STAT461 Undergraduate Song Knowledge

Neil Hatfield and E. A. Student

March 22, 2023

Introduction and Background

Trivia contests are a popular type of event in American bars/pubs that came from the 1970s British culture[1]. While increasing in popularity, pub trivia is emblematic of a common phenomenon of trivia and game nights. A common round of pub trivia is a music round where short snippets of songs are played over the public address (PA) system. Trivia contestants then try to identify the artist and title of each song.

Dr. Hatfield (a big fan of pub trivia) noticed an apparent relationship between what age group a person belonged to and how well they people did at identifying songs. This noticing served as the basis for a set of in-class activities for his inquiry-based ANOVA (STAT461) course. The first activity had students design a study to investigate this noticing. The second, to implement a version of that designed study. The final activity sequence involved using the collected data to develop the One-way ANOVA toolkit and practice writing up a report.

To this end, we want to explore whether a STAT461 undergraduate student's year in school (junior, senior, or other) impacts how well they can identify song title and artist (i.e., their song knowledge)?

Study Design and Methods

In order to investigate our comparative/relationship research question, we've designed a quasi-experimental study by synthesizing the various versions of studies STAT461 students created during class. Due to enrollments, we did reduce the number of levels to our factor (year in school) from 4 levels to 3 (junior, senior, and other). Our population of interest is all students enrolled in STAT461 during the Spring 2023 semester.

To collect our data, Dr. Hatfield constructed a set of 20-second snippets of 10 different songs. The songs came from submissions that members of the class made as part of an introductory assignment in the course. Only student submissions that included valid YouTube links to the songs were included in the pool of potential songs. Dr. Hatfield then placed the snippets into an Apple Keynote presentation and prepared an answer sheet with lines for artist and title for each song along with which year in school group individuals were in.

On the day of data collection, every enrolled student took part in the activity of attempting to identify the song titles and artists. Nine enrolled individuals either did not consent to be part of the data pool or were not present this day. After class, Dr. Hatfield sorted the response sheets filtering out individuals who did not clearly consent to be part of the research data from those who did. He then scored each sheet. For each correctly identified primary artist and full title (baring spelling issues), Dr. Hatfield gave a point and then summed the points earned.

Dr. Hatfield then placed all scores and the associated years in school into an Excel table and used stratified sampling (via `dplyr`'s `group_by` and `slice_sample` functions) to randomly select four students from each of the three year in school categories. This sample size is a reflection of the effect sizes detectable from past

[1] https://www.independent.co.uk/life-style/food-and-drink/features/q-what-is-one-of-britain-s-fastgrowing-pastimes-a-the-pub-quizzes-that-are-seeing-big-screens-switched-for-answer-sheets-8507761.html

semesters as well as limitations on how many individuals were available in each stratum. We wanted to design the study to be balanced.

The final data collection is publicly available[2].

Our primary response is a student's level of song knowledge. We've operationalized this as the total score they earned based upon their correct identification of song titles and primary artists. Our only factor of interest is their categorical year in school. Taken together and with our research question, ANOVA methods appear to be appropriate. Figure 1 shows the Hasse diagram for this study. We can see that an additive model will work and that we have sufficient degrees of freedom to estimate effects and residuals/errors.

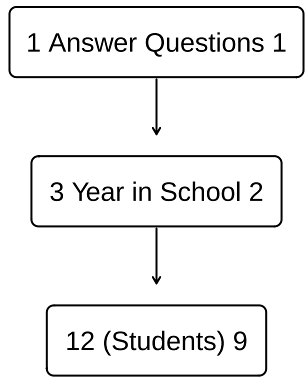


Figure 1: Hasse Diagram for the Song Knowledge Study (Sp'23)

Thus, we will adopt the following null hypothesis: there is no statistically significant impact of year in school on song knowledge score. Our alternative hypothesis is then: there is a statistically significant impact of year in school on song knowledge score. We may express these hypothesized models as

$$\begin{aligned} H_0: &\ y_{ij} = \mu_{\bullet\bullet} + \epsilon_{ij} \\ H_A: &\ y_{ij} = \mu_{\bullet\bullet} + \alpha_i + \epsilon_{ij} \end{aligned}$$

where y_{ij} represents a student's score, $\mu_{\bullet\bullet}$ the baseline performance of STAT461 students in answering song trivia questions, α_i the additional performance due to being in factor level i, and ϵ_{ij} the residuals for each student j in factor level group i.

We made the choice to control our overall Type I risk at 6%. For multiple comparisons, we will control the Simultaneous Confidence Interval error rate at this level by using Tukey's HSD. We will set our unusualness threshold at 4%[3].

[2] Data available at https://raw.github.com/neilhatfield/STAT461/master/dataFiles/songKnowledge_Sp23.csv

[3] We encourage our readers to select their own unusualness thresholds when examining our results.

Exploration of the Data

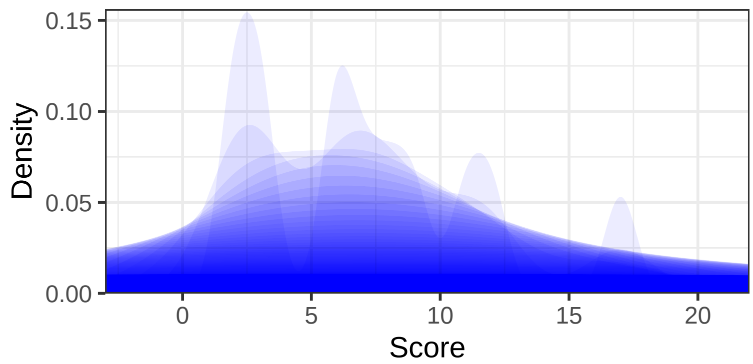


Figure 2: Shadowgram of Song Knowledge Scores

Figure 2 provides the shadowgram for our 12 song knowledge scores. In examining the shadowgram, we can that there is one dominant modal clump with some faint separation in the background. While we know that we have three groups based upon year in school, Figure 2 suggests that there may not be significant differences in terms of performance between them.

Table 1: Summary Statistics for Song Knowledge Scores

	n	Min	Q1	Median	Q3	Max	MAD	SAM	SASD	Sample Skew	Sample Ex. Kurtosis
Junior	4	3	5.25	7	9.00	12	3.707	7.25	3.775	0.134	-1.968
Senior	4	3	6.00	9	12.50	17	5.930	9.50	5.972	0.158	-2.019
Other	4	2	2.00	4	6.75	9	2.965	4.75	3.403	0.235	-2.174

Table 1 shows the values of various descriptive statistics broken out by year in school. Visually, we can see that the seniors tend to have scores just as good or better than juniors. The Other group appears to be slightly lower than the other two groups but not by much. Visually, we can see this in the box plots of Figure 3. There does appear to be differences in the performance of each year when we look at values of the *Sample Arithmetic Mean* (*SAM*). There also appears to be different amounts of variation with each year as evidenced by the values of the *Sample Arithmetic Standard Deviation* (*SASD*).

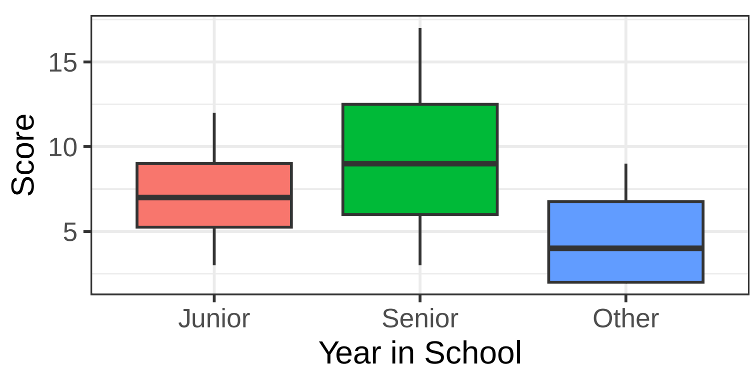
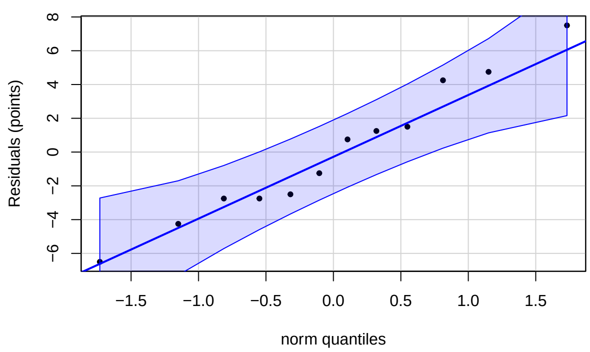


Figure 3: Side-by-side Box Plots of Score by Year

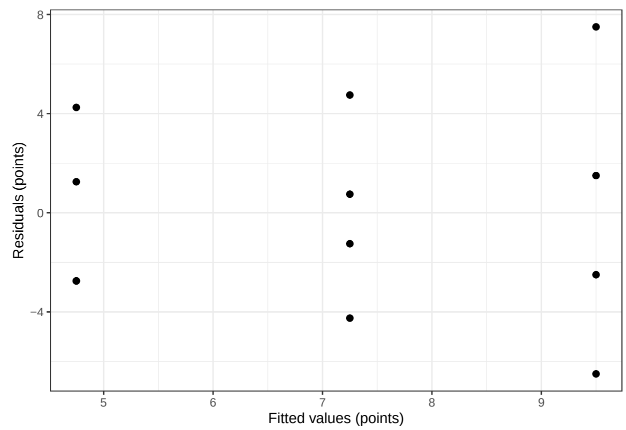
Results

To answer our research question, we will seek to use the parametric shortcut known as the ANOVA F test. There are three assumptions that our data must satisfy to use this approach: residuals follow a Gaussian distribution, homoscedasticity, and independence of observations.

Assumptions



(a) QQ Plot



(b) Srip Chart

Figure 4: Assessing Assumptions for Song Knowledge Study

Let us first turn towards the Gaussian assumption. Figure 4a shows the QQ plot for our residuals with a 90% confidence envelope. No observations fall outside of this envelope. Additionally from Table 1, while all groups have positive skewness, the skewness values are all fairly small and in the same direction. Further, all three groups have similar levels of excess kurtosis, pointing to there being fewer potential outliers in the data than we would expect for a Gaussian distribution. This is not surprising given the relatively small sample size used in this study. We find strength in the fact that all of the groups have (roughly) the same skewness and kurtosis issues. Taking these values together with the QQ plot, we will decide that our residuals (response data) satisfy this assumption.

In Figure 4b we can also see the strip chart for assessing the homoscedasticity assumption. The leftmost group uses just about half the vertical space as the rightmost group. There is additionally a faint fanning as we move from lower fitted values to higher ones. Taken together this suggests we should have some care as to whether we've satisfied this assumption. We will proceed with caution.

For the issue of independence of observations, we know that each student completed the quizzes individually. Further, Dr. Hatfield used a computerized system to carry out the selection of individuals from the broader pools. Thus, we can be relatively assured that we have the independence of observations.

Given that we have a balanced design, we will cautiously proceed with the parametric ANOVA F test and our planned post hoc analysis (as needed).

Omnibus Results

Table 2: ANOVA Table for Song Knowledge Study

Source	SS	df	MS	F	p-value	Eta Sq.	Omega Sq.	Epsilon Sq.
Year	45.1667	2	22.5833	1.1016	0.3733	0.1967	0.0167	0.0181
Residuals	184.5000	9	20.5000					

As we can see from Table 2, a STAT461 undergraduate's year in school accounts for ~1.1 times as much variation as the residuals. Under the null model that year in school does not statistically impact song knowledge, we would anticipate observing a value for the F ratio at least as large as we did ~37% of the time we repeat the study. Since our p-value is greater than our unusualness threshold ($0.3733 > 0.04$), we will fail to reject the null hypothesis and decide to act as if a STAT461 undergraduate student's year in college does not impact their song knowledge score. This statistical significance is supported by the two of the effect size metrics, ω^2 and ϵ^2, which suggest that only about 2% of variation in score could be attributable to a students' year in school. Given that η^2 is known to overestimate the effect size, we will attend to the other two metrics.

Table 3: Point Estimates from the Song Knowledge Study

	Estimate
Grand Mean	7.17
Juniors	0.08
Seniors	2.33
Others	-2.42

In general a group of students will accumulate 7.17 times as many points as students; this is our estimate for baseline performance ($\widehat{\mu_{\bullet\bullet}}$). We can also see the factor level (treatment) effects ($\widehat{\alpha_i}$) estimates. For Seniors, they accumulated an additional 2.33 points per student where as the Juniors only accumulated 0.08 points per student and the Others accumulated -2.42 points per student. This suggests that Others were performing worse that baseline ($GSAM$).

Post Hoc

Given that we failed to reject our null hypothesis, we will not conduct any post hoc analyses.

Discussion and Limitations

We explored our research question of whether a STAT461 undergraduate student's year in school impacted their song knowledge. From our data, we found that year in school does not appear to influence the score they got on the trivia quiz. However, this effect appears to be limited to just juniors and seniors.

Our study has several limitations. First, we are looking at a niche population of just students taking STAT461 in the Spring 2023 semester. In future work, we may want to broaden this to a larger population. Second, our total sample size was only 12 students. We may want to increase this sample size along with the broadening of the population. Additionally, we may want to incorporate additional attributes that will allow us to more accurately investigate what might be going on. For example, we might collect information on what a student's most commonly listened to genre of music might be.

References and Materials Consulted

Hatfield, N. J. (2023). Unit 2: Study Design. In STAT 461: Analysis of Variance (ANOVA) Course Materials. Pennsylvania State University.

Hatfield, N. J. (2023). Unit 3: Oneway ANOVA. In STAT 461: Analysis of Variance (ANOVA) Course Materials. Pennsylvania State University.

Peck, T. (2013, February 23). Q: What is one of Britain's fast-growing pastimes? A: The pub. The Independent. https://www.independent.co.uk/life-style/food-and-drink/features/q-what-is-one-of-britain-s-fastgrowing-pastimes-a-the-pub-quizzes-that-are-seeing-big-screens-switched-for-answer-sheets-8507761.html

Author Contributions

The authors of this report would like to acknowledge their individual contributions to the report. Both authors contributed to ongoing discussions about study design and analysis.

- Dr. Hatfield contributed to the design of the study, collection of data, analysis of data, coding, and writing of the report.
- E.A. Student contributed to the design of the study, and participated in the study.

Code Appendix

```r
# Setting Document Options ----
knitr::opts_chunk$set(
  echo = FALSE,
  warning = FALSE,
  message = FALSE,
  fig.align = "center",
  dpi = 300 # helps create higher quality graphics
)

# Add additional packages by name to the following list ----
packages <- c(
  "tidyverse", "knitr", "kableExtra", "hasseDiagram",
  "psych", "car", "parameters"
  )
lapply(X = packages, FUN = library, character.only = TRUE, quietly = TRUE)

# Loading Helper Files and Setting Global Options ----
options(knitr.kable.NA = "")
options("contrasts" = c("contr.sum", "contr.poly"))
source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/ANOVATools.R")

source("https://raw.github.com/neilhatfield/STAT461/master/rScripts/shadowgram.R")

songData <- read.table(
  file = "https://raw.github.com/neilhatfield/STAT461/master/dataFiles/songKnowledge_Sp23.csv",
  header = TRUE,
  sep = ","
)

songData$Year <- factor(
  x = songData$Year,
  levels = c("Junior", "Senior", "Other")
)
# Create a Hasse diagram for the study ----
# Feel free to use the Hasse diagram wizard app to generate the code for you:
## https://psu-eberly.shinyapps.io/Hasse_Diagrams/

modelLabels <- c("1 Answer Questions 1", "3 Year in School 2", "12 (Students) 9")
modelMatrix <- matrix(
  data = c(FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, TRUE, TRUE, FALSE),
  nrow = 3,
  ncol = 3,
  byrow = FALSE
)
hasseDiagram::hasse(
 data = modelMatrix,
 labels = modelLabels
)

# Creating a shadowgram of scores ----
# Note: you do not have to use shadowgrams.
```

```r
# You can use a histogram or any other kind of data visualization.
shadowgram(
  dataVec = songData$Score,
  label = "Score",
  layers = 50,
  color = "blue",
  aStep = 4
)

# Descriptive statistics on score by year in school ----
scoreStats <- psych::describeBy(
  x = songData$Score,
  group = songData$Year,
  na.rm = TRUE,
  skew = TRUE,
  ranges = TRUE,
  quant = c(0.25, 0.75),
  IQR = FALSE,
  mat = TRUE,
  digits = 4
)

scoreStats %>%
  tibble::remove_rownames() %>%
  tibble::column_to_rownames(
    var = "group1"
  ) %>%
  dplyr::select(
    n, min, Q0.25, median, Q0.75, max, mad, mean, sd, skew, kurtosis
  ) %>%
  knitr::kable(
    caption = "Summary Statistics for Song Knowledge Scores",
    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")
  )

# Side by side box plots ----
ggplot(
  data = songData,
  mapping = aes(x = Year, y = Score, fill = Year)
) +
  geom_boxplot() +
  theme_bw() +
  xlab("Year in School") +
  ylab("Score") +
```

```r
  theme(
    legend.position = "none",
    text = element_text(size = 12)
  )

# Fit the model and parametric shortcut ----
songModel <- aov(
  formula = Score ~ Year,
  data = songData,
  na.action = "na.omit"
)

# Assumption Assessment Visualizations ----
## Gaussian Residuals Assumption
car::qqPlot(
  x = songModel$residuals,
  distribution = "norm",
  envelope = 0.90,
  id = FALSE,
  pch = 20,
  ylab = "Residuals (points)"
)

## Strip Chart for Homoscedasticity ----
ggplot(
  data = data.frame(
    residuals = songModel$residuals,
    fitted = songModel$fitted.values
  ),
  mapping = aes(x = fitted, y = residuals)
) +
  geom_point(size = 2) +
  theme_bw() +
  xlab("Fitted values (points)") +
  ylab("Residuals (points)")

# Modern ANOVA Table ----
parameters::model_parameters(
  model = songModel,
  effectsize_type = c("eta", "omega", "epsilon")
) %>%
  knitr::kable(
  digits = 4,
  col.names = c(
    "Source", "SS", "df", "MS", "F", "p-value",
    "Eta Sq.", "Omega Sq.", "Epsilon Sq."),
  caption = "ANOVA Table for Song Knowledge Study",
  booktabs = TRUE,
  align = c("l", rep("c", 8))
  ) %>%
  kableExtra::kable_styling(
    font_size = 10,
    latex_options = c("HOLD_position")
```

```r
  )

# Point Estimates for Parametric Shortcut ----
pointEst <- dummy.coef(songModel)
pointEst <- unlist(pointEst)
names(pointEst) <- c("Grand Mean", "Juniors", "Seniors",
                     "Others")

data.frame("Estimate" = pointEst) %>%
  knitr::kable(
  digits = 2,
  caption = "Point Estimates from the Song Knowledge Study",
  format = "latex",
  booktabs = TRUE,
  align = "c"
  ) %>%
  kableExtra::kable_styling(
    font_size = 12,
    latex_options = c("HOLD_position")
  )

# Post Hoc via Tukey HSD ----
hsdSong <- TukeyHSD(
  x = songModel, # Your aov/lm object
  conf.level = 0.94 # 1 -- Your overall Type I Error level
)

## Kable Code for Tukey HSD
knitr::kable(
  x = hsdSong$Year, # Notice the factor's name
  digits = 3,
  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"
  )

# Post Hoc Effect Sizes ----
anova.PostHoc(songModel) %>%
  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
  ) %>%
```

```r
kableExtra::kable_styling(
  bootstrap_options = c("condensed", "boardered"),
  font_size = 12,
  latex_options = "HOLD_position"
)
```