ProblemSheet1a

Problem Sheet 1a

Measures of Location

1. Describe each of the following measures of location including their the pros and cons

i Mean

To get the mean (or average), you have to sum all the elements of a dataset and divide by the number of elements on it. The mean can be represented by the following formula:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

An example could be used to calculate the mean of the salary hour from the Merit Market company editors

```
editorsSalaries=c(12, 10,11, 12,9,13,12)


mean(editorsSalaries)
```

```
## [1] 11.28571
```

Pros:
1. Easy to calculate (for small sets, you can do the maths in your head)
2. Is well understood, you can probably say to most of people: I sleep 7 hours per day on average.

Cons:
It's sensitive to extreme values. Look what happens with the mean salary from Merit Market editors when mister Merit Jr. becomes 18 and starts working on the department:

```
meritJrGenerousSalary <- 90
editorsSalaries=c(12, 10,11, 12,9,13,12,meritJrGenerousSalary)

mean(editorsSalaries)
```

```
## [1] 21.125
```

Suddenly, you can't rely on the mean to answer questions like: how much me, Mr. Norman NoMerit could make per hour working on that market agency

ii Median

Median is the value in the middle of a list of values. If the list has an even number of values, it's the mean between the two in the middle.

Pros: not as sensitive as mean for extremes. Using the previos salaries example, median still a pretty good measure to answer how much mister Mr. Norman NoMerit would expect to make joining Merit Market editors group:

```
editorsSalaries=c(12, 10,11, 12,9,13,12)


median(editorsSalaries)
```

```
## [1] 12
```

```
meritJrGenerousSalary <- 90
editorsSalaries=c(12, 10,11, 12,9,13,12,meritJrGenerousSalary)

median(editorsSalaries)
```

```
## [1] 12
```

Cons: The set must be ordered, you can't make a calculation up in your head that easily. Also, It doesn't use all the data set so, if your distribution is left or right skewed, it might be misleading used alone

iii Variance

The variance is, how much the data spread around the mean. It's calculated with the following formula:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1},$$

It's frequently square rooted to become the stantad deviation and be used on other calculations.

Pros: It uses all data and it's basis to calculate other stable measures largely used on statistics.
Cons: It's not that easy to interpret by using it alone and can be influenced by extreme values on the dataset. Good look saying on your sleep tracking phone app: the standard deviation of your sleep time is 1.23...

iv Skewness

Is the measure of symmetry on a given distribution. It can indicate whether a distribution is symmetric, left or right skewed. It's formula basically says: get how far the mean is from the mode, then divide it by the standard deviation:

$$sk_1 = \frac{\bar{x} - Mode(x)}{\sigma}.$$

Zero/near zero means the distribution is symmetric, high positive or high negative values means the distribution is right or left skewed respectevely.

Pros: Gives an overview of how your data is spread, can be used to determine how to work with you data also.

Cons: sensitive to cons from the mode like more than one mode in the distribution which can hide a even or bi-modal distribution.

Counting

2. How many different combinations of 4 cards can be made for a 52 card deck.

The order doesn't matter here so that will be:

$$\frac{52!}{4! * (52-4)!} -> \frac{52!}{4! * 48!} -> \frac{52 * 51 * 50 * 49}{4!} -> \frac{6497400}{24} -> 270725$$

3. A bank issues bank cards with PINs consisting of 4 digits, each one {0,1,2,. . . ,9}. How many unique PINs are there if

i. Any 4-digit code can be used.

That is 10 possibilities for each digit so, it's `10*10*10*10` = 10.000 or

$$10^4$$

ii. The digits must be different.

That is 10 possibilities for the first digit and one less for each subsequent `10*9*8*7` = 5040 or

$$\frac{10!}{(10-4)!} -> \frac{10 * 9 * 8 * 7 * 6!}{6!} -> 10 * 9 * 8 * 7 -> 5040$$

4. In a lottery, each ticket has 5 one-digit numbers 0-9 which is not repeated on it.

i You win if your ticket has the digits in any order. What are the total number of possible combinations?

$$\frac{10!}{5! * (10-5)!} -> \frac{10 * 9 * 8 * 7 * 6 * 5!}{5! * 5!} -> \frac{10 * 9 * 8 * 7 * 6}{5!} -> \frac{30240}{120} -> 252$$

##ii You would win only if your ticket has the digits in the required order. What are the total number of combinations?

$$\frac{10!}{(10-5)!} -> \frac{10 * 9 * 8 * 7 * 6 * 5!}{5!} -> 10 * 9 * 8 * 7 * 6 -> 30240$$

5. How many different combinations of 6 cards can be made for a 52 card deck if

i) order matters.

$$\frac{52!}{(52-6)!} -> \frac{52 * 51 * 50 * 49 * 48 * 47 * 46!}{46!} -> 52 * 51 * 50 * 49 * 48 * 47 -> 14.658.134.400$$

ii) order does not matter

$$\frac{52!}{6! * (52-6)!} -> \frac{52 * 51 * 50 * 49 * 48 * 47 * 46!}{6! * 46!} -> \frac{52 * 51 * 50 * 49 * 48 * 47}{6!} -> 20358520$$

6. A poker hand consists of 7 cards:

i) How many different hands are possible, if order does not matter,

ii) How many hands can be made with at least one king and one queen.

7. In a game of 5 card poker what are the number of different possible hands are there?

a) A hand with a pair

b) A hand with two pair

c) A hand with Three of a kind

d) A hand with a Flush (all the same suit)