

# Lab 3 Activity

In the Lab 3 slides, we loaded in and manipulated the `Airlines.csv` data. Run all of these commands first.

```
library(tidyverse)

flights <- read_csv("Airlines.csv")

flights <- flights %>%
  mutate(Delay = factor(Delay,
                        levels = c("0", "1"), # values observed in the data
                        labels = c("On Time", "Delay"))) # meaningful labels
```

## 1

Use the  $\chi^2$  test of independence to test whether there is an association between whether a flights gets delayed depending on whether Delta Airlines (DL) or Southwest Airlines (WN). Do this only for flights originating from Los Angeles (LAX) and flying to Salt Lake City (SLC).

- Step 1. Use `filter` to retain only those flights that we want.
- Step 2. Create a contingency table.
- Step 3. Run a  $\chi^2$  test of independence with the continuity correction

## 2

Compute the  $W$  coefficient.

**3**

Write  $\chi^2(1, N = 242) = 9.71$ ,  $p = .002$ ,  $W = .2$  using Quarto or R Markdown (the outputted document should render the mathematical notation properly).

**4**

Run Fisher's exact test on these data. Which test do you think is more appropriate?

**5**

Based on your results, should you choose Delta or Southwest if you were to fly from Los Angeles to Salt Lake city? Feel free to calculate other types of effects to inform your answer