**STAT581 Project Proposal (10 pts)**

**Name(s): Molly Hischke & Kyle Hancock**

1. **Response variable.** Identify as binary, nominal (identify levels), ordinal (identify levels), discrete (give range) or continuous (give range).

We will look at the response variable as binary and continuous (count data). The binary response (claim\_filed) is whether or not a workers’ compensation claim was filed (0 = no claim filed, 1 = claim filed). The continuous response is the total number of claims filed (total\_claims, ranging from 0 to 7).

1. **Predictor variables.** Identify each variable as continuous or categorical.

For analysis #1 (Binary Logistic Regression), the predictor variable will be continuous (adj\_payroll which is proxy for the size of the brewery).

For analysis #2 (Poisson Regression), the predictor variables will be continuous (adj\_payroll which is proxy for the size of the brewery and years\_with\_policy which is the number of years the brewery held a workers’ compensation policy).

1. **Sample size**

The dataset includes the workers’ compensation data for all craft breweries in Colorado who held policy with a specific Colorado based insurance company between 2013 and 2018. There is a total of 975 observations encompassing a total of 130 breweries.

1. **Design.** Is there any blocking or repeated measures?

No blocking or repeated measures.

1. **Brief description of proposed analysis.** This needs to include a generalized linear model!

Prior to any model generation, we will run descriptive statistics and exploratory graphs.

For analysis #1, we anticipate running a Binary Logistic Regression with the predictor being adj\_payroll (proxy to the size of the brewery) and response being claim\_filed.

Model1 < - glm(claim\_filed ~ adj\_payroll, family = binomial, data = wc2)

For analysis #2, we anticipate running a Poisson Regression with the predictor being adj\_payroll (proxy to the size of the brewery) and years\_with\_policy (or the number of years the brewery held a workers’ compensation policy), and the response being total\_claims (the number of claims filed).

Model2 < - glm(total\_claims ~ adj\_payroll + years\_with\_policy, family = poisson(link = “log”), data = wc2)

**Other comments?**