Week 5 R Exercise

Calum Webb

21/10/2021

Part 1

Running statistical inference tests

This .Rmd file will take you through three statistical inference tests that can be run in `R` using data from different kinds of research studies.

Your first task will be to run the analyses (in `R`), interpret the output, and write what kind of causal inference *could* be made based on the description of the study.

Your second task will be to write the code to perform some additional statistical tests (ones you will have already seen the code for), and practice interpreting the output.

Efficacy of Small Class Sizes in Early Education (from Imai, 2017)

You are not expected to write any original code in this example! You are just expected to interpret it.

The Student-Teacher Achievement Ratio Project (STAR) was a four year longitudinal study on the effect of class size in early US grade levels on educational performance and personal development. It was run between 1985 and 1989.

During the study, students were randomly assigned to either small classes, regular classes, or regular classes with teaching aids. It followed 11,601 students from Tennessee, USA.

```r
# Run this code to read in and tidy up the data - you don't need to write anything else in here!

star_data <- read_csv("star.csv")
```

```
## Rows: 6325 Columns: 6
## -- Column specification --------------------------------------------------
## Delimiter: ","
## dbl (6): race, classtype, yearssmall, hsgrad, g4math, g4reading
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
star_data <- star_data %>%
  mutate(
    race = case_when(race == 1 ~ "white",
                     race == 2 ~ "black",
                     race == 3 ~ "asian",
                     race == 4 ~ "hispanic",
                     race == 5 ~ "native american",
                     race == 6 ~ "other",
                     TRUE ~ NA_character_),
    classtype = case_when(classtype == 1 ~ "small",
```

```
                             classtype == 2 ~ "regular",
                             classtype == 3 ~ "regular+aid",
                             TRUE ~ NA_character_),
    hsgrad = case_when(hsgrad == 1 ~ "graduated",
                       hsgrad == 0 ~ "did not graduate",
                       TRUE ~ NA_character_)
 )



# Additional data filtering only students in the small classes for either 4 years or not at all
star_data_4years <- star_data %>%
    filter(yearssmall == 4 | yearssmall == 0)
```

Research Question 1: Did students who were in different types of classes for all four years have different outcomes on their 4th Grade standardised Maths score.

Variables:

- classtype: The randomised condition of class that students were in
- g4math: The student's score on a standardised maths test.

To answer this research question, the researchers used an ANOVA test and a Tukey Honest Significant Difference test.

```
# Analysis of Variance Test - Continuous outcome goes on the left hand side of the ~ sign. Grouping var

aov_star_maths <- aov(g4math ~ classtype, data = star_data_4years)
summary(aov_star_maths)


##                Df   Sum Sq Mean Sq F value Pr(>F)
## classtype       2     2812    1406   0.764  0.466
## Residuals    2053  3780350    1841
## 2758 observations deleted due to missingness
```

Interpret the output from the ANOVA test above and answer the following questions:

---

1. Why did the researchers use an ANOVA test for this particular combination of variables?

The researchers used ANOVA because they were analysing the association between a categorical variable (with more than two categories) — classtype — and a continuous variable — g4math.

2. What does the p-value for this test indicate about the effect of the small class sizes? Assume a critical value of 0.05.

The p-value for this test was 0.466. This is greater than our critical value of 0.05, therefore we are not able to reject the null hypothesis. This suggests that there was no significant relationship between class type and grade 4 maths score.

3. What can we infer about the causal effect of small class sizes on Grade 4 Maths test outcomes from this?

Because the class types were randomly assigned, we can infer that there was no causal effect of small classes on Grade 4 maths scores.

4. On what basis are we making this causal inference (robust dependence and/or consequential manipulation and why)?

This inference is largely based on consequential manipulation — the researchers randomly allocated the class type and therefore were, in theory, able to randomise other potentially confounding variables across the groups.

The researchers then used a Tukey Honest Significant Difference test to look specifically at the differences between the three classtype groups.

```
TukeyHSD(aov_star_maths)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = g4math ~ classtype, data = star_data_4years)
##
## $classtype
##                          diff       lwr      upr     p adj
## regular+aid-regular -1.598286 -6.976527 3.779955 0.7652664
## small-regular        1.284303 -4.168325 6.736931 0.8452409
## small-regular+aid    2.882589 -2.605893 8.371071 0.4344784
```

---

1. What was the mean difference between the small class sizes and the regular class sizes?

The small class sizes had standardised maths scores that were on average 1.284 points higher than the regular size classes.

Research Question 2: Did students who were in different types of classes for all four years have different outcomes in their chances of graduating from High School?

Next, the researchers wanted to test whether those children who had been in the small or regular class sizes for four years had different likelihoods of graduating high school. They used a chi-squared test of association to test whether there was an association between class size type and high school graduation outcome.

Variables:

- classtype: The randomised condition of class that students were in
- hsgrad: Whether the student graduated high school or not

```
# Create a crosstab of hsgrad and classtype

star_data_4years %>%
  janitor::tabyl(classtype, hsgrad, show_na = FALSE) %>%
  adorn_percentages("row") %>%
  adorn_pct_formatting()
```

```
##    classtype did not graduate graduated
##      regular           17.6%     82.4%
##   regular+aid           16.7%     83.3%
##        small           12.2%     87.8%
```

```
# Run a chi-squared test of association between classtype and hsgrad

star_data_4years %>%
  janitor::tabyl(classtype, hsgrad, show_na = FALSE) %>%
  janitor::chisq.test()
```

```
## 
##  Pearson's Chi-squared test
## 
## data:   .
## X-squared = 7.6838, df = 2, p-value = 0.02145
```

---

1. Why did the researchers choose to use a chi-squared test of association for this analysis?

The researchers used a chi-squared test of association for this analysis because they were exploring differences between two categorical variables (whether the student graduated high school or not, and whether the class type they were in in grade school).

2. Using the bivariate descriptive statistics from the tabyl above, describe the differences in overall graduation rates between students in the different class sizes.

The regular class size students had the lowest proportion of students graduating high school (82.4%), the regular class sizes with teaching aids had slightly higher rates of students graduating high school (83.3%). The small class size students had the highest proportion of high school graduates — 87.8% — around 5.4 percentage points higher than the regular class size students.

3. What was the result of this chi-squared test of association? Interpret the p-value — was the association between high school graduation rates and class type significant? Assume a critical value of 0.05.

This result was statistically significant at a critical value of 0.05 ($p < 0.05$); the p-value (0.02145) was lower than this value. This means that the differences between the class size's graduation rates was much greater than we might have expected if, in reality, graduation rates across class sizes were approximately equal in the wider population.

[NB: I don't necessarily expect you to write more than the first sentence in any assessments — I have just expanded a little here to help you revise.]

4. What can we infer about the causal effect of class size on high school graduation on this study, and what basis for causality can we use?

As the class sizes were randomly assigned to students we can base our causal inference on consequential manipulation. From this study design we would be justified claiming that smaller class sizes caused some change in high school graduation rates (though we should keep in mind that there was quite a large gap between the treatment - the class size - and the outcome - graduating high school).

Research Question 3: Did the students in the small class sizes have better 4th Grade Maths scores if they were in the small sized classes for a greater number of years?

```
# Filter only students who were allocated to the small class sizes
star_data_small <- star_data %>%
  filter(classtype == "small")

# Run a correlation test (with p-value) for the correlation between the number of years in the small cl
cor.test(star_data_small$yearssmall, star_data_small$g4math, use = "complete.obs")
```

```
## 
##  Pearson's product-moment correlation
## 
## data:  star_data_small$yearssmall and star_data_small$g4math
## t = 1.4211, df = 738, p-value = 0.1557
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.01990748  0.12384327
## sample estimates:
##       cor
## 0.0522385
```

---

1. Report the correlation between the number of years the children spent in small classes and the 4th Grade Maths scores. How strong was the correlation? What direction was it in and what does this mean with regards to answering our research question.

There was a very weak positive correlation (0.0522) between the number of years children were in the small classes and the children's 4th Grade maths scores.

2. Report the outcome of the statistical significance test on this correlation. Assume a critical value of 0.05.

The p-value for the correlation was 0.1557. This is greater than the critical value, therefore our data does not suggest that we can reject the idea that there is no association between years in the small class size and Grade 4 maths score.

3. If this result was significant, what would our basis for causal inference be and why?

In this case, we may be able to use the basis of robust dependence as an argument for causality (as being in the class for a given number of years comes before the test was taken), however, the time the children were in the classes for was not randomly assigned. This means we would not be able to claim causality on the basis of consequential manipulation.

Part 2

You have now seen the code for running a chi-squared test of association, a correlation test (with hypothesis test), and an ANOVA test. This part will see you writing the code to conduct these tests on additional data.

Racial discrimination in the labour market (adapted from Imai, 2017)

Based on the work of Marianne Bertrand and Sendhil Mullainathan (2004), two social resarchers conducted an experiment where they sent CVs to job advertisements and randomly varied the name on the CV so that it was stereotypically African-American sounding or stereotypically white European-American sounding names. They then recorded which CVs received a call back from the employer and which did not.

```r
# Reading and tidying the data - You don't need to add code here!

cv_data <- read_csv("resume.csv")
```

```
## Rows: 4870 Columns: 4
## -- Column specification ---------------------------------------------------
## Delimiter: ","
## chr (3): firstname, sex, race
## dbl (1): call
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# Create new variable of race and sex combination of name chosen and change call variable to be a chara
cv_data <- cv_data %>%
  mutate(race_and_sex = paste(race, sex),
         call = ifelse(call == 1, "Called back", "Did not call back"))
```

- Task 1: Get descriptive statistics for the proportion of which types of stereotypical names received calls back and which did not.

```
# Frequency counts
cv_data %>%
  janitor::tabyl(race_and_sex, call, na_show = FALSE)
```

```
##  race_and_sex Called back Did not call back
##  black female         125              1761
##    black male          32               517
##  white female         184              1676
##    white male          51               524
```

```
# Row percentages
cv_data %>%
  janitor::tabyl(race_and_sex, call, na_show = FALSE) %>%
  janitor::adorn_percentages("row") %>%
  janitor::adorn_pct_formatting()
```

```
##  race_and_sex Called back Did not call back
##  black female        6.6%             93.4%
##    black male        5.8%             94.2%
##  white female        9.9%             90.1%
##    white male        8.9%             91.1%
```

- Task 2: Write the code below to conduct **the appropriate** hypothesis test on the relationship between these two variables

```
cv_data %>%
  janitor::tabyl(race_and_sex, call, na_show = FALSE) %>%
  janitor::chisq.test()
```

```
##
##  Pearson's Chi-squared test
##
## data:  .
## X-squared = 17.867, df = 3, p-value = 0.0004686
```

- Task 3: Interpret the output of the above two pieces of code and write about your findings, remember to cover the following: describe the differences found; report whether they were statistically significant (use a critical value of 0.05); report what we can infer about causality in this example.

The experiment recorded that the highest proportion of call backs was found among stereotypically 'white female' sounding names (9.9%, N = 184) and stereotypically 'white male' sounding names (8.9%, N = 51). Stereotypically 'black male' sounding names received the lowest proportion of call backs (5.8%, N = 32), followed by stereotypically 'black female' sounding names (6.6%, N = 125). This difference was found to be statistically significant (X-squared = 17.867, df = 3, $p < 0.05$). As the CVs were identical, with names randomly manipulated, we can infer that there was a causal association between CV name and the likelihood of being called back by an employer advertising a job. The research found that people with stereotypically black names may be discriminated against in employment practices.

Local authority deprivation and numbers of children in the care system (adapted from Webb, et al. 2020)

This example uses data on the Indices of Multiple Deprivation score for 150 local authorities in England (exluding the City of London and the Isles of Scilly) and the rates of children in care per 10,000. Researchers across several universities in the UK were interested in identifying the association between poverty and child welfare interventions (such as children being taken into care), to test whether there was a similar relationship to what is found in outcomes to do with health. This was called the Child Welfare Inequalities Project. All data was collected around the years 2011-2012 by the Department for Education and the Department for Communities and Local Government.

The relevant variables in this dataset are:

- imd_score: The level of deprivation in the local authorities. A higher score means more deprivation.
- cla_rate: The rate per 10,000 children in the local authority currently living in the care system.

```
# Read in the CWIP data - You don't need to add any extra code here!
cwip_data <- read_csv("cwip_ex.csv")
```

```
## Rows: 150 Columns: 5
## -- Column specification -------------------------------------------------
## Delimiter: ","
## chr (2): new_la_code, la_name
## dbl (3): year, cla_rate, imd_score
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

- Task 1: Identify the most appropriate descriptive statistic for summarising the relationship between imd_score and cla_rate.

The most appropriate descriptive statistic for summarising this relationship would be the Pearson's R correlation coefficient. This is because both variables are continuous and approximately normally distributed.

- Task 2: Use R to calculate this descriptive statistic and interpret it for a general audience: what is the relationship between deprivation and children looked after rate and how strong or weak is this relationship?

```
cor(cwip_data$imd_score, cwip_data$cla_rate, use = "complete.obs")
```

```
## [1] 0.7449934
```

The Pearson's R correlation coefficient for the relationship between IMD score and children looked after rate was 0.745. This suggests that there is a strong relationship between local authority deprivation and rates of children in the care system.

- Task 3: Use R to test a hypothesis associated with descriptive statistic. Interpret the p-value associated with the test statistic using a critical value of 0.05. Is it meaningful to use inferential statistics in this case? Why/why not?

```
cor.test(cwip_data$imd_score, cwip_data$cla_rate, use = "complete.obs")
```

```
##
##  Pearson's product-moment correlation
##
## data:  cwip_data$imd_score and cwip_data$cla_rate
## t = 13.587, df = 148, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  0.6640107 0.8087011
## sample estimates:
##       cor
## 0.7449934
```

The relationship between these two variables was statistically significant (t = 13.59, df = 148, p < 0.05). In this case, our data includes 150 of 152 local authorities. As such, it may not necessarily be very meaningful to generalise to a wider population as the 150 local authorities in England is likely our entire population of interest and does not necessarily represent a meaningful sample of any other larger population.

[NB: In this case, our p-value is so small that R has switched to scientific notation (2.2e-16). This means the p-value is 2.2 times by 10 to the power of minus 16. This means the p-value is actually 0.00000000000000022. You can think of scientific notation as either moving the decimal place to the left 16 times (when the power is a negative number, e.g. -16) or to the right (when the power is a positive number, e.g. 3.5e10 = 35,000,000,000).]

- Task 4: What can be said about the causal association between poverty and rates of children in care based on this data and analysis? Explain your answer.

This data does not meet the criteria for causal inference either by robust dependence or consequential manipulation; in this case, the causation could be in either direction, or there could be no causation at all (the association could be caused by a third variable).

Week 5 Challenge

I have included two tidied datasets from Open Data sources from the Leibniz Institute for Social Sciences GESIS database (https://www.gesis.org/home). These are:

- `eurosceptic.rds`: Data from the 'Eurosceptic Cues and Citizen Attitudes' study by Galina Za-pryanova, University of Mannheim. (https://data.gesis.org/sharing/#!Detail/10.7802/1318)
- `nonprof_simple.rds`: Data from the Quality certifications for nonprofits, charitable giving, and donor's trust study by Maja Adena et al. (https://data.gesis.org/sharing/#!Detail/10.7802/1.2121)

Eurosceptic Cues and Citizen Attitudes

This dataset and its associated questionnaire file (`eurosceptic-questionnaire.pdf`, which can act as a codebook) is from a survey experiment study where the researchers randomly presented survey respondents in Germany, France, and the UK with different eurosceptic statements by elites.

They then asked the respondents to answer several questions about support for EU integration, support for national contributions to EU bailouts, attitudes towards immigration, and beliefs about democracy. They also asked them how they would vote in there were (another) referendum on EU membership.

Possible questions:

- Did exposure to different eurosceptic messages from elites result in significantly different responses about how the respondent felt about supporting EU integration?

- Was satisfaction with national democracy associated with satisfaction of EU democracy?

- Did respondents from different countries have significantly different views on the statement that immigration enriches culture?

- Was there an association between treatment group and respondent's view on how they would vote if there were a second Brexit referendum?

- Or, come up with your own!

Quality certifications for nonprofits, charitable giving, and donor's trust

This data is taken from an experiment by Adena et al. (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2961882) where participants were subject to an experiment where they were either presented with information about a charity (Bjorn Schulz Foundation) that included that this charity had a quality certificate (treat1 = B), or were not presented with information about the quality certificate (treat1 = A).

The participant was awarded prize money in the experiment and, before finding out if they had been awarded the prize, they were required to say what amount they would like to donate to the charity.

They were also given a questionnaire about the trust they had in the charity and about the trust they had in people in general.

Possible research questions
- Did exposure to the quality certification result in people donating more of their prize money to the foundation?
- Did exposure to the quality certification result in higher ratings of trust in the foundation?
- Was trust in the work of nonprofits associated with general trust?
- Or come up with your own!