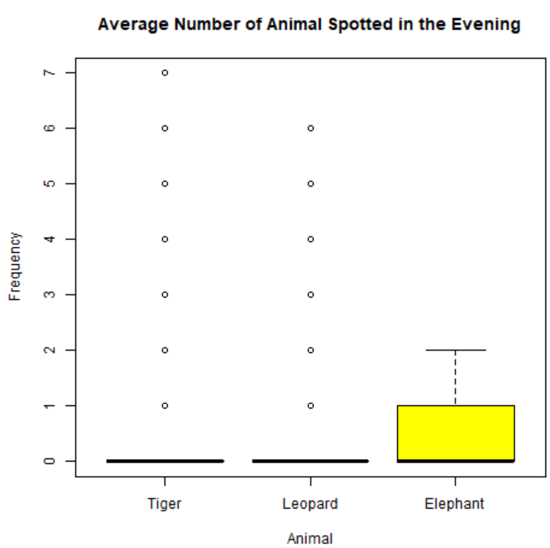
Spotting Tigers in the Wild

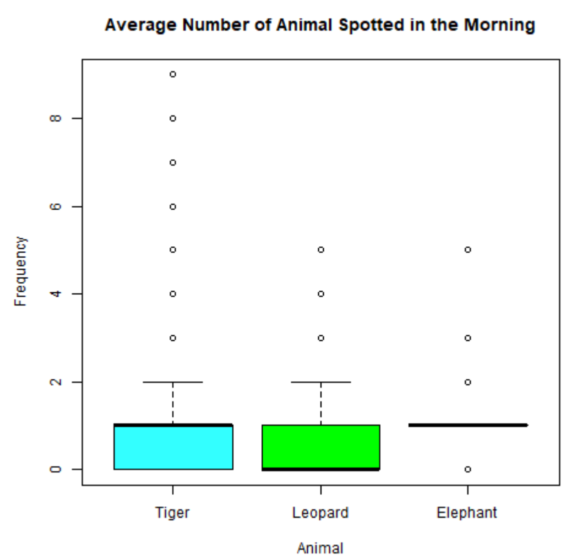
By: Rohit Manjunath

Hello everyone,

My name is Rohit, and I am a freshman majoring in computer science. Wildtrails is a website that can be used to book wildlife trips and has a lot of wildlife data, including a sighting index for animals. As a wildlife enthusiast, my main goal is to see as many big cats as possible and as often as possible without too much effort. Not only do these big cats such as tigers, leopards, lions, etc., make the safari a pleasurable experience, but it also gives us the opportunity to take amazing pictures of them in their habitat. Using the data provided by Wildtrails, I figured out what time slot is the best to maximize our chances of seeing tigers, leopards, and elephants. Now, most wildlife sanctuaries let you either go for an early morning safari or an evening safari. If you're as lazy as me, you probably would hate the early morning safaris.

The wildlife spottings data has the date and animals(birds, mammals, and reptiles) spotted each day since 2015. I then decided to add extra columns to store the number of tigers, leopards, and elephants spotted each day. I did this because these are the top three animals that everyone wants to see on a safari in southern India(No lions present). Now looking at this, I immediately thought to find the average number of tigers, leopards, and elephants spotted in the morning and the evening. The below graph is a box plot of the animals spotted during the morning and evening safaris.





You can easily see that tigers and leopards are spotted a lot more during the morning, while the median of the number of elephants spotted each day is higher in the morning. We can infer that tigers, leopards, and elephants love making us wake up earlier to spot them. Now, these plots could be a result of random data, which means nothing. We will have to find the p-value to reject the null hypothesis, which states that the average number of tigers, leopards, and elephants spotted in the morning is the same as in the evening. For this dataset, a p-value lesser than 0.05 or 5% will reject the null hypothesis and tell us that our hypothesis isn't due to random data.

The Code for the Tiger Subset:

```r
tigerMorMean <- mean(morning$tigercount)
tigerEveMean <- mean(evening$tigercount)
tigerMorSD <- sd(morning$tigercount)
tigerEveSD <- sd(evening$tigercount)
tigerMorLength <- length(morning$tigercount)
tigerEveLength <- length(evening$tigercount)
bothSD <- sqrt((tigerMorSD^2) / tigerMorLength + (tigerEveSD^2) / tigerEveLength)
Zscore <- (tigerMorMean - tigerEveMean) / bothSD
p <- (1 - pnorm(Zscore)) / 2
```

Output:

```
[1] 1.070898
[1] 0.3610134
[1] 1.422682
[1] 0.8614597
[1] 1481
[1] 1421
[1] 0.04346156
[1] 16.33362
[1] 0
```

You can see that the p-value is 0 for the tiger hypothesis. Hence we can successfully reject the null hypothesis.

The Code for the Leopard Subset:

```
leopardMorMean <- mean(morning$leopardcount)
leopardEveMean <- mean(evening$leopardcount)

leopardMorSD <- sd(morning$leopardcount)
leopardEveSD <- sd(evening$leopardcount)

leopardMorLength <- length(morning$leopardcount)
leopardEveLength <- length(evening$leopardcount)

bothSD <- sqrt((leopardMorSD^2) / leopardMorLength + (leopardEveSD^2) / leopardEveLength)

Zscore <- (leopardMorMean - leopardEveMean) / bothSD

p <- (1 - pnorm(Zscore)) / 2
```

Output:

```
[1] 0.6130993
[1] 0.2892329
[1] 0.8918951
[1] 0.7160756
[1] 1481
[1] 1421
[1] 0.02996613
[1] 10.80775
[1] 0
```

You can see that the p-value is 0 for the leopard hypothesis. Hence we can successfully reject the null hypothesis.

The Code for the Elephant Subset:

```
elephantMorMean <- mean(morning$elephantcount)
elephantEveMean <- mean(evening$elephantcount)

elephantMorSD <- sd(morning$elephantcount)
elephantEveSD <- sd(evening$elephantcount)

elephantMorLength <- length(morning$elephantcount)
elephantEveLength <- length(evening$elephantcount)

bothSD <- sqrt((elephantMorSD^2) / elephantMorLength + (elephantEveSD^2) / elephantEveLength)

Zscore <- (elephantMorMean - elephantEveMean) / bothSD

p <- (1 - pnorm(Zscore)) / 2
```

Output:

```
[1] 0.8642809
[1] 0.3413089
[1] 0.3764346
[1] 0.483142
[1] 1481
[1] 1421
[1] 0.01612295
[1] 32.43649
[1] 0
```

You can see that the p-value is 0 for the elephant hypothesis. Hence we can successfully reject the null hypothesis.

Since we successfully rejected all null hypotheses with a p-value of 0, which is lesser than 5%, we can successfully say the average number of tigers/leopards/elephants seen is higher during morning safaris than evening safaris. However, this data is only from the Kabini wildlife sanctuary and not any other wildlife sanctuary. Hence, this may not be the same for other wildlife sanctuaries.