

Foundations of Statistics – Laboratory Class 3

Summarising Data

Learning Objectives

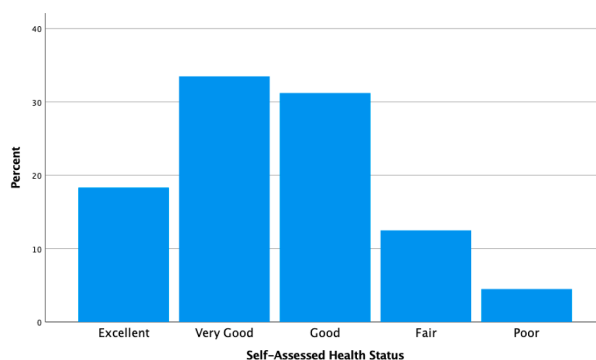
- Be able to produce and interpret the SPSS outputs used to summarise data
- Be able to choose the most appropriate output to describe a distribution
- Be able to write a rudimentary report to describe a distribution

OUTPUT FOR Fruit_Veg_&_Exercise.sav

FREQUENCIES OUTPUT: [categorical variables]

Self-Assessed Health Status

Self-Assessed Health Status					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Excellent	364	18.2	18.3	18.3
	Very Good	665	33.3	33.5	51.8
	Good	620	31.0	31.2	83.0
	Fair	248	12.4	12.5	95.5
	Poor	89	4.5	4.5	100.0
	Total	1986	99.3	100.0	
Missing	System	14	.7		
Total		2000	100.0		



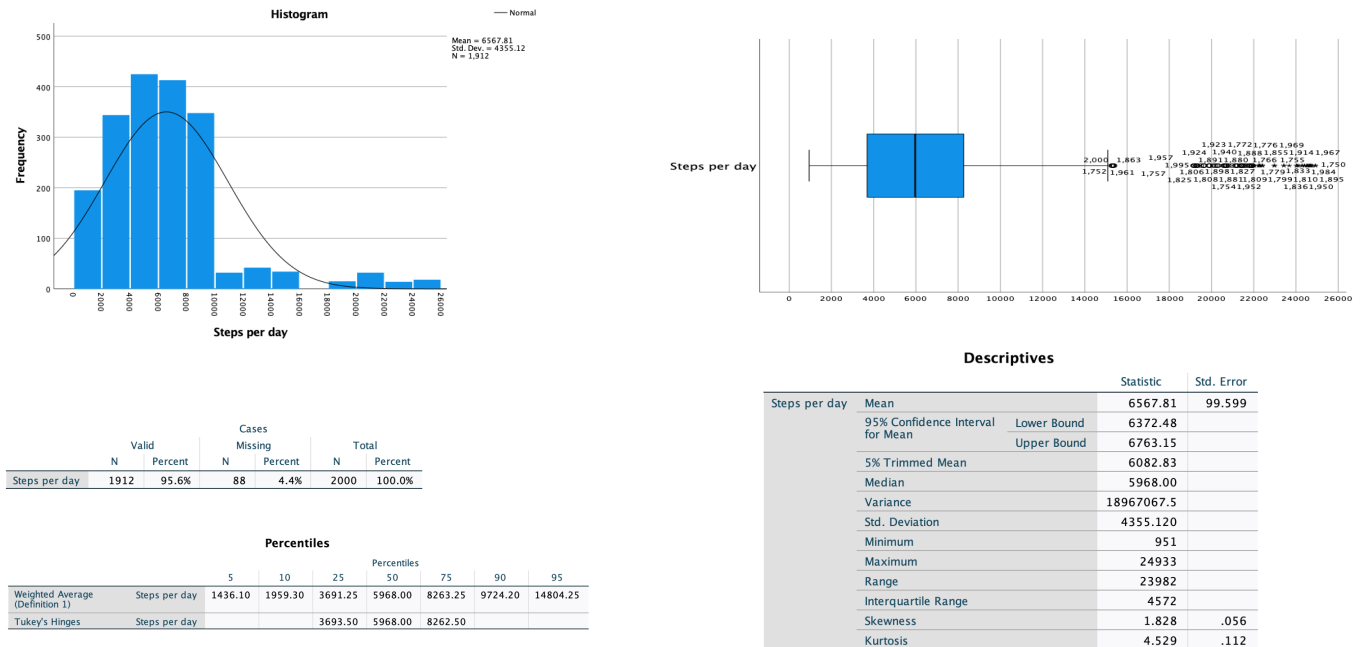
Reporting the distribution of a categorical variable:

The distribution of Self-Assessed Health Status for a sample of _____ respondents is displayed in Figure 1. The most typical reported health status in this sample was _____ (____ %), followed by _____ (____ %). While ____ % of respondents reported excellent health, only ____ % reported fair health, and very few respondents reported being in poor health.

Foundations of Statistics – Laboratory Class 3

EXPLORE OUTPUT: [metric variables]

Steps per day



Reporting the distribution of a metric variable:

As this distribution is skewed, we need to report the median not the mean.

The distribution of the number of steps per day in a sample of _____ respondents is displayed in Figure 1. The distribution is _____ with 50% of steps per day being _____ or more (or less). **Typically**, steps per day were between _____ and _____, with half of the steps per day falling within this range. Some respondents reported exceptionally high numbers of steps per day, of over _____, with the highest being _____ steps per day.

Foundations of Statistics – Laboratory Class 3

Complete the table below, indicating if the variable listed is Categorical or Metric, as well as the procedure required [Frequencies or Explore] to produce the most appropriate output:

Variable		Categorical OR Metric?c	Frequencies OR Explore?
a.	Test Result: [score out of 100]		
b.	Music Preference: 1. Jazz 2. Classical 3. Popular 4. Grunge 5. Heavy Metal 6. Other		
c.	Wellbeing: Scale of 0 to 25		
d.	Test Result: 1. Pass 2. Credit 3. Distinction 4. High Distinction		
e.	Weight [kgs]		
f.	Height [cm]		
g.	Weight: 1. Underweight 2. Normal range 3. Overweight 4. Obese		
h.	Use of Public Transport 0. Never 1. Sometimes 2. Always		
i.	Length of Stay [days]		
j.	Coffee Consumption [cups per day]		

Foundations of Statistics – Laboratory Class 3

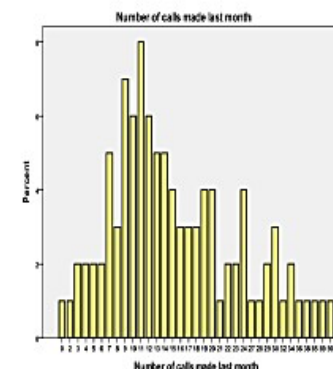
SCD: Which Output?

A mobile phone company 'Expencitel', wants to encourage people who already own one of their mobile phones to upgrade to a new model. To inform this marketing campaign, they decided to investigate how many people use their existing mobile phones, and how satisfied they are with them. They have accurate records of who has purchased one of their mobile phones over the past three years when they commenced operating. From this list they randomly select 100 people. As a first step in their study, they access the billing records of these 100 customers and record how many phone calls they made with the mobile phone over the past month and how many text messages they sent.

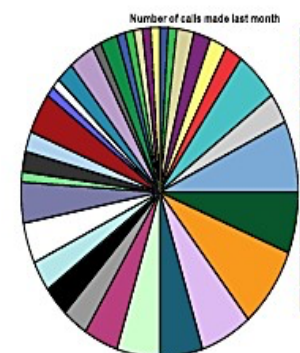
Output from FREQUENCIES procedure:

Number of calls made last month				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 0	1	1.0	1.0	1.0
1	1	1.0	1.0	2.0
2	2	2.0	2.0	4.0
3	2	2.0	2.0	6.0
4	2	2.0	2.0	8.0
5	2	2.0	2.0	10.0
6	2	2.0	2.0	12.0
7	5	5.0	5.0	17.0
8	3	3.0	3.0	20.0
9	7	7.0	7.0	27.0
10	6	6.0	6.0	33.0
11	8	8.0	8.0	41.0
12	6	6.0	6.0	47.0
13	5	5.0	5.0	52.0
14	5	5.0	5.0	57.0
15	4	4.0	4.0	61.0
16	3	3.0	3.0	64.0
17	3	3.0	3.0	67.0
18	3	3.0	3.0	70.0
19	4	4.0	4.0	74.0
20	4	4.0	4.0	78.0
21	1	1.0	1.0	79.0
22	2	2.0	2.0	81.0
23	2	2.0	2.0	83.0
24	4	4.0	4.0	87.0
27	1	1.0	1.0	88.0
28	1	1.0	1.0	89.0
29	2	2.0	2.0	91.0
30	3	3.0	3.0	94.0
32	1	1.0	1.0	95.0
34	2	2.0	2.0	97.0
36	1	1.0	1.0	98.0
38	1	1.0	1.0	99.0
85	1	1.0	1.0	100.0
90	1	1.0	1.0	100.0
Total	100	100.0	100.0	

10



11

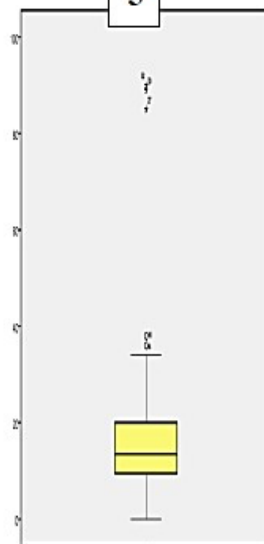


- What is the population of this study?
- Was the sample taken in a biased or unbiased way?
- Which part of this output gives us the best information on how many phone calls the people in this sample typically made?
- Which of the graphs gives the best picture of the distribution of number of phone calls?
- Which procedure gave us the most useful information for describing the distribution – EXPLORE or FREQUENCIES?

8

9

5



1

2

3

4

6

The following SPSS output was produced for the number of phone calls.

Output from EXPLORE procedure:

Percentiles							
	5	10	25	50	75	90	95
Number of calls made last month	4.00	6.10	9.50	13.50	20.00	30.00	35.90
Number of calls made last month			9.50	13.50	20.00		

Descriptives			
	Statistic	Std. Error	
Number of calls made last month	Mean	17.32	1.490
	95% Confidence Interval for Mean	Lower Bound	14.36
		Upper Bound	20.28
	5% Trimmed Mean		15.36
	Median		13.50
	Variance		221.998
	Std. Deviation		14.900
	Minimum		0
	Maximum		90
	Range		90
	Interquartile Range		11
	Skewness	3.301	.241
	Kurtosis	13.557	.478

7

