

Contrasts and Multiple Comparisons

Statistical Experimental Design

Kris Sankaran | UW Madison | 16 September 2021

Today

- Book Sections: 3.5
- Online Notes: Week 3 [3] and [4]

Limitation of ANOVA F -test

Recall the ANOVA model,

$$y_{ij} = \mu + \tau_i + \epsilon_{ij}$$
$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$$

and the associated hypothesis test,

$$H_0 : \tau_1 = \dots = \tau_a = 0$$
$$H_1 : \tau_i \neq 0 \text{ for at least one } i.$$

Note that it *does not* allow us to conclude which group(s) are responsible for a rejection.

Follow-up Tests

We may instead define a follow-up hypothesis test.

For example, imagine we had 4 groups,

- Are the first two means different from the last two? $H_0 : \tau_1 + \tau_2 = \tau_3 + \tau_4$
- Are the first two means equal to each other? $H_0 : \tau_1 = \tau_2$
- ...

Contrasts

- Fortunately, the same framework covers all these cases.
- Let $\mu_i = \mu + \tau_i$ be the mean of group i .
- A contrast is a linear combination of the means,

$$\Gamma(c) = \sum_{i=1}^a c_i \mu_i$$

- The previous examples reduce to,

$$H_0 : \Gamma(c) = 0$$

$$H_1 : \Gamma(c) \neq 0$$

for $c = (1, 1, -1, -1)$ and $c = (-1, -1, 0, 0)$, respectively.

Plug-In Approximations

Since the sample means $\bar{y}_i \approx \mu_i$, we can approximate,

$$\hat{\Gamma}(c) := \sum_i c_i \bar{y}_i \approx \sum_i c_i \mu_i = \Gamma(c)$$

Moreover, since $\hat{\sigma}^2 \approx \sigma^2$,

$$\widehat{\text{Var}}(\hat{\Gamma}(c)) := \left[\sum_i c_i^2 \right] \frac{\hat{\sigma}^2}{n} \approx \left[\sum_i c_i^2 \right] \frac{\sigma^2}{n} = \text{Var}(\hat{\Gamma}(c))$$

Reference Distribution

If the null hypothesis is true, the statistic is close to 0, because,

$$\hat{\Gamma}(c) \approx \Gamma(c) = 0$$

In fact, under the null, it's possible to derive that

$$\frac{\hat{\Gamma}(c)}{\sqrt{\widehat{\text{Var}}(\hat{\Gamma}(c))}}$$

is t -distributed with $N - a$ degrees of freedom (the proof is unimportant in this class).

Confidence Interval

Given this reference distribution, it's also possible to derive a confidence interval,

$$\left[\hat{\Gamma}(c) - t_{\text{left}} \sqrt{\widehat{\text{Var}}(\hat{\Gamma}(c))}, \hat{\Gamma}(c) + t_{\text{left}} \sqrt{\widehat{\text{Var}}(\hat{\Gamma}(c))} \right]$$

- This quantifies the uncertainty in our estimate of $\hat{\Gamma}(c)$.
- In practice, we would never compute these statistics by hand.

Multiple Comparisons

- If we knew the interest contrasts in advance, we will be fine
- But what if we go in search for significant results using different contrasts?

Multiple Comparisons

- We need to adapt our methodology to account for the search over contrasts
- The quantity of interest is the experimentwise error rate, the probability that we get at least one false positive across the entire experiment
- Three methods, each with different properties,
 - Scheffe's method
 - Tukey's Honest Significant Difference
 - Fisher's Least Significant Difference

Scheffe's Method

- Suppose we are interested in m contrasts c_1, \dots, c_m
- We can widen the confidence intervals for each to control the experimentwise error
- It's not obvious, but the appropriate scaling factor is

$$\sqrt{(a - 1) F_{\frac{\alpha}{2}, a-1, N-a}}$$

Tukey's Honest Significant Difference

- A common special case is when we're interest in all pairwise comparisons,

$$\Gamma(c) = \mu_i - \mu_j$$

- If we want to make confidence intervals, we should center them around,

$$\hat{\Gamma}(c) = \bar{y}_i - \bar{y}_j.$$

but how wide should they be?



Tukey's Honest Significant Difference

- Let \bar{y}_{\max} and \bar{y}_{\min} be the largest and smallest of the group means
- Notice that

$$|\bar{y}_i - \bar{y}_j| \leq \bar{y}_{\max} - \bar{y}_{\min}$$

- Therefore, we can rescale our confidence intervals based on the reference distribution for the difference $\bar{y}_{\max} - \bar{y}_{\min}$

Fisher's Least Significant Difference

Notice that the variance of the differences between two group's means is,

$$\begin{aligned}\text{Var}(\bar{y}_i - \bar{y}_j) &= \text{Var}(\bar{y}_i) + \text{Var}(\bar{y}_j) \\ &= \frac{\sigma^2}{n_i} + \frac{\sigma^2}{n_j} \\ &\approx \hat{\sigma}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)\end{aligned}$$

Fisher's LSD

Fisher's LSD compares each difference $|y_i - y_j|$ to the cutoff,

$$t_{\text{right}} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

and rejects the null that the pairs have equal means if the difference is larger.

Though, like Tukey's method, it considers all pairs of means, it *does not* control the experimentwise error.

Implementation

ANOVA Estimates

Before we construct any contrasts, we need to have an underlying ANOVA fit.

```
library(readr)
etch_rate <- read_csv("https://uwmadison.box.com/shared/static/vw3ldbvgvn7rupt4tz3ditl1mpupw")
etch_rate$power <- as.factor(etch_rate$power)
fit <- lm(rate ~ power, data = etch_rate)
aov_fit <- aov(fit)
```

Defining and Testing a Contrast

- To fit a contrast, we can use the `fit.contrast` function from the `gmodels` package.
- Which power levels are we comparing with this contrast?

```
library(gmodels)
contrast <- c(1, -1, 0, 0)
fit.contrast(aov_fit, "power", contrast)
```

```
##               Estimate Std. Error   t value    Pr(>|t|)
## power c=( 1 -1 0 0 )    -36.2    11.55335  -3.133289 0.006416224
## attr(,"class")
## [1] "fit_contrast"
```

Confidence Intervals

We can get a confidence interval for the contrast using the `confi.int` parameter.

```
fit.contrast(aov_fit, "power", contrast, conf.int = 0.95)
```

```
##               Estimate Std. Error  t value    Pr(>|t|)  lower CI  upper CI
## power c=( 1 -1 0 0 )    -36.2    11.55335 -3.133289 0.006416224 -60.69202 -11.70798
## attr(,"class")
## [1] "fit_contrast"
```

Defining Many Contrasts

In a multiple testing setting, we can specify each contrast as a separate row in a matrix.

```
contrasts <- matrix(
  c(1, -1, 0, 0,
    1, 1, -1, -1,
    0, 0, 1, -1),
  nrow = 3, byrow = TRUE
)

fit.contrast(aov_fit, "power", contrasts, conf.int = 0.95)
```

```
##              Estimate Std. Error   t value    Pr(>|t|)   lower CI   upper CI
## power c=( 1 -1 0 0 )    -36.2    11.55335   -3.133289 6.416224e-03 -60.69202 -11.70798
## power c=( 1 1 -1 -1 )  -193.8    16.33891  -11.861256 2.434581e-09 -228.43694 -159.16306
## power c=( 0 0 1 -1 )   -81.6    11.55335   -7.062884 2.683834e-06 -106.09202 -57.10798
## attr(,"class")
## [1] "fit_contrast"
```

Scheffe's Method

- The Scheffe adjusted confidence intervals can be found using `PostHocTest` from the `DescTools` package
- Make sure to set `method = "scheffe"`

```
library(DescTools)
PostHocTest(aov_fit, method = "scheffe", conf.level = 0.95, contrast = t(contrasts))
```

```
##
## Posthoc multiple comparisons of means: Scheffe Test
## 95% family-wise confidence level
##
## $power
##          diff      lwr.ci      upr.ci    pval
## 160-180      -36.2   -72.21352   -0.1864788  0.0486 *
## 160,180-200,220 -193.8 -244.73081 -142.8691899 3.8e-08 ***
## 200-220      -81.6  -117.61352   -45.5864788 3.6e-05 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tukey's Method

- All the intervals for Tukey's method can be found using the [TukeyHSD](#) function.
- Each row gives an interval for a $\mu_i - \mu_j$.

```
TukeyHSD(aov_fit)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = fit)
##
## $power
##      diff      lwr      upr      p adj
## 180-160  36.2    3.145624  69.25438 0.0294279
## 200-160  74.2   41.145624 107.25438 0.0000455
## 220-160 155.8 122.745624 188.85438 0.0000000
## 200-180  38.0    4.945624  71.05438 0.0215995
## 220-180 119.6   86.545624 152.65438 0.0000001
## 220-200  81.6   48.545624 114.65438 0.0000146
```

Fisher's Method

- The `PostHocTest` function can also be used for Fisher's method
- Notice that the intervals are all narrower: we are not controlling the experimentwise error

```
PostHocTest(aov_fit, method = "lsd", conf.level = 0.95)
```

```
##
##   Posthoc multiple comparisons of means : Fisher LSD
##     95% family-wise confidence level
##
## $power
##           diff      lwr.ci      upr.ci      pval
## 180-160   36.2   11.70798   60.69202   0.0064 **
## 200-160   74.2   49.70798   98.69202  8.4e-06 ***
## 220-160  155.8  131.30798  180.29202  3.7e-10 ***
## 200-180   38.0   13.50798   62.49202   0.0046 **
## 220-180  119.6   95.10798  144.09202  1.7e-08 ***
## 220-200   81.6   57.10798  106.09202  2.7e-06 ***
##
```

Exercise

This walks through parts of problem 3.9.

The tensile strength of Portland cement is being studied. Four different mixing techniques can be used economically. An experiment was conducted and the following data were collected.

