---
title: 'One-Way ANOVA: Shapesplosion'
output: html_document

Background: *Shapesplosion* was an online game modeled on the classic (physical) game "Perfection". The point was to drag different shapes into their corresponding holes. (Look up "perfection game" on Google Images if you want a visual.) Unfortunately, Shapesplosion is now defunct, but over the next few weeks we will analyze data collected by me over the last several years of student experiments.

When playing Shapesplosion, one can vary the sensitivity to which one can match one object to another. That is, as you drag the shape towards its hole, how close does the shape have to be before you can "drop" it in the hole? Do the edges need to perfectly line up, or can you just be "near" the hole? We will call this sensitivity the “match proximity” (or MP). If the MP is too restricted, players will spend more time in aligning the two objects than if the MP is less stringent.

Over the past several years, I have conducted experiments with students in Applied Stat Methods. Using randomly-drawn poker chips, each student was assigned to one of four MP groups (exact, small, medium, large).   Each student then played one round of “Shapesplosion” with that MP and recorded his/her time. The data is in "Shapesplosion-MP", which is on Moodle (load it now).

A. CONTEXT.
----
1.   Explain why this study is an experiment.

2.   Clearly define an individual in this study. In an experiment, an individual is called an *experimental unit* or a *subject* (if a person).

3. Clearly define the explanatory/predictor variable in this study. In an experiment, an explanatory variable is called a *factor*.

4. How many categories does the factor have? In an experiment, the categories of a categorical explanatory variable (factor) are called *levels* or *factor levels*.

5. Clearly define the response variable in this study. What is the unit of measure?

RESEARCH QUESTION: Is there a significant difference in average time to "beat" Shapeplosion based on the MP level? That is, does MP has a significant effect on performance? (We'll try to answer this question in this activity.)

B. EXPLORATORY DATA ANALYSIS
----
1. Construct an appropriate plot(s) to visually compare the distributions of time by MP.

2. Use R to obtain the numerical summaries of response time by MP level.

3. **Comparing the means of four groups: By inspection**
Examine the visual display and numerical summaries. Is there evidence that MP has an effect on performance? Be specific about why or why not, talking about both the plot and the statistics.

C. USING THE ANOVA MODEL
----
1. CHOOSE
Write the theoretical model symbolically:

2. FIT

  a. Use R to calculate the grand mean and the treatment means.

  b. Compute estimates for the treatment effects.

You can fit the ANOVA model in R using the function `aov()`. The syntax is exactly like `lm()`:
```{r}
#shapes.aov <- aov(response ~ explanatory, data=data)
```
If you save this object (as `shapes.aov`, for example), all the same functions work on this object as worked on lm objects: `summary()` to see the ANOVA table (although that is dangerous, as you'll see below), `plot()` to look at the residual plots, etc.

3. ASSESS: Diagnostics and Residual Analysis
Just like in the simple linear regression model, we can use the residuals to determine if the model assumptions are reasonably met.

  a. State the implicit model assumptions and indicate how you will use the residuals to determine if the model assumptions are met.

  b. Verify the model assumptions using the residuals.
```{r}
par(mfrow=c(2,2))
#plot(shapes.aov)
```

  c. Does it appear that the model conditions are satisfied? Do you think we need to take any transformations? If so, take them now and re-fit the model.

4. USE: Inference
Assuming the model conditions are satisfied...

  a. The hypotheses of interest are:
H_0:

H_a:

  b. The ANOVA table is given by:
```{r}
library(car)
#Anova(shapes.aov)
```
NOTE: You can also use `summary()` or `anova()` (instead of `Anova()`), but those will *not* work when we add more variables! (So it's good to get into the habit of using `Anova()` now.)

  c. Use the ANOVA table to conduct the appropriate test of significance and make a conclusion.

D. COMPARISON TO LINEAR REGRESSION.

Of course, there's another way to analyze this data.

We learned back in Unit A that categorical variables can be analyzed using linear regression, as well. Let's investigate the similarities and differences of ANOVA and a linear regression model in this case...

1. CHOOSE
Write the theoretical linear regresson model symbolically:

2. FIT
Fit the linear regression model, and call it `shapes.lm`.

3. ASSESS: Diagnostics and Residual Analysis
Verify the *regression* model assumptions using the residuals. What do you notice about the residual plots for `shapes.aov` compared to `shapes.lm`?

````
```{r}
par(mfrow=c(2,2))
#plot(shapes.lm)
```
````

4. USE: Inference

a. Compare the ANOVA table from Section C to `anova(shapes.lm)`. What do you notice?

b. What conclusion can you make about the variable `MP` using the `anova(shapes.lm)`?

c. What additional information do you get from `summary(shapes.lm)` that you didn't get from `Anova(shapes.aov)`?

d. Make a conclusion about the differences between MP levels (large, medium, small, exact) from `summary(shapes.lm)`. How does this compare to what we learned using the Fisher's comparisons?

e. Interpret the Intercept, and the coefficients of the Medium, Small, and Xact levels.