PS0700 Descriptive Statistics

1. File Description

File is "job_training.teaching.dta". Random sample of 474 public employees in a midwestern city, some treated in a training program to improve skills (a quasi-experimental data set)

2. Load Data

```
# install.packages("haven")
# Install Haven library to read .dta file if you haven't
library(haven)
data <- read_dta("job_training.teaching.dta")
# Save .dta file to "data"
```

3. Frequency Distribution of a Norminal Variable

```
sex <- factor(data$sex,
              levels = c(0,1),
              labels = c("Female", "Male"))
              # Label 0 = Female & 1 = Male
Freq <- table(sex) # Frequency table
Percent <- prop.table(Freq) # Percentage table
CumPer <- cumsum(Percent) # Cumulative Percentage
cbind(Freq, Percent, CumPer) # Combine Three tables
```

```
##        Freq  Percent   CumPer
## Female  227 0.478903 0.478903
## Male    247 0.521097 1.000000
```

4. Frequency Distribution of an Ordnial Variable

```
experience <- factor(data$experience,
                     levels = c(1,2,3,4),
                     labels = c("none","lt 6 months",
                                "6 months-2 years",
                                "more than 2 years"))
                    # Labeling Experience
Freq <- table(experience) # Frequency table
Percent <- prop.table(Freq) # Percentage table
CumPer <- cumsum(Percent) # Cumulative Percentage table
cbind(Freq, Percent, CumPer) # Combine Three tables
```

```
##                    Freq   Percent    CumPer
## none                114 0.2405063 0.2405063
## lt 6 months          76 0.1603376 0.4008439
## 6 months-2 years    100 0.2109705 0.6118143
## more than 2 years   184 0.3881857 1.0000000
```

NOTE

1. "Prior work experience" is an ordinal variable. (right?)
2. "Frequency" is the raw number of cases in a given category.
3. "Percent" is the raw number of cases in a given category divided by the total number of cases (in the case, 474), multiplied by 100 (Formula: pct(i)=(f(i)/N)*100)
4. "Proportion" is a percentage divided by 100 (or f(i)/N). So the proportion with more than two years of work experience is 0.3882, or 0.39.
5. CANNOT compare raw frequencies across samples unless sample size is identical, so use percentages or proportions instead.
6. "Cumulative percent" is the cumulative percentage of cases in that category or below (e.g. 40% of all individuals have 6 months or less of work experience, 61% individuals have 2 years of work experience or less.)
7. CAN calculate cumulative percemtage *ABOVE* a given category as well with formula: 100-(Cumulative Percentage Below)