Olympic Figure Skating: One-ANOVA (Solutions)

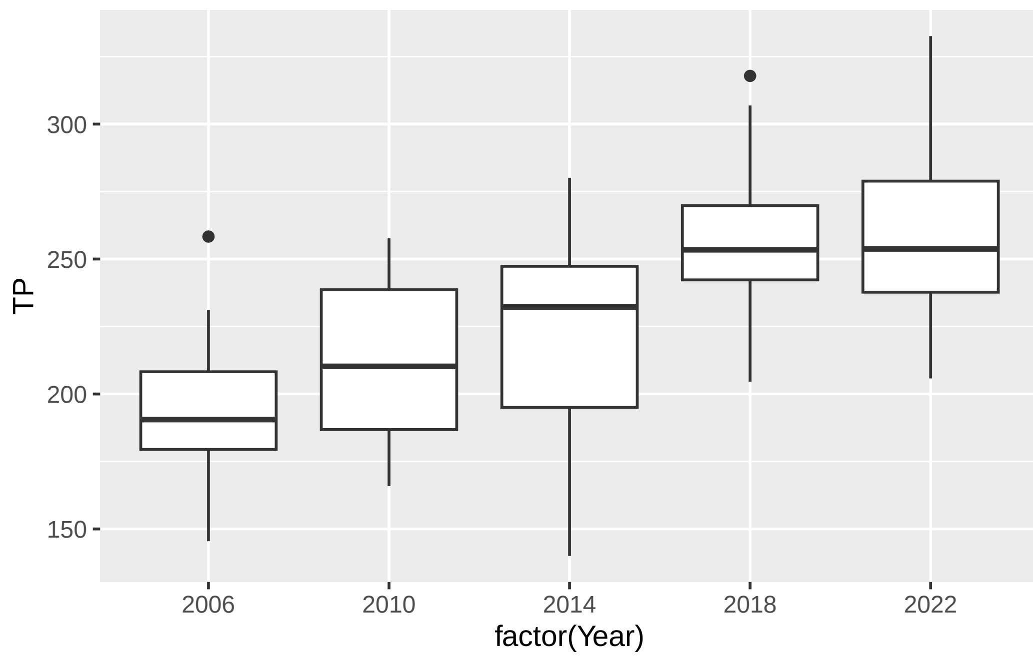
Figure skating is a sport where athletes perform choreographed routines on ice to music, combining athletic skill with artistic expression. Skaters compete in different categories like men's singles, women's singles, pairs (a man and a woman skating together), and ice dance (which focuses more on rhythm and movement). Each skater or team performs two routines: the short program and the free skate. The short program is a shorter routine with a list of required moves that every skater must include. Judges rate how well each move is done and how smoothly everything fits with the music. The free skate routine is longer and gives skaters more freedom to show their personal style and creativity. They still need to include certain types of moves, but they can choose how to put them together. Judges give scores for both routines based on how difficult the moves are, how well they're done, and how artistic the performance is. The scores from both routines are added together, and the skater or team with the highest total wins.

Read in the dataset:

```r
library(readr)
library(ggplot2)
library(dplyr)
menSkating <- read_csv("MensFigureSkating.csv")
```

1) Create a box plot with `Year` on the x-axis and `TP` (Total Points) on the y-axis. Group by Year to get a box plot for each year. Describe what you see.

```r
ggplot(data = menSkating,
       aes(x = factor(Year),
           y = TP,
           group = Year)
       ) +
  geom_boxplot()
```



We would like to determine if these means have any statistically significant differences.

2) Explain why **One-Way ANOVA** may be a good candidate method to use here.
3) Before using ANOVA, we will first need to check for assumptions.

The 3 assumptions are:

- **Independence:** if the samples are collected independently from each other
- **Normality:** if the data is normally distributed (or we have sufficiently large sample sizes such that the Central Limit Theorem holds)
- **Constant Variance:** if the difference between observed and predicted values is equal among data points.

Evaluate the three assumptions. Use your boxplots to assist. You may also want to investigate the data directly (paying close attention to the names of the skaters).

Solution: - Constant Variance: The boxplots all look to have approximately the same variation. Could also check standard deviations for each group. Here is the code to calculate them. (Note the the ratio of largest to smallest is $36.6/25.3 \approx 1.44$)

```r
menSkating |>
  group_by(Year) |>
  summarize(n = n(), sd = sd(TP))
```

```
# A tibble: 5 x 3
   Year     n    sd
  <dbl> <int> <dbl>
1  2006    24  25.3
2  2010    24  28.9
3  2014    24  36.6
4  2018    24  30.3
5  2022    24  31.4
```

- Normality: 2005 and 2018 have outliers and several years have a small amount of skew. However, with 24 observations there is likely enough observations that these issues won't have a major impact.
- Independence: This one fails for two reasons. (1) Olympic Skaters are not a random sample (however, one might be able to justify using as a representative sample of extremely competitive skaters) and (2) multiple skaters appear in multiple years. (It is beyond the scope of this module as to measuring the impact of this.)

This could lead to a nice discussion with students on the tradeoff between weak assumptions and value gained from the analysis.

---

4. Regardless of your evaluation of the assumptions, we will perform the ANOVA hypothesis test - recognizing that it may not be perfect, but it is still potentially informative.

a. State the null hypothesis and alternative hypotheses

Null - There is no difference in the mean total points throughout the years

Alternative - Not all years have equal mean total points

b. Perform an ANOVA hypothesis test to determine if there are any statistically significant differences in `TP` (Total Points) between the years using the aov() function. You will need to change `Year` to a factor.

```r
anova_result <- aov(TP ~ as.factor(Year), data = menSkating)
summary(anova_result)
```

```
                 Df Sum Sq Mean Sq F value   Pr(>F)    
as.factor(Year)   4  79109   19777   20.95 5.39e-13 ***
Residuals       115 108546     944                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c. Report the F-Statistic, degrees of freedom, and p-value.

F-Statistic = 20.95 P-value = 5.39e-13 df = 4 and 115

d. Use the p-value to make a conclusion.

There is clear statistical evidence to conclude that there is a least one pair of years with different mean total points (F-Statistic = 20.95, P-value = 5.39e-13, df = 4 and 115).

---

5. In order to find out what years are statistically discernable from each other we can perform a Post Hoc Tukey Test.

Perform a Post Hoc Tukey test using the `TukeyHSD()` function to determine between which years there is a statistically significant difference in scoring. Report those years. Also report the years that are not significant.

```r
tukey_result <- TukeyHSD(anova_result)
tukey_result
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = TP ~ as.factor(Year), data = menSkating)

$`as.factor(Year)`
              diff        lwr      upr     p adj
2010-2006 16.85125  -7.729027 41.43153 0.3231979
2014-2006 27.15375   2.573473 51.73403 0.0225473
2018-2006 63.92542  39.345139 88.50569 0.0000000
2022-2006 64.07542  39.495139 88.65569 0.0000000
2014-2010 10.30250 -14.277777 34.88278 0.7730498
2018-2010 47.07417  22.493889 71.65444 0.0000054
2022-2010 47.22417  22.643889 71.80444 0.0000050
2018-2014 36.77167  12.191389 61.35194 0.0006121
2022-2014 36.92167  12.341389 61.50194 0.0005747
2022-2018  0.15000 -24.430277 24.73028 1.0000000
```

Significant:

- 2014 and 2006
- 2018 and 2006
- 2022 and 2006
- 2018 and 2010
- 2022 and 2010
- 2018 and 2014
- 2022 and 2014

Not significant: - 2010 and 2006 - 2014 and 2010 - 2022 and 2018

6. The Tukey Test also provides a confidence interval.

Interpret the confidence interval for 2018 and 2014 comparison in the context of this data situation.

With 95% confidence, on average, competitive figure skaters scored between 12.2 and 61.4 more (total) points in 2018 than 2014.

---

7. Optional: Compact Letter Displays

A **pairwise comparison** is used to test whether differences between group means are statistically significant. In our case, each **year** represents a group. After running a post-hoc test like `TukeyHSD()`, we often summarize the results using **letters**:

- **Groups that share a letter have means that are *not* statistically significantly different.**
- **Groups that do *not* share a letter have means that are statistically significantly different.**

This system of labeling is called a **compact letter display**. It helps us quickly interpret which groups are different based on the test results.

Ordering the Groups

Compact letter displays are typically presented in a logical order, most often by ordering groups according to the **sample mean response** in **descending order**.

a. Start by filling in the first two columns of this table. Remember to order the groups according to the sample mean Total Points (`TP`) in descending order.

Year	Mean Total Points (`TP`)	Group (Letter)
2022	258	a
2018	258	a
2014	221	b
2010	211	b c
2006	194	c

One could also use R to obtain the CLD. Here is a quick example.

```
library(multcomp)
glht(anova_result, linfct=mcp(`as.factor(Year)`="Tukey")) |> cld(decreasing = TRUE)
```

```
2006 2010 2014 2018 2022
 "c" "bc"  "b"  "a"  "a"
```

b. To make the rest of the table, we will compare the statistical significance between each of pair of years. If it is not significant, both years gets a letter (following abcd… order). Use a 5% level of significance.

For example, starting with 2022… Any year that is not significant with 2022 gets the letter "a". Place a letter "a" at 2022 and 2018 since they are not significant with each other. No other year will have an "a" because they are significant with 2022.

Then we move to 2018 with the letter "b" and so on.

c. After assigning all letters, compare your results to what you found in the previous question when using `TukeyHSD()`? They should match. If not, double-check your letters!

---

8. At the 2018 Winter Olympics, many figure skaters began placing their jumps toward the end of their free skate programs to take advantage of the scoring bonus awarded for second-half elements. This strategy led to an overall increase in scores. In response, the International Skating Union (ISU) introduced the "Zagitova Rule," which limits the number of jumps that can be performed in the second half of a free skate. Additionally, the ISU expanded the Grade of Execution (GOE) scoring range from -3 to +3 to a broader scale of -5 to +5. Despite these changes, skaters continued to find ways to maximize their scores by attempting more difficult jumps and leveraging the new scoring range to their advantage.

Do the results from Tukey's HSD support this trend?

Yes the tukey test does support this trend showing the significant difference between years with the rule changes and without the rule changes.