

Stat 140 - Quiz 1 Sample

What's Your Name? _____

Which section are you in? _____

This is a sample quiz. For the real quiz, I will use a different data set, but will ask the questions below about the new data set with minimal modification.

Below are the first few rows of a data frame named NHANES. NHANES stands for "National Health and Nutrition Examination Surveys", and the data frame contains information about the health of randomly sampled Americans.

```
##      ID Gender Age Weight Height  BMI BPSysAve BPDiaAve Diabetes
## 1 51624  male  34  87.4  164.7 32.22    113      85      No
## 2 51625  male   4  17.0  105.4 15.30     NA     NA      No
## 3 51630 female  49  86.7  168.4 30.57    112     75      No
## 4 51638  male   9  29.8  133.1 16.82     86     47      No
## 5 51646  male   8  35.2  130.6 20.64    107     37      No
## 6 51647 female  45  75.7  166.7 27.24    118     64      No
```

1. What is each observational unit in this data set?

2. For each of the following variables, is that variable categorical or quantitative? If it is categorical, is it ordinal or nominal?

- Gender
- Height
- Diabetes

3. The following command counts how many observational units are in each combination of levels of the gender and diabetes variables.

```
NHANES %>%
  count(Diabetes, Gender) %>%
  spread(Gender, n)
```

```
## # A tibble: 2 x 3
##   Diabetes female  male
##   <fct>      <int> <int>
## 1 No           3088  3013
## 2 Yes           269   283
```

a. Calculate the joint distribution of Diabetes and Gender

b. Calculate the marginal distribution of Diabetes

c. Calculate the conditional distribution of Diabetes given that the subject's Gender is male

d. Calculate the conditional distribution of Diabetes given that the subject's Gender is female

e. Is a person's Diabetes status independent of their Gender?