

Problem Set 1

Applied Stats/Quant Methods 1

Jack Merriman

Question 1

Part One

I begin by reading in the IQ data for y

```
1 y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113, 112, 98,  
      80, 97, 95, 111, 114, 89, 95, 126, 98)
```

I assign relevant calculations on the IQ data to objects for use in later functions

```
1 n <- length(y) #sample size  
2 ybar <- mean(y) #sample mean  
3 s <- sd(y) #sample standard deviation  
4 cc <- 0.9 #confidence coefficient
```

Now I verify that my sample size is not greater than 30, therefore the Central Limits Theorem does not apply, and I will use a t-test rather than a z-test

```
1 n>30  
2 [1]FALSE
```

I follow by using the t-distribution to extract a score by which to structure the confidence interval, I multiply it by the sample standard error to find my confidence intervals by adding and subtracting the product to the sample mean.

```
1 z <- qt((1+cc)/2, df=n-1) #z-score  
2 se <- (s/sqrt(n)) #sample standard error  
3 md <- z*se #distance of mean (md) from interval boundary  
4 ybar - md #lower interval  
5 [1] 93.95993  
6 ybar + md #upper interval  
7 [1] 102.9201
```

So my confidence interval is:

$$93.96 \leq x \leq 102.92$$

Part Two

As established in Part One, the sample size is less than 30, so a t-test will be used to test hypotheses. The school counselor is wishing to test whether the students' IQ is higher on average, so we will use a right-tailed t-test, and therefore our hypotheses are as follows:

$$H_O : \mu \leq 100$$

$$H_A : \mu > 100$$

I assign more relevant values to objects and calculate a t-score for these set of specific hypotheses

```
1 mu0 <- 100 #null hypothesis population mean
2 alp <- 0.05 #alpha
3 dgfr <- n-1 #degrees of freedom
4 t <- (ybar-mu0)/se #t-score
```

Now I have all the information I need to calculate the p-value for this t-score using the `pt()` function of R. Then I verify whether the p-value is greater than α .

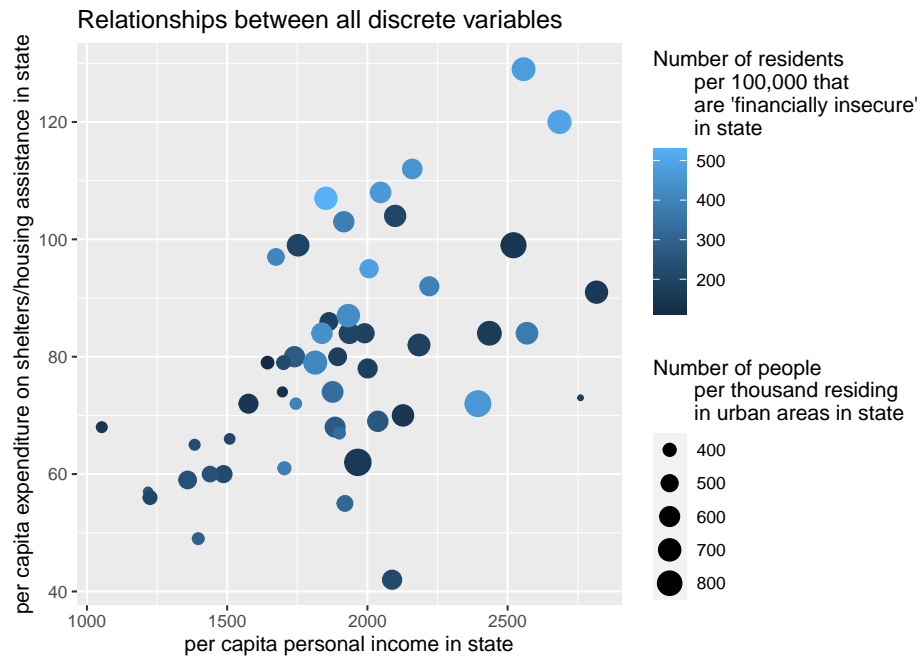
```
1 p <- pt(t, dgfr, lower.tail = FALSE) #p value for right tailed t-test
2 p > alp
```

The value of $p = 0.7125$, as $p > \alpha$, we have insufficient evidence to reject the null hypothesis.

Question 2

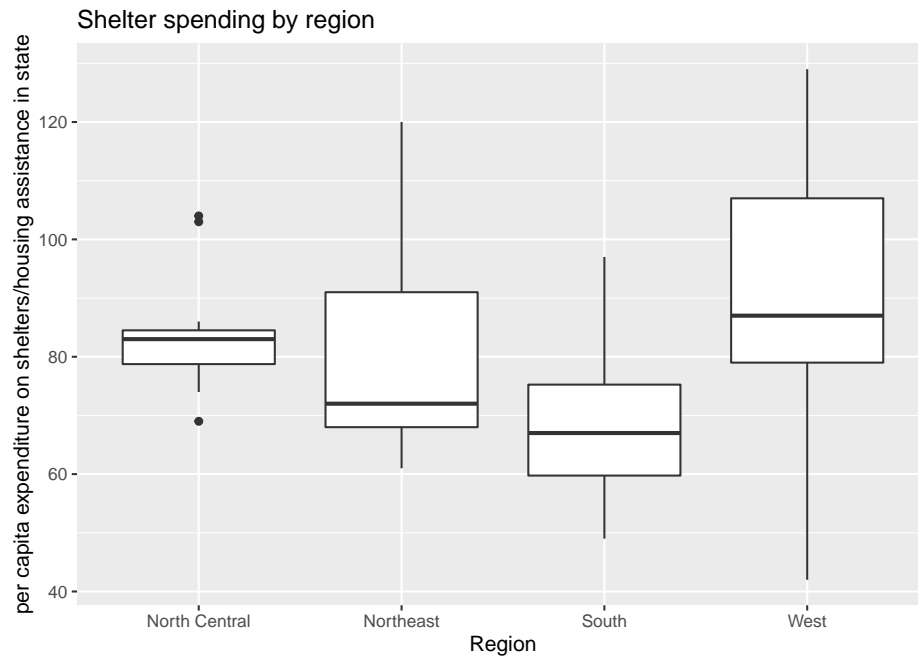
Part One

I plot the relationships using the `ggplot()` function of the `tidyverse` package in R, producing the following multivariate scatterplot.



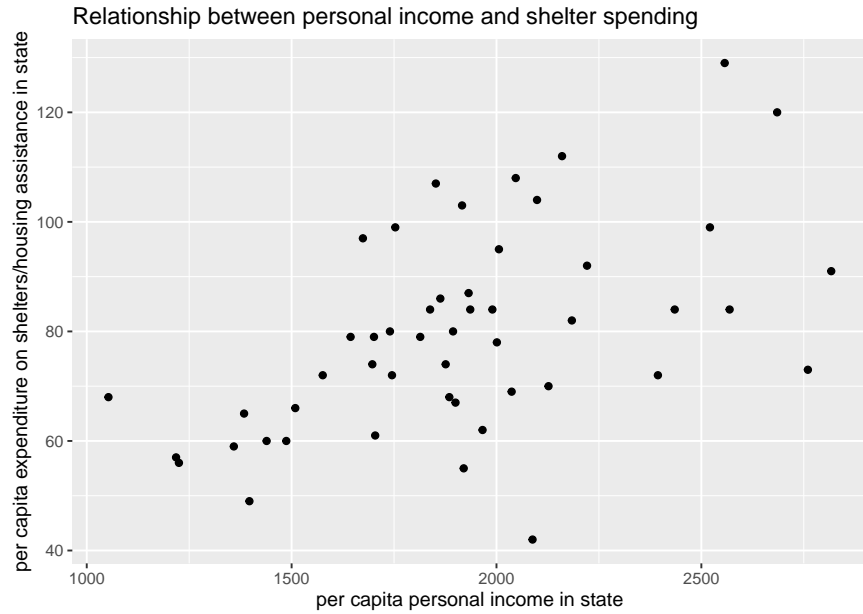
At a glance, there seems to be a positive correlation between per capita personal income in state, and per capita expenditure on shelter/housing assistance in state, which we can see through the diagonal shape of the axial variables. There does not seem to be any correlation between shelter spending and the other two variables, as the size and colour of the points are vertically distributed fairly randomly. There also appears to be no relationship between the number of financially insecure residents and per capita personal income, as the colour of the points is distributed randomly across the x axis. There does seem to be a positive correlation between per capita personal income and the number of people per capita residing in urban areas, as the size of the points grow along the x-axis.

Part Two



Taking the average to be the mean of each subset of data, the region with the highest average per capita shelter spending is the West, as the horizontal black lines in the box plot denote the mean.

Part Three



There appears to be a slightly positive relationship between per capita personal income and per capita expenditure on shelters/housing assistance across the states in the dataset. The points on the scatter plot slope diagonally upwards but do not closely cluster in one line, with other factors likely influencing the spending on shelters and housing assistance. The plot below shows the data by region.

