Exercise: Facebook Statistics

James M. Murray, Ph.D.
University of Wisconsin - La Crosse

PDF file location: http://www.murraylax.org/rtutorials/facebook-univariate.pdf

HTML file location: http://www.murraylax.org/rtutorials/facebook-univariate.html

---

Note on required packages: The following code requires the packages in the `tidyverse`. The `tidyverse` actually contains many packages that allow you to organize, summarize, and plot data. The package `psych` is used to perform statistics related to the median. If you have not already done so, download and install the libraries (needed only once per computer), and load the libraries (need to do every time you start R) with the following code:

```
install.packages("psych") # This only needs to be executed once for your machine
install.packages("tidyverse") # This only needs to be executed once for your machine
library("psych") # This needs to be executed every time you load R
library("tidyverse") # This needs to be executed every time you load R
```

Data Set

The following data set comes from the following study on Facebook marketing and performance metrics:

Moro, S., Rita, P. and Vala, B., (2016) "Predicting Social Media Performance Metrics and Evaluation of the Impact on Brand Building: A Data Mining Approach" *Journal of Business Research*, Vol. 68, pp. 3341-3351. Available at http://www.sciencedirect.com/science/article/pii/S0148296316000813

Download and load into memory the data set (see other formats at end of document):

```
load(url("http://murraylax.org/datasets/facebook.RData"))
```

The data set includes statistics from 500 Facebook posts in 2014 related to the marketing of a globally known cosmetic brand. Facebook marketing is an important part of many businesses marketing strategy. Facebook interaction can help businesses build their brand and market new products. Marketing executives such statistics to better understand the effectiveness of their Facebook marketing.

The data set includes the following variables:

1. **Type**: Scale / Class: Nominal / Factor. Type of post. Possible outcomes are "Link", "Photo", "Status", and "Video"
2. **Month**: Scale / Class: Ordinal / Ordered factor. Month of the year for the post.
3. **Weekday**: Scale / Class: Ordinal / Ordered factor. Day of the week for the post.
4. **Hour**: Scale / Class: Ratio / Integer. Hour of the day - between 0 (12:00AM) and 23 (11:00PM)
5. **Paid**: Scale / Class: Binary / Integer. Dummy variable equal to 1 if a paid post, 0 if a free or unsolicited post.
6. **Reach**: Scale / Class: Ratio / Integer. Number of unique individuals who saw the post appear on their news feeds.

7. **Impressions**: Scale / Class: Ratio / Integer. Number of times the post appeared on people's news feeds (some individuals may have had the post appear more than once)
8. **EngagedUsers**: Scale / Class: Ratio / Integer. Number of unique individuals that clicked anywhere in the post.
9. **Comments**: Scale / Class: Ratio / Integer. Number of comments on the post.
10. **Likes**: Scale / Class: Ratio / Integer. Number of likes for the post
11. **Shares**: Scale / Class: Ratio / Integer. Number of shares for the post
12. **Interactions**: Scale / Class: Ratio / Integer. The sum, Comments + Likes + Shares.
13. **Weekday.Int**: Scale / Class: Ordinal / Integer: Number associated with day of the week in **Weekday**
14. **Month.Int**: Scale / Class: Ordinal / Integer: Number associated with month in **Month**

Exercises

1. Descriptive Statistics

A. Compute the mean and standard deviation for the number of engaged users per post.

B. Compute the median and interpolated median for the number of comments per post.

C. Compute the frequencies of Facebook posts by day of the week. What two days have the most number of Facebook posts?

D. Compute the median number of comments by day of week. What day results in the most number of comments? What three days result in the least number of comments?

E. Compute the mean number of engaged users by post *type*. What type of post leads to the most engaged users? What type of post leads to least number of engaged users? Is this what you expected?

2. Inferential Statistics

A. Compute and interpret a 95% confidence interval for the mean number of shares of a Facebook post

B. Compute the mean and interpret a 95% confidence interval for the mean for the number of shares generated by a Facebook post created on a Monday.

C. Compute the median for the number of shares for Facebook posts. Compute and interpret 95% confidence interval for the median.

D. Compute the interpolated median for the number of comments for Facebook posts. Compute and interpret 95% confidence interval for the median.

E. Compute the mean number of engaged users for paid posts. Test the hypothesis that the average number of engaged users per paid post is greater than 1,000.