ANOVA and Multiple Comparisons

Lecture 3 Handout

Statistics 139

Topics

- Multiple Comparisons
- Bonferroni
- ANOVA
- Simulation

The material in this lab corresponds to the Lecture 3 Notes.

Note: when performing a hypothesis test, be sure to explicitly state (1) hypotheses, (2) the calculated test statistic (and degrees of freedom if appropriate), (3) the calculated p-value or critical value, and (4) the conclusion in context of the problem along with the scope of inference. Use Type I error rates of $\alpha = 0.05$ and confidence levels of 95% unless explicitly stated otherwise. You can assume all tests are two-sided unless otherwise specified.

In this lab we will use the General Social Survey (GSS) from 2018, in the data file 'gss18.csv', to answer the question:

1. Is the amount of education associated with religious affiliation?
2. If an association does exist, how is it affected by the sexes?

The following variables will be of interest:

- `educ`: years of education of respondent (beyond kindergarten). Respondents with 20 or more years of education were recorded as 20 years.
- `relig`: religious preference of respondent, with 13 options, see here.
- `sex`: sex of respondent: 1 for males and 2 for females.

Concept Checks

a) What happens to power as the mean under the alternative is further from the null hypothesis value (called the 'effect size')? What happens to power as standard deviation increases? What happens to power as sample size increases?

- Power increases as effect size increases (easier to detect effect)
- Power decreases as standard dev increases (harder to get precise measurements and infer effect)
- Power increases as sample size increases (easier to get precise measurements and infer effect)

b) In what way is the Bonferroni correction a 'conservative' one? If 6 groups were compared via the $\binom{6}{2} = 15$ pairwise tests, how would the results of these 15 tests be related? What would this do to Type I error?

-

c) Provide an example when the ANOVA F-test would be insignificant while there would exist a significant [unadjusted] pairwise two-sample t-test. Provide an example of the reverse: when the ANOVA F-test would be significant while all the [unadjusted] pairwise two-sample t-tests would not be.

Question 1: pairwise comparisons

a) The variable `relig` has 13 categories, many of which have only a handful of respondents. Create a variable `relig.factor` that is a factor that has only 5 categories: 1 = Protestant, 2 = Catholic, 3 = Jewish, 4 = None, and 5 = Other (in essence, combine all original categories 5 through 13 into group 5). Check your results to make sure the variable was created properly (Hint: the `table` function will be useful for this). Note: we also turn `sex` into a factor for you as well.

```
gss18 = read.csv("data/gss18.csv")
gss18$sex = factor(gss18$sex,labels=c("M","F"))

# do some work (define 'x')
x <- gss18$relig
x[x>5] = 5

# below we use x and turn into a well-labelled factor
gss18$relig.factor = factor(x,
            labels = c("Protestant","Catholic","Jewish","None","Other"))
```

b) Determine the Bonferroni-corrected α^* for comparing the five religion groups pairwise. What confidence level would each pairwise confidence interval be calculated at? Roughly how much of an increase in CI width does this result in?

c) Use the function `pairwise.t.test` to compare education across all 5 religion categories, with a valid adjustment for multiple comparisons.

```
pairwise.t.test(y ~ x, p.adjust.method = "bonferroni")
```

d) Summarize your findings.

Question 2: ANOVA

a) Provide a visual to compare education across the 5 religion groups. Calculate the 5 group means. Comment on what you notice and how this relates to the results from Question 1.

b) Use the functions `aov` and `summary` to perform the ANOVA and print out the ANOVA table. Perform a formal hypothesis test for this analysis.

c) Comment on the assumptions for this ANOVA F test.

Question 3: a first simulation The code below uses a `for` loop to repeatedly create samples of size $n = 50$ from normal distributions for 5 different groups. It then compares the groups through 3 approaches: (i) ANOVA, (ii) unadjusted pairwise t-tests, and (iii) Bonferroni corrected pairwise t-tests, and stores the p-values.

```r
set.seed(12345) # Space Balls!
nsims = 500

# set up the parameters for the data generating process
mu1 = 0
mu2 = 0
mu3 = 0
mu4 = 0
mu5 = 0
sigma = 1
n = 50

# create blank storage vectors for results
pairwise.pvalue.min = rep(NA,nsims)
anova.pvalue = rep(NA,nsims)
bonf.pvalue.min = rep(NA,nsims)

#the for loop does the bulk of the work
for(i in 1:nsims){
  # generate the data
  x1 = rnorm(n, mu1, sigma)
  x2 = rnorm(n, mu2, sigma)
  x3 = rnorm(n, mu3, sigma)
  x4 = rnorm(n, mu4, sigma)
  x5 = rnorm(n, mu5, sigma)

  # combine them into a single response vector
  x = c(x1,x2,x3,x4,x5)
  # create a vector to define the 5 groups
  groups = c(rep(1,n),rep(2,n),rep(3,n),rep(4,n),rep(5,n))

  # perform ANOVA F tests and store its pvalue
  anovatest = aov(x~groups)
  anova.pvalue[i] = anova(anovatest)[1,"Pr(>F)"]

  # perform pairwise t tests, and store the smallest pvalue
  pairtests = pairwise.t.test(x,groups,p.adjust.method = "none")
  pairwise.pvalue.min[i] = min(pairtests$p.value,na.rm=T)

  # perform Bonferroni-corrected pairwise t tests,
  # and store the smallest pvalue
  bonftests = pairwise.t.test(x,groups,p.adjust.method = "bonferroni")
  bonf.pvalue.min[i] = min(bonftests$p.value,na.rm=T)

}
```

a) Determine the rejection rate for each of the 3 approaches (assume $\alpha = 0.05$).

```
#Determine the rejection rate for each of the 3 approaches
```

b) Play around with the code a little bit: change the sample size, the true standard deviation of observations, and the true population means. How does each affect the rejection rate?

c) Bonus: sample instead from different distributions, and see how the rejection rate is affected.

Question 4 (Extra Problem): how does sex play a role?

a) Compare education between males and females visually and inferentially. Describe what you see.

b) Split the data into males and females separately, and run an ANOVA model within each subgroup. Interpret the results.

c) The boxplot created in the R chunk below looks at education split across combinations of religion group and sex (Gray is for males, Orange is for females). Comment on any differences you see.

```r
boxplot(educ~sex+relig.factor,data=gss18,col=c("gray","orange"),cex.axis=0.5)
```

d) A two-way ANOVA takes into account two grouping variables at once, and essentially measures the effect of one grouping variable controlling for the other. The code for a two-way ANOVA model (without interaction) is provided below. Is there any evidence of confounding between these 2 variables? Is this surprising?

```
aov2 = aov(educ~sex+relig.factor,data=gss18)
summary(aov2)
```

e) A two-way ANOVA model **with interaction** incorporates a third source of variability: the interaction between the 2 grouping variables. This model essentially allows for "the effect of one grouping variable on response to vary depending on the levels of the other grouping variable." Interpret what that means for the results of the model below, and link it to the separate ANOVA models in part (b) and boxplot in part (c).

```
aov3 = aov(educ~sex*relig.factor,data=gss18)
summary(aov3)
```