tennis_ball_launcher_ANOVA

Introduction

Research question

The following question guides this experiment:

- What angles and levels of PSI are associated with the greatest distance that a pneumatic tennis ball launcher can launch a tennis ball?

Factor definitions and replicates

Angle is defined as the angle that the launcher barrel is pointed at when fired and PSI is defined as the PSI level that is used to fire a given tennis ball from the launcher. Each combination of angle and PSI will be tested five times, therefore there will be five replicates and 45 total experimental runs.

Randomization

During the running of the experiment itself, the experimental runs will be performed in a random order to ensure that the ANOVA assumption of independence of residuals is not violated. Therefore, this experiment will use a balanced, replicated two-factor ANOVA.

Measurement evaluation

The response variable distance will be measured using the following procedure:

1. A table will be set up on one end of an open field with an angle measure fastened to it.
2. Orange sports cones will be laid out in a line down the field every 25 feet up to 175 feet, making seven cones in total.
3. Volunteer 1 will stand next to the table, filling the tennis ball launcher to the specifiedPSI level with a portable air compressor.
4. Volunteer 2 will then load the tennis ball into the launcher and fire it down the field at the specified angle.
5. The experimenter will run down the field to where the tennis ball first hits the ground and stand in that spot.
6. Volunteer 3 will use a measuring wheel to measure the distance from the nearest cone to where the tennis ball hit the ground and the experimenter is now standing.
7. The experimenter will use the distance from the cone to calculate the total distance traveled by the ball down the field and record it.

Loading packages

```r
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2

## Warning: package 'ggplot2' was built under R version 4.3.2

## Warning: package 'tibble' was built under R version 4.3.2

## Warning: package 'tidyr' was built under R version 4.3.2

## Warning: package 'readr' was built under R version 4.3.2

## Warning: package 'purrr' was built under R version 4.3.2

## Warning: package 'dplyr' was built under R version 4.3.2

## Warning: package 'stringr' was built under R version 4.3.2

## Warning: package 'forcats' was built under R version 4.3.2

## Warning: package 'lubridate' was built under R version 4.3.2

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 4.3.3
```

```r
library(rstatix)
```

```
## Warning: package 'rstatix' was built under R version 4.3.3

##
## Attaching package: 'rstatix'
##
## The following object is masked from 'package:stats':
##
##     filter
```

```r
library(nortest)
```

```
## Warning: package 'nortest' was built under R version 4.3.1
```

Importing data

```r
data = read.csv("C:\\Users\\scwag\\OneDrive\\Desktop\\R_Workspace\\Statistics\\tennis_ball_data.csv")
head(data)
```

```
##   run angle psi distance
## 1   1    30  30   124.50
## 2   2    45  30   153.50
## 3   3    60  30   137.63
## 4   4    45  20    93.19
## 5   5    30  20    61.63
## 6   6    60  40   171.31
```

Setting seed

```r
set.seed(99)
```

Converting to factors

```r
data$angle = as.factor(data$angle)
levels(data$angle)
```

```
## [1] "30" "45" "60"
```

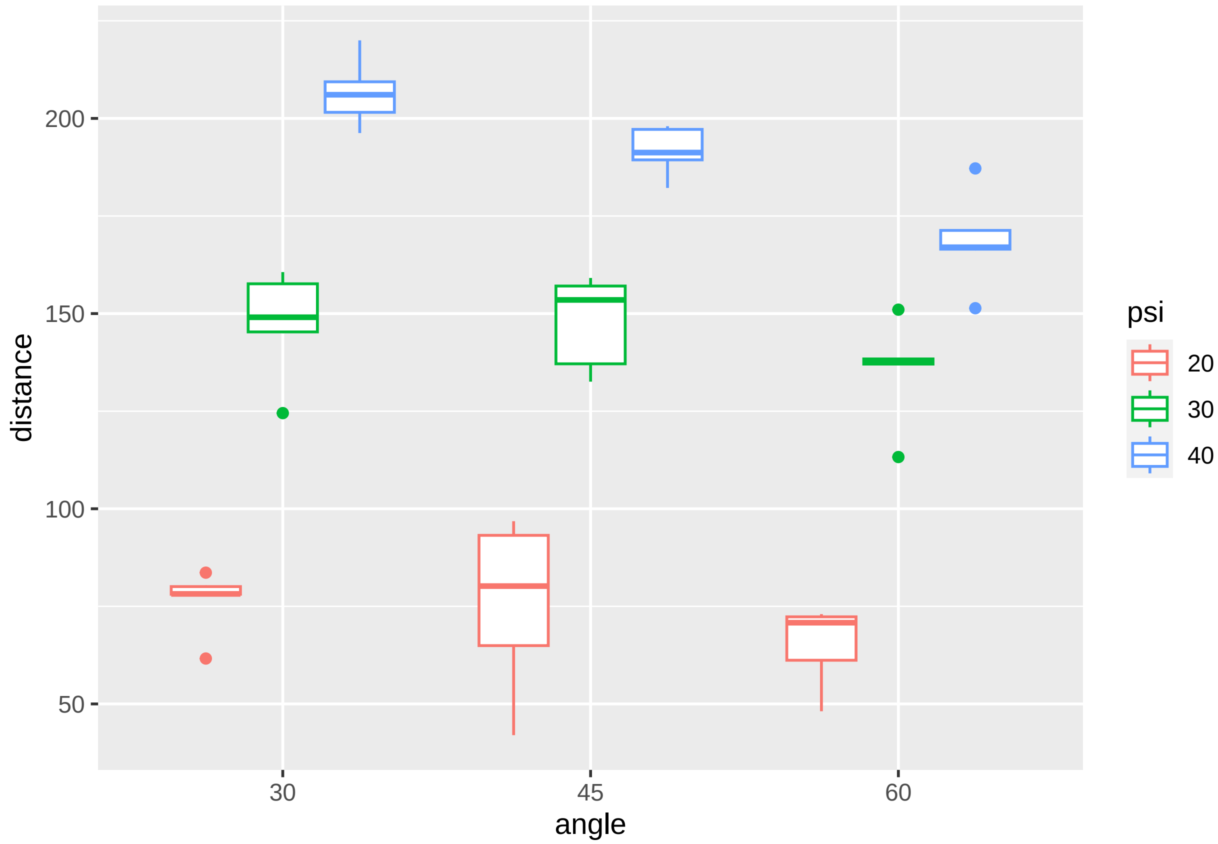
```r
data$psi = as.factor(data$psi)
levels(data$psi)
```

```
## [1] "20" "30" "40"
```

Boxplot of data

```r
box_plot = ggplot(data = data, aes(x = angle, y = distance, color = psi)) +
  geom_boxplot()

box_plot
```



2-way ANOVA

Angle and PSI are statistically significant, while their interaction term is not. Therefore, there is a statistically significant difference in mean distance between at least one angle and the other angles and a statistically significant difference in mean distance between at least one PSI setting and the other PSI settings. However, there is no statistically significant difference in mean distance for the interaction of angle and PSI.

```
anova_1 = aov(distance ~ psi + angle + angle * psi, data = data)
anova_summary = summary(anova_1)
anova_summary
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## psi           2 103817   51908 308.382  < 2e-16 ***
## angle         2   3371    1686  10.014 0.000348 ***
## psi:angle     4   1171     293   1.739 0.162792
## Residuals    36   6060     168
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Getting standardized residuals

```r
data$std_residuals = rstandard(anova_1)
head(data$std_residuals)
```

```
## [1] -1.9756441  0.4846472  0.1866547  1.5308096 -1.2660806  0.2269845
```

Getting adjusted mean squared error

```r
adj_mse = anova_summary[[1]][["Mean Sq"]][4]
adj_mse
```

```
## [1] 168.3252
```

Calculating estimated standard deviation

```r
est_sd = sqrt(adj_mse)
est_sd
```

```
## [1] 12.97402
```

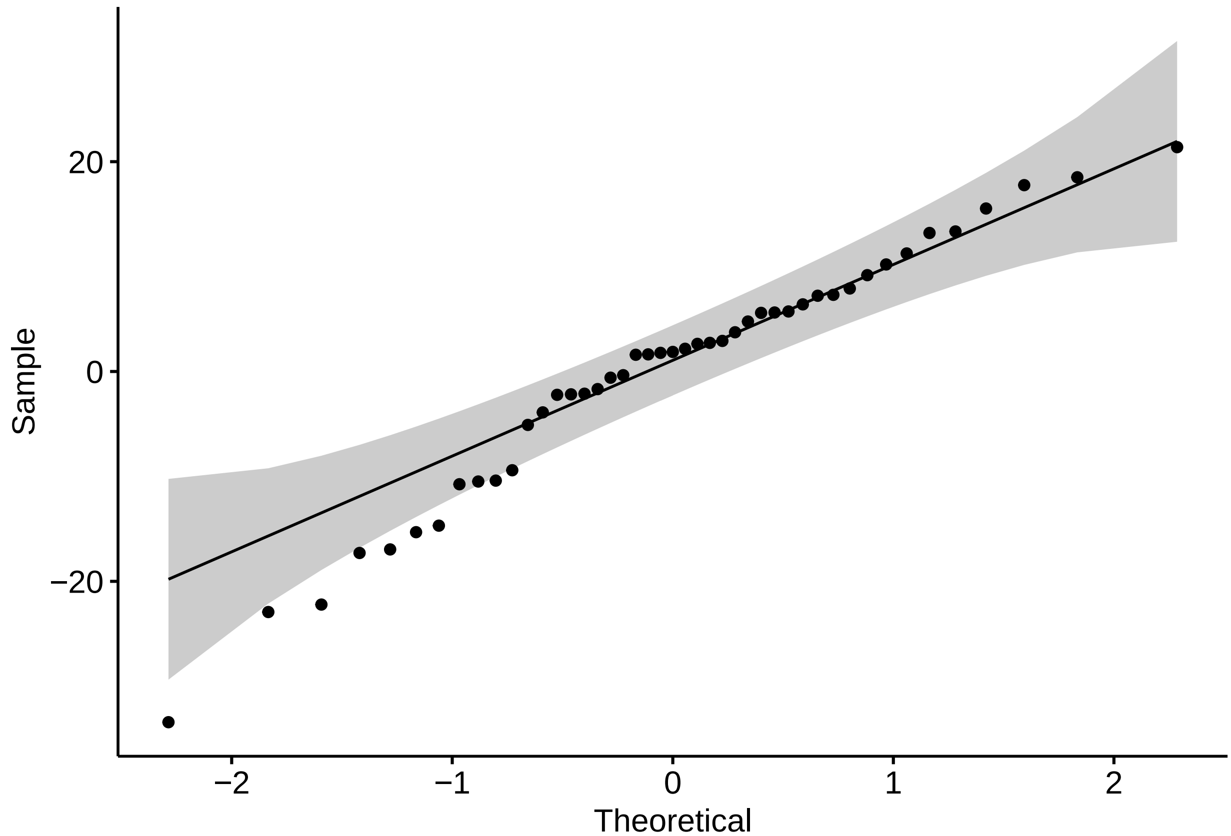
Checking ANOVA assumptions

Normality

QQ plot

The data points form a roughly straight line with some deviation near the lower extreme.

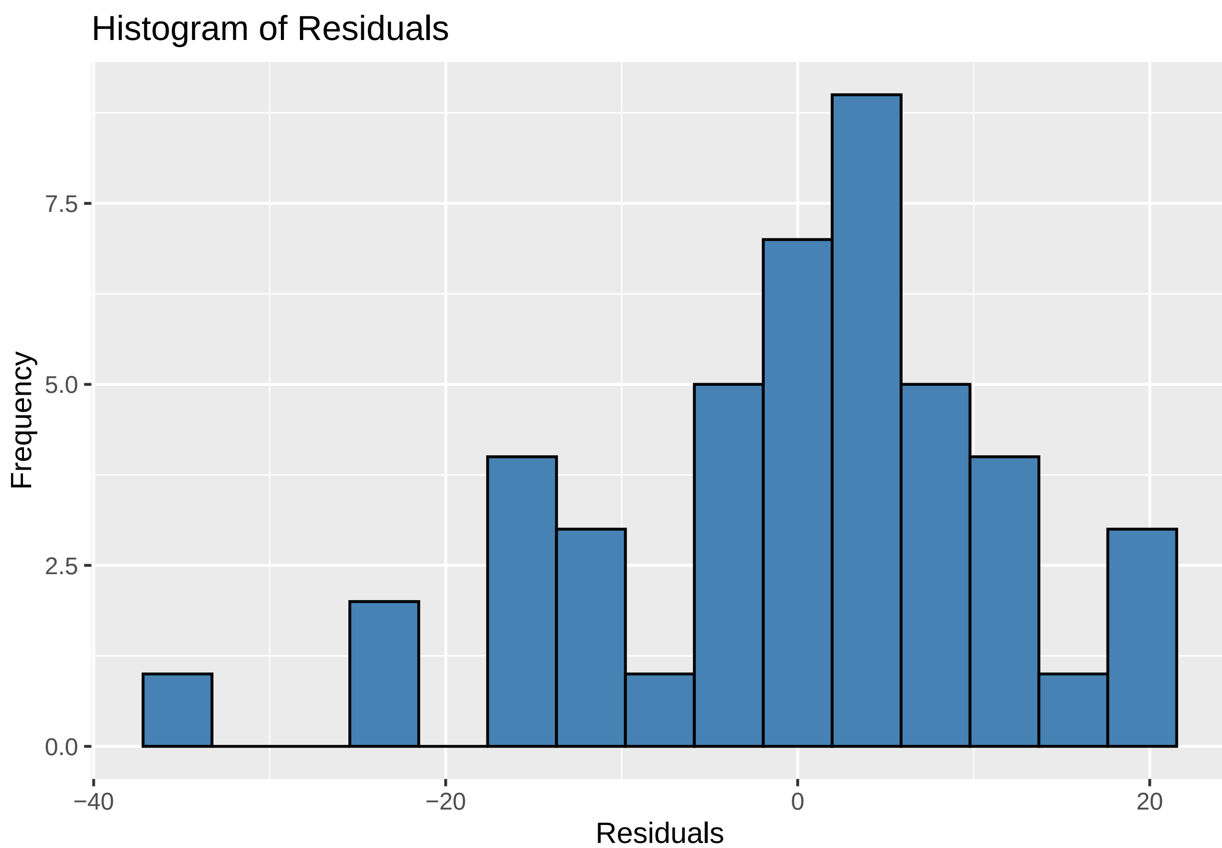
```r
ggqqplot(residuals(anova_1))
```



Histogram of residuals

The histogram of residuals appears to be roughly the shape of a normal curve.

```
ggplot(data = data, aes(x = anova_1$residuals)) +
  geom_histogram(bins = 15, fill = "steelblue", color = "black") +
  labs(title = "Histogram of Residuals",
       x = "Residuals",
       y = "Frequency")
```



Anderson-Darling Test

The Anderson-Darling Test p-value is greater than the alpha of 0.05; therefore, the assumption of normality of residuals is not violated.

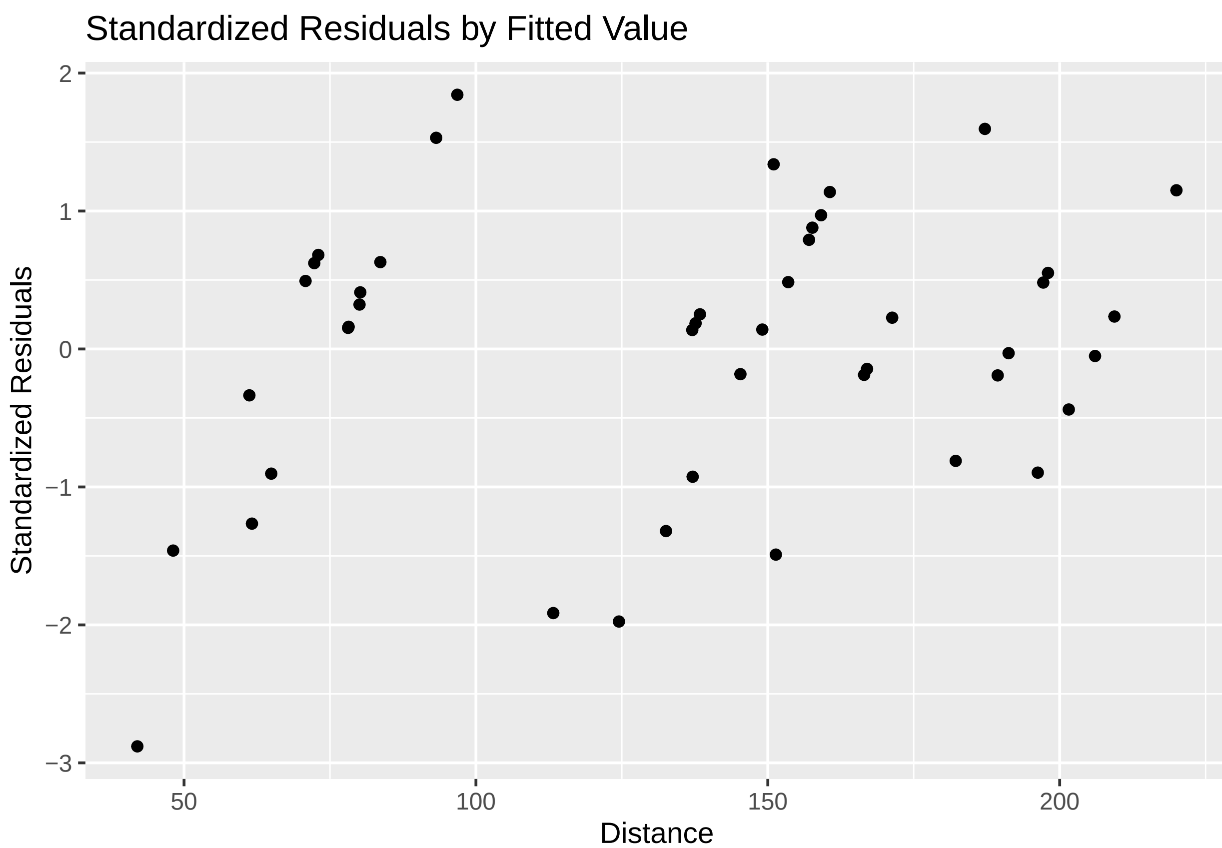
```
ad.test(residuals(anova_1))
```

```
##
##  Anderson-Darling normality test
##
## data:  residuals(anova_1)
## A = 0.63866, p-value = 0.08987
```

Homogeneity of variance

Standardized residuals by fitted value (distance)

```
ggplot(data = data, aes(x = distance, y = std_residuals)) +
  geom_point() +
  labs(title = "Standardized Residuals by Fitted Value",
       x = "Distance",
       y = "Standardized Residuals")
```



Standardized residuals by PSI

```
ggplot(data = data, aes(x = as.numeric(psi), y = std_residuals)) +
  geom_point() +
  xlim(min(as.numeric(data$psi)), max(as.numeric(data$psi))) +
  labs(title = "Standardized Residuals by PSI",
       x = "PSI",
       y = "Standardized Residuals")
```

Standardized residuals by angle

The residuals vs. fitted values, residuals vs. PSI, and residuals vs. angle plots are randomly distributed with no discernible skewing or patterns; therefore the assumption of homogeneity of variance of residuals is not violated

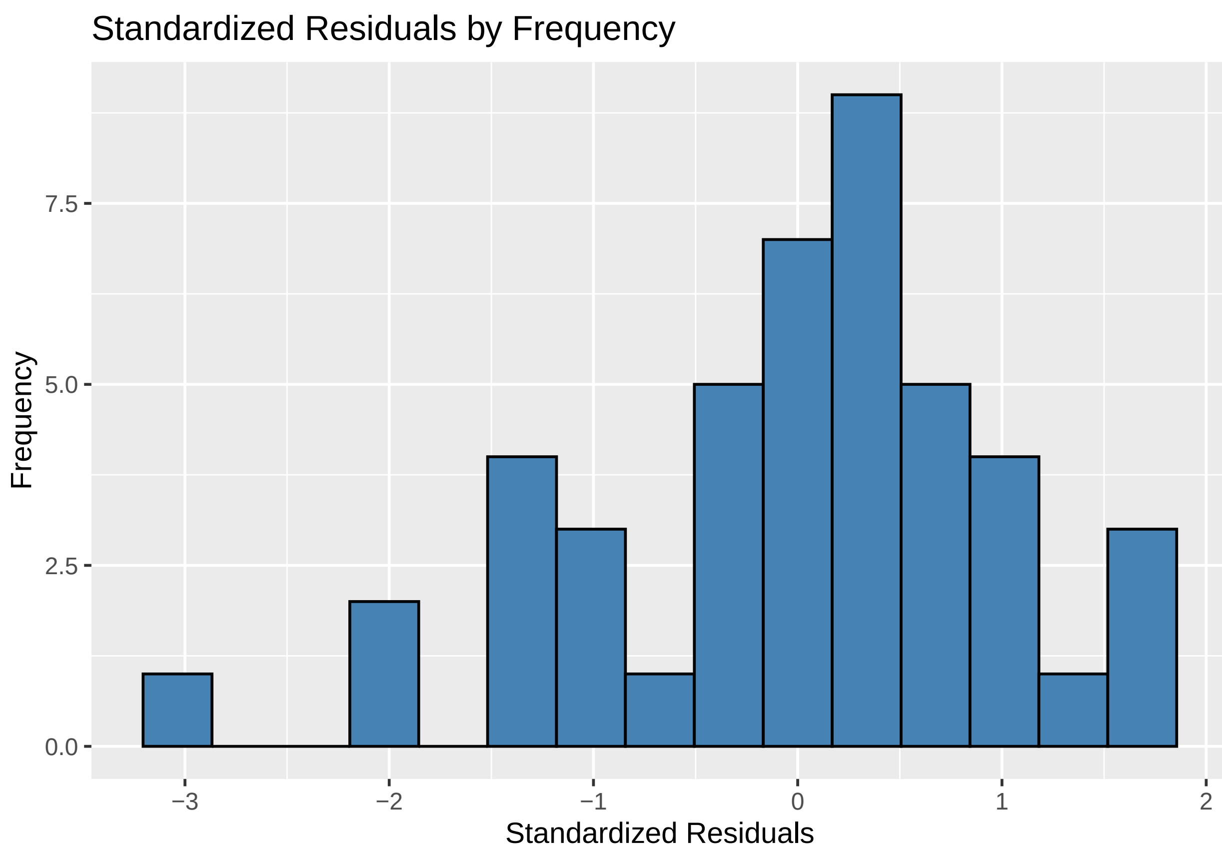
```
ggplot(data = data, aes(x = as.numeric(angle), y = std_residuals)) +
  geom_point() +
  xlim(min(as.numeric(data$angle)), max(as.numeric(data$angle))) +
  labs(title = "Standardized Residuals by Angle",
       x = "Angle",
       y = "Standardized Residuals")
```

No outliers

Standardized residuals by leverage

While there is one observation with a standardized residual of around -3, this is the most extreme value and none of the observations have standardized residuals with values beyond positive or negative 3. Therefore, the assumption of no outliers is not violated.

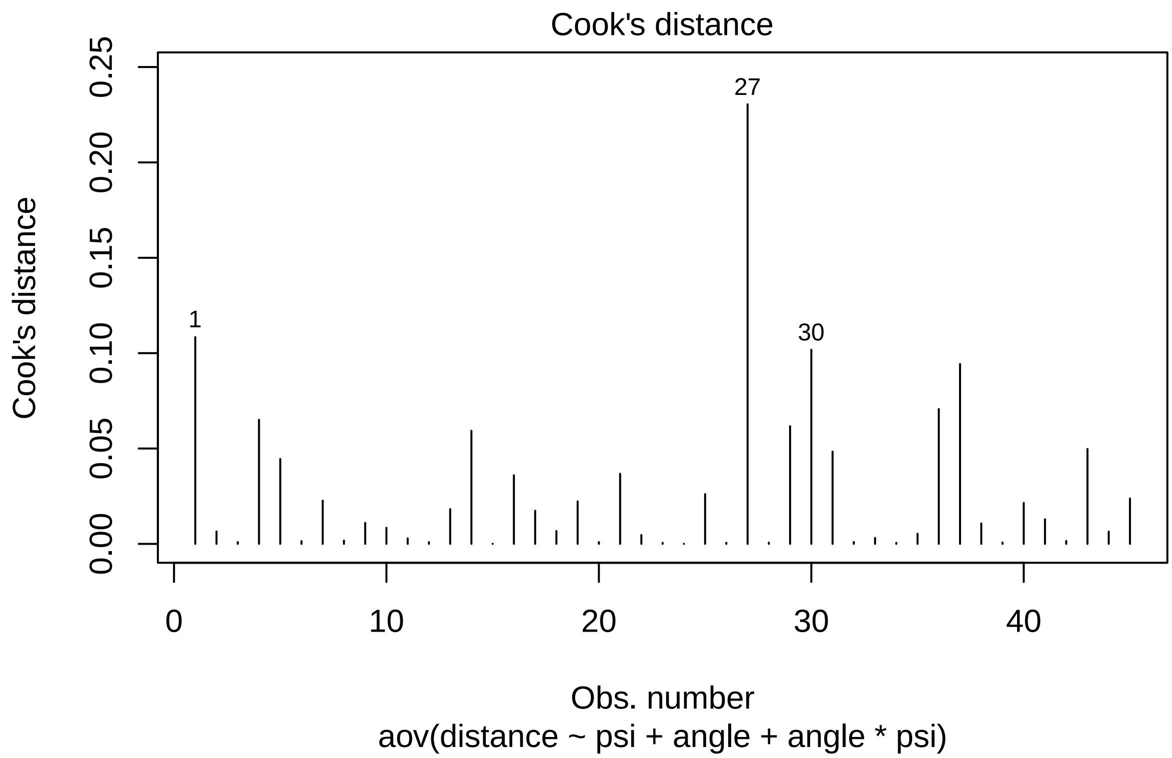
```
ggplot(data = data, aes(x = std_residuals)) +
  geom_histogram(bins = 15, fill = "steelblue", color = "black") +
  labs(title = "Standardized Residuals by Frequency",
       x = "Standardized Residuals",
       y = "Frequency")
```



Cook's distance

While observations, 1, 27, and 30 are influential based on Cook's Distance, no observation has a Cook's Distance greater than 0.5; therefore, no observation exerts undue influence on the data and analysis.

```
plot(anova_1, 4)
```



Independence of residuals

The residuals vs. order plot is randomly distributed with no discernible skewing or patterns; therefore, the assumption of independence of residuals is not violated.

```
ggplot(data = data, aes(x = seq_along(std_residuals), y = std_residuals)) +
  geom_point() +
  labs(title = "Standardized Residuals by Order",
       x = "Order of Observations",
       y = "Standardized Residuals")
```

Post-hoc Tests

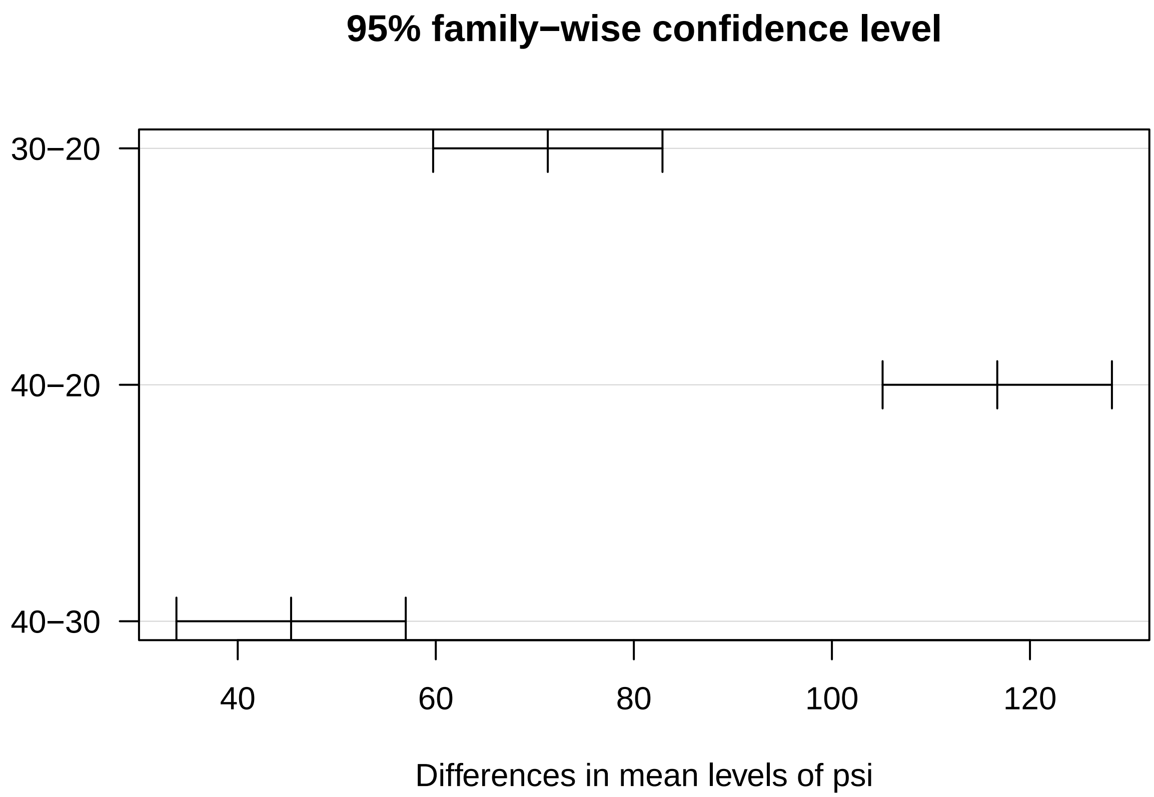
PSI

```
psi_tukey = TukeyHSD(x = anova_1, which = "psi")
psi_tukey
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = distance ~ psi + angle + angle * psi, data = data)
##
## $psi
##             diff       lwr       upr p adj
## 30-20  71.31000  59.73028  82.88972     0
## 40-20 116.69733 105.11762 128.27705     0
## 40-30  45.38733  33.80762  56.96705     0
```

All levels of PSI differ from each other in a statistically significant way.

```
plot(psi_tukey, las = 1)
```



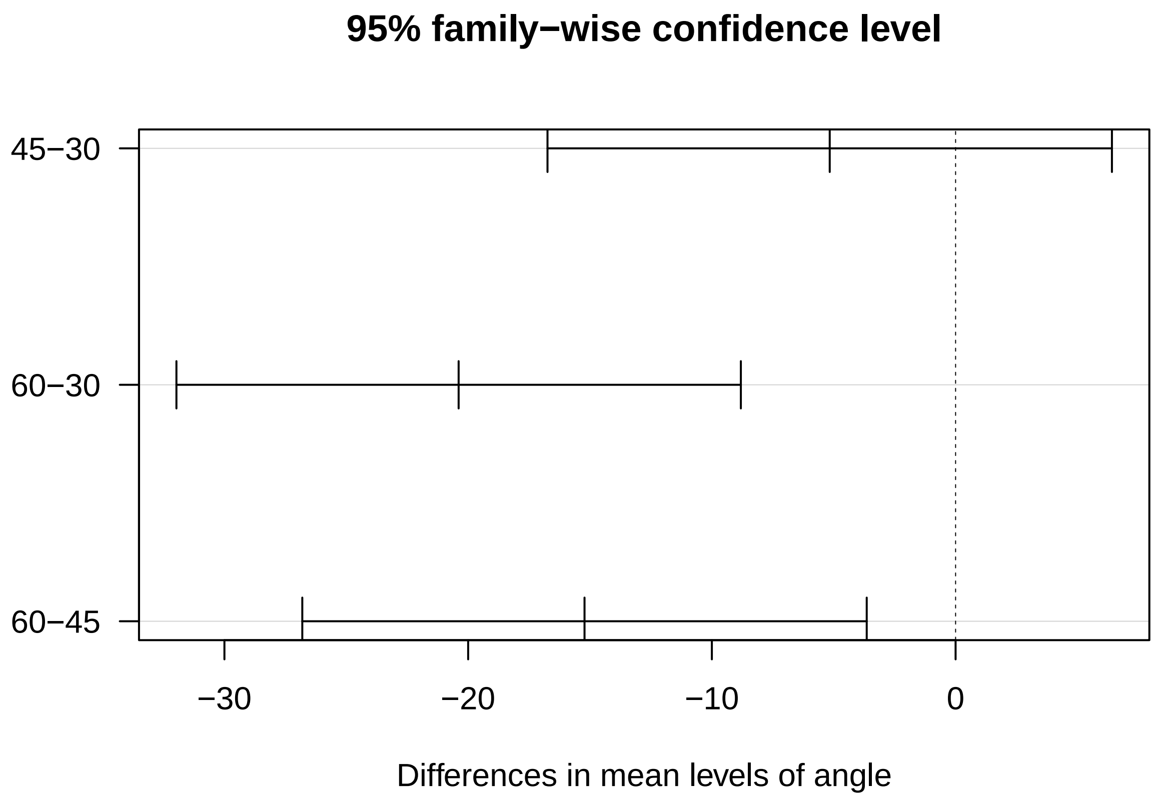
Angle

```r
angle_tukey = TukeyHSD(x = anova_1, which = "angle")
angle_tukey
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = distance ~ psi + angle + angle * psi, data = data)
##
## $angle
##             diff       lwr       upr     p adj
## 45-30  -5.164667 -16.74438  6.415050 0.5262315
## 60-30 -20.390000 -31.96972 -8.810283 0.0003546
## 60-45 -15.225333 -26.80505 -3.645616 0.0075968
```

An angle of 60 degrees differs significantly from both 45 degrees and 30 degrees, while the 45 and 30 degree angles do not differ from each other in a statistically significant way.

```r
angle_tukey = TukeyHSD(x = anova_1, which = "angle")
plot(angle_tukey, las = 1)
```



Angle group means

```
angle_means = data %>%
  group_by(angle) %>%
  dplyr::summarize(Mean = mean(distance, na.rm = TRUE))
angle_means
```

```
## # A tibble: 3 x 2
##   angle  Mean
##   <fct> <dbl>
## 1 30     143.
## 2 45     138.
## 3 60     123.
```

- An angle of 30 degrees is associated with the highest mean distance, differing from an angle of 45 degrees by just 0.4 sigma.
- An angle of 60 degrees is associated with the lowest mean distance, differing from an angle of 45 degrees by 1.17 sigma.

```
(143.466 - 138.3013) / est_sd
```

```
## [1] 0.3980801
```

```r
(138.3013 - 123.0760) / est_sd
```

```
## [1] 1.173522
```

PSI group means

```r
psi_means = data %>%
  group_by(psi) %>%
  dplyr::summarize(Mean = mean(distance, na.rm = TRUE))
psi_means
```

```
## # A tibble: 3 x 2
##   psi    Mean
##   <fct> <dbl>
## 1 20     72.3
## 2 30    144.
## 3 40    189.
```

- 40 PSI is associated with the highest mean distance, differing from 30 PSI by 4.3 sigma.
- 20 PSI is associated with the lowest mean distance, differing from 30 PSI by 5.5 sigma.

```r
(199.976 - 143.58867) / est_sd
```

```
## [1] 4.346172
```

```r
(143.58867 - 72.27867) / est_sd
```

```
## [1] 5.496369
```

Conclusions

A balanced, replicated two factor ANOVA was performed to determine what levels of angle and PSI were associated with the greatest mean distance of a tennis ball fired from a pneumatic launcher. The following conclusions can be drawn from the experiment:

- There is a statistically significant difference in mean distance between at least one angle and the others.
- There is a statistically significant difference in mean distance between at least one level of PSI and the others.
- While both angle and PSI were significant factors in the ANOVA, their interaction term was not.
- The 30° angle is associated with the greatest mean distance, while the 60° angle is associated with the lowest mean distance.
- The 30° and 45° angles do not differ from each other in a statistically significant way, while both differ from 60° in a statistically significant way.
- 40 PSI is associated with the greatest mean distance, while 20 PSI is associated with the lowest mean distance.
- All PSI levels differ from each other in a statistically significant way.