

Homework for Lecture 11

The following homework assignment should be completed as an RMarkdown Document. Only show your working code and place comments/explanations in the “white space” to help me evaluate your work. Please submit both the knitted RMarkdown document (.doc) and the .Rmd file. You may discuss the assignment with your classmates, but your write-up and work should be done independently of your classmates. The use of AI is considered a violation of academic policy for this assignment. When in doubt, ask me what is and is not allowed.

Every example will use the exact same code, I want you to interpret the results. We will want to ask ourselves, what is the mean and median and what impression does this give us about the spread/skew of the data and what are our initial impressions related to the skew/tails of the data? For each example write how your mean/median/mode are related and the skew of the distribution.

```
# This is an example of the code that you can use for all of the problems. The only
thing you will have to adjust is the variable that you are wanting to investigate.
```

```
# Keep in mind that all of the output will not align to the same conclusion. To see an
example of this you can look at the variable homeruns in the “mlb_teams” dataset. So
just because one piece of information in the output shows something does not mean
that all of the output will support that claim.
```

```
library(openintro)
x <- mlb_teams$wins

center_stats(x)
position_stats(x)
spread_stats(x)
shape_stats(x)

hist(x, freq=FALSE)
lines(density(x, na.rm=TRUE), col="red", lwd=2)

xfit <- seq(min(x, na.rm=TRUE), max(x, na.rm=TRUE), length=100)
yfit <- dnorm(xfit, mean=mean(x, na.rm=TRUE), sd=sd(x, na.rm=TRUE))
lines(xfit, yfit, col="blue", lwd=2)
```

1. **Note:** In order to use these functions in RMarkdown, you will need to call the “MSMU” library in your RMarkdown document. If you get an error that says: could not find function “center_stats” then you will need to check that you have the library called in your first chunk.

****After running the code and looking at the results you might say the following****

- The center stats output indicates that the distribution might be slightly left-skewed. This is because the mean is less than the median which is less than the mode. Additionally, the trimmed mean rises, indicating that outliers on the left-side might be influencing the calculation
- The position stats similarly indicates that the distribution might be left-skewed. Doing Quartile Analysis, we can see that we are more spread out on the left-hand side. This is because $Q2 - \min = 65$ and $\max - Q2 = 38$ (we would also want to look at $Q2 - Q1$ and $Q3 - Q2$)
- The spread stats show the lowest z-score value is -4.22 while the highest is 2.5. This indicates that we may be more spread out on the left-hand side of the distribution or we have an outlier on that side. This output also shows that the distribution has more data in the tails than a normal distribution, as we only expect 0.3% of the data to be in the tails but we have 1.15% of the data in the tails
- All of this information lines up with the output of the shape stats as we can see we have a negative skew (left-skewed) and a positive kurtosis (thicker tails)
- The Histogram overlayed with the density line and the normal distribution all support these claims. We can see this with the left-side being “pulled” a little to the left and the positive kurtosis since the data’s distribution line is above the normal distributions density line in the left-hand tail.
- Therefore, the distribution of this data is left-skewed with thicker tails than a normal distribution

1. Describe the “gained” variable in the “births14” dataset
2. Describe the “weeks” variable in the “births14” dataset
3. Describe the ‘course_grade’ variable in the ‘exam_grades’ dataset.
4. Describe the “homeruns” column in the “mlb_teams” dataset

After all of this, we should be able to see that if $\text{mean} = \text{median}$ it does not guarantee symmetry, if $\text{mean} < \text{median}$ it does not guarantee negative skew, and if $\text{mean} > \text{median}$ it does not guarantee positive skew. Likewise a skew of zero does not guarantee symmetry.