Problem 1) Suppose you sample $n=8$ household incomes from a specific neighborhood and obtain the following values (in thousands of dollars):

$X = 6, 19, 14, 33, 8, 22, 9$ and $9$

a) What is the median household income of this sample?

The scores in order are 6, 8, 9, 9, 14, 19, 22 and 33
The median annual income is 11.5 thousand dollars.

b) Calculate the mean:

$$\bar{X} = \frac{\sum X}{n}$$

$$\sum X = 6 + 19 + 14 + 33 + 8 + 22 + 9 + 9 = 120$$

The mean is $\frac{120}{8} = 15$ thousand dollars

c) Compare the median and mean and make a comment about the shape of the distribution of incomes for this example. Which is the most appropriate measure of central tendency? (mean median or mode).

If a distribution is positively skewed then the mean will be greater than the median.
If a distribution is negatively skewed then the mean will be smaller than the median.
If a distribution is skewed, then the median is most appropriate.
d) Calculate the deviation from the mean for each score:

\[(X - \bar{X})\]

-9, 4, -1, 18, -7, 7, -6 and -6

e) Calculate \(SS_X\), the sum of squared deviations from the mean:

\[SS_X = \sum (X - \bar{X})^2\]

\[SS_X = (6 - 15)^2 + (19 - 15)^2 + (14 - 15)^2 + (8 - 15)^2 + (22 - 15)^2 + (9 - 15)^2 + (9 - 15)^2\]

\[SS_X = 81 + 16 + 1 + 324 + 49 + 49 + 36 + 36 = 592\]

f) Calculate the variance using:

\[MS = \frac{\sum(X - \bar{X})^2}{n}\]

Round to two decimal points.

\[MS = \frac{592}{8} = 74.00\]

g) Calculate the standard deviation by taking the square root of the variance:

\[s = \sqrt{s^2} = \sqrt{\frac{\sum(X - \bar{X})^2}{n}}\]

Round to two decimal points.

\[s = \sqrt{74.00} = 8.60\]
h) Calculate the sum of the squares:
\[ \sum X^2 \]
\[ 36 + 361 + 196 + 1089 + 64 + 484 + 81 + 81 = 2392 \]

i) Calculate the square of the sum:
\[ (\sum X)^2 \]
\[ 6 + 19 + 14 + 33 + 8 + 22 + 9 + 9 = 120 \]
\[ 120^2 = 14400 \]

j) Calculate the sum of squared deviations $SS_X$ using this formula:
\[ SS_X = \sum X^2 - \frac{(\sum X)^2}{n} \]

Check that you got the same answer as for part e above.
\[ SS_X = 2392 - \frac{14400}{8} = 592 \]

k) Suppose there was a recession and all of the incomes are cut in half. What is the new mean and the new standard deviation?

The new mean is the old mean divided by 2:
\[ \frac{15.00}{2} = 7.50 \]

The new sd is the old sd divided by 2
\[ \frac{8.60}{2} = 4.30 \]

l) Suppose there was an economic boom and all of the incomes increased by 10 thousand dollars. What is the new mean and the new standard deviation?

The new mean is the old mean plus 10:
\[ 15+10 = 25 \text{ thousand dollars} \]

The new sd is the same as the old sd: 8.60 thousand dollars.
**Problem 2** Suppose that the preferred outdoor temperature of the population is normally distributed with a mean of 68 degrees and a standard deviation of 8 degrees.

a) On the graph below, label the x-axis so that the bell-curve represents this distribution.

b) Shade the area under the curve for temperatures above 72 degrees.

c) Find the z score for 72

\[ z = \frac{72 - 68}{8} = 0.5 \]

d) Use the table to determine what percent of the population prefers a temperature above 72.

Using the table, \( Pr(z > 0.50) = 0.3085 \)

30.85 percent of the population prefers a temperature above 72 degrees.
Problem 3) Assume that a population of men is normally distributed with a mean height of 69 inches and a standard deviation of 3 inches.

a) Draw a smooth bell curve below and label the axes so that it matches the probability distribution of heights.

![Bell Curve](image)

b) Mark the heights of 61, 65, 69, 73 and 77 as vertical lines on this curve.

c) Calculate the z scores for each of these 5 heights.

61 inches: \( z = \frac{61 - 69}{3} = -2.67 \)
65 inches: \( z = \frac{65 - 69}{3} = -1.33 \)
69 inches: \( z = \frac{69 - 69}{3} = 0.00 \)
73 inches: \( z = \frac{73 - 69}{3} = 1.33 \)
77 inches: \( z = \frac{77 - 69}{3} = 2.67 \)

c) What proportion of men fall below each of these 5 heights?

\[ Pr(X < 61) = Pr(z < -2.67) = 0.0038 \]
\[ Pr(X < 65) = Pr(z < -1.33) = 0.0918 \]
\[ Pr(X < 69) = Pr(z < 0.00) = 0.5000 \]
\[ Pr(X < 73) = Pr(z < 1.33) = 0.9082 \]
\[ Pr(X < 77) = Pr(z < 2.67) = 0.9962 \]
d) These 5 heights divide the range of heights into 6 sections. Find the proportion of heights that fall within each section. Check that the proportions all add up to 1.

\[ Pr(X < 61) = Pr(z < -2.67) = 0.0038 \]
\[ Pr(61 < X < 65) = Pr(z < -2.67) - Pr(z < -1.33) = 0.0918 - 0.0038 = 0.0880 \]
\[ Pr(65 < X < 69) = Pr(z < -1.33) - Pr(z < 0.00) = 0.5000 - 0.0918 = 0.4082 \]
\[ Pr(69 < X < 73) = Pr(z < 0.00) - Pr(z < 1.33) = 0.9082 - 0.5000 = 0.4082 \]
\[ Pr(73 < X < 77) = Pr(z < 1.33) - Pr(z < 2.67) = 0.9962 - 0.9082 = 0.0880 \]
\[ Pr(X > 77) = Pr(z > 2.67) = 1 - Pr(z < 2.67) = 0.0038 \]

Check: \(0.0038 + 0.088 + 0.4082 + 0.4082 + 0.088 + 0.0038 = 1\)

e) What is the height for which 75% of the heights fall below? (Round to the nearest tenth of an inch)

Using the table, the z-score for \(p = 0.7500\) is \(z = 0.6745\).
Converting from z to heights:
\[ X = 69.00 + (0.6745)(3.00) = 71.0 \text{ inches} \]
Problem 4) It possible for more than half of the students at UW to have a greater than the average GPA? If so, how could this happen?

This could happen if we assume that ‘average’ means ‘mean’. If the median is greater than the mean, then more than half of the distribution is greater than the mean, which can happen if the distribution is negatively skewed.

Problem 5) Do the following 6 problems using R. For all examples, assume that the population is normally distributed. Provide your answers as a printout.

Very helpful hints can be found in the R files associated with the tutorials:

NormalDistribution.R

The easiest way to submit your answers is to copy the output in the 'Console' window and paste into a document to print. Your printout could look something like this:

```r
pnorm(1)
[1] 0.8413447
qnorm(0.5)
[1] 0
pnorm(115,100,15)
[1] 0.8413447
```

a) A Google search shows that the music of teams has a mean of 62 and a standard deviation of 3.7, find value for which 81.38 percent of the music fall below.

```r
qnorm(0.8138,62,3.7)
[1] 65.30035
```
Answer: 65.30

b) A friend tells you that the peace of undergraduates has a mean of 41 and a standard deviation of 3.9, find the proportion of the peace below 34.4

```r
pnorm(34.4,41,3.9)
[1] 0.04529366
```
Answer: 0.0453

c) A friend tells you that the jewelry of elbows has a mean of 83 and a standard deviation of 2.3, find the proportion of the jewelry above 80.9

```r
1-pnorm(-0.913043,83,2.3)
[1] 1
```
Answer: 0.8194

d) A Google search shows that the response time of beer has a mean of 78 and a standard deviation of 3.6, find the proportion of the response time between 75 and 78.5

```r
pnorm(78.5,78,3.6) - pnorm(75,78,3.6)
[1] 0.3529026
```
Answer: 0.3529

e) Suppose the depth of response times has a mean of 72 and a standard deviation of 6.8, find value for which 41.27 percent of the depth fall below.

\[ qnorm(0.4127,72,6.8) \]

\[ [1] \ 70.49989 \]

Answer: 70.50

f) If you know that the experience of statistics problems has a mean of 30 and a standard deviation of 6.5, find the proportion of the experience between 16 and 33.4

\[ pnorm(33.4,30,6.5) - pnorm(16,30,6.5) \]

\[ [1] \ 0.6839135 \]

Answer: 0.6839

Problem 6

Use R to calculate some statistics based on your survey. Very helpful hints can be found in the R files associated with the tutorials:

DescriptiveStatistics.R

a) Find the mean of your father’s heights:

Big hint: Here's how to calculate the mean of your mother’s heights:

```r
# load in the survey data
survey <- read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W20survey.csv")
# calculate the mean of 'mheight', ignoring missing data:
mean(survey$mheight, na.rm = TRUE)
[1] 63.34454

Father's heights can be found in:

survey$pheight

[1] 72 NA 72 67 76 75 72 68 72 67 73 64 70 65 72 69 66 66 66 67 65 70 66 68 70
[26] 69 66 68 71 70 67 74 73 70 72 72 71 74 68 71 65 66 76 66 68 69 70 69 70 72
[51] 68 71 67 68 68 76 75 60 71 77 70 70 72 73 72 75 73 70 61 72 72 71 69 68 70
[76] 70 NA 70 69 66 67 71 73 56 74 70 73 62 69 68 63 65 70 68 NA 72 69 67 75 74
[101] 71 NA 67 68 74 67 72 73 68 70 68 73 71 65 67 72 71 63 72 67

# answer:
mean(survey$pheight, na.rm = TRUE)
[1] 69.51724
```

b) Find the standard deviation of the heights of the male students.

Big hint: Here’s how to calculate the standard deviation of the heights of the female students:

```r
# load in the survey data
survey <- read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W20survey.csv")
# find the heights of the female students
female.height <- survey$height[survey$gender=="Female"]
# calculate the standard deviation
```
The heights of the men can be found with:

```r
male.height <- survey$height[survey$gender=="Male"]
```

Big hint: Here’s how to count where people went to high school:

```r
# load in the survey data
survey <- read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W20survey.csv")
# use 'table' to sort the entries in 'school'
school <- table(survey$school)
print(school)
```

The voting preferences can be found in:

```r
# load in the survey data
data.set <- read.csv("http://www.courses.washington.edu/psy315/datasets/Psych315W20survey.csv")
# use 'table' to sort the entries in 'school'
school <- table(data.set$school)
print(school)
```

```r
table(data.set$vote)
```

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>I never (or can’t) vote</th>
<th>I never (or can’t) vote</th>
<th>Independent</th>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>71</td>
<td>32</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# answer:

```r
table(data.set$vote)
```

<table>
<thead>
<tr>
<th></th>
<th>Democrat</th>
<th>I never (or can’t) vote</th>
<th>I never (or can’t) vote</th>
<th>Independent</th>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>71</td>
<td>32</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>