# Contrast Analysis for Treatment Levels

Yuanrong Liu

2024-09-13

## Packages Loaded

```
library(ggplot2)
```

## Introduction

In this question, we assess the impact of three increasing treatment levels (I, II, III) compared to a control group (C). Our analysis focuses on two main aspects:

1. The overall effect of the treatments by comparing the control group with the average response of the treatment groups.
2. The incremental effects between the treatment levels, specifically between I and II, and II and III.

## Data

```
# Create a data frame with one observation per treatment level
df <- data.frame(
  Group = factor(c("C", "I", "II", "III"), levels = c("C", "I", "II", "III")),
  Response = c(10, 12, 15, 22)  # Hypothetical response values
)
```

## Methodology

In this analysis, we use two types of contrasts:

1. **Control vs. Average of Treatments**: A *custom contrast* is used to compare the control group (C) to the average of the treatment groups (I, II, III).
2. **Successive Differences Among Treatments**: *Successive-difference contrasts* are applied to compare Treatment II with Treatment I, and Treatment III with Treatment II.

```
C <- matrix(c(
  -1,  1/3,  1/3,  1/3,  # Control vs Average of I, II, III (custom contrast)
   0,  1,   -1,    0,    # I vs II (Successive-difference contrast)
   0,  0,    1,   -1     # II vs III (Successive-difference contrast)
), nrow = 4, byrow = TRUE)

contrasts(df$Group) <- C
```

## Analysis

```r
# Fit the linear model
model <- lm(Response ~ Group, data = df)
summary(model)
```

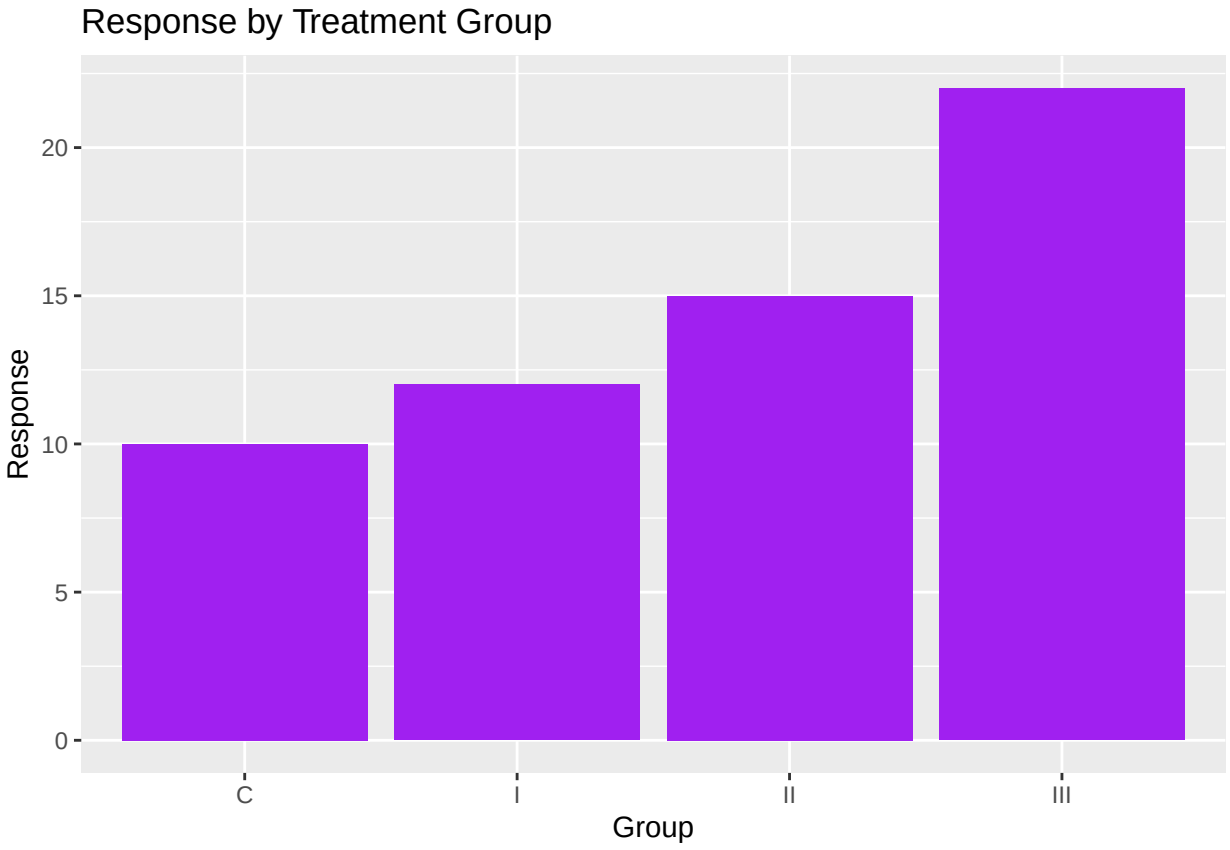
```
##
## Call:
## lm(formula = Response ~ Group, data = df)
##
## Residuals:
## ALL 4 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   19.364        NaN     NaN      NaN
## Group1         4.364        NaN     NaN      NaN
## Group2        -6.182        NaN     NaN      NaN
## Group3        -8.818        NaN     NaN      NaN
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 3 and 0 DF,  p-value: NA
```

**Explanation:**

1. **NaN Values**: There are `NaN` values in the model summary, for the standard errors, t-values, and p-values, due to the lack of residual degrees of freedom.
2. **Perfect Fit**: There are 4 observations (one for each level: C, I, II, III), and the linear model is trying to estimate 4 parameters (the intercept and three group contrasts), leaving no room for residual variation. This causes the standard errors of the coefficients to be `NaN`, as there is no uncertainty left to estimate.
3. **Focus on Coefficients**: The goal of the analysis is to interpret the **contrast estimates** (the model's coefficients), rather than relying on p-values or standard errors, which cannot be calculated here.

## Visualization

```r
# Visualize the response for each group using a bar plot
ggplot(df, aes(x = Group, y = Response)) +
  geom_bar(stat = "identity", fill = "purple") +
  labs(title = "Response by Treatment Group", x = "Group", y = "Response")
```




## Reference


- OpenAI, ChatGPT (2024). Assistance with R programming and contrast analysis. Accessed on September 13, 2024.
- Dr. Ben Bolker, Lecture Notes for "Statistical Modeling", McMaster University. Accessed on September 10, 2024.