anova

Table of contents

1 ANOVAs

1.1 Introduction to ANOVA

In statistics, when we want to compare **more than two group means**, we use **Analysis of Variance (ANOVA)**. While a t-test can compare two groups, running multiple t-tests inflates the Type I error rate. ANOVA allows us to test all groups simultaneously while controlling for error.

- **ANOVA compares the variance *between groups* to the variance *within groups*.**
- We use the **F-ratio** to evaluate whether observed group differences are larger than we'd expect due to random chance.

1.1.1 Hypotheses for a One-Way ANOVA:

- **Null hypothesis (H):** All group means are equal.
  1= 2= 3
- **Alternative hypothesis (H):** At least one group mean is different.
  (Not all s are equal, but doesn't specify which ones.)

This sets the stage for using statistical modeling to determine whether group membership explains a meaningful portion of the variability in the outcome.

1.2 Descriptive Statistics and Visuals

Before running ANOVA, it's important to **explore the data**.

Descriptive Tools

- xtabs(): Cross-tabulation to check how many observations are in each group.
- aggregate(): Calculates group means (or other summaries like SD).
- plotmeans() from the gplots package: Plots means with confidence intervals.

2 Setup

```
file_path <- file.path("..", "data", "data.rda")
df<-load(file_path)
df <- da04256.0001
remove(da04256.0001)
```

I will be examining the different focuses that facilities were doing when they received patients on drug abuse.

3 Descriptives

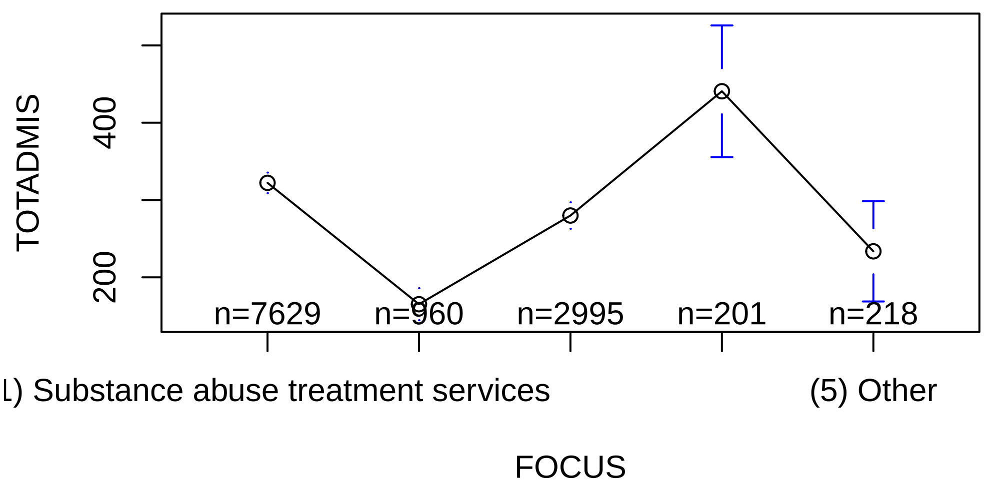
```
# Check number of cases in each group
xtabs(~FOCUS, data = df)
```

```
FOCUS
      (1) Substance abuse treatment services
                                        8340
                  (2) Mental health services
                                        1104
(3) Mix of mental health and substance abuse
                                        3558
                     (4) General health care
                                         219
                                   (5) Other
                                         233
```

```r
# Calculate group means
aggregate(TOTADMIS ~ FOCUS, data = df, mean)
```

```
                                         FOCUS TOTADMIS
1       (1) Substance abuse treatment services 322.2160
2                   (2) Mental health services 165.2427
3 (3) Mix of mental health and substance abuse 279.9559
4                      (4) General health care 440.7114
5                                    (5) Other 233.6376
```

```r
# Plot means with confidence intervals
plotmeans(TOTADMIS ~ FOCUS, data = df)
```



4 ANOVA

4.1 Create the model

```r
# Run the one-way ANOVA
model <- aov(TOTADMIS ~ FOCUS, data = df)
```

4.2 Assumptions

```r
leveneTest(TOTADMIS ~ FOCUS, data = df)
```

```
Levene's Test for Homogeneity of Variance (center = median)
         Df F value    Pr(>F)
group     4  13.042 1.341e-10 ***
      11998
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#sample size must be between 3 and 5000; I can't run this in my case.
#shapiro.test(residuals(model))
```

```r
# Summary of ANOVA table
summary(model)
```

```
               Df    Sum Sq Mean Sq F value Pr(>F)
FOCUS           4 2.734e+07 6834611   22.88 <2e-16 ***
Residuals   11998 3.584e+09  298685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
1451 observations deleted due to missingness
```

```r
# Calculate eta squared effect size
etaSquared(model)
```

```
           eta.sq eta.sq.part
FOCUS 0.007570966 0.007570966
```

4.3 Post-Hoc Anaylsis

```
# Post-hoc comparisons (pairwise t-tests)
posthocPairwiseT(model)
```

```
    Pairwise comparisons using t tests with pooled SD

data:  TOTADMIS and FOCUS

                                             (1) Substance abuse treatment services
(2) Mental health services                   5.5e-16
(3) Mix of mental health and substance abuse 0.00169
(4) General health care                      0.00967
(5) Other                                    0.05494
                                             (2) Mental health services
(2) Mental health services                   -
(3) Mix of mental health and substance abuse 1.2e-07
(4) General health care                      7.6e-10
(5) Other                                    0.19066
                                             (3) Mix of mental health and substance abuse
(2) Mental health services                   -
(3) Mix of mental health and substance abuse -
(4) General health care                      0.00038
(5) Other                                    0.22702
                                             (4) General health care
(2) Mental health services                   -
(3) Mix of mental health and substance abuse -
(4) General health care                      -
(5) Other                                    0.00064

P value adjustment method: holm
```