practice-lab-anova

Kevin George

***##a)***  
*# Data for the marks based on the bench*  
front = **c**(82, 83, 97, 93, 55, 67, 53, 54, 58)  
middle = **c**(83, 78, 68, 61, 77, 54, 69, 51, 63)  
back = **c**(38, 59, 55, 66, 45, 52, 52, 61, 60)  
  
*# Check Normality*  
total = **cbind**(front,middle,back);total

## front middle back  
## [1,] 82 83 38  
## [2,] 83 78 59  
## [3,] 97 68 55  
## [4,] 93 61 66  
## [5,] 55 77 45  
## [6,] 67 54 52  
## [7,] 53 69 52  
## [8,] 54 51 61  
## [9,] 58 63 60

**shapiro.test**(total)

##   
## Shapiro-Wilk normality test  
##   
## data: total  
## W = 0.94787, p-value = 0.1903

Since p-value < 0.05, We reject H0 and conclude that the data follows normal distribution

*# Combine into a single dataset*  
marks = **c**(front, middle, back)  
bench = **factor**(**rep**(**c**("Front", "Middle", "Back"), each = 9))  
  
*# Perform ANOVA*  
anova\_result = **aov**(marks **~** bench)  
**summary**(anova\_result)

## Df Sum Sq Mean Sq F value Pr(>F)   
## bench 2 1430 715.1 4.252 0.0263 \*  
## Residuals 24 4036 168.2   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Since p-value for ‘bench’ < 0.05, we reject H0 and conclude that there is a difference in marks between benches

*# If significant, perform Post-hoc analysis (Tukey's HSD)*  
tukey\_result = **TukeyHSD**(anova\_result)  
**print**(tukey\_result)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = marks ~ bench)  
##   
## $bench  
## diff lwr upr p adj  
## Front-Back 17.111111 1.844027 32.37819 0.0259743  
## Middle-Back 12.888889 -2.378195 28.15597 0.1093033  
## Middle-Front -4.222222 -19.489306 11.04486 0.7711567

From Tukey’s Test we conclude that means values of Front and Back bench differ significantly, since their p-value < 0.05

***##b)***  
*# Shift the origin (e.g., subtract 50 from all marks)*  
marks\_shifted = marks **-** 50  
  
*# Perform ANOVA on shifted data*  
anova\_result\_shifted = **aov**(marks\_shifted **~** bench)  
**summary**(anova\_result\_shifted)

## Df Sum Sq Mean Sq F value Pr(>F)   
## bench 2 1430 715.1 4.252 0.0263 \*  
## Residuals 24 4036 168.2   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

*# Tukey's HSD on shifted data*  
tukey\_shifted = **TukeyHSD**(anova\_result\_shifted)  
**print**(tukey\_shifted)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = marks\_shifted ~ bench)  
##   
## $bench  
## diff lwr upr p adj  
## Front-Back 17.111111 1.844027 32.37819 0.0259743  
## Middle-Back 12.888889 -2.378195 28.15597 0.1093033  
## Middle-Front -4.222222 -19.489306 11.04486 0.7711567

Since ANOVA measures the variation in groups from their means, it is not affected by a change in origin.

***##c)***  
*# Marks data*  
front = **c**(82, 83, 97, 93, 55, 67, 53, 54, 58)  
middle = **c**(83, 78, 68, 61, 77, 54, 69, 51, 63)  
back = **c**(38, 59, 55, 66, 45, 52, 52, 61, 60)  
  
*# Combine data into a single vector*  
marks = **c**(front, middle, back)  
  
*# Define the factors*  
bench = **factor**(**rep**(**c**("Front", "Middle", "Back"), each = 9))  
batch = **factor**(**rep**(1**:**9, times = 3))  
  
*# Create a data frame*  
data = **data.frame**(marks, bench, batch)  
  
*# Two-way ANOVA with interaction*  
anova\_result <- **aov**(marks **~** bench **+** batch, data = data)  
**summary**(anova\_result)

## Df Sum Sq Mean Sq F value Pr(>F)   
## bench 2 1430 715.1 4.327 0.0315 \*  
## batch 8 1392 174.0 1.053 0.4399   
## Residuals 16 2644 165.3   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

In a 2 way ANOVA test, we include batches as another factor and perform the test.

Here too, only there is a significant difference between benches and not batches, since only p-value of bench < 0.05

d)

#### Purpose of Post-hoc Analysis

Post-hoc analysis is used to identify specific group differences after finding a statistically significant result in a hypothesis test, such as ANOVA. While ANOVA indicates whether at least one group mean is different, post-hoc tests pinpoint which groups are significantly different from each other.

**Why It’s Needed:**

1. ANOVA is an omnibus test and doesn't specify pairwise differences.
2. Post-hoc tests compare all possible pairs of groups while controlling for Type I error (false positives).

#### Common Post-hoc Tests

1. **Tukey’s Honestly Significant Difference (HSD) Test**
   * **Purpose:** Compares all possible pairs of group means.
   * **Key Features:**
     + Assumes equal variances.
     + Suitable for balanced designs (equal group sizes).
     + Controls the family-wise error rate (FWER).
   * **Advantages:**
     + Powerful when assumptions are met.
     + Easy to interpret.

* **Output:** Identifies which pairs of means are significantly different.

1. **Scheffé’s Test**

* **Purpose:** Tests all possible contrasts (linear combinations) between group means.
* **Key Features:**
  + More conservative than Tukey’s HSD.
  + Controls Type I error but has less power.
  + Suitable for unbalanced designs (unequal group sizes).
* **Advantages:**
  + Broad applicability, even when group sizes are unequal or variances differ.
* **Output:** Identifies whether a general contrast between groups is significant.

e)

We can perform Post-hoc analysis in objective 1 since we can observe a significant difference among benches after performing an ANOVA test.

### Why Post-hoc Analysis is Possible

1. **ANOVA Objective**: ANOVA tests whether there is a significant difference in means among the groups (here, "Front", "Middle", and "Back").
2. **Significant Result**: If ANOVA indicates a significant difference (p-value < 0.05), it only tells us that at least one group mean differs but does not specify which groups are different.
3. **Post-hoc Analysis Purpose**: Post-hoc tests are used to pinpoint the specific group(s) contributing to the observed difference.

A Tukey’s Test has been performed above under section a)