Solutions for Assignment Unit 4 -- MATH1280

(Element 1, Yes/No scale)

1. The median of this data is: **2.8**. (Correct answer = "Yes", incorrect answer = "No") Explanation: You should understand two methods of finding the answer.

A) A manual calculation: First load the values into an R object (programmers sometimes say "variable" instead of "object"). I called the R object *x*, but you can use a different name.

> x <- c(7.2, 1.2, 1.8, 2.8, 18, -1.9, -0.1, -1.5, 13.0, 3.2, -1.1, 7.0, 0.5, 3.9, 2.1, 4.1, 6.5)

The sorted observations are:

> sort(x)[1] -1.9 -1.5 -1.1 -0.1 0.5 1.2 1.8 2.1 2.8 3.2 3.9 4.1 6.5 7.0 7.2

[16] 13.0 18.0
>
There are 17 observations: > length(x)[1] 17>

Find the middle value by counting. The 9th observation is the middle observation (there are 8 observations on the left and 8 on the right). The value of the 9th observation is 2.8

>

B) Using R:

> median(x)
[1] 2.8
(Element 2, Yes/No scale) (Correct answer = "Yes", incorrect answer = "No") The answer is: **True**. The example above shows the sorting technique.

(Element 3, Yes/No scale) (Correct answer = "Yes", incorrect answer = "No") The answer is: **6.00**

Short Explanation: The first quartile (Q1) in this data is equal to 0.5. The third quartile (Q3) in this data is equal to 6.5. The interquartile range for this data is: **6.00**

(Element 4, 4-point scale) (Rate the answer on a 4-point scale. The correct answer gets the maximum number of stars**)**

4 . The equation for the interquartile range is **Q3 - Q1**, where Q3 is the third quartile and Q1 is the first quartile. You can also say "**the third quartile minus the first quartile**."

(Element 5, 3-point scale) (Rate the answer on a 3-point scale. The correct answer gets the maximum number of stars**)**5. The lower cutoff point (criterion value) used to identify outliers is [- 8.5] **and the upper cutoff (criterion value) is 15.50**

Do the calculations manually. Do not use R:

(Element 6, 2-point scale) **(**Correct answer = "Yes", incorrect answer = "No"**)**The answer is: **FALSE**. One of the example solutions above shows the summary() command. It shows the quartiles but does not show the outliers.

(Element 7, 3-point scale)7. The list of outlier values is: **18**. (Rate the answer on a 3-point scale. The correct answer gets the maximum number of stars. If the student included the correct answer and added either one or two additional answers, you may award partial credit.)

Here is one way to find the answer in R... and this answer uses the calculated values from above for lower.cutoff and upper.cutoff (so you would have to run the code above first):

```
> x[x<lower.cutoff | x >upper.cutoff]
[1] 18
```

(Element 8, Yes/No scale)

```
> sd(x)
```

[1] **5.249 or you can show more decimals: 5.249468 A Random**

Variable:

(Element 9, Yes/No scale)
The missing probability value is: **0.15**. (Correct answer = "Yes", incorrect answer = "No")The key to finding the missing value is that the sum of probabilities must equal 1 (Yakir, 2011, p. 66).

The sum of the probabilities in the table was: .1 + .15 + .25 + .10 + .10 + .15 = .85 The missing value was 1 - .85 = .15

(Element 10, 4-point scale) (Rate the answer on a 4-point scale. The correct answer gets the maximum number of stars)

The sum of probabilities must equal **1** (Yakir, 2011, p. 66).
(Element 11, Yes/No scale)

The answer is: **NO**. (Correct answer = "Yes", incorrect answer = "No")

Explanation: The data table was described as a random variable. Samples of data contain real observations (Yakir, 2011, p. 16), but random variables contain no observations and only describe the probability of obtaining values if a value were selected (Yakir, 2011, p. 53).

The following quote helps to answer the question, but is not required as part of the response for grading purposes:

"A random variable is the future outcome of a measurement, before the measurement is taken. It does not have a specific value, but rather a collection of potential values with a distribution over these values. After the measurement is taken and the specific value is revealed then the random variable ceases to be a random variable! Instead, it becomes data." (Yakir, 2011, p. 53).

(Element 12, Yes/No scale)

The answer is: **FALSE (or NO). (**Correct answer = "Yes", incorrect answer = "No"**)**Explanation: The values in the second row of this table describe the probability of selecting the value in the same column in the first row. They are not cunulative probabilities.

(Element 13, Yes/No scale)The answer is **..60**. (Correct answer = "Yes", incorrect answer = "No")

To answer the question, first complete the missing probability value and confirm that the sum of the probability is 1:

```
>v<-c(0, 1, 2, 3, 4, 5, 6) > p <- c(.1, .15, .25, .10, .15, .10, .15) >> # verify that p sums to 1:

> sum(p)
[1] 1
>
```

Look at the table of probabilities for the random variable. The probability of obtaining 0 is .1 and the probability of obtaining 1 is .15, and so on, upt to the culminating value of 3. The sum of those numbers is exactly .60. There are no other values in the sample space that are less than or equal to one other than those four values. The answer is .60.

(Element 14, YES/NO scale) (Correct answer = "Yes", incorrect answer = "No"**)**

The answer is: **0**. The list of values in the first row of the table is the sample space. The number 1.5 is not in the sample space, so it is not in the population and the probability of selecting it is zero. Stated another way, the probabilities in the second row (after you complete the table) sums to 1, and there is no “left-over” probability to assign to the value 1.5.

(Element 15, YES/NO scale) (Correct answer = "Yes", incorrect answer = "No"**)** The answer is **2.95.**

Using the probability table after completing the missing value:

```
>> # Long Way (following Yakir, 2011, p. 57-59)> E = 0*.1 + 1*.15 + 2*.25 + 3*.1 + 4*.15 + 5*.1 + 6*.15 > round(E, 3)[1] 2.95>> # Quick Way (quicker when
```

there are many observations): >v<-c(0, 1, 2, 3, 4, 5, 6)> p <- c(.1, .15, .25, .10, .15, .10, .15)> # E = the expectation:> E <- sum(v*p)

> round(E, 3)
[1] 2.95
>

(Element 16, YES/NO scale) (Correct answer = "Yes", incorrect answer = "No") The answer is: **FALSE**.

The question describes exactly what NOT to do. When you are given values in the first row and relative frequencies in the second row, you must "weight" the values in the first row corresponding to the values in the second row.

Think of an extreme situation: A random variable has a sample space of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, and the probabilities for those values are .99, .002, .001, .001, .001, .001, .001, .001, .001, .001. In that situations, you will pick a 1 99% of the time, so the mean of any sample would be very close to 1. If you do not use the relative frequencies as "weights," you would mistakenly say that that the expectation is 5.5.

(Element 17, yes/no scale) (Correct answer = "Yes", incorrect answer = "No"**)** The answer is **1.910** (or something that rounds to that answer)

The Long Way (following Yakir, 2011, p. 57-59). First find the expectation:> E = 0*.1 + 1*.15 + 2*.25 + 3*.1 + 4*.15 + 5*.1 + 6*.15> round(E, 3)

[1] 2.95

Remember that "E" is the same as typing its value: 3.95
>variance=(0-E)^2*.1 +(1-E)^2*.15+(2-E)^2*.25+(3-E)^2*.1+(4- E)^2 * .15 + (5-E)^2 * .1 + (6-E)^2 * .15> # show the variance:> variance

[1] 3.6475> std.dev = sqrt(variance) > std.dev[1] 1.909843> round(std.dev, 3)[1] 1.91>

Quick Way (following 2011, p. 61, solutions 4.1.6-4.1.8). First find the expectation

```
> v <- c( 0, 1, 2, 3, 4, 5,  6)
> p <- c(.1, .15, .25, .10, .15, .10, .15)
> E <- sum(v*p)
> round(E, 3)
[1] 2.95
> variance <- sum((v-E)^2 * p)
> # show the variance:
> variance
[1] 3.6475
> std.dev <- sqrt(variance)
> std.dev
[1] 1.909843
> round(std.dev, 3)
[1] 1.91
>
```

(Element 18, 4-point scale) (Rate the answer on a 4-point scale. The correct answer gets the maximum number of stars**)**

Answer: To find the variance when you know the standard deviation, square the standard deviation.

A Population:

(Element 19, 4-point scale) (Rate the answer on a 4-point scale. The correct answer gets the maximum number of stars**)**Answers: **Count of a = 49,949.**Here is some example R code:

```
> pppp3 <- read.csv("pop3.csv")
> colnames(pppp3)
[1] "type" "time"
>
> length(pppp3$type[ pppp3$type == 'a'])
[1] 49949
>
```

(Element 20, 5-point scale) (Rate the answer on a 5-point scale. The correct answer gets the maximum number of stars**)**The answer is **4.473 (or you can show more decimal places).**> round(median (pppp3$time), 3)

[1] 4.473

(Element 21, 5-point scale) (Rate the answer on a 5-point scale. The correct answer gets the maximum number of stars**)**The answer is: **54.916 (or you can show more decimal places)**> round(var(pppp3$time), 3)

[1] 54.916 >