Practice Examples

The median is more robust (less sensitive to outliers in the data)

Can be more robust in comparison to the mean when looking at extremely large or small values

1. Median is robust to extreme values

   TRUE / FALSE

Skewed right: mean > median (+)

Long right tail (large observations on the right of the chart)

Skewed left: mean < median (-)

Long left tail (small observations on the left of the chart)

2. For right-skewed data, median > mean

   TRUE / FALSE

Zero Skewness (mean = median)

Normal distribution is symmetric, bell shaped

3. When p-value is 0.04, we always reject the null hypothesis

   TRUE / FALSE

Need to Know Significance Level

It depends on alpha (significance level)

4. When performing unbalanced 2-way ANOVA, inferential conclusions from Type 1 and Type 3 SS tests will be always same

   TRUE / FALSE

5. If the ANOVA model is significant with very small p-value like 0.00001, this model will always show very large R-square

   TRUE / FALSE

The size of the data may have an impact

There is no established association/relationship between p-value and R-square. This all depends on the data (i.e.; contextual).

6. When using the Type 3 SS for unbalanced ANOVAs: if we change the order of the variables in the model, then the Type 3 SS will change

   TRUE / FALSE

Type 3 SS

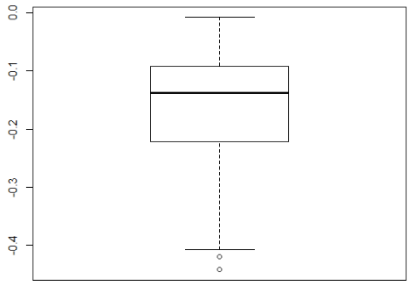
If we change the order of the variables, Type 3 SS will NOT change.

7. In forward selection, if we increase the p-value cut-off, the final model will get larger

   TRUE / FALSE

8. In 1-way ANOVA, although model turns out to be insignificant, we need to perform post-hoc test

   TRUE / FALSE



9. how the data is skewed?

Left Skewed

mean < median (-)

Long left tail (small observations on the left of the chart)

10. Below is a result for the equal variance test to perform two-sample t-test. State which test you choose (pooled t-test? Or Satterthwaith test?)

```
        F test to compare two variances

data:  July$Wind and Aug$Wind
F = 0.8857, num df = 30, denom df = 30, p-value = 0.7418
alternative hypothesis: true ratio of variances is not equal to 1
```

Result

• After conducting an F test (var.test), the results indicated a p-value of 0.7418. This high p-value means we can't reject the null hypothesis, indicating that the two groups have an equal variance.

• Based on these findings (equal variance), the next step is to conduct a Pooled T-test

11. Below is a result for the equal variance check (Levene's test) for ANOVA. State which test you choose between ANOVA and Welch's ANOV.

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.6457 0.003
      57
```

Result

Small p-value means we reject the null, therefore, there does not exist an equal variance. There is heteroscedascitiy of variance.

Alternate Result

Large p-value means we can't reject the null, therefore, there exists an equal variance. There is homogeneity of variance.

12. 4 assumptions of ANOVA

Model Assumption

1. Y = continuous ; X = Categorical

2. Data follows the normal distribution (each group follows the normal distribution)

3. Equal variance (Homoscedasticity)

a. No equal variance means we conduct a Welch's ANOVA

4. Independent samples

1) The response (dependent) variable is **continuous**
2) Populations from which samples were drawn **follow normal distribution**
   - i.e., Each group should be normally distributed
   - ✓ Note: ANOVA relatively *robust* to violations of normality
3) Populations from which samples were drawn must have **equal variances** (Homogeneity of Variance)
   - ✓ Need to perform equal variance test before applying ANOVA
4) Observations must be **independent of one another**

13. What is the goal of ANOVA?

Goal

Analyze the difference among groups and study the behaviors of response variable depending on grouping variable

14. Why we perform post-hoc test?

- Making many comparisons at once!!
- Need to account for increased probability of making wrong decision
  Because making multiple comparisons increases the probability of making the wrong decision, we need to make a correction in calculating p-value
- Need correction in calculating p-value from t-test
  - **Scheffe method, Tukey's Method, etc.**

15. Interpretation of post-hoc test result

• Should know how to interpret the result. What is null hypothesis and what kind of conclusion can we make?

16. Check quiz questions

17. Able to Interpret R outputs in HW.

Hypothesis is for the first comparison (Dose 1 and 0.5)

One-way ANOVA example:

Pairwise t-test with modified p-values:
H0: $\mu_1 = \mu_{0.5}$ vs. H1: $\mu_1 \neq \mu_{0.5}$

```
ScheffeTest(aov.res)

##
##   Posthoc multiple comparisons of means: Scheffe Test
##     95% family-wise confidence level
##
## $Dose
##         diff    lwr.ci    upr.ci    pval
## 1-0.5  9.130  5.758155 12.501845 4.3e-08 ***
## 2-0.5 15.495 12.123155 18.866845 1.2e-15 ***
## 2-1    6.365  2.993155  9.736845 7.6e-05 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

'>' means significantly different

Dose 1 mean > Dose 0.5 mean ; Dose 2 mean > Dose 0.5 mean ; Dose 2 mean > Dose 1 mean

Ho: μ2 = μ0.5 vs. H1: μ2 != μ0.5

Ho: μ2 = μ1 vs. H1: μ2 != μ1

diff: the estimated difference between the mean values of first group and second group

All dose groups have significantly different mean values of tooth length.

Able to get information which pairs are significantly different
Final conclusion:
All three different Dose have different effect on tooth length and specifically, Dose 2 > Dose 1 > Dose 0.5