OI 7.5: Post Hoc Tests for ANOVA

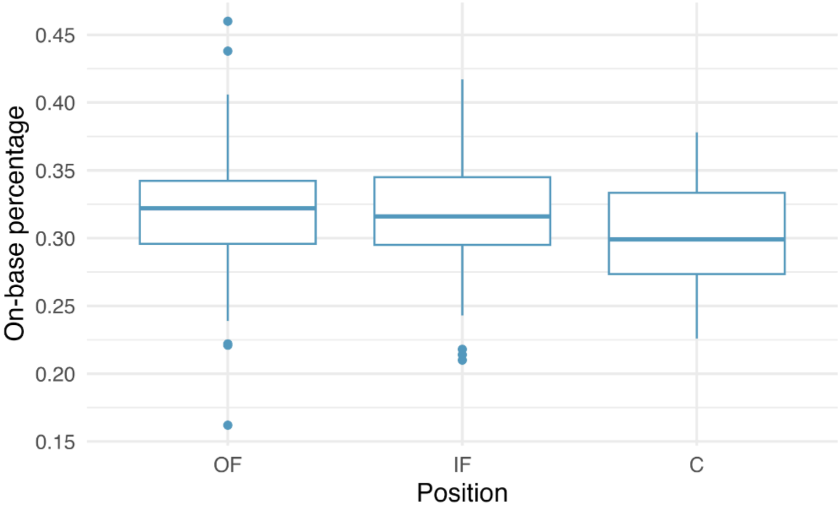
Packages

```r
library(openintro)
library(tidyverse)
```

Example

We would like to discern whether there are real differences between the on-base percentage (OBP) of baseball players according to their position: outfielder (OF), infielder (IF), and catcher (C). We will use a dataset called `mlb_players_18`, which includes batting records of over 400 Major League Baseball (MLB) players from the 2018 season who had at least 100 at bats. The on-base percentage roughly represents the fraction of the time a player successfully gets on base or hits a home run.

Example


On-base percentage
0.45
0.40
0.35
0.30
0.25
0.20
0.15
OF
IF
C
Position

Example

$$H_0 : \mu_{IF} = \mu_{OF} = \mu_C$$

H_A : At least one mean is different from another.

We can get the p-value and other relevant statistics from an ANOVA table.

term	df	sumsq	meansq	statistic	p.value
position	2	0.0161	0.0080	5.08	0.0066
Residuals	426	0.6740	0.0016		

Post-Hoc Tests

- There is strong evidence to conclude against (reject) H_0.
- We might wonder, which of the classes are actually different?
- We might be tempted to preform a t-test for differences in each possible pair of groups.

What are all the different pairs of groups for this problem?

Post-Hoc Tests

- If we preform multiple t-test for differences in each possible pair of groups using the standard method we learned earlier we increast Type 1 error rate (which means the test is invalid!)
- **post-hoc**: comes from latin, means *after this*
- If we conclude against (reject) H_0 can preform post-hoc tests (after an ANOVA F-test) to see which pairs of means are different from each other.
- Let $k =$ the number of means in the ANOVA tests.
- Let $K = \frac{k(k-1)}{2} =$ the number of all possible pairs are compared.

Post-Hoc Tests

- Post-Hoc ANOVA tests are *very* similar to standard t-tests for difference in means.
- The only change is we need to adjust the p-value (*or* we can correct α)

Bonferroni Correction for Multiple Comparisons

Conduct a t-test for the difference in means for all K possible pairs and collect the p-values (the pairwise p-values). The corrected the pairwise p-value for multiple comparisons is calculated using

$$\text{p-value}^* = \text{p-value} \times K$$

for each pairwise p-value.

- Use p-value* compared to α to assess which groups are different.

Post-Hoc Tests

Recall our baseball data set comparing OB, IF, and C. Using the summary statistics below, calculate the all pairwise p-values, and make the adjusted p-value.

```
mlb_data <- mlb_players_18 |>
  filter(AB>100) |>
  rename(position1 = position)|>
  mutate(position = ifelse(position1 == "C", "C",
                          ifelse(position1 == "RF"| position1 == "LF"|
                                 position1=="CF", "OF", "IF")))
mlb_data |>
  group_by(position) |>
  summarize(n = n(), mean = mean(OBP), sd = sd(OBP))
```

```
# A tibble: 3 x 4
  position     n  mean     sd
  <chr>    <int> <dbl>  <dbl>
1 C           64 0.302 0.0382
2 IF         210 0.319 0.0375
3 OF         160 0.320 0.0426
```

Post-Hoc Tests in R

R will do this for us!

```
pairwise.t.test(mlb_data$OBP,
                mlb_data$position,
                pool.sd = F, paired = F,
                p.adjust.method = "bonferroni")
```

```

    Pairwise comparisons using t tests with non-pooled SD

data:  mlb_data$OBP and mlb_data$position

   C      IF
IF 0.0065 -
OF 0.0080 1.0000

P value adjustment method: bonferroni
```

Post-Hoc Tests in R

```
pairwise.t.test(mlb_data$OBP,
                mlb_data$position,
                pool.sd = F, paired = F,
                p.adjust.method = "bonferroni")
```

Outcome Variable

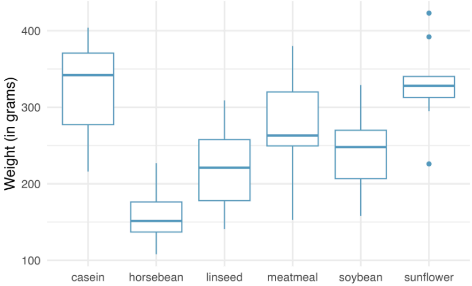
Grouping Variable

Post-Hoc Tests

- It is possible to reject the null hypothesis using ANOVA and then to not subsequently identify differences in the pairwise comparisons.
- This does not necessarily mean something wrong!
- Post-Hoc tests are not perfect, there are more sophisticated versions that might be able to better find the difference.
- The Bonferroni correction is very conservative (this is good), but sometimes it over-corrects the p-value.

Practice Problem

*Recall the problem **[IMS 22.5]** from our last set of slides*. An experiment was conducted to measure and compare the effectiveness of various feed supplements on the growth rate of chickens. Newly hatched chicks were randomly allocated into six groups, and each group was given a different feed supplement.



The F-test test produced a small p-value. Conduct post-hoc pairwise tests for the different feed groups and determine which feeds are different from each other. This data set is stored in R and is called `chickwts`.