CWT Case Study Report

Dylan DeJean, Rose Shui, Simiao Dong

2023-3-3

Introduction

Educational television programs have existed for decades, aiming at turning what parents would otherwise consider a waste of time into something beneficial to children while still enjoyable. Many of these shows cover basic subjects also taught in school, exposing children to a variety of topics such as math, language, and critical thinking.

Sesame Street started airing on television networks in the 1970s, and quickly gained popularity in primarily toddlers who were still not quite old enough to go to school yet. The purpose of this analysis is to determine whether watching Sesame Street is actually beneficial to children, and if they were able to improve their test scores on various subjects.

Project Description

This study was conducted on 240 children ages 3 to 5 years. Tests on body parts, letters, geometric forms, numeracy, relational comparisons, and classification skills were administered before and after a period of watching the first season Sesame Street. A Peabody Picture Vocabulary Test was also administered before the children watched the show. The goal of this study is, for the use of Sesame Street producers, to determine if Sesame Street was effective in increasing cognitive ability between five populations of interest similarly, meaning that no population of interest is falling behind in test improvement compared to the other populations of interest. These populations cover multiple socioeconomic backgrounds which will be explored later. Unfortunately, not much is known of the methods in which the data was collected. This means that we can not determine causation about any statistical discovery since we do not know if the study was a randomized experiment.

Research Questions

- Question 1: Do children, after watching Sesame Street, perform similarly on cognitive tests regardless of their socioeconomic background?
- Question 2: Is there a significant difference in cognitive test improvement between disadvantaged children from rural areas (site 4) and advantaged children from rural areas (site 3)?

Variables

No.	Variable Name	Description
1	id	Subject identification number
2	site	Sampling sites corresponding to the five populations of interest
3	sex	Male (1), female (2)
4	age	Subject age in months
5	viewcat	Viewing frequency: (1) rarely; (4) more than 5 times per week
6	setting	Setting in which program was typically viewed: home (1) or school (2)
7	viewenc	A treatment condition in which some children were to be actively encouraged to watch Sesame Street (1) and others were not (2)
8	prebody	pretest on naming and functions of body parts (max score: 32)
9	prelet	pretest on letter recognition, naming, and sounds (max score: 58)
10	preform	pretest on recognition and name of geometric forms (max score: 20)
11	prenumb	pretest on numbers, counting, addition, and subtraction (max score: 54)
12	prelat	pretest on relational terms, e.g., amount, size, position (max score: 17)
13	preclasf	pretest on classification skills, e.g., size, form, number, etc (max score: 24)
14	postbody	posttest on body parts
15	postlet	posttest on letters
16	postform	posttest on geometric forms
17	postnumb	posttest on numeracy
18	postrelat	posttest on relational comparisons
19	postclasf	posttest on classification skills
20	peabody	Mental age scores obtained from administration of the Peabody Picture Vocabulary Test as a pretest measure of vocabulary maturity

For the variable site, these are the populations of interest as follows:

1. Three to five year old disadvantaged children from inner city areas
2. Four year old advantaged children from suburban areas
3. Advantaged children from rural areas
4. Disadvantaged children from rural areas
5. Disadvantaged children for whom Spanish is their first language

Response Variable

Improvement between pretest and postttest composite scores should be the response to measure Sesame Street's effectiveness in cognitive development. Using a composite score that is the mean proportion of correct answers across all of the cognitive tests would be the most interpretable and standardized option. This is because each test has its own scale which makes the mean proportion the most uniform interpretation of test scores. We will also be multiplying the composite scores by 100 to present them as a percent in an attempt to make the analysis as easily interpretable as possible.

Exploratory Data Analysis

Outliers and Incorrect Data Entry

There are multiple issues that have been found with the data that was collected. The subject with the id number 198 most likely did not participate in the posttests and their scores were recorded as 0. This is an issue for analysis and this individual's data should be removed. The 100th row in the data set has an id of

0. It can be reasonably assumed that this is a typo and should actually have an id of 100. Id 213 and 57 are out of bounds for post relational comparisons test, Id 5 is out of bounds for post letter test, and Id 163 is out of bounds for post body parts test. Since these are most likely typos and are incorrect data, the best course of action would be to remove these cases from the dataset. If we had a way to correct the mistakes that would be great, but that is not possible with the resources provided.

Correlation Structure

There are several issues that can impact the correlation structure of the collected data. Since children in this study can be siblings, that would mean that there can be correlation between cases for viewing frequency and viewing encouragement when watching Sesame Street at home. Also, if multiple children in this study happen to be in the same class with the same teacher, viewing encouragement and frequency can be correlated between those children if they are watching Sesame Street at school. All of this can cause problems with the ANOVA and regression model assumption of independence. We have no way of knowing if any of these proposed problems are true within the data, so interpretation of analysis has to be cautious since some cases can be dependent on each other.

Composite Score Plots

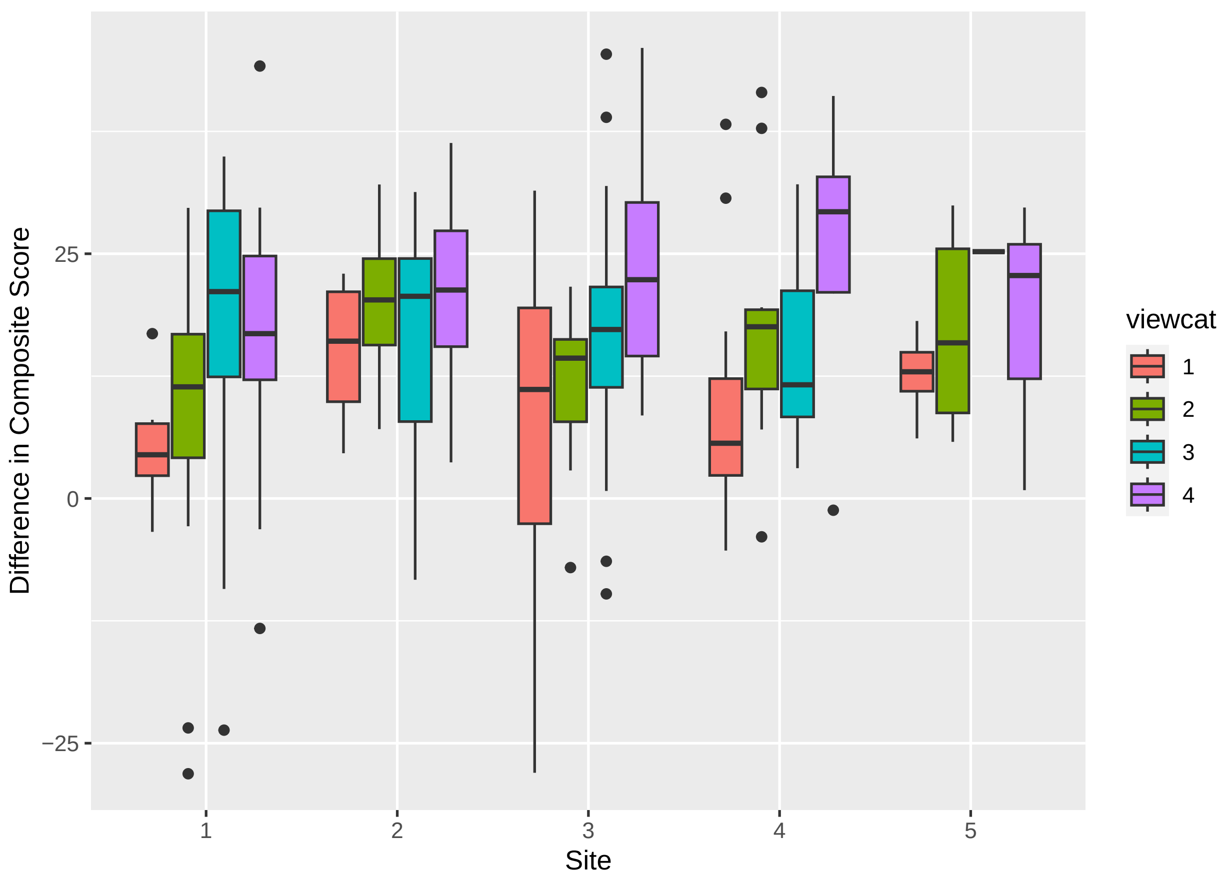


Figure 1: Difference in Composite Scores by Site

As we see in figure **??**, site 2 had the highest average difference in composite score, meaning that four year old advantaged children from suburban areas saw the highest amount of development. Site 1 and 4 had the least amount of development. Overall, it looks like all populations of interest showed similar development on the tests, however site 2 could be significantly higher. This is definitely something we should be looking for in analysis. Looking at advantaged vs. disadvantaged children in rural areas (sites 3 and 4 respectively), we see that advantaged children see a slightly higher level of average development. Whether that difference is statistically significant is yet to be seen.

Looking at the box plots in 1, we see a lot of the same information that is in the bar charts above. However it is important to note that there are some large outliers, specifically in the negative direction with site 1 and 3. Since we do not have a lot of information about data collection, we cannot conclude what might have caused these large drops in scores for those children or if they are data entry errors. Because of our lack of knowledge, it would unsafe to remove these data points for analysis purposes.

Analysis

For this study, we will be using an unusualness threshold of 0.1. Post-hoc analysis will be conducted by controlling FDR at the same error rate of 0.1.

Assumptions

Note: All plots can be found in the appendix.

Based on the QQ Plot of Residuals, the assumption that the residuals follow a Gaussian distribution appears to be satisfied. Although there are a couple of residuals that lie outside the envelope, it is not enough to violate the assumption. Taking a look at a strip chart of the residuals vs. fitted values, there doesn't seem to be a problem with homoscedasticity. There appears to be no pattern among the plots, meaning the homoscedasticity assumption is also satisfied. For this study, we will assume that there is independence of observations. There are some issues with correlation structure that may impact the independence of each response as we stated before, however for the purpose of building our model we will proceed cautiously under the assumption that the responses are independent. Based on our assumptions, we will proceed cautiously with the parametric ANOVA F=test.

Model

We will be fitting our ANOVA model with the composite difference score as a response of site, sex, setting, viewing frequency, age, and the possible interaction between all of these variables. To narrow down the model to the most important variables, a model will be fit with all possible predictors, and then step-by-step predictors will be removed until only the most important variables remain. Based on our research question, we will be adopting the following hypotheses for testing:

- H_0: Site is not a statistically significant variable in determining the composite difference score.
- H_A: Site is a statistically significant variable in determining the composite difference score.

For this study, we will adopt an unusualness threshold of 0.10.

Table 2: ANOVA Table for Difference in Composite Score Model

Source	SS	df	MS	F	p-value	Partial Omega Sq.	Partial Eta Sq.	Partial Epsilon Sq.
site	1335.8888	4	333.9722	2.3336	0.0566	0.0400	0.0221	0.0229
sex	17.2159	1	17.2159	0.1203	0.7290	0.0005	0.0000	0.0000
setting	211.8196	1	211.8196	1.4801	0.2250	0.0066	0.0020	0.0021
viewenc	591.0819	1	591.0819	4.1302	0.0433	0.0181	0.0131	0.0137
site:setting	1671.9411	3	557.3137	3.8942	0.0097	0.0496	0.0355	0.0368
sex:setting	1030.4281	1	1030.4281	7.2001	0.0078	0.0311	0.0256	0.0268
Residuals	32057.2118	224	143.1126					

The final model that we decided on predicted difference in composite test scores by site, sex, viewing setting, viewing encouragement, the interaction between site and viewing setting as well as the interaction between sex and viewing setting. As we see in table 2, the p-values for both site and viewing frequency are less than our unusualness threshold of 0.10, so we can reject our null hypothesis and say that these are statistically significant in determining the composite difference scores. We also find that the interaction between site and setting is significant as well as the interaction between sex and setting. This means that there is an effect on scores in which site and setting depend on each other as well as a separate effect in which sex and setting depend on each other. We can also see that Site accounts for about 2.334 times as much variation as our residuals. Despite site being significant, the effect sizes (partial Omega, Eta, and Epsilon squared) are very small. This indicates that the site groupings most likely do not affect score improvement very much.

Based on the coefficients of the model, which can be found in the appendix, we find that for sites 2 and 3, children show less development when viewing from home, while sites 1, 4, and 5 see greater development when viewing from home. Specifically, home viewers in site 2 improved by 2.73% less on average than in-school viewers and home viewers in site 3 improved by 0.56% less than in-school viewers on average. For sites 1, 4, and 5, home viewers improved 0.95%, 5.55%, and 5.55% more than their in-school counterparts on average respectively. Both males and females overall show more development when viewing from home compared to in school, but we find that males show less development than females in this setting by 2.39% on average.

Post-hoc

When examining the post-hoc pairwise comparisons, which can be found in the appendix, we are looking at the site 3 and 4 comparison mainly although other significant comparisons should be noted. The site 3 and 4 comparison is between advantaged children from rural areas and disadvantaged children from rural areas respectively. Since these are the only sites that have the same geographic setting, we can use this to gain insight to see if Sesame Street is affecting test scores based on economic status specifically. However, we see that in both viewing settings, the comparison between sites 3 and 4 are not significant, as their p-value is above the threshold of 0.1. This means that we fail to make a statistical discovery in which one site improves more or less on the tests than the other.

Other than the comparisons of sites 3 and 4, there are a few that were statistically significant. In site 4, students who watched Sesame Street at home improved test scores significantly more than their in-school counterparts by an average of 11.1%. Also, for in-school viewing, site 2 did significantly better than site 4 by an average of 12.3%.

Limitations

The nature in which the data for this study was collected presents some problems when trying to answer our research questions. Firstly, there may be problems with correlation structure and independence of observations. For example, if multiple subjects had the same teacher, or maybe took the tests in the same room as each other, or were siblings, that would cause problems with correlation structure and independence of observations. For the sake of this study, we assumed independence of observations because we have no other way of determining if there really was.

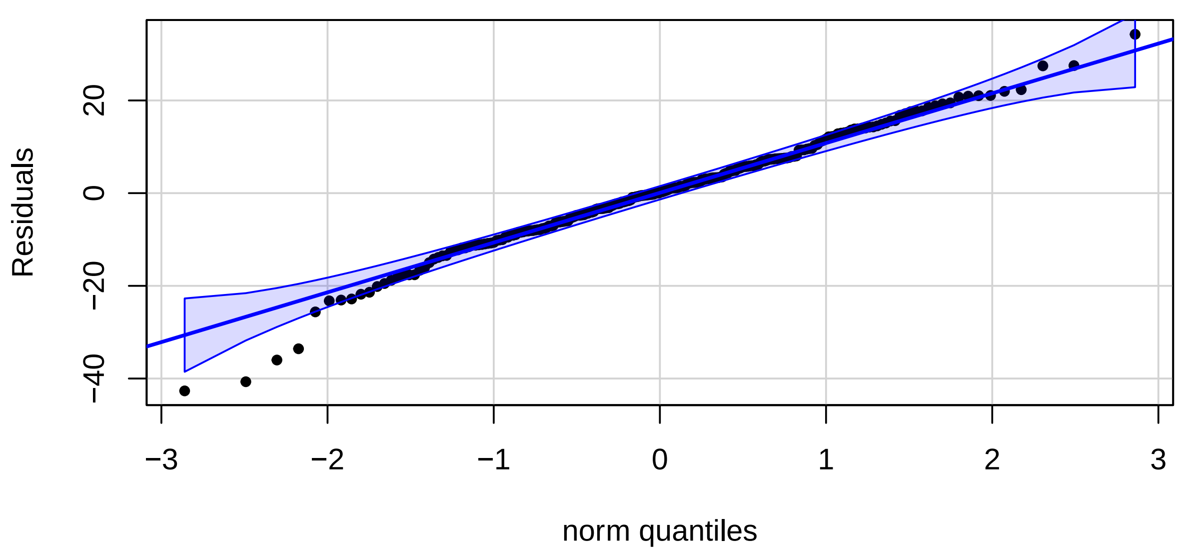
Another problem was with the data for the various test scores. The tests were all a different number of points, with the arithmetic test having the largest number of points possible. This may imply that there is more of an emphasis on arithmetic, or that perhaps the show was more catered towards these topics, but there is no way of telling. This makes comparing scores across different subjects difficult and inconsistent, forcing us to mutate the data and create a composite difference score.

Finally, when looking at our original ANOVA model for the composite difference scores, we found that site was a statistically significant variable. However looking at our Post-hoc analysis, we did not find any statistically significant discoveries for any pairs between the sampled sites. This may be because viewing frequency was another statistically significant variable in our model, and possibly affected scores more so than site. The two variables may be correlated, since it would be logical that different sites have different viewing frequencies based on socioeconomic status. However, there really is no way to tell.

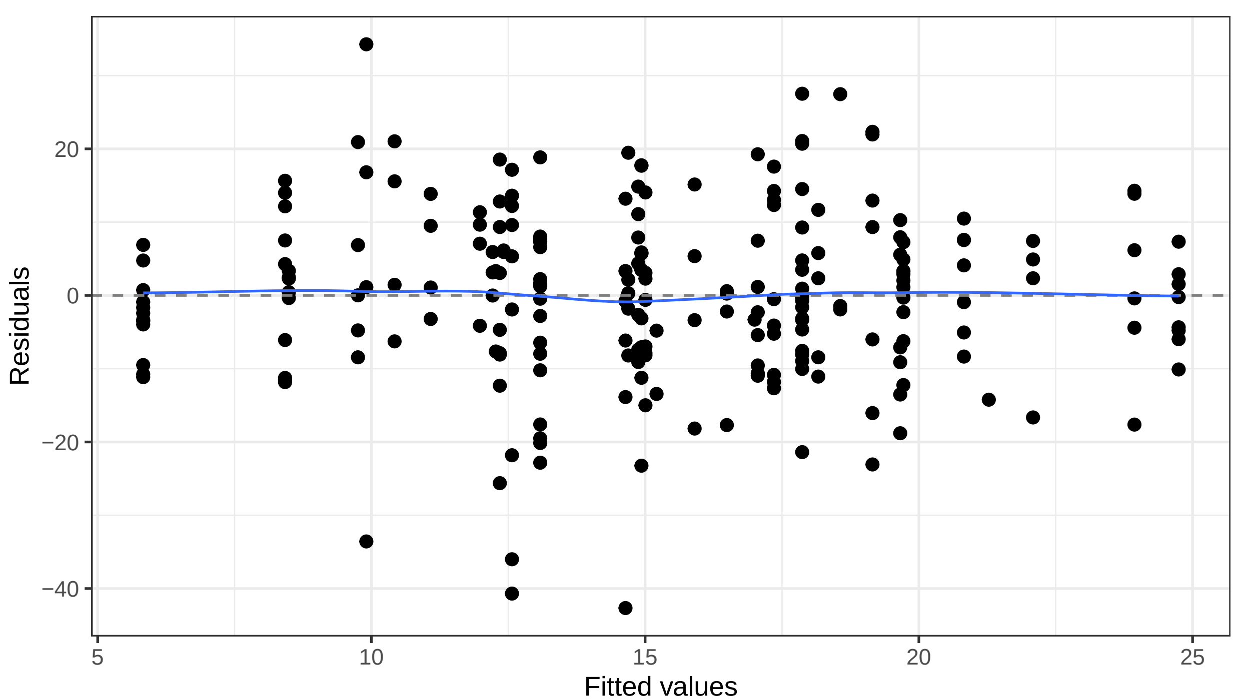
Conclusion

Despite the F-test showing that site is a significant variable in our ANOVA model, none of the site comparisons were significant and the effect sizes were almost negligible. Given this we can reasonably assume that socioeconomic status is not a significant factor in the improvement of test scores after viewing Sesame Street. We can also reasonably assume that there is no significant difference in test score improvement between advantaged children from rural areas and disadvantaged children in rural areas. This is good news, however, because it supports the claim that Sesame Street improves test scores regardless of a child's socioeconomic status.

Appendix



Residuals
20
0
−20
−40
−3
−2
−1
0
1
2
3
norm quantiles



Residuals
20
0
−20
−40
5
10
15
20
25
Fitted values

Comparison	Estimate	SE	DF	t Statistic	p-value
site1 setting1 - site2 setting1	-2.3648	3.1229	224	-0.7572	0.8095
site1 setting1 - site3 setting1	-0.5161	2.8879	224	-0.1787	1.0000
site1 setting1 - site4 setting1	-6.5853	3.6274	224	-1.8154	0.3540
site1 setting1 - site5 setting1	-2.3070	3.6232	224	-0.6367	0.8749
site1 setting1 - site1 setting2	1.9158	3.1869	224	0.6011	0.8813
site1 setting1 - site2 setting2	-7.8244	3.3299	224	-2.3498	0.1488
site1 setting1 - site3 setting2	-1.6425	3.6944	224	-0.4446	0.9856
site1 setting1 - site4 setting2	4.5069	3.4339	224	1.3124	0.5048
site1 setting1 - site5 setting2					
site2 setting1 - site3 setting1	1.8487	2.8100	224	0.6579	0.8749
site2 setting1 - site4 setting1	-4.2205	3.5753	224	-1.1805	0.5546
site2 setting1 - site5 setting1	0.0578	3.5662	224	0.0162	1.0000
site2 setting1 - site1 setting2	4.2805	3.1152	224	1.3741	0.4804
site2 setting1 - site2 setting2	-5.4596	3.2619	224	-1.6737	0.3787
site2 setting1 - site3 setting2	0.7223	3.6331	224	0.1988	1.0000
site2 setting1 - site4 setting2	6.8716	3.3673	224	2.0407	0.2729
site2 setting1 - site5 setting2					
site3 setting1 - site4 setting1	-6.0692	3.3433	224	-1.8154	0.3540
site3 setting1 - site5 setting1	-1.7909	3.3316	224	-0.5376	0.9177
site3 setting1 - site1 setting2	2.4318	2.9313	224	0.8296	0.7976
site3 setting1 - site2 setting2	-7.3083	3.0426	224	-2.4020	0.1488
site3 setting1 - site3 setting2	-1.1265	3.4533	224	-0.3262	1.0000
site3 setting1 - site4 setting2	5.0229	3.2158	224	1.5619	0.3848
site3 setting1 - site5 setting2					
site4 setting1 - site5 setting1	4.2783	3.9904	224	1.0721	0.6103
site4 setting1 - site1 setting2	8.5010	3.6233	224	2.3462	0.1488
site4 setting1 - site2 setting2	-1.2391	3.7410	224	-0.3312	1.0000
site4 setting1 - site3 setting2	4.9428	4.0721	224	1.2138	0.5546
site4 setting1 - site4 setting2	11.0921	3.8459	224	2.8841	0.0646
site4 setting1 - site5 setting2					
site5 setting1 - site1 setting2	4.2228	3.6343	224	1.1619	0.5546
site5 setting1 - site2 setting2	-5.5174	3.7415	224	-1.4746	0.4251
site5 setting1 - site3 setting2	0.6645	4.0763	224	0.1630	1.0000
site5 setting1 - site4 setting2	6.8139	3.8605	224	1.7650	0.3551
site5 setting1 - site5 setting2					
site1 setting2 - site2 setting2	-9.7401	3.3046	224	-2.9474	0.0646
site1 setting2 - site3 setting2	-3.5583	3.6604	224	-0.9721	0.6792
site1 setting2 - site4 setting2	2.5911	3.3532	224	0.7727	0.8095
site1 setting2 - site5 setting2					
site2 setting2 - site3 setting2	6.1818	3.8022	224	1.6259	0.3787
site2 setting2 - site4 setting2	12.3312	3.5391	224	3.4843	0.0267
site2 setting2 - site5 setting2					
site3 setting2 - site4 setting2	6.1494	3.8261	224	1.6072	0.3787
site3 setting2 - site5 setting2					
site4 setting2 - site5 setting2					

	Point Estimate
(Intercept)	10.4388
site1	2.2336
site2	8.2860
site3	4.2707
site4	4.2307
sex1	-0.2156
setting1	5.5461
viewenc1	1.3301
site1:setting1	-4.5882
site2:setting1	-8.2759
site3:setting1	-6.1093
site4:setting1	
sex1:setting1	-2.1770

```
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center",
  fig.height = 5, fig.width = 7
)

options(knitr.kable.NA = "")
options("contrasts" = c("contr.sum", "contr.poly"))

# clean up & set default chunk options
rm(list = ls())
knitr::opts_chunk$set(echo = FALSE)

# packages
library(tidyverse)
library(knitr)
library(kableExtra)

# inputs
data <- read.csv('./sesame.csv')
data$site <- as.factor(data$site)
data$viewcat <- as.factor(data$viewcat)
data$sex <- as.factor(data$sex)
data$setting <- as.factor(data$setting)
data$viewenc <- as.factor(data$viewenc)
colnames(data)[1] <- gsub('^...','',colnames(data)[1])
# Percentages
data$percentPrebody <- data$prebody/32
data$percentPrelet <- data$prelet/58
data$percentPreform <- data$preform/20
data$percentPrenumb <- data$prenumb/54
data$percentPrerelat <- data$prerelat/17
data$percentPreclasf <- data$preclasf/24
data$percentPostbody <- data$postbody/32
data$percentPostlet <- data$postlet/58
```

```r
data$percentPostform <- data$postform/20
data$percentPostnumb <- data$postnumb/54
data$percentPostrelat <- data$postrelat/17
data$percentPostclasf <- data$postclasf/24

# Difference of percentages
data$diffBody <- (data$percentPostbody-data$percentPrebody)
data$diffLet <- (data$percentPostlet-data$percentPrelet)
data$diffForm <- (data$percentPostform-data$percentPreform)
data$diffNumb <- (data$percentPostnumb-data$percentPrenumb)
data$diffRelat <- (data$percentPostrelat-data$percentPrerelat)
data$diffClasf <- (data$percentPostclasf-data$percentPreclasf)

# Composite Scores
data$compositePre <- ((data$percentPrebody+data$percentPrelet+data$percentPreform+
                        data$percentPrenumb+data$percentPrerelat+data$percentPreclasf)/6)*100
data$compositePost <- ((data$percentPostbody+data$percentPostlet+data$percentPostform+
                         data$percentPostnumb+data$percentPostrelat+
                          data$percentPostclasf)/6)*100
data$compositeDiff <- (data$compositePost - data$compositePre)
# Fixing errors in data
data$id[data$id == 0] <- 100
data <- data[-c(5, 57, 163, 213),]
data%>%
  ggplot(aes(x=site, y=compositeDiff))+
  geom_boxplot(aes(fill = viewcat))+
  ylab("Difference in Composite Score")+
  xlab("Site")
step1 <- aov(compositeDiff ~ site + sex + setting + viewenc + age
             + site:setting + site:viewenc + setting:viewenc + site:sex + setting:sex
             + viewenc:sex + age:site + age:sex + age:setting + age:viewenc, data=data)
step2 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               site:viewenc + site:sex + setting:sex + viewenc:sex + age:site +
               age:sex + age:setting + age:viewenc, data=data)
step3 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               site:viewenc + site:sex + setting:sex + viewenc:sex + age:site +
               age:setting + age:viewenc, data=data)
step4 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               site:viewenc + setting:sex + viewenc:sex + age:site + age:setting +
               age:viewenc, data=data)
step5 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               site:viewenc + setting:sex + viewenc:sex + age:setting +
               age:viewenc, data=data)
step6 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               setting:sex + viewenc:sex + age:setting + age:viewenc, data=data)
step7 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               setting:sex + viewenc:sex + age:setting, data=data)
step8 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               setting:sex + age:setting, data=data)
step9 <- aov(compositeDiff ~ site + sex + setting + viewenc + age + site:setting +
               setting:sex, data=data)
step10 <- aov(compositeDiff ~ site + sex + setting + viewenc + site:setting +
                setting:sex, data=data)
```

```r
model <- aov(compositeDiff ~ site + sex + setting + viewenc + site:setting + setting:sex, data=data)
parameters::model_parameters(
  model = model,
  effectsize_type = c("eta", "omega", "epsilon"),
  type = 1,
  drop = "(Intercept)",
  verbose = FALSE
  ) %>%
  dplyr::mutate(
  p = ifelse(
  test = is.na(p),
  yes = NA,
  no = p
  )
) %>%
  knitr::kable(
  digits = 4,
  col.names = c("Source", "SS", "df", "MS", "F", "p-value",
                "Partial Omega Sq.", "Partial Eta Sq.", "Partial Epsilon Sq."),
  caption = "ANOVA Table for Difference in Composite Score Model",
  align = c('l',rep('c',8)),
  booktab = TRUE,
) %>%
  kableExtra::kable_styling(
  bootstrap_options = c("striped", "condensed"),
  font_size = 12,
  latex_options = c("scale_down", "HOLD_position")
)
car::qqPlot(
  x = residuals(model),
  distribution = "norm",
  envelope = 0.90,
  id = FALSE,
  pch = 20,
  ylab = "Residuals"
)
ggplot(
  data = data.frame(
    residuals = residuals(model),
    fitted = fitted.values(model)
  ),
  mapping = aes(x = fitted, y = residuals)
) +
  geom_point(size = 2) +
  geom_hline(
    yintercept = 0,
    linetype = "dashed",
    color = "grey50"
  ) +
  geom_smooth(
    formula = y ~ x,
    method = stats::loess,
    method.args = list(degree = 1),
```

```
    se = FALSE,
    linewidth = 0.5
  ) +
  theme_bw() +
  xlab("Fitted values") +
  ylab("Residuals")
sitePost <- emmeans::emmeans(
  object = model,
  specs = pairwise ~ site:setting,
  adjust = "bh",
  level = 0.9
)

as.data.frame(sitePost$contrasts) %>%
  knitr::kable(
  digits = 4,
  col.names = c("Comparison", "Estimate","SE", "DF",
                "t Statistic", "p-value"),
  #align = rep("c", 6),
  booktabs = TRUE
) #%>%
 # kableExtra::kable_styling(
 # latex_options = c("striped","scale_down")
#)
as.data.frame(model$coefficients)%>%
  knitr::kable(
    digits=4,
    col.names=c('Point Estimate'),
    #caption="Point Estimates for Model",
    #align = c('l',rep('c',8)),
    booktab = TRUE#)%>%
  #kableExtra::kable_styling(
  #bootstrap_options = c("striped", "condensed"),
  #font_size = 12
  #latex_options = c("scale_down", "HOLD_position")
)
```