STAT9006: Multi-Variable Data Analysis with *R*
Part I

Outline

Outline

Outline

Outline

Statistical tests

Goal	Normally distributed	Not normally distributed
Describe one group	Mean, SD	Median, IQR
Compare one group to a hypothetical value	One sample t-test	Wilcoxon test
Compare two independent groups	Independent t-test	Mann-Whitney test
Compare two dependent groups	Paired t-test	Wilcoxon test
Compare three or more independent groups	One-way ANOVA	Kruskal-Wallis test
Compare three or more dependent groups	Repeated-measures ANOVA	Friedman test

Statistical significance

When a statistic is significant:

- It simply means that you are very sure that the statistic is **reliable**.
- It does not allude to the importance of the finding.
- It does not mean the finding has any decision-making utility.

Statistical significance

It does not matter what type of statistic you are calculating (e.g., a t-test, a χ^2-test, a F-test etc.), the procedure to test for significance is always the same. If the p-value is **lower** than the level of significance:

1. The finding is **significant**.
2. **Reject** the null hypothesis.
3. The probability is **small** and the relationship/difference is meaningful, $(p \leq \alpha)$.

If the p-value is **higher** than the level of significance:

1. The finding is not significant.
2. **Fail to reject** the null hypothesis.
3. The probability is **high** and the relationship/difference happened by chance, $(p > \alpha)$.

After finding a significant relationship/difference, it is important to evaluate the **strength** of the relationship/difference.

Outline

Explore

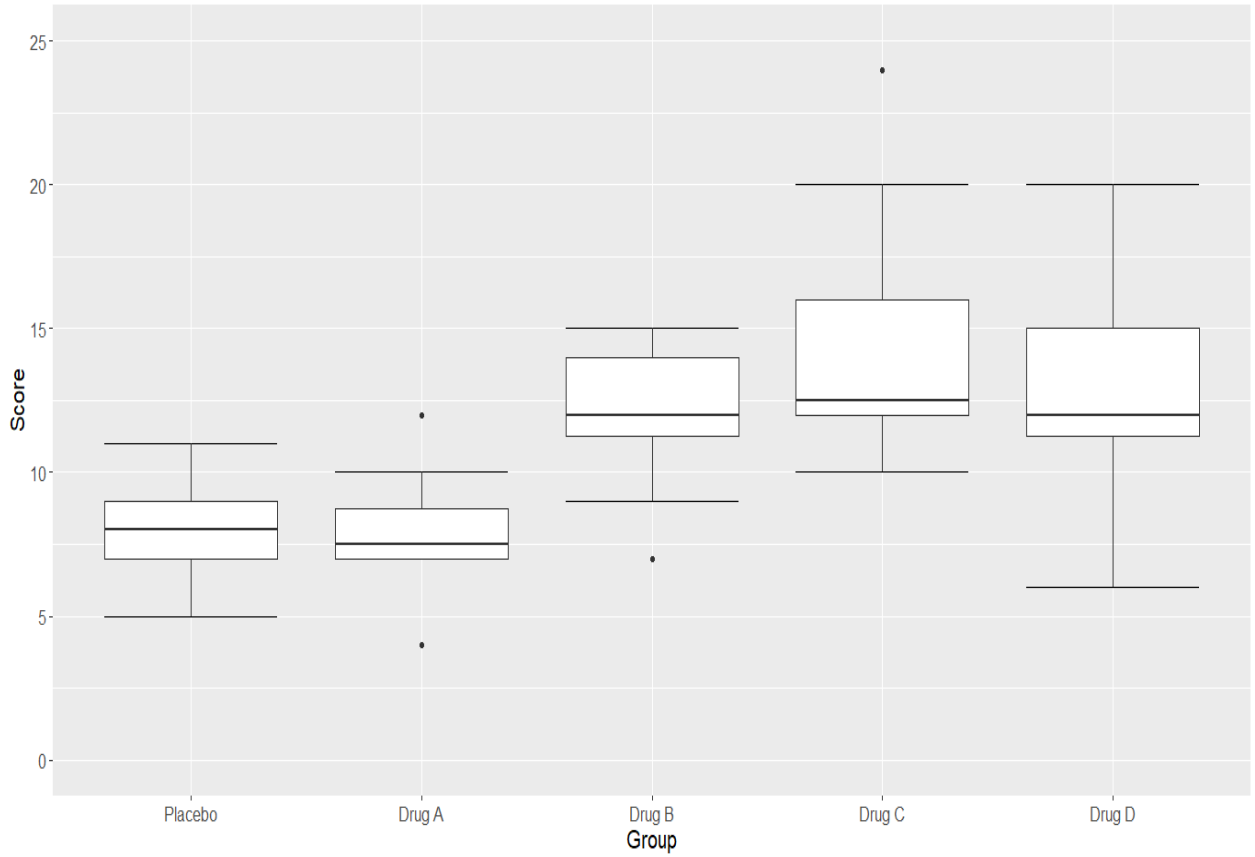
Consider the experiment (ANOVA.xlsx) in which participants were tested on shooting accuracy after ingesting one of four drug types; the remaining participants took a placebo and took the same test.

1. Check the data for anomalies such as extreme values or skewed distributions.

```
# FIRST explore the data
# Step 01: check propoerties
is.factor(ANOVA$Group)
ANOVA$Group<-factor(ANOVA$Group,levels=c("Placebo", "Drug A","Drug B","Drug C","Drug D"))
is.numeric(ANOVA$Score)

# Step 02: numerical descriptive statistics
library(dplyr)
(Stats<-ANOVA %>% group_by(Group) %>% summarise("Sample size"=n(),Mean = mean(Score),
                                                "Standard deviation"=sd(Score),
                                                Median = median(Score), "1st quartile"=quantile(Score, 0.25),
                                                "3rd quartile"=quantile(Score, 0.75), Min=min(Score),
                                                Max=max(Score)))

t(Stats)
```

Graphical descriptive statistics

```
# Step 03: graphical descriptive statistics
library(ggplot2)
(plot1<-ggplot(ANOVA,aes(x=ANOVA$Group, y=ANOVA$Score))+stat_boxplot(geom = "errorbar")+geom_boxplot()+
    coord_cartesian(ylim=c(0,25))+labs(x = "Group", y = "Score")+scale_y_continuous(breaks=seq(0,25,5))+
    theme(text = element_text(size=15)))
```



Assumptions

2. Check that the data does not violate the assumptions of normality and homogeneity of variances:

```
# SECOND check that the assumptions are not violated
# Step 01: tests of normality
library(psych)
(Norm<-ANOVA %>% group_by(Group) %>% summarise("Sample size"=n(),Mean = mean(Score),
                                               Median = median(Score), Skewness=skew(Score),
                                               "Normally distributed"=ifelse(
                                                 shapiro.test(Score)$p.value>0.05,"Yes","No")))
t(Norm)

# Step 02: test of homogeneity of variances
library(car)
leveneTest(ANOVA$Score ~ ANOVA$Group,center=mean)
```

One-way ANOVA

- H_0: **No difference** exists between the means.
- H_1: **A difference** exists between *at least* two means.

- If $p \leq \alpha$, then reject H_0 - i.e., reject the claim that no difference exists between the overall means;
- If $p > \alpha$, then fail to reject H_0 - i.e., fail to reject the claim that no difference exists between the overall means.

ANOVA

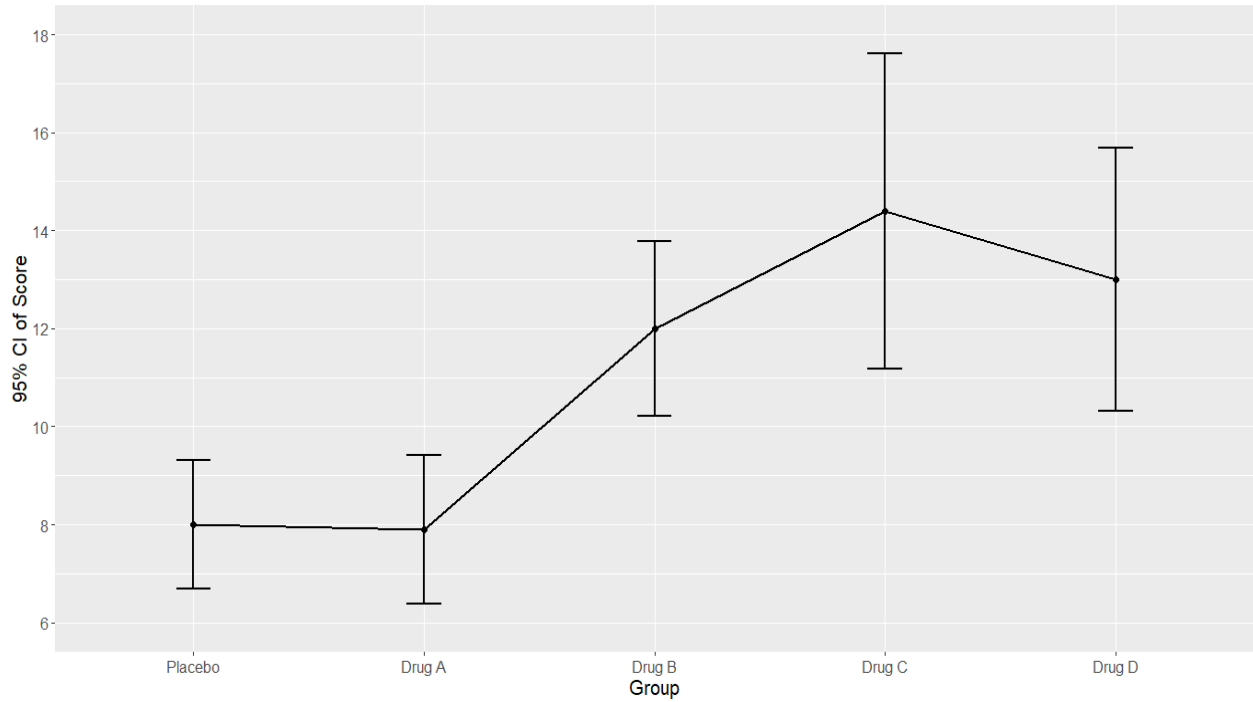
3. Depending on whether the assumptions are violated, apply the appropriate test:

```
> # THIRD apply the appropriate test
> # Option 01: ANOVA if conditions are not violated
> result<-aov(Score ~ Group, data=ANOVA)
> summary(result)
            Df Sum Sq Mean Sq F value   Pr(>F)
Group        4  351.5   87.88   9.085 1.82e-05 ***
Residuals   45  435.3    9.67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> # Option 02: If equal variances are not assumed
> library(car)
> oneway.test(ANOVA$Score ~ ANOVA$Group,var.equal=F)

        One-way analysis of means (not assuming equal variances)

data:  ANOVA$Score and ANOVA$Group
F = 9.8373, num df = 4.000, denom df = 22.036, p-value = 0.000101

> # Option 03: Kruskal Wallis if the condition of normality is violated
> kruskal.test(ANOVA$Score ~ ANOVA$Group)

        Kruskal-Wallis rank sum test

data:  ANOVA$Score by ANOVA$Group
Kruskal-Wallis chi-squared = 25.376, df = 4, p-value = 4.227e-05
```

Main effects plot

4. Plot the means:

```
# FOURTH Main effects plot (plot of the means)
library(ggpubr)
(mp<-ggline(ANOVA, y ="Score", x = "Group",add =c("mean_ci"),size=1)+theme_gray())
(plot2<-mp+  theme(text = element_text(size=15))+
    labs(y="95% CI of Score", x="Group")+
    coord_cartesian(ylim=c(6, 18)) + scale_y_continuous(breaks=seq(6,18,2)))
```



Post-hoc test

The purpose of post hoc testing is to determine exactly **which groups differ** from which others in terms of mean differences. It is possible to do the post-hoc test yourself by implementing an independent *t*-test (or Mann-Whitney *U* test) numerous times. But you have to be careful with the *p*-values.

- Suppose there are k null hypotheses to be tested and the overall Type *I* error rate (significance level) is α.
- Start by ordering the *p*-values and comparing the smallest *p*-value to α/k.
- If that *p*-value is less than α/k, then reject that hypothesis.
- Compare the remaining smallest *p*-value to $\alpha/(k-1)$.
- Continue doing this until the hypothesis with the smallest *p*-value cannot be rejected. At that point, stop.
- None of the remaining hypotheses can be rejected.

Note: For k groups, the total number of possible paired comparisons is $k(k-1)/2$.

Post-hoc test

For example:

- Four null hypotheses are tested with $\alpha = 0.05$. The four unadjusted p-values are 0.01, 0.03, 0.04, and 0.005.
- The smallest of these is 0.005. Since this is less than 0.05/4, null hypothesis four is rejected (meaning some alternative hypothesis likely explains the data).
- The next smallest p-value is 0.01, which is smaller than 0.05/3. So, null hypothesis one is also rejected.
- The next smallest p-value is 0.03. This is not smaller than 0.05/2, so you fail to reject this hypothesis (meaning you have not seen evidence to conclude an alternative hypothesis is preferable to the level of $\alpha = 0.05$).
- As soon as that happens, you stop, and therefore, also fail to reject the remaining hypothesis that has a p-value of 0.04.
- Therefore, hypotheses one and four are rejected while hypotheses two and three are not rejected.

Post-hoc test

5.

- H_0: **No difference** exists between the individual means.
- H_1: **A difference** exists between the individual means.

```
> # FIFTH Posthoc analysis
> pairwise.t.test(ANOVA$Score,ANOVA$Group,p.adj="holm") #Holm

        Pairwise comparisons using t tests with pooled SD

data:  ANOVA$Score and ANOVA$Group

       Placebo Drug A  Drug B  Drug C
Drug A 0.95863 -        -       -
Drug B 0.03069 0.03036  -       -
Drug C 0.00031 0.00027  0.36522 -
Drug D 0.00561 0.00517  0.95863 0.95863

P value adjustment method: holm
> pairwise.t.test(ANOVA$Score,ANOVA$Group,p.adj="bonferroni") #Bonferroni

        Pairwise comparisons using t tests with pooled SD

data:  ANOVA$Score and ANOVA$Group

       Placebo Drug A  Drug B  Drug C
Drug A 1.00000 -        -       -
Drug B 0.06138 0.05060  -       -
Drug C 0.00034 0.00027  0.91304 -
Drug D 0.00802 0.00647  1.00000 1.00000

P value adjustment method: bonferroni
```

Post-hoc test

```
> TukeyHSD(result) #Tukey...result is the name of the aov() output
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Score ~ Group, data = ANOVA)

$Group
                 diff         lwr       upr     p adj
Drug A-Placebo -0.1 -4.05223799  3.852238 0.9999937
Drug B-Placebo  4.0  0.04776201  7.952238 0.0460196
Drug C-Placebo  6.4  2.44776201 10.352238 0.0003180
Drug D-Placebo  5.0  1.04776201  8.952238 0.0068354
Drug B-Drug A   4.1  0.14776201  8.052238 0.0385792
Drug C-Drug A   6.5  2.54776201 10.452238 0.0002524
Drug D-Drug A   5.1  1.14776201  9.052238 0.0055623
Drug C-Drug B   2.4 -1.55223799  6.352238 0.4291513
Drug D-Drug B   1.0 -2.95223799  4.952238 0.9510451
Drug D-Drug C  -1.4 -5.35223799  2.552238 0.8510119

>
> # you might only want to test against a control - i.e., Dunnett's test is required
> library(DescTools)
> ?DunnettTest
> DunnettTest(ANOVA$Score~ANOVA$Group,control="Placebo")

  Dunnett's test for comparing several treatments with a control :
    95% family-wise confidence level

$Placebo
                 diff     lwr.ci   upr.ci   pval
Drug A-Placebo -0.1 -3.6214429 3.421443 1.0000
Drug B-Placebo  4.0  0.4785571 7.521443 0.0215 *
Drug C-Placebo  6.4  2.8785571 9.921443 0.0002 ***
Drug D-Placebo  5.0  1.4785571 8.521443 0.0029 **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Effect size

6. Determine the strength of the result.

The following table offers a rough guide to the classification of effect size in relation to values of eta-squared.

Size of effect	Eta squared
Small	$0.01 \leq \eta^2 < 0.06$
Medium	$0.06 \leq \eta^2 < 0.14$
Large	$\eta^2 \geq 0.14$

```
> # SIXTH Effect size of overall ANOVA result
> library(DescTools)
> EtaSq(result) # again this function uses output from aov()
         eta.sq eta.sq.part
Group 0.4467604   0.4467604
```

Outline

Exercise

- A manufacturer of high-performance automobiles produces disc brakes that must measure 322 millimeters in diameter. Quality control randomly draws 16 discs made by each of 8 production machines and measures their diameters.

- Use the brakes.xlsx dataset to determine whether or not the mean diameters of the brakes from the eight machines differ significantly from each other.

- This exercise should be answered using the 6 steps outlined in the above slides.