ExampleWelch

2024-01-15

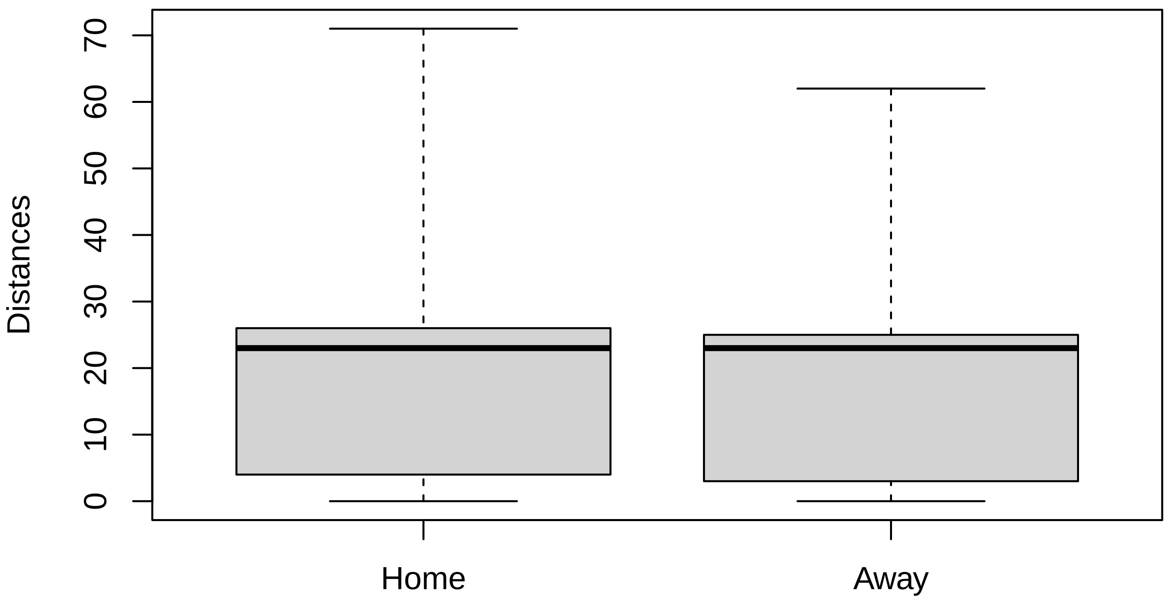
Steph Curry Shot Distances

Using data provided, we can determine if the average shot distance Curry makes differs between home and away games. The data being used is the shot distance, whether or not he made the shot, and if the shot was done at an away or home game during a season.

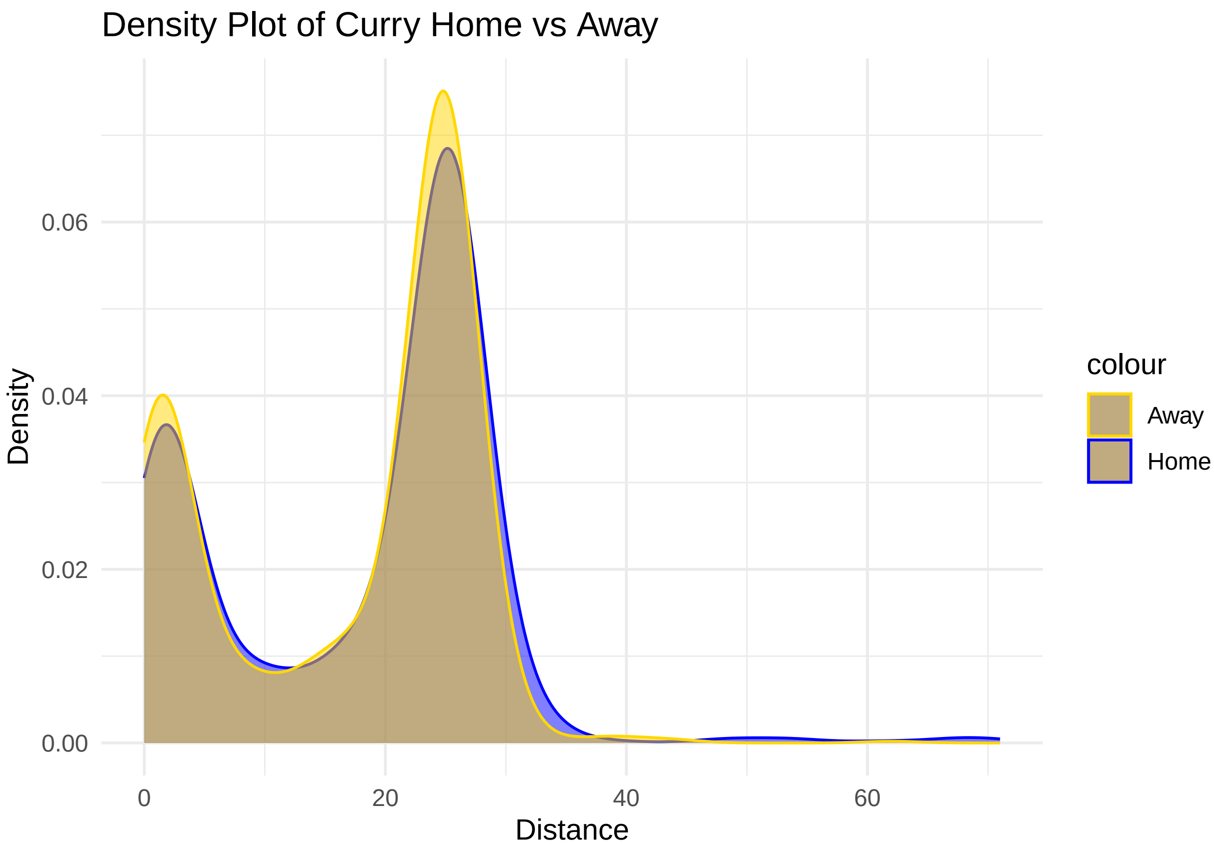
Plots

```r
library(ggplot2)
currydist <- read.table("~/Desktop/consulting examples/currydist.txt", header=TRUE)
home=currydist$distance[currydist$venue=="Home"]
away=currydist$distance[currydist$venue=="Away"]

boxplot(home, away, range=0,names=c("Home","Away"),ylab="Distances")
```




```r
ggplot() +
  geom_density(aes(x=home, color="Home"), fill="blue", alpha=0.5) +
  geom_density(aes(x=away, color="Away"), fill="gold", alpha=0.5) +
  labs(title = "Density Plot of Curry Home vs Away", x="Distance", y="Density") +
  scale_color_manual(values=c("Home"="blue", "Away"="gold"))+
  theme_minimal()
```



Plot analysis

Based on the above Box plots and Density plots, we can observe that the most likely answer to our question is no, there is not a noticeable difference in shot distance for Curry between home and away games. First, the box plots show us that the median value of both data sets are relatively close, and since the size of each box is similar, this tells us that the IQR is similar between the home and away shots as well. Then, in the density plots, we can see that both distributions are bimodal and somewhat right-skewed but at both spikes the away game line is slightly higher than the home game line. We can also note that both distributions are extremely similar. In order to help with the assumption that the difference between the two is insignificant, we will test the hypothesis of the average distance being approximately the same by using the method of hypothesis testing. That is, let $\Delta = \mu_1 - \mu_2$ where μ_1 is the mean of distances for home games and μ_2 is the mean of distances for away games. The null and alternative hypothesis are as follows:

$$H_0 : \Delta = 0$$
$$H_1 : \Delta \neq 0$$

Welch's Two Sample t-test

The Welch's two-sample t-test requires us to have two IID samples from two independent normal populations. In our case, looking at the above density plot as well as the Box plot, we can see that the distributions do not look normal. However since the sample size is large enough we can still use the t-test since the Central Limit Theorem tells us that even though the population is not normally distributed, the distribution of the sample means will tend to be normal for a large enough sample size.

Then, using Welch's test on the above hypothesis,

```
t.test(home,away)
```

```
##
##  Welch Two Sample t-test
##
## data:  home and away
## t = 1.5953, df = 1567.8, p-value = 0.1108
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2046775  1.9882702
## sample estimates:
## mean of x mean of y
##  17.83226  16.94046
```

Analysis of Welch's

We can say with 95% confidence that the difference in shot distances is in the interval $(-0.205, 1.99)$

Since $p = 0.1120364 > \alpha = 0.05$ we fail to reject the null hypothesis. In other words, there is not enough evidence to support that average shot distance Curry makes at home is different than the average shot distance he makes away.