

# Analyzing golf ball carry distance.

Ved Piyush

10/12/2020

We will use the concepts learned from Sections 2, 5, and 7 to analyze golf ball carry distances. Our ultimate goal is to determine if a less expensive golf ball will perform similarly to a more expensive golf ball. Carry distance is the distance that the golf ball travels in the air after being hit by the golfer. Golfers would prefer a golf ball that covers more distance on average. Comparing the average carry distance across golf ball types can help a golfer make an informed decision as to which ball they should use.

The data in this study was collected by Mark Crossfield, who is a golf instructor. He used a launch monitor to determine carry distance. Let's first look at a plot comparing the carry distances for the five different golf balls used by Mark.

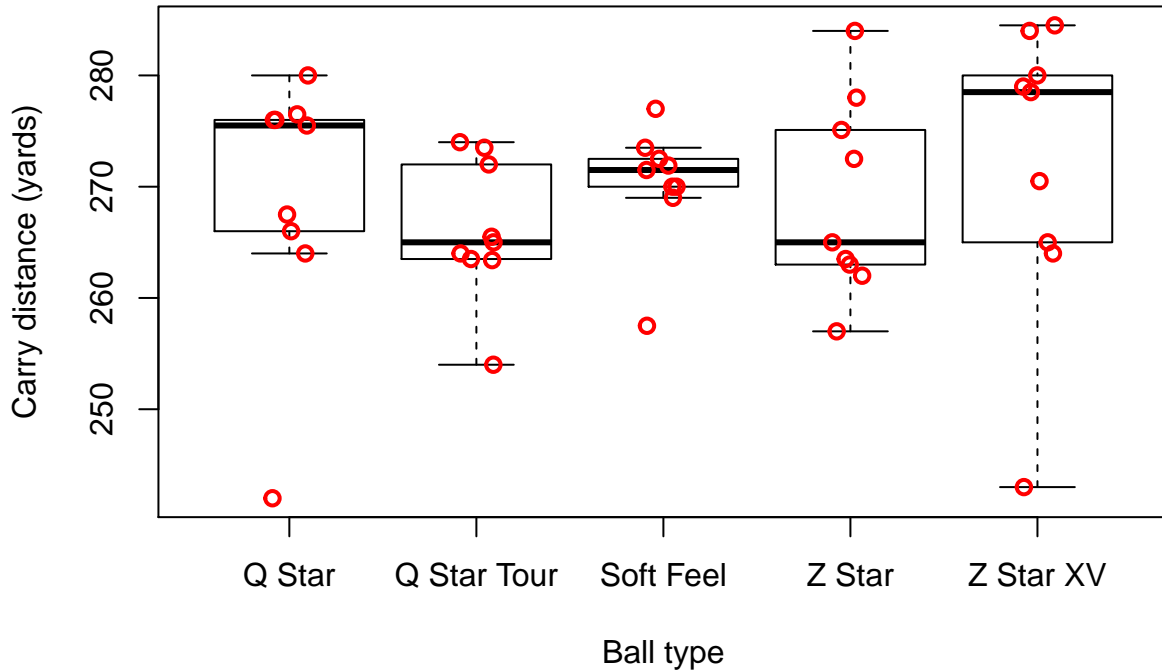
**Question** - If we want to compare the mean carry distance for two golf balls, then do we have independent samples or dependent samples?

```
> # read the csv file comprising the carry distance data make
> # sure you have this file in your working directory first
> carry_distance <- read.csv("Carry Distance.csv")
> # let's look at the data
> head(carry_distance)
```

	Ball	Carry.Distance
1	Q Star	242.0
2	Q Star	264.0
3	Q Star	266.0
4	Q Star	267.5
5	Q Star	275.5
6	Q Star	276.0

```
> # let's make side by side boxplots for the carry distances
> par(mfrow = c(1, 1))
> # col = NA removes gray background in box
> boxplot(formula = Carry.Distance ~ Ball, data = carry_distance,
  main = "Box and dot plot", ylab = "Carry distance (yards)",
  xlab = "Ball type", pars = list(outpch = NA), col = NA)
> stripchart(x = carry_distance$Carry.Distance ~ carry_distance$Ball,
  lwd = 2, col = "red", method = "jitter", vertical = TRUE,
  pch = 1, add = TRUE)
```

## Box and dot plot



**Question** - Do you think there are differences among the means or variances? Remember a plot like this is meant to give an initial impression of the data relative to the research hypotheses. Focus on shifts of points in addition to the variability seen in the points and the location of the medians.

We will compare the average carry distances of the Soft Feel golf balls to the Z Star golf balls. The Soft Feel golf balls cost around 20 dollars while the Z Star golf balls cost around 40 dollars. This comparison is interesting for a golfer because if there is no significant difference between the mean carry distance of these two balls, then the golfer can purchase the Soft Feel balls which are 20 dollars cheaper.

Let's first construct the confidence interval for  $\mu_1 - \mu_2$  where  $\mu_1$  represent the population means carry distance for the Soft Feel ball and  $\mu_2$  represents the population means carry distance for the Z Star ball. We will use  $\alpha = 0.05$ .

We know that the  $(1 - \alpha)100\%$  CI for  $\mu_1 - \mu_2$  with  $\sigma_1^2$  and  $\sigma_2^2$  unknown and possibly unequal is given by

$$\bar{y}_1 - \bar{y}_2 - t_{\frac{\alpha}{2}, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \bar{y}_1 - \bar{y}_2 + t_{\frac{\alpha}{2}, v} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where  $t_{\frac{\alpha}{2}, v}$  is the  $1 - \frac{\alpha}{2}$  quantile from a t distribution with

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

degrees of freedom.

**Question** - Why does our assumption of unequal  $\sigma_1^2$  and  $\sigma_2^2$  make sense? What visual evidence can we present to justify this assumption?

```

> alpha <- 0.05
> # Let 1 = Soft feel and 2 = Z star
> save.mean <- aggregate(formula = Carry.Distance ~ Ball, data = carry_distance,
  FUN = mean)
> ybar1 <- save.mean$Carry.Distance[3]
> ybar2 <- save.mean$Carry.Distance[4]
> save.var <- aggregate(formula = Carry.Distance ~ Ball, data = carry_distance,
  FUN = var)
> s.sq1 <- save.var$Carry.Distance[3]
> s.sq2 <- save.var$Carry.Distance[4]
> save.n <- aggregate(formula = Carry.Distance ~ Ball, data = carry_distance,
  FUN = length)
> n1 <- save.n$Carry.Distance[3]
> n2 <- save.n$Carry.Distance[4]
> # Variance unequal
> nu <- (s.sq1/n1 + s.sq2/n2)^2/((s.sq1/n1)^2/(n1 - 1) + (s.sq2/n2)^2/(n2 -
  1))
> data.frame(ybar1, ybar2, s.sq1, s.sq2, n1, n2, nu, t.quant = qt(p = 1 -
  alpha/2, df = nu))

```

```

      ybar1 ybar2   s.sq1   s.sq2 n1 n2      nu t.quant
1 270.3222 268.9 28.70944 78.8275  9  9 13.14485 2.157951

```

```

> lower <- ybar1 - ybar2 - qt(p = 1 - alpha/2, df = nu) * sqrt(s.sq1/n1 +
  s.sq2/n2)
> upper <- ybar1 - ybar2 + qt(p = 1 - alpha/2, df = nu) * sqrt(s.sq1/n1 +
  s.sq2/n2)
> data.frame(lower, upper)

```

```

      lower  upper
1 -6.037097 8.881542

```

```

> # easier way to verify the calculations
> t.test(formula = Carry.Distance ~ Ball, data = carry_distance[carry_distance$Ball %in%
  c("Soft Feel", "Z Star"), ], var.equal = FALSE, conf.level = 0.95)

```

#### Welch Two Sample t-test

```

data: Carry.Distance by Ball
t = 0.41144, df = 13.145, p-value = 0.6874
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.037097  8.881542
sample estimates:
mean in group Soft Feel   mean in group Z Star
      270.3222             268.9000

```

From our calculations, we find that the 95% CI for  $\mu_1 - \mu_2$  is (-6.03, 8.88). Since 0 is not inside this interval, we do not have sufficient evidence to reject the null hypothesis. Therefore we do not have sufficient evidence to indicate a difference between the mean carry distance for the two golf balls.

**Question** - Interpret the confidence interval calculated above.

**Question** - How can your conclusion above help a golfer save money?

**Question** - Using the confidence intervals calculated above, conduct the following hypothesis test and interpret the results.

$$H_o : \mu_1 - \mu_2 = 0$$

$$H_a : \mu_1 - \mu_2 \neq 0$$

**Exploration** - Please conduct a similar analysis for the golf balls Q Star and Z Star XV. Interpret your results in context of the golf problem.

**Question** - How many possible pairwise comparisons can be made for the golf ball data? Suppose you are conducting a hypothesis test for each possible pairwise comparison and using a level of  $\alpha$  for the type I error rate in each test. How does the probability of making at least one type I error for all multiple comparisons compare with  $\alpha$ ?

**Question** - Apart from the inference for the difference in means, why might the inferences for the ratio of variances for the carry distance also make sense?

Now, we will do the inference for the ratio of variances.

We know that a  $(1 - \alpha)100\%$  CI for the ratio of variances is given by

$$\frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2}, v_2, v_1} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} F_{\frac{\alpha}{2}, v_2, v_1}$$

where  $v_1 = n_1 - 1$  and  $v_2 = n_2 - 1$ .

```
> alpha <- 0.05
> data.frame(n1, n2)
```

```
  n1 n2
1  9  9
```

```
> qf(p = alpha/2, df1 = n2 - 1, df2 = n1 - 1)
```

```
[1] 0.2255676
```

```
> qf(p = 1 - alpha/2, df1 = n2 - 1, df2 = n1 - 1)
```

```
[1] 4.43326
```

```
> lower <- s.sq1/s.sq2 * qf(p = alpha/2, df1 = n2 - 1, df2 = n1 -
  1)
> upper <- s.sq1/s.sq2 * qf(p = 1 - alpha/2, df1 = n2 - 1, df2 = n1 -
  1)
> data.frame(lower, upper)
```

```
  lower  upper
1 0.08215308 1.61462
```

```

> # easier way to verify the calculations
> var.test(x = carry_distance[carry_distance$Ball == "Soft Feel",
  ]$Carry.Distance, y = carry_distance[carry_distance$Ball ==
  "Z Star", ]$Carry.Distance, conf.level = 0.95)

```

F test to compare two variances

```

data: carry_distance[carry_distance$Ball == "Soft Feel", ]$Carry.Distance and carry_distance[carry_dis
F = 0.36421, num df = 8, denom df = 8, p-value = 0.1746
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.08215308 1.61461963
sample estimates:
ratio of variances
 0.3642059

```

From our calculations, we find that the 95% CI for  $\frac{\sigma_1^2}{\sigma_2^2}$  is (0.08, 1.61). Since 1 is inside this interval, we do not have sufficient evidence to reject the null hypothesis. Therefore we do not have sufficient evidence to indicate that the ratio of variances of carry distance for the two golf balls is different from 1.

**Question** - Interpret the confidence interval calculated above.

**Question** - Using the confidence intervals calculated above, conduct the following hypothesis test and interpret the results.

$$H_o : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

**Exploration** - Please conduct a similar analysis for the golf balls Q Star and Z Star XV. Interpret your results in context of the golf problem.