

Putting Units 1-4 Into Practice

Neil J. Hatfield

March 19, 2021

We are going to work through the following situation to put into practice everything that we've covered in Units 1-4. To assist us, I've created this R Markdown document with placeholders for R code and narrative interpretations.

The Context

We've been hired again, this time by the Kenton Food Company. They wish to test four different package designs for a new breakfast cereal. Using a lottery, they have identified twenty stores, all with approximately equal sales volumes and comparable locations, to serve as test sites. Outside of the package design, all twenty stores have agreed to keep other conditions which might impact sales (e.g., price, amount and location of shelf space, special promotions, etc.) at pre-determined fixed states. They have come to us for assistance to ensure that they can find out which design leads to the higher sales (measured in number of cases sold by each store), including whether 3-color printing (Designs A and B) is different from 5-color printing (Designs C and D) and whether including cartoons (Designs A and C) vs. no cartoons (Design B and D) impact sales figures.

What are the Statistical Research Questions?

1. Does the cereal box design impact the sales?
2. Which designs are statistically different from the others in terms of sales?
3. Is there a statistically significant difference in sales between designs which use 3-color printing and those that use 5-color printing?
4. Is there a statistically significant difference in sales between designs which contain cartoons and those that do not?

Study Design

With our SRQ in mind, let us design a study which will allow us to answer the SRQ.

Key Elements

- Response: Sales of cereal
 - Unit of Measure: cases sold
- Measurement Unit: stores
- Factor: Box design (fixed, 4 levels)
- Experimental Design:
 - 1) Experimental Units: stores
 - 2) Treatments: Designs A, B, C, and D
 - 3) Treatment Assignment Method: we'll use a lottery method
- What kind of design? Completely Randomized Design/One-way [Layout] ANOVA

Study Design Description

From the context, we will note that Kenton has already used one lottery to select 20 stores which are comparable in terms their typical sales volumes and locations. Further, they have all agreed to use the same, fixed conditions to ensure our study of the impact of box design on sales remains unconfounded with other attributes.

Given that we have four particular designs to look at, we will use the following lottery system to assign a design to a store. Place the names of all 20 stores onto 20 otherwise identical index cards. Turn the cards over and thoroughly shuffle the deck. Deal out the cards into a four by five grid. The first row will receive Design A, second B, third C, and the fourth will get Design B.

Exploratory Data Analysis

Table 1: Summary Statistics for Cereal Sales by Design

	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
A	5	11	14.00	15	16.00	17	1.483	14.6	2.302	-0.496	-1.539
B	5	10	11.00	12	15.00	19	2.965	13.4	3.647	0.511	-1.688
C	4	17	17.75	19	20.75	23	2.224	19.5	2.646	0.324	-2.009
D	5	22	26.00	27	28.00	33	1.483	27.2	3.962	0.171	-1.486

Table 1 shows the values of several descriptive statistics, broken down by the four cereal box designs. One of the first things to notice, is even though we used 20 stores, we only have the data for 19. One store (Store 15) is missing sales figures. Upon investigating, we learned that Store 15 suffered an unfortunate fire and could not continue in the study. Fortunately, we still have sufficient degrees of freedom to continue even without this store.

The second thing to notice is that the sales figures for Designs C and D are markedly higher than Designs A and B. The values of the *Sample Minimum* for C and D are essentially the same as the values of the *Sample Maximum* for A and B. Further, Design D has just over twice the sales performance of Design B; 27.2 cases/store vs. 13.4 cases/store, respectively. While all stores have slight skewness, Store A possess a negative skewness while the others are positively skewed. This might indicate a problem with the Gaussian assumption later one.

Figure 1 shows the box plots for the sales figures broken out by design. We can see that Designs C and D appear to be out performing Designs A and B. Another interesting feature we can see in the box plots is that Designs B and C are more dense in the lower halves than their upper halves. Design A is more dense in the upper half while Design D is most dense in the middle half. This might be a quirk of the data we have or might be a first sign of the impact that design has on sales.

The shadowgram (Figure 2) shows us that while there is a fair bit of variation in the data (the solid band of blue along the bottom), there are some clusters of values: a large set between 10 and 25 cases sold, and another between 25 and 30 cases sold. This would suggest that there might be two distinct groupings of sales.

Statistical Inference Methods

Given this design, the CRD/One-way Layout is an appropriate method to analyze the data. Figure 3 shows a Hasse diagram for this study design. The single factor (design) fits well within the additive model and we have sufficient degrees of freedom to estimate our parameters.

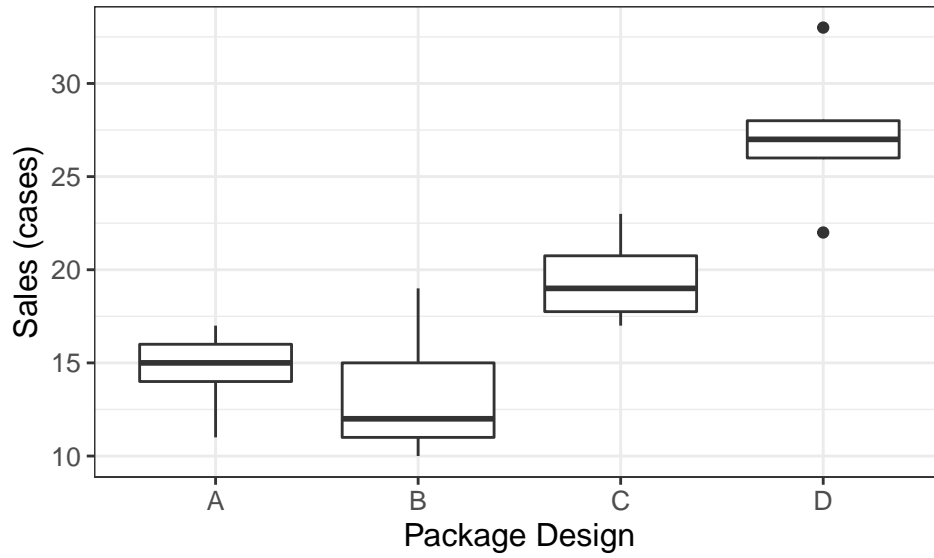


Figure 1: Box plots of Sales by Design

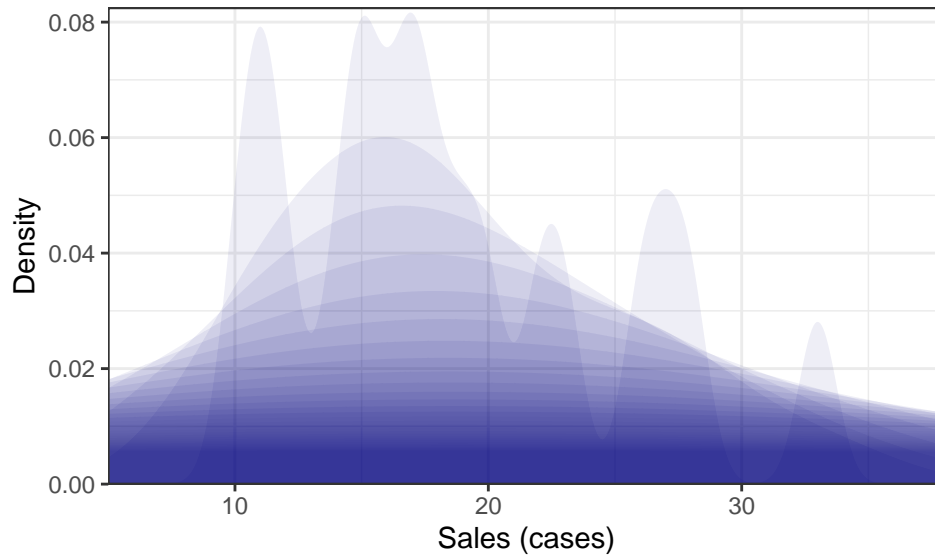


Figure 2: Shadowgram of Sales

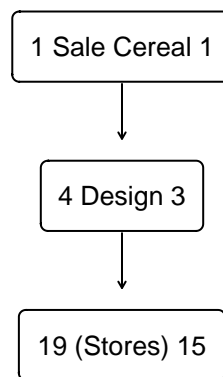


Figure 3: Hasse Diagram for Cereal Box Design Study

Hypotheses

Given that we have four statistical research questions, we will have multiple hypotheses.

First, for our omnibus question (“Does the cereal box design impact the sales?”), we will test the hypotheses

$$\begin{aligned}H_{1,0} : y_{ij} &= \mu_{..} + \epsilon_{ij} \\ H_{1,A} : y_{ij} &= \mu_{..} + \alpha_i + \epsilon_{ij}\end{aligned}$$

where there is at least one α_i not equal to zero.

For our pairwise comparisons (question 2), we will six sets of hypotheses of the form $H[0]$: there is no statistical difference between the two designs versus $H[1]$: there is a statistical difference between the two designs.

For the third question (“Is there a statistically significant difference in sales between designs which use 3-color printing and those that use 5-color printing?”), we will use

$$\begin{aligned}H_{3,0} : \mu_{3 \text{ color}} &= \mu_{5 \text{ color}} \\ H_{3,1} : \mu_{3 \text{ color}} &\neq \mu_{5 \text{ color}}\end{aligned}$$

The fourth question (“Is there a statistically significant difference in sales between designs which contain cartoons and those that do not?”) will be similar with hypotheses of

$$\begin{aligned}H_{4,0} : \mu_{\text{cartoon}} &= \mu_{\text{no}} \\ H_{4,1} : \mu_{\text{cartoon}} &\neq \mu_{\text{no}}\end{aligned}$$

Method

To decide between the hypotheses and answer our research question, we will use the parametric shortcut, the One-way ANOVA F test. To account for the multiple comparison problem, we will conceptualize two testing families: the omnibus and the pairwise comparisons; and two contrasts. We make this choice for easy of use of R.

Further, we will control the Simultaneous Confidence Intervals Type I Error Rate, with an overall Type I risk of 0.1. We will achieve this by using the Tukey HSD method for the pairwise comparisons and the Scheffé method for the contrasts. We will use an overall Unusualness Threshold of 0.1.

Assumptions

There are three main assumptions that we need to examine for using the parametric shortcut: residuals follow a Gaussian distribution, homoscedasticity, and independence of observations.

While do not know the measurement order, there is little in the design that suggests that the 20 stores who participated in the study directly impacted one another’s sales. Thus, we will decide to act as if the independence of observations assumption is met.

The QQ plot (Figure 4) reveals that while there is one point beyond the envelope, ~5% of the observations, there is little to be concerned about for the Gaussian assumption.

For the homoscedasticity assumption, the strip chart in Figure 5, shows that we have pretty similar amounts of variation (vertical length of each strip of dots) for each of the fitted values (levels of our factor). Thus, we may state that the homoscedasticity assumption is satisfied.

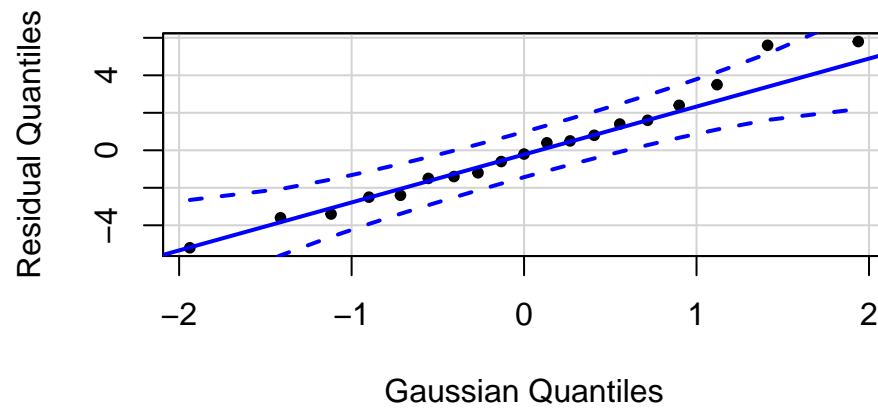


Figure 4: QQ Plot for Residuals from the Cereal Design Study

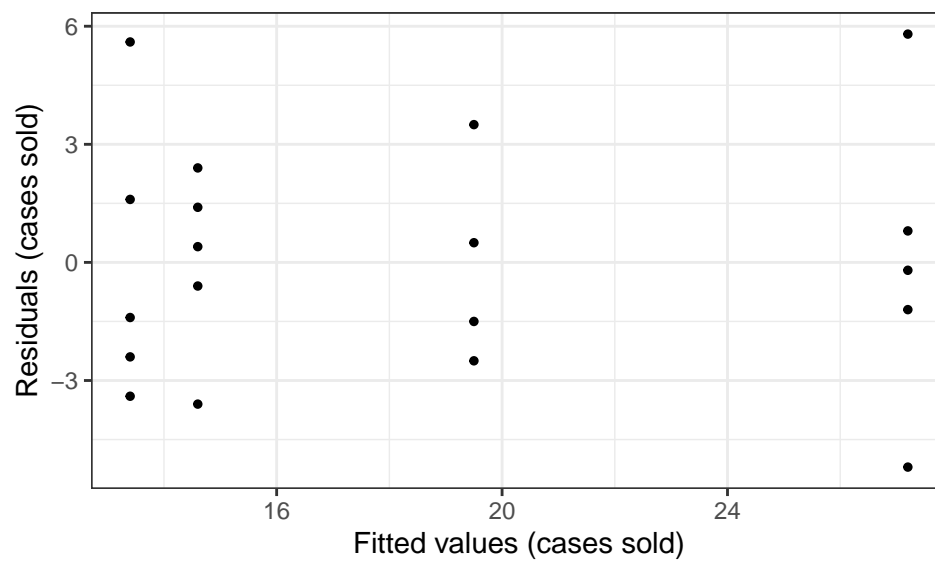


Figure 5: Strip Chart of the Residuals from the Cereal Design Study

Results

Table 2: ANOVA Table for Cereal Package Design Study

Source	SS	df	MS	F	p-value	Omega Sq.	Eta Sq.	Epsilon Sq.
design	588.2211	3	196.0737	18.5911	< 0.0001	0.7353	0.7881	0.7457
Residuals	158.2000	15	10.5467					

Table 2 provides a modern ANOVA table for our omnibus test. We can see that we have an F ratio of 18.59, which indicates that the design captured approximately 18.59 times as much variation as our residuals. The p -value is less than our Unusualness Threshold (0.1), so we will reject the null hypothesis and decide to act as if the design does impact sales. Further, our model explains a larger proportion of the variation in sales at around 75% of the total.

Figure 6 shows confidence intervals for questions 2-4; quickly looking through, we can see that 0 (the under the null hypothesis for each test, vertical red line) is contained in three of the intervals (A,C – B,D; B-A; C-A), indicating that those three comparisons result in non-statistically significant differences. With our adjustment for the multiple comparison problem, we can be confident that if we repeat the experiment again and again, then 90% of the time we will capture all 8 of the true differences for the post hoc comparisons (6 pairwise and two contrasts).

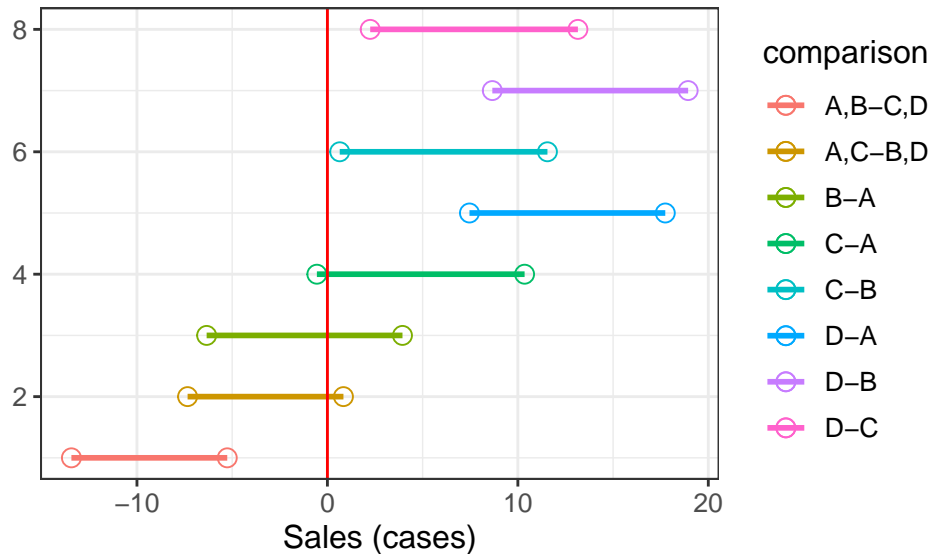


Figure 6: 90% Simultaneous Confidence Intervals for Cereal Box Design Study

Table 3: Post Hoc Tukey HSD Comparisons

	Difference	Lower Bound	Upper Bound	Adj. p-Value
B-A	-1.2	-6.3412	3.9412	0.9353
C-A	4.9	-0.5530	10.3530	0.1549
D-A	12.6	7.4588	17.7412	0.0001
C-B	6.1	0.6470	11.5530	0.0583
D-B	13.8	8.6588	18.9412	< 0.0001
D-C	7.7	2.2470	13.1530	0.0142

For our second research question, we can examine the pairwise comparisons listed in Table 3. Comparing the six adjusted p -values to our Unusualness Threshold of 0.1, we can see that the last four comparisons are statistically significant. This indicates that Design D is statistically different from the others (more sales, given the values of the difference column). Design C is statistically different from Design B (more sales) as well as D (previously mentioned). However, Designs A and B are not statistically different from one another, nor are Designs A and C.

We can see rather large effects sizes at play in these pairwise comparisons looking at Table 4. The last four rows all indicate more than 1.5 standard deviations of separation (Cohen's d and Hedge's g) as well as extreme Probabilities of Superiority. For example, for B vs. D, the Probability of Superiority indicates that if we were to pick a store that got design B and a store that got Design D via a lottery, then the store with B will produce the higher sales just half of one percent of the time. Taken together, we can take this as indicating that Design D has the best observed performance of the four designs.

Table 4: Post Hoc Comparison Effect Sizes

Pairwise Comparison	Cohen's d	Hedge's g	Prob. Superiority
A vs. B	0.393	0.355	0.610
A vs. C	-1.996	-1.774	0.079
A vs. D	-3.888	-3.512	0.003
B vs. C	-1.874	-1.665	0.093
B vs. D	-3.624	-3.273	0.005
C vs. D	-2.225	-1.978	0.058

For our last two questions, we made use of linear contrasts. Recall that Designs A and B used 3-color printing while Designs C and D used 5-color printing. From Table 5, the first row indicates we should reject the null hypothesis and decide to act as if there is a sales difference between 3- and 5-color printing.

For including cartoons (Designs A and C) or not (Designs B and D), we will fail to reject the null and decide to act as if there is not a difference in sales due to using cartoons. (Second row of Table 5)

Table 5: Scheffe Test Results

	Difference	Lower Bound	Upper Bound	p-value
A,B-C,D	-9.35	-13.4415	-5.2585	0.0002
A,C-B,D	-3.25	-7.3415	0.8415	0.2379

Discussion

From our work, we can confirm that the design used on the cereal box impacts the sales figures. In particular Design D is statistically different from the other designs, and appears to lead to higher sales figures. Further research would be needed to state clearly that Design D is the best. Further, we found out that in general, 5-color printing tends to lead to higher sales but the inclusion of cartoons does not appear to matter. Future work should aim to confirm that Design D is the best as well as attempt to establish the underlying mechanism for why Design D is the best.

Code Appendix

```
# Setting Document Options
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center"
)

# Add additional packages by name to the following list
packages <- c("tidyverse", "knitr", "kableExtra",
             "hasseDiagram", "car", "psych", "parameters",
             "DescTools")

lapply(
  X = packages,
  FUN = library,
  character.only = TRUE
)

# Set constraint
options("contrasts" = c("contr.sum", "contr.poly"))

options(knitr.kable.NA = "")
source("https://raw.githubusercontent.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")
source("https://raw.githubusercontent.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")

# Load the data
packageSales <- read.table(
  file = "https://raw.githubusercontent.com/neilhatfield/STAT461/master/dataFiles/packageSales.csv",
  header = TRUE,
  sep = ",",
)

packageSales$design <- as.factor(packageSales$design)

# Get values of descriptive statistics
groupStats <- psych::describeBy(
  x = packageSales$sales,
  group = packageSales$design,
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = TRUE,
  mat = TRUE,
  digits = 4
)

# Set row names as type of cookie; select useful columns
groupStats <- groupStats %>%
  tibble::remove_rownames() %>%
  tibble::column_to_rownames(
    var = "group1"
```

```

) %>%
dplyr::select(
  n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
)

groupStats %>%
knitr::kable(
  caption = "Summary Statistics for Cereal Sales by Design",
  digits = 3,
  format.args = list(big.mark = ","),
  align = rep('c', 11),
  col.names = c("n", "Min", "Q1", "Median", "Q3", "Max", "MAD", "SAM", "SASD",
    "Sample Skew", "Sample Ex. Kurtosis"),
  booktabs = TRUE
) %>%
kableExtra::kable_styling(
  font_size = 12,
  latex_options = c("scale_down", "HOLD_position")
)

# Create side-by-side box plots
ggplot(
  data = packageSales,
  mapping = aes(x = design, y = sales)
) +
  geom_boxplot(na.rm = TRUE) +
  theme_bw() +
  xlab("Package Design") +
  ylab("Sales (cases)") +
  theme(
    text = element_text(size = 12)
  )

# Create a shadowgram of sales
shadowgram(
  dataVec = packageSales$sales,
  label = "Sales (cases)",
  layers = 30,
  color = "navyblue",
  aStep = 2
)

# Hasse Diagram
modellLabels <- c("1 Sale Cereal 1", "4 Design 3", "19 (Stores) 15")
modelMatrix <- matrix(
  data = c(FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, TRUE, TRUE, FALSE),
  nrow = 3,
  ncol = 3,
  byrow = FALSE
)
hasseDiagram::hasse(
  data = modelMatrix,
  labels = modellLabels
)

```

```

)

# Applying parametric shortcut
c1 <- c(1/2, 1/2, -1/2, -1/2)
c2 <- c(1/2, -1/2, 1/2, -1/2)

contrasts(packageSales$design) <- cbind(c1, c2)

salesModel <- aov(
  formula = sales ~ design,
  data = packageSales,
  na.action = "na.omit"
)

# Checking Gaussian assumption
car::qqPlot(
  x = salesModel$residuals,
  distribution = "norm",
  envelope = 0.90,
  id = FALSE,
  pch = 20,
  ylab = "Residual Quantiles",
  xlab = "Gaussian Quantiles"
)

# Checking homoscedasticity
ggplot(
  data = data.frame(
    residuals = salesModel$residuals,
    fitted = salesModel$fitted.values
  ),
  mapping = aes(x = fitted, y = residuals)
) +
  geom_point(size = 1) +
  theme_bw() +
  xlab("Fitted values (cases sold)") +
  ylab("Residuals (cases sold)")

# Display the anova table
tableResults <- parameters::model_parameters(
  model = salesModel,
  omega_squared = "raw",
  eta_squared = "raw",
  epsilon_square = "raw"
)

tableResults$p <- sapply(
  X = tableResults$p,
  FUN = function(x){
    ifelse(
      test = is.na(x),
      yes = NA,
      no = pvalRound(x)
    )
  }
)

```

```

    )
  }
)

knitr::kable(
  x = tableResults,
  digits = 4,
  row.names = FALSE,
  col.names = c(
    "Source", "SS", "df", "MS", "F", "p-value",
    "Omega Sq.", "Eta Sq.", "Epsilon Sq."),
  caption = "ANOVA Table for Cereal Package Design Study",
  booktabs = TRUE,
  align = c("l", rep("c", 8))
) %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("HOLD_position", "scale_down")
  )

# Conduct Pairwise tests
hsdPH <- as.data.frame(TukeyHSD(
  x = salesModel,
  conf.level = 0.9
)$design)

names(hsdPH) <- c("diff", "lwr", "upr", "p")

# Conduct contrast tests
scheffeSales <- DescTools::ScheffeTest(
  x = salesModel,
  contrasts = cbind(c1, c2),
  conf.level = 0.9
)

# Create data frame of confidence intervals
temp1 <- as.data.frame(scheffeSales$design)

temp1 <- temp1[, c(2,3)]
names(temp1) <- c("lwr", "upr")

temp1 <- rbind(
  temp1,
  hsdPH[, c(2,3)]
) %>%
  tibble::rownames_to_column(
    var = "comparison"
  )
temp1 <- temp1[1:8,]
temp1$row <- c(1:8)

# Plot post hoc confidence intervals
ggplot(

```

```

data = temp1,
mapping = aes(y = row, yend = row, x = lwr, xend = upr, color = comparison)
) +
geom_segment(size = 1) +
geom_point(
  mapping = aes(x = lwr, y = row),
  shape = 1,
  size = 3
) +
geom_point(
  mapping = aes(x = upr, y = row),
  shape = 1,
  size = 3
) +
geom_vline(
  xintercept = 0,
  color = "red"
) +
theme_bw() +
theme(
  text = element_text(size = 12),
) +
xlab("Sales (cases)") +
ylab(NULL)

# Display pairwise results
hsdPH$p <- sapply(
  X = hsdPH$p,
  FUN = function(x){
    ifelse(
      test = is.na(x),
      yes = NA,
      no = pvalRound(x)
    )
  }
)

knitr::kable(
  x = hsdPH,
  digits = 4,
  caption = "Post Hoc Tukey HSD Comparisons",
  col.names = c("Difference", "Lower Bound",
                "Upper Bound", "Adj. p-Value"),
  align = 'cccc',
  booktabs = TRUE,
) %>%
kableExtra::kable_styling(
  bootstrap_options = c("condensed", "boardered"),
  font_size = 12,
  latex_options = "HOLD_position"
)

# Display pairwise effect sizes

```

```

anova.PostHoc(salesModel) %>%
  knitr::kable(
    digits = 3,
    caption = "Post Hoc Comparison Effect Sizes",
    col.names = c("Pairwise Comparison", "Cohen's d", "Hedge's g",
                  "Prob. Superiority"),
    align = 'lccc',
    booktabs = TRUE
  ) %>%
  kableExtra::kable_styling(
    bootstrap_options = c("condensed", "boardered"),
    font_size = 12,
    latex_options = "HOLD_position"
  )

# Display contrast results
knitr::kable(
  x = scheffeSales[[1]], # Grab the output
  digits = 4,
  col.names = c(
    "Difference", "Lower Bound", "Upper Bound", "p-value"),
  caption = "Scheffe Test Results",
  booktabs = TRUE,
  align = rep("c", 4)
) %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("HOLD_position")
  )

```