How to participate?
joinpd.com
code: see above

Statistics for Data Science-I

Week 2 Solve with Instructor (graded)

How to participate?
joinpd.com
code: see above

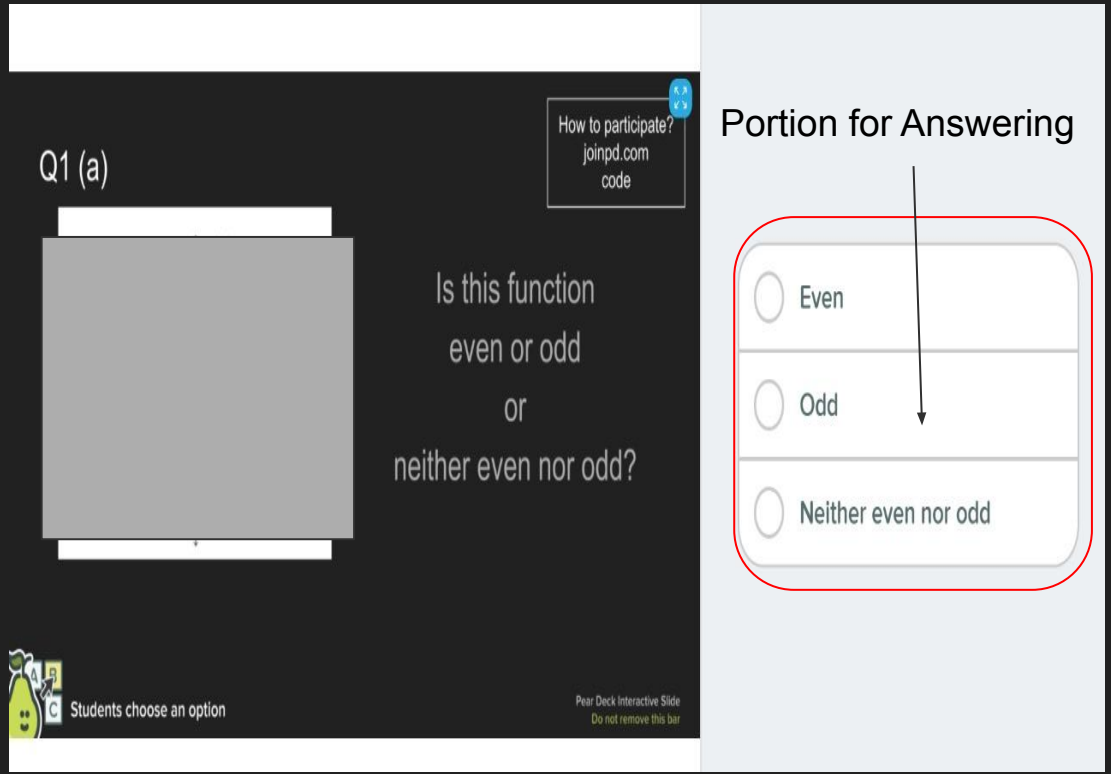
Statistics I: Week 2 Solve with Instructor

- Keep a notebook and pen ready for solving problems
- How to join?
  - Audio/screenshare on webex - click on link sent to you
    - Doubts? Use webex chat. Do not answer questions on webex chat.
  - Join on pear deck - joinpd.com (enter code seen on top right)
    - Answer questions only here
- For every question - 5 to 15 minutes allotted
  - Question will be shown in a slide for solving
  - If you are done solving, enter your answer at joinpd.com
  - Presenter will provide a solution
  - Questions and discussion

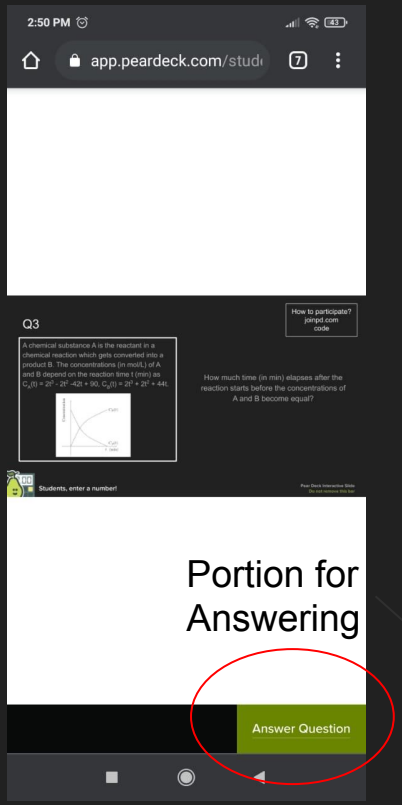
Example Screenshots

How to participate?
joinpd.com
code: see above

Laptop/Desktop



Mobile



Frequency and Relative frequency

- What is frequency distribution?

Frequency distribution of qualitative data is a listing of the distinct categories and their frequencies.

All the categories of categorical data are made in tabular form with their respective frequencies.

- What is relative frequency?

Relative frequency of category is ratio of frequency of category to the total frequency of categorical variable.

Prelude 1 to Q1

Amount invested by a company in 4 different types of products in the fiscal year 2020 is given in the table given below.

Products	Investment (in crores)	Relative frequency
P1		
P2		
P3	20	0.40
P4	10	

Q. Find the total amount (in crores) invested by the company in 4 products in year 2020.

Students, enter a number!

Solution:

Given, Amount invested in product P3 is 20 crores.

Amount invested in product P4 is 10 crores.

Relative frequency for product P3 is 0.40.

As we know that,

$$Relative\ frequency = \frac{frequency}{Total\ frequency}$$

$$\Rightarrow Total\ frequency = \frac{frequency}{Relative\ frequency}$$

For product P3,

$$Relative\ frequency\ for\ product\ P3 = 0.40$$
$$\Rightarrow \frac{Frequency\ of\ product\ P3}{Total\ frequency} = 0.40$$
$$\Rightarrow \frac{20}{Total\ frequency} = 0.40$$
$$\Rightarrow Total\ frequency = \frac{20}{0.40} = 50$$

Therefore, total amount invested by company in 4 products is 50 crores.

Prelude 2 to Q1

Amount invested by a company in 4 different types of products in the fiscal year 2020 is given in the table given below.

Products	Investment (in crores)	Relative frequency
P1		
P2		
P3	20	0.40
P4	10	

The company has invested thrice as much amount in product P1 as it has invested in product P2.

1. Find the amount (in crores) invested by company in product P1.
2. Find the amount (in crores) invested by company in product P2.

Enter the answers separated by comma.

Students, write your response!

Solution:

Let the amount invested by a company in the product P2 be x.

Therefore, the amount invested by a company in the product P1 is 3x.

Since the total amount invested by a company in products P1, P2, P3, and P4 is 50 crores, therefore,

$$x + 3x + 20 + 10 = 50$$
$$\Rightarrow 4x + 30 = 50$$
$$\Rightarrow 4x = 20$$
$$\Rightarrow x = 5$$

Therefore, amount invested by company in products P1 and P2 respectively is 15 and 5 crores.

Q.1

Amount invested by a company in 4 different types of products in the fiscal year 2020 is given in the table given below.

Products	Investment (in crores)	Relative frequency
P1		
P2		
P3	20	0.40
[illegible]	10	

If the company has invested thrice as much amount in product P1 as it has invested in product P2, find the relative frequency of product P1.

(Enter your answer up to 1 decimal places.)

Students, enter a number!

Solution:

Since,

$$Relative\ frequency = \frac{frequency}{Total\ frequency}$$

Relative frequency of product P1 is

$$\frac{15}{50} = 0.3$$

Prelude 1 to Q2

Consider a factory where 2 different types of objects namely, A and B are produced. Each object undergoes through 4 stages: drilling, welding, grinding, and fitting. At a given time T, 200 objects of type A, 100 objects of type B are being produced (i.e., they are in these 4 stages).

Stages	Type A
Drilling	0.30
Welding	0.30
Grinding	0.15
Fitting	0.25

Stages	Type B
Drilling	0.25
Welding	0.30
Grinding	0.20
Fitting	0.25

Q1. Find the number of objects of type A that is in grinding stage at time T.

Q2. Find the number of objects of type B that is in grinding stage at time T.

(Enter the answers separated by comma)

Students, write your response!

Solution: Q1

Total number of objects of type A that are being produced is 200.

15% of the objects of type A are in grinding stage at time T.

Therefore,

$$200 \times \frac{15}{100} = 30$$

Total of 30 objects of type A are in grinding stage at time T.

Solution: Q2

Total number of objects of type B that are being produced is 100.

20% of the objects of type B are in grinding stage at time T.

Therefore,

$$100 \times \frac{20}{100} = 20$$

Total of 20 objects of type B are in grinding stage at time T.

Prelude 2 to Q2

Consider a factory where 2 different types of objects namely, A and B are produced. Each object undergoes through 4 stages: drilling, welding, grinding, and fitting. At a given time T, 200 objects of type A, 100 objects of type B are being produced (i.e., they are in these 4 stages).

Stages	Type A
Drilling	0.30
Welding	0.30
Grinding	0.15
Fitting	0.25

Stages	Type B
Drilling	0.25
Welding	0.30
Grinding	0.20
Fitting	0.25

Q. Find the total number of objects that are in grinding stage at time T.

Students, enter a number!

Solution:

Number of objects of type A that are in grinding stage at time T is 200 * 0.15 = 30.

Number of objects of type B that are in grinding stage at time T is 100 * 0.20 = 20.

Therefore, the total number of objects that are in grinding stage at time T is 30 + 20 = 50.

(Since, only objects of type A and type B are being produced).

Q.2

Consider a factory where 2 different types of objects namely, A and B are produced. Each object undergoes through 4 stages: drilling, welding, grinding, and fitting. At a given time T, 200 objects of type A, 100 objects of type B are being produced (i.e., they are in these 4 stages).

Stages	Type A
Drilling	0.30
Welding	0.30
Grinding	0.15
Fitting	0.25

Stages	Type B
Drilling	0.25
Welding	0.30
Grinding	0.20
Fitting	0.25

Q. What proportion (in %) of objects of both type A and B are in grinding stage at time T.

(Enter your answer correct up to 2 decimal places).

Students, enter a number!

Solution:

Total of 300 objects of both type are being produced at time T.

Number of objects that are in grinding stage at time T is 50.

Therefore, proportion of objects of both type that are in grinding stage at time T is

$$\left(\frac{50}{300} \times 100\right)\% = 16.66\%$$

Prelude 1 to Q3

Frequency table for dataset 1

Categories	Frequency
A	2
B	5
C	3

Frequency table for dataset 2

Categories	Frequency
A	20
B	50
C	30

Find the relative frequency for categories listed in dataset 1 and 2.
(Enter your answer separated by comma.)

Students, write your response!

Solution:

Categories	Frequency	Relative frequency
A	2	0.2
B	5	0.5
C	3	0.3

Categories	Frequency	Relative frequency
A	20	0.2
B	50	0.5
C	30	0.3

The relative frequency for both the dataset 1 and 2 are same even though the frequency distribution for both the dataset are different.

Prelude 2 to Q3

Dataset 1: A, B, A, B, B, C, B, C, C, B.

Categories	Frequency
A	2
B	5
C	3

Dataset 2: A, B, C, B, B, C, B, A, C B, A, B, C, B, B, C, B, A, C B.

Categories	Frequency
A	4
B	10
C	6

Will the relative frequencies be same for dataset 1 and 2?

Students, write your response!

Solution:

Categories	Frequency	Relative frequency
A	2	0.2
B	5	0.5
C	3	0.3

Categories	Frequency	Relative frequency
A	4	0.2
B	10	0.5
C	6	0.3

The relative frequency for both the dataset 1 and 2 are same corresponding to the frequencies of categories in the dataset but the frequency distributions are not the same.

Q.3

Choose the correct statements from the following:

1. Two dataset having the same frequency distribution will always have the same relative frequency.
2. Two dataset having the same frequency distribution cannot have the same relative frequency.
3. Two dataset having the same relative frequency will always have the same frequency distribution.

Students choose an option

Bar chart

What is bar chart?

It is the graphical representation of frequency distribution of categorical variable. Each bin represents a category. The height of the bar represents the frequency/ relative frequency of that category.

Pareto chart

What is Pareto chart?

It is same as the bar chart, except that here the bins are arranged in either descending order or ascending order of the frequencies of categories.

It is not preferred to use pareto chart for variable with ordinal scale as the order in categories should be maintained.

Pie chart

What is Pie chart?

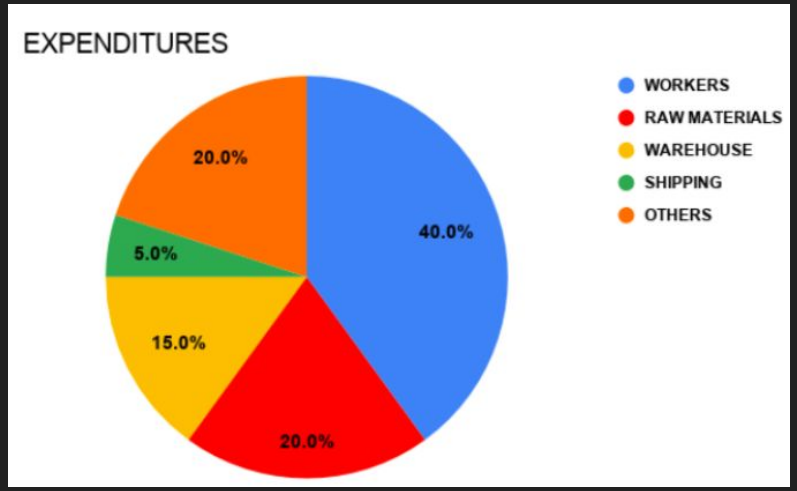
A pie chart is a circle divided into pieces proportional to the relative frequencies of the qualitative data.

The angle occupied by a particular category is r*360 degrees, where r is the relative frequency of that category.

- It is best to use pie chart if we are representing relative frequencies.
- f ranking should be in order of frequencies then pareto chart is best preferred.
- If frequencies of categories are required to be plotted, then bar chart is used.

Q4.

The expenditure for the month of October of a manufacturing company is given in Figure 2.1.A.



If 50,000 was spent on shipping, then what was the total expenditure of October?

1. 50,000
2. 1,00,000
3. 10,00,000
4. 50,00,000

Students choose an option

Solution:

Out of the total expenditure, 5% has been spent on shipping.

50,000 is the amount that is spent on shipping.

Expenditures	Relative frequency
Raw materials	0.20
Shipping	0.05
Warehouse	0.15
Workers	0.40
Others	0.20

$$Total\,expenditure = \frac{50,000}{0.05} = 10,00,000$$

Mode

What is mode of a categorical variable?

Mode is defined as the category of categorical variable which has the highest frequency.

Mode is defined for both ordinal and nominal scale of measurements.

Median

What is median of categorical variable?

Median is defined as the middle value in the ordered data (Ascending/Descending order).

If the data has odd number of observations, then the (n+1)/2 th observation is median.

If the data has even number of observations, then n/2, n/2 +1 th observations are median.

Median is not defined for the categorical variable with nominal scale of measurement as they do not have any order.

Q.5

The responses given by students in the class for a True/False question asked by a teacher is given below:

T, F, T, F, T, T, T, F, T, F, F, F, F, T, F.

1. Identify the mode.

   (a) Mode of the responses is True.

   (b) Mode of the responses is False.

   (c) Responses are bimodal.

Students choose an option

Solution:

There are a total of 15 students who has given the responses.

- 7 students have answered True.
- 8 students have answered False.

Since, more students have answered False for the question, the mode of the responses is False.

Therefore, the correct answer is (b).

Q5.

What can be said about the median of the responses?

1. Median of the responses is True.
2. Median of the responses is False.
3. Median of the responses can be either True or False.
4. Median is not defined for the given dataset.

Students choose an option

Solution:

The responses here have a nominal scale of measurement. Therefore, it is not possible to arrange the responses (True/False) in either ascending or descending order.

Hence, median is not defined for the responses.

Tips to solve the question

What is area Principle?

Displays of data must obey a fundamental rule called the area principle.

The area principle says that the area occupied by a part of the graph should correspond to the amount of data it represents.

What are misleading graphs?

Graphs that do not follow area principle are called misleading graphs.

Missing baseline (baseline does not start at zero). Also called as truncated graphs.

Tips to solve the question

What is round off errors?

When table entries are percentages or proportions, the total may sum to a value slightly different from 100% or 1. If percentages do not sum up to 100, then pie chart is said to have round off errors and thereby does not follow area principle.

Q6.

Which of the following statements are true?

1. Area principle is only applicable for pie chart not for bar chart.
2. Rounding off the error values of categories leads to misleading of pie chart.
3. Changing the radius of pie chart leads to loss of information.
4. Graphs which violate the area principle are called truncated graphs.
5. If the graphs do not have a baseline, the graph will be truncated.

Students choose an option

Thank You