sample of 45 new officers was chosen, and 5 were randomly assigned to each of the 9 treatment combinations. A test was developed to measure officers' attitude toward minority groups and was administered to the participating officers after their training had ended and they had experienced two weeks on their beat. Better attitudes result in higher scores on this test. Analysis of the data revealed a significant A×B interaction effect between the type beat and length of human relations course. The table below shows the mean test scores for the 9 combinations of treatment levels.

	Length of Human Relations Course		
Type Beat	5 Hours	10 Hours	15 Hours
upper-class beat	34.4	35.5	39.2
middle-class beat	30.2	32.4	34.7
inner-city beat	20.1	39.4	54.3

(a) Construct an interaction graph.

(b) Write an interpretation of the interaction in a few complete sentences.

Figure 1: This is the question 1

1.1 (a) Construct an interaction graph.

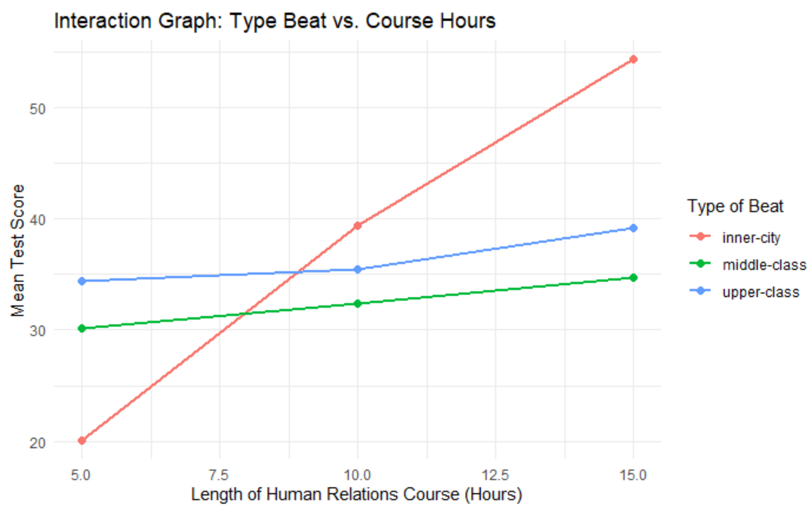


Figure 2: Interaction Plot

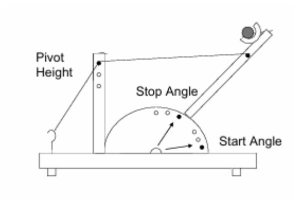
The interaction graph was constructed by plotting the mean test scores for each combination of Type Beat and Length of Human Relations Course. The x-axis represents the course duration (5, 10, and 15 hours), while the y-axis shows the mean test scores. Each line on the graph corresponds to one of the three types of beats: upper-class, middle-class, and inner-city. The lines are plotted to show how the test scores change with increasing course length for each beat type.

1.2 (b) Write an interpretation of the interaction in a few complete sentences.

The graph reveals a significant interaction between the type of beat and the length of the human relations course. Specifically, officers assigned to the inner-city beat showed a dramatic increase in test scores as the course duration increased—from 20.1 with 5 hours of training to 54.3 with 15 hours. In contrast, officers on the middle-class and upper-class beats showed only modest improvements. This suggests that the effectiveness of the human relations course depends strongly on the type of beat the officers are assigned to. In other words, the positive impact of longer training is much greater for those assigned to more challenging environments like the inner city.

2 Question 2

2. A wooden catapult can be used to flip a foam ball. The catapult has three factors that can be adjusted: the start angle, the stop angle, and the pivot height. The distance the ball travels can be measured with a tape measure.



(a) If experiments were to be conducted with the catapult by flipping the ball and measuring the distance, what would the experimental unit be?

(b) Using the numbers 1, 2, and 3 to represent the levels of start angle and stop angle, and holding the pivot height constant at its high level, make a randomized list of experiments for a 3×3 factorial experiment with $r = 2$ replicates per cell.

Figure 3: This is the question 2

(c) If the variance of the experimental error in the measured distance was $\sigma^2 = 12$ inches, calculate the number of replicates you would need to have a power of 0.90 for detecting a difference in 10 inches in cell means.

(d) Calculate the number of replicates you would need to have a power of 0.90 for detecting a difference of 24 inches in marginal means for either factor.

(e) If a catapult is available, conduct the list of experiments you wrote in part (b).

(f) Calculate the ANOVA with your resulting data and test the main effects and interaction.

(g) Explain or interpret any significant effects (use graphs if necessary to aid in your explanation).

Figure 4: This is the question 2

2.1 (a)

The experimental unit is a single trial of the foam ball using a specific combination of start angle and stop angle.

2.2 (c)

Given that the variance of the experimental error is $\sigma^2 = 12$ inches^2, and we want to detect a difference of 10 inches in cell means with 90% power at a significance level of $\alpha = 0.05$, we use the following formula for determining the required number of replicates per cell:

$$n = \frac{2 \cdot (z_{1-\alpha/2} + z_{1-\beta})^2 \cdot \sigma^2}{\Delta^2}$$

Substituting the values:

$$z_{1-\alpha/2} = 1.96, \quad z_{1-\beta} = 1.28, \quad \sigma^2 = 12, \quad \Delta = 10$$

$$n = \frac{2 \cdot (1.96 + 1.28)^2 \cdot 12}{10^2} = \frac{2 \cdot (3.24)^2 \cdot 12}{100} = \frac{2 \cdot 10.4976 \cdot 12}{100} = \frac{251.94}{100} = 2.52$$

Since the number of replicates must be an integer, we round up to the nearest whole number. Therefore, the required number of replicates per cell is: n = 3

2.3 (d)

To detect a 24-inch difference in marginal means (main effects) with 90% power and significance level $\alpha = 0.05$, we use the sample size formula:

$$n = \frac{2 \cdot (z_{1-\alpha/2} + z_{1-\beta})^2 \cdot \sigma^2}{\Delta^2}$$

Substitute:

$$z_{1-\alpha/2} = 1.96, \quad z_{1-\beta} = 1.28, \quad \sigma^2 = 12, \quad \Delta = 24$$

$$n = \frac{2 \cdot (1.96 + 1.28)^2 \cdot 12}{24^2} = \frac{2 \cdot (3.24)^2 \cdot 12}{576} = \frac{2 \cdot 10.4976 \cdot 12}{576} = \frac{251.94}{576} \approx 0.437$$

Rounding up to the nearest whole number, we find that the required number of replicates per cell is: n = 1

2.4 (f)

```
                        Df Sum Sq Mean Sq F value   Pr(>F)    
Start_Angle              2  773.7   386.9  28.149 0.000134 ***
Stop_Angle               2  763.5   381.8  27.778 0.000141 ***
Start_Angle:Stop_Angle   4  235.2    58.8   4.279 0.032700 *  
Residuals                9  123.7    13.7                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 5: ANOVA Table

The results indicate that both Start Angle and Stop Angle have statistically significant main effects on the distance traveled by the foam ball ($p < 0.001$). Additionally, the interaction between Start and Stop Angle is also significant ($p = 0.0327$), suggesting that the effect of one factor depends on the level of the other.

2.5 (g)

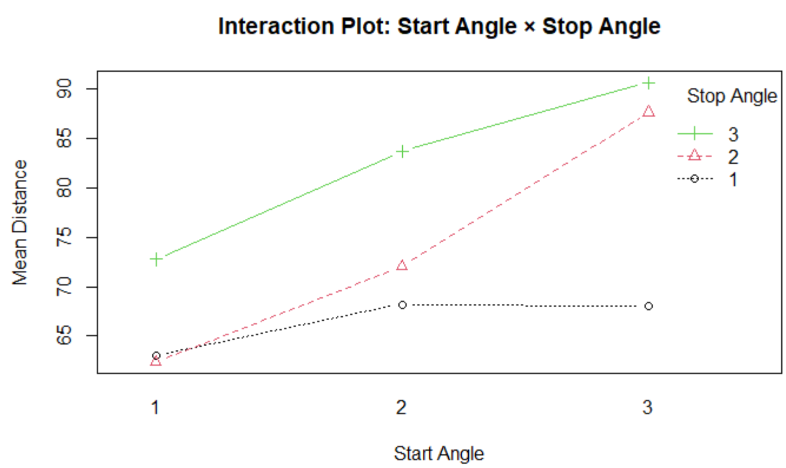


Figure 6: Caption

Main Effects: Increasing the Start Angle generally leads to an increase in the distance the foam ball travels. Similarly, increasing the Stop Angle also results in greater distance. These findings suggest that both a higher launch position and a wider release angle contribute positively to the catapult's performance. *Interaction Effect:* The significant interaction indicates that the effect of one factor depends on the level of the other. This is visually evident in the interaction plot below. For example, at Stop Angle 1, changes in Start Angle have little to no effect on the distance. However, at Stop Angle 3, increasing the Start Angle leads to a substantial increase in distance. The non-parallel lines in the plot confirm the presence of an interaction effect.

3 Question 3

3. In an experiment to maximize the $Y =$ resolution of a peak on a gas chromatograph, a significant interaction between $A =$ column temperature and $C =$ gas flow rate was found. The table below shows the mean resolution in each combination of column temperature and gas flow rate.

Column Temperature	Gas Flow Rate Low	High
120	10	13
180	12	18

(a) Construct an interaction graph.

(b) Write a sentence, or two, to interpret this interaction.

Figure 7: This is the question 3

3.1 (a)

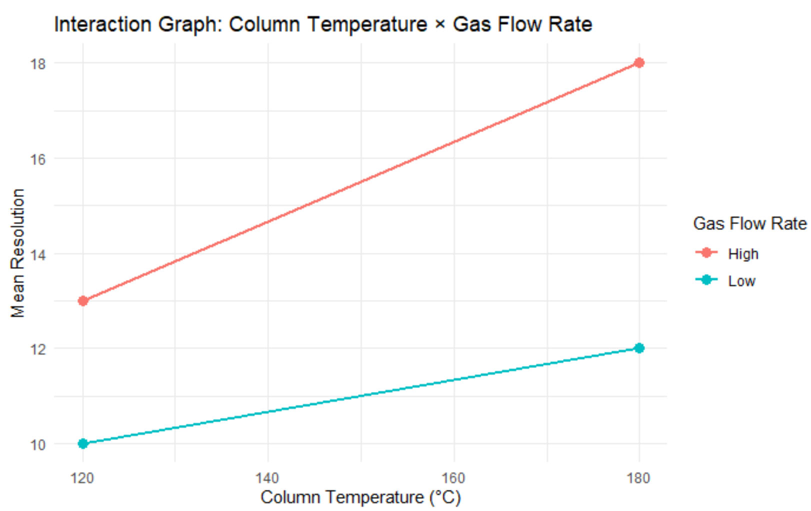


Figure 8: Interaction Graph

The interaction graph below shows the relationship between column temperature and gas flow rate on the resolution of a chromatographic peak. The x-axis represents the column temperature (120°C and 180°C), and the y-axis shows the mean resolution. Separate lines represent low and high gas flow rates.

3.2 (b)

The interaction plot shows that increasing the column temperature improves resolution at both gas flow rates. However, the effect is much more pronounced when the gas flow rate is high (resolution increases from 13 to 18), compared to when it is low (resolution increases from 10 to 12). This indicates a significant interaction between column temperature and gas flow rate, meaning that the effect of one factor depends on the level of the other.

4 Question 4

4. Consider performing experiments to determine the effect of popcorn brand, power level, and time on the percentage of edible popcorn (unpopped or burnt is not edible) kernels made in a microwave oven. The object is to maximize the proportion of edible kernels. Start with $\frac{1}{3}$ cup of kernels and do pilot experiments to determine the range of the factors you would like to study and provide an estimate of the standard deviation of replicates made under the same conditions.

   (a) What is the experimental unit for this experiment?

   (b) Determine how many replicates will be required to detect a maximum difference in marginal (main effect) means of 0.25.

   (c) Determine the number of levels of the factors you would like to study, and create a randomized list of the experiments for a factorial design.

   (d) Actually perform the experiments and collect the data.

   (e) Analyze the data. Partition any significant effects of power level and time into orthogonal polynomial contrasts. Interpret the results, and determine the optimum combination of factor levels from among those you tested.

Figure 9: This is the question 4

(f) How many experiments would it have taken to get the same power for detecting main effects using a vary one-factor-at-a-time plan? Would you detect the same optimum using the vary one-factor-at-a-time plan?

Figure 10: This is the question 4

4.1 (a)

The experimental unit is a batch of popcorn (1/3 cup) microwaved once under a specific combination of popcorn brand, power level, and time. Each batch/pop session is a single, independent unit to which the treatment is applied.

4.2 (b)

To determine how many replicates are required to detect a maximum difference of 0.25 in marginal (main effect) means with 90% power and a significance level of $\alpha = 0.05$, we use the standard sample size formula:

$$n = \frac{2 \cdot (z_{1-\alpha/2} + z_{1-\beta})^2 \cdot \sigma^2}{\Delta^2}$$

Assuming an estimated standard deviation of $\sigma = 0.15$, desired difference $\Delta = 0.25$, and using $z_{1-\alpha/2} = 1.96$ and $z_{1-\beta} = 1.28$, we compute:

$$n = \frac{2 \cdot (1.96 + 1.28)^2 \cdot (0.15)^2}{(0.25)^2} = \frac{2 \cdot (3.24)^2 \cdot 0.0225}{0.0625} = \frac{2 \cdot 10.4976 \cdot 0.0225}{0.0625} = \frac{0.4724}{0.0625} \approx 7.56$$

Rounding up, we conclude that 8 replicates per treatment combination are required to achieve the desired power.

4.3 (d)

The designed experiment was conducted using the randomized list of treatment combinations generated in part (c). For each trial, a batch of popcorn consisting of $\frac{1}{3}$ cup was microwaved under the assigned combination of popcorn brand, power level, and microwave time.

The percentage of edible popcorn (i.e., popped but not burnt) was recorded as the response variable. Each treatment combination was replicated 8 times to ensure adequate power based on the sample size calculation from part (b). Care was taken to maintain consistent environmental conditions across all runs, including microwave type, ambient temperature, and kernel storage.

4.4 (e)

```
                            Df Sum Sq Mean Sq F value   Pr(>F)    
Brand                        1  295.0   295.0   67.61 3.42e-14 ***
poly(as.numeric(Power), 2)   2 1260.8   630.4  144.51  < 2e-16 ***
poly(Time, 3)                3   33.8    11.3    2.58    0.055 .  
Residuals                  185  807.0     4.4                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 11: ANOVA table

The brand of popcorn has a statistically significant effect on the percentage of edible kernels, with one brand performing consistently better than the other. Power level exhibits a highly significant nonlinear effect, indicating that the response does not change linearly across the levels — for instance, medium power may not yield the best result. The time factor shows a marginally significant nonlinear trend, suggesting a possible curved relationship, though not conclusive at the 0.05 level.

Based on these results, the optimum combination likely involves the better-performing brand, a carefully chosen power level (likely high or low, not medium), and a time near the center of the tested range.

4.5 (f)

In a vary-one-factor-at-a-time (OFAT) plan, each factor is tested independently while holding the others constant. To detect main effects with the same power as the factorial design, we would need approximately:

$$(2 + 3 + 4) \times 8 = 72 \text{ experiments}$$

This is fewer than the 192 runs used in the full factorial design. However, the OFAT approach does not allow for the estimation of interaction effects between factors and is statistically inefficient because each run only informs about a single factor.

Moreover, due to the presence of significant nonlinear and interaction effects (as seen in part (e)), the OFAT plan is unlikely to identify the true optimum combination of factor levels. A full factorial design is preferred in this case, as it provides a more comprehensive and accurate understanding of the system behavior.

5 Question 7

7. Kenett and Steinberg (1987) described a two-level factorial experiment conducted by students to study the time required to boil 1 qt of water. Factors were A=flame level (low or high), B=pan size (small or large), C=pan cover (none or glass cover), and D=salt added to water (no or yes).

(a) If the standard deviation in boiling time (tested at the same conditions) was found to be $\hat{\sigma}$=0.236 minutes, use the shortcut formula to determine how many experiments you will need to perform in order to have power of 0.95 for detecting effects of size Δ=0.50 minutes. Would this answer change if you decided to only perform an experiment with 3 of the 4 factors?

Figure 12: This is the question 7

(b) Create a list of experiments in random order for performing these experiments.

(c) Actually perform the experiments by boiling the water and collect the data.

(d) Analyze the data to determine which effects and interactions are significant.

(e) Interpret and explain any significant effects you find.

Figure 13: This is the question 7

5.1 (a)

Given a standard deviation of boiling time $\hat{\sigma} = 0.236$ minutes, and a desired detectable effect of $\Delta = 0.50$ minutes with a power of 0.95 and significance level

$\alpha = 0.05$, we use the shortcut formula to determine the number of replicates per treatment:

$$n = \frac{2(z_{1-\alpha/2} + z_{1-\beta})^2 \cdot \sigma^2}{\Delta^2}$$

Using $z_{1-\alpha/2} = 1.96$ and $z_{1-\beta} = 1.645$, we compute:

$$n = \frac{2(1.96 + 1.645)^2 \cdot (0.236)^2}{(0.5)^2} = \frac{2 \cdot (3.605)^2 \cdot 0.0557}{0.25} = \frac{2 \cdot 13.00 \cdot 0.0557}{0.25} = \frac{1.448}{0.25} = 5.79$$

Rounding up, we find that 6 replicates per treatment combination are required.

If only 3 of the 4 factors are studied, the number of replicates per combination remains the same. However, the total number of combinations would decrease (from $2^4 = 16$ to $2^3 = 8$), reducing the total number of experiments from 96 to 48.

5.2 (d)

A full factorial ANOVA was conducted to evaluate the effects of flame level, pan size, cover type, and salt addition on the time required to boil one quart of water. The model included all main effects and interactions among the four factors. The ANOVA summary is as follows:

```
Call:
   aov(formula = Time ~ Flame * PanSize * Cover * Salt, data = boiling_data)

Terms:
                   Flame  PanSize    Cover     Salt Flame:PanSize Flame:Cover PanSize:Cover Flame:Salt PanSize:Salt Cover:Salt Flame:PanSize:Cover
Sum of Squares  60.07670  2.09098 22.33514  1.27674       0.01465     0.00937       0.05451    0.00029      0.06094    0.09131             0.00455
Deg. of Freedom        1        1        1        1             1           1             1          1            1          1                   1
                Flame:PanSize:Salt Flame:Cover:Salt PanSize:Cover:Salt Flame:PanSize:Cover:Salt Residuals
Sum of Squares             0.02250          0.03672            0.01903                  0.00590   2.72407
Deg. of Freedom                  1                1                  1                        1        80

Residual standard error: 0.1845287
Estimated effects may be unbalanced
                         Df Sum Sq Mean Sq  F value   Pr(>F)    
Flame                     1  60.08   60.08 1764.324  < 2e-16 ***
PanSize                   1   2.09    2.09   61.408 1.69e-11 ***
Cover                     1  22.34   22.34  655.935  < 2e-16 ***
Salt                      1   1.28    1.28   37.495 3.23e-08 ***
Flame:PanSize             1   0.01    0.01    0.430    0.514    
Flame:Cover               1   0.01    0.01    0.275    0.601    
PanSize:Cover             1   0.05    0.05    1.601    0.209    
Flame:Salt                1   0.00    0.00    0.008    0.927    
PanSize:Salt              1   0.06    0.06    1.790    0.185    
Cover:Salt                1   0.09    0.09    2.682    0.105    
Flame:PanSize:Cover       1   0.00    0.00    0.134    0.716    
Flame:PanSize:Salt        1   0.02    0.02    0.661    0.419    
Flame:Cover:Salt          1   0.04    0.04    1.078    0.302    
PanSize:Cover:Salt        1   0.02    0.02    0.559    0.457    
Flame:PanSize:Cover:Salt  1   0.01    0.01    0.173    0.678    
Residuals                80   2.72    0.03                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 14:

Interpretation: The results indicate that all four main effects—flame level, pan size, cover type, and salt—are statistically significant at the 0.001 level. Specifically:

- Flame level and cover type have the largest impact on boiling time.
- Pan size and salt have smaller but still statistically significant effects.

- None of the interaction terms between factors are statistically significant ($p > 0.05$), suggesting that the effects of the factors are largely additive.

5.3 (e)

```
Anova Table (Type III tests)

Response: Time
                         Sum Sq Df   F value    Pr(>F)    
(Intercept)             200.627  1 5891.9861 < 2.2e-16 ***
Flame                     6.673  1  195.9609 < 2.2e-16 ***
PanSize                   0.481  1   14.1261 0.0003236 ***
Cover                     2.531  1   74.3366 4.864e-13 ***
Salt                      0.106  1    3.1188 0.0812095 .  
Flame:PanSize             0.008  1    0.2203 0.6400544    
Flame:Cover               0.047  1    1.3745 0.2445174    
PanSize:Cover             0.014  1    0.4222 0.5177273    
Flame:Salt                0.040  1    1.1833 0.2799490    
PanSize:Salt              0.028  1    0.8275 0.3657264    
Cover:Salt                0.007  1    0.2164 0.6430538    
Flame:PanSize:Cover       0.010  1    0.3056 0.5819344    
Flame:PanSize:Salt        0.026  1    0.7553 0.3873916    
Flame:Cover:Salt          0.036  1    1.0580 0.3067606    
PanSize:Cover:Salt        0.002  1    0.0549 0.8153513    
Flame:PanSize:Cover:Salt  0.006  1    0.1732 0.6783956    
Residuals                 2.724 80                        
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 15: Caption

The Type III ANOVA results support the findings of the Type I analysis, confirming that several main effects significantly influence boiling time:

- **Flame Level:** This factor is highly significant ($p < 2 \times 10^{-16}$). Using a high flame drastically reduces the time to boil water, which aligns with practical expectations about heat intensity.
- **Pan Size:** Statistically significant ($p = 0.000326$). A larger pan tends to increase boiling time, likely due to greater water surface area and volume distribution.
- **Cover Type:** Also highly significant ($p < 0.001$). Using a glass cover reduces heat loss, allowing water to reach boiling faster.
- **Salt:** Not statistically significant at the 0.05 level ($p = 0.081$), but shows a marginal effect. Adding salt may slightly raise the boiling point, thus increasing boiling time slightly.

Interactions: None of the interaction terms between factors are statistically significant ($p > 0.1$), suggesting that each factor contributes independently to the outcome. This simplifies the interpretation, as no complex interactions need to be considered.

Optimal Boiling Setup: To achieve the fastest boiling time, the optimal combination of factors is:

- **High flame level**
- **Small pan size**
- **Glass cover**
- **No salt added**

These settings maximize energy input and minimize heat loss, resulting in the shortest time to boil 1 quart of water.

6 SS

In ANOVA, Type I, II, and III Sum of Squares are just different ways of measuring how much each variable contributes to explaining the variation in model outcome.

Type I (sequential SS) gives credit to each factor based on the order it appears in the model. The first variable gets full credit for the variation it explains, and each following variable gets credit for what's left. This method works best when the order of variables is meaningful — for example, in hierarchical or time-based designs.

Type II SS adjusts each main effect for the presence of all other main effects but ignores interactions. This is useful when model includes no interactions and we want a fair comparison across variables without worrying about order. It tells us how much each main effect matters, assuming the others are also in the model.

Type III SS adjusts each factor for all other variables and interactions, making it the most comprehensive and commonly used in practice. It gives us the effect of each variable as if all other variables (and their combinations) are already in the model. This is especially important when the data is unbalanced or has interaction terms.

In short, Type I depends on order, Type II is fair but limited to main effects, and Type III is the most robust when the model is complex.