11-1 Graphical Tools in Sta?s?cs

Dec 12-10:59 AM

- **Population** - an entire collection of individuals about which we want to draw conclusions
- **Census** - the collection of information from the **whole population**
- **Sample** - a subset of the population which should be chosen at **random** to avoid **bias** in the results
- **Survey** - the collection of information from a **sample**

Statistic and parameter

A statistic is a descriptive measure computed from a sample of data. A parameter is a descriptive measure computed from an entire population of data.

Oct 30-8:17 AM

Classification of variables

Numerical or categorical

NUMERICAL (QUANTITATIVE) DATA – Quantitative variables measure a numerical quantity or amount on each experimental unit. Quantitative data yields a numerical response.

CATEGORICAL (QUALITATIVE) DATA – Qualitative variables measure a quality or characteristic of the experimental unit. Categorical data yields a qualitative response, i.e. data is kind or type rather than quantity.

Oct 29-8:28 AM

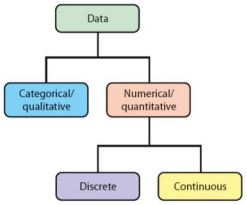
Classification of Numerical Data: Discrete or continuous

DISCRETE – responses which arise from counting.

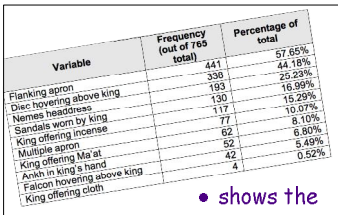
Example: Number of courses students take in a day.

CONTINUOUS – responses which arise from measuring.

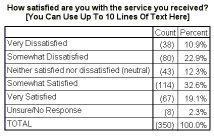
Example: Time it takes a student to travel from home to school.



Oct 29-8:33 AM



Variable	Frequency (out of 765 total)	Percentage of total
Flanking apron	441	57.65%
Diec hovering above king	338	44.18%
Nemes headdress	193	25.23%
Sandals worn by king	130	16.99%
King offering incense	117	15.29%
Multiple apron	77	10.07%
King offering Ma'at	62	8.10%
Ankh in king's hand	52	6.80%
Falcon hovering above king	42	5.49%
King offering cloth	4	0.52%

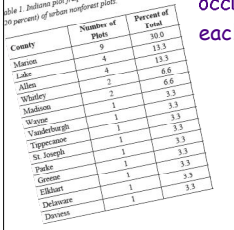


How satisfied are you with the service you received? [You Can Use Up To 10 Lines Of Text Here]

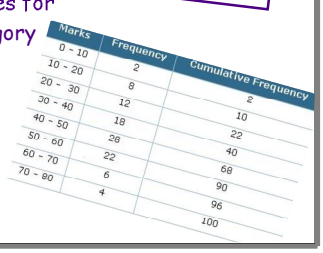
	Count	Percent
Very Dissatisfied	(38)	10.9%
Somewhat Dissatisfied	(80)	22.9%
Neither satisfied nor dissatisfied (neutral)	(43)	12.3%
Somewhat Satisfied	(114)	32.6%
Very Satisfied	(67)	19.1%
Unsure/No Response	(8)	2.3%
TOTAL	(350)	100.0%

FREQUENCY TABLES

- shows the number of occurrences for each category

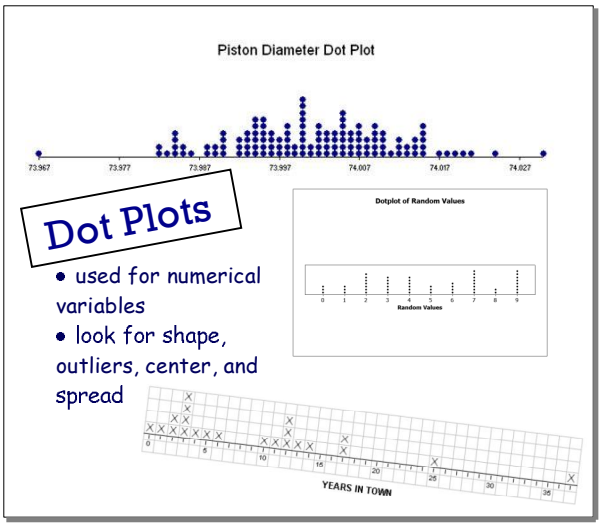


County	Number of Plots	Percent of Total
Marion	9	30.0
Lake	4	13.3
Allen	4	13.3
Whitley	2	6.6
Madison	2	6.6
Wayne	1	3.3
Vanderburgh	1	3.3
Tippecanoe	1	3.3
St. Joseph	1	3.3
Porter	1	3.3
Greene	1	3.3
Elkhart	1	3.3
Delaware	1	3.3
Daviess	1	3.3



Marks	Frequency	Cumulative Frequency
0 - 10	2	2
10 - 20	8	10
20 - 30	12	22
30 - 40	18	40
40 - 50	28	68
50 - 60	22	90
60 - 70	6	96
70 - 80	4	100

Oct 24-6:44 AM



Jan 30-3:27 PM

Chapter 5 Analyzing Univariate Data
Section 5.3 - Numerical Data: Dot Plots & Stem Plots
Section 5.6 - Numerical Data: Comparing Data Sets

STEMPLOTS

```
4 | 1 3 9
5 | 2 7
6 | 7 8 9
7 | 0 2 3 3 3 3 3 4 4 5 6 7
8 | 1 2 2
9 | 2 6
10 | 1 8
11 | 2 4
12 | 3
```

4 | 3 represents 43

- use for numerical variables
- describe with SOCCS

Stem-and-leaf plot of 'Skinfold Thickness'

N = 40
Leaf Unit = 0.10

```
 5 02468
 6 246688
 7 0000226
 8 2444666
 9 44666
10 488
11 088
12
13 22
14 0
15
16
17
18 2
```

4th. Grade Test Scores

```
5 | 0
6 | 4
7 | 58
8 | 00011122234456678 8
9 | 6789
```

7 | 5 represents 75

Feb 6-11:19 PM

225	250	213	216	183
211	200	246	243	231
209	209	225	200	217
224	230	237	185	235
258	225	232	216	227
216	256	226	271	217
196	243	232	230	246
228	200	216	219	
200	224	209	191	

Stem-and-leaf display	
18	3 5
19	1 6
20	0 0 0 0 9 9 9
21	1 3 6 6 6 6 7 7 9
22	4 4 5 5 5 6 7 8
23	0 0 1 2 2 5 7
24	3 3 6 6
25	0 6 8
26	
27	1

Oct 29-8:39 AM

```
                96 | 2
           9887540 | 3 2589
   9988888887665320 | 4 2679
  8754444322221100 | 5 00111458899
            986410 | 6 02679
                21 | 7 233346799
                   | 8 001223445
                   | 9 2469
                   | 10 25
                   | 11 02
```

- use when your want to compare two sets of numerical data
- describe with SOCCS

```
                72 | 48
                54 | 49
                 9 | 50
                 9 | 51
                 9 | 52 6
               321 | 53 7
               986 | 54 5
        7775544322 | 55 556
               977 | 56 024789
            96220 | 57 23466679
     7665444222110000 | 58 0003444455667888 9
   99988877643333210 | 59 0122355567788899
          743211110 | 60 0001223344445677 89
    8876555443221 10 | 61 2222444455556777789
 999988875544322110 | 62 0011222333444466777788999
    9966554333321000 | 63 001122345556688 8999
9888877665543322111 10 | 64 000001122455666 89
          985544200 | 65 12334677789
       99866422220 | 66 00478
            77664 | 67 00136
             9331 | 68 8
             8210 | 69
                  | 70
                9 | 71
                3 | 72
                  | 73
```

BACK-TO-BACK STEMPLOTS

```
99 | 0
   | 1
   | 2
992 | 3
 82 | 4 1166679
 82 | 5 449
  0 | 6 0
    | 7
```

Feb 6-11:41 PM

Suppose a consumer organization was interested in studying weekly food and living expenses of college students. A survey of 80 students yielded the following expenses to the nearest euro:

38	50	55	60	46	51	58	64	50	49	48	65	58	61	65	53
39	51	56	61	48	53	59	65	54	54	54	59	65	66	47	49
40	51	56	62	47	55	60	63	60	59	59	50	46	45	54	47
41	52	57	64	50	53	58	67	67	66	65	58	54	52	55	52
44	52	57	64	51	55	61	68	67	54	55	48	57	57	66	66

Oct 29-8:46 AM

Construction of a frequency distribution (table)

Rule 1: Intervals (classes) must be inclusive and non-overlapping; each observation must belong to one and only one class interval. Consider a frequency distribution for the living expenses of the 80 college students. If the frequency distribution contains the intervals '35–40' and '40–45', to which of these two classes would a person spending €40 belong?

The boundaries, or endpoints, of each class must be clearly defined. For our example, appropriate intervals would be '35 but less than 40' and '40 but less than 45'.

Oct 30-8:34 AM

Rule 2: Determine *k*, the number of classes. Practice and experience are the best guidelines for deciding on the number of classes. In general, the number of classes could be between 5 and 10. But this is not an absolute rule. Practitioners use their judgement in these issues. If the number of classes is too few, some characteristics of the distribution will be hidden, and if too many, some characteristics will be lost with the detail.

Oct 30-8:34 AM

Rule 3: Intervals should be the same width, w. The width is determined by the following:

$$\text{interval width} = \frac{\text{largest number} - \text{smallest number}}{\text{number of intervals}}$$

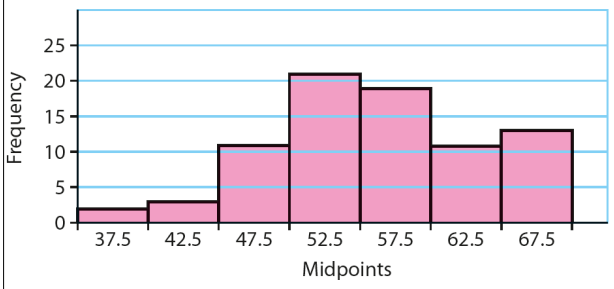
Both the number of intervals and the interval width should be rounded upward, possibly to the next largest integer. The above formula can be used when there are no natural ways of grouping the data. If this formula is used, the interval width is generally rounded to a convenient whole number to provide for easy interpretation.

In the example of the weekly living expenses of students, a reasonable grouping with nice round numbers was that of '35 but less than 40' and '40 but less than 45', etc.

Oct 30-8:34 AM

Living expenses (€)	Number of students	Percentage of students
35 but < 40	2	2.50
40 but < 45	3	3.75
45 but < 50	11	13.75
50 but < 55	21	26.25
55 but < 60	19	23.75
60 but < 65	11	13.75
65 but < 70	13	16.25
Total	**80**	**100.00**

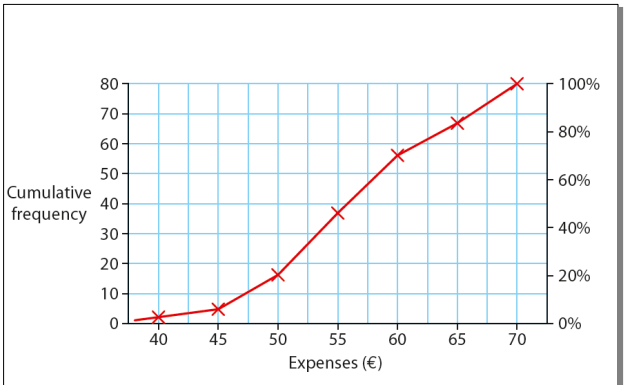
Oct 30-8:30 AM



Oct 30-8:30 AM

Living expenses (€)	Number of students	Cumulative number of students	Percentage of students	Cumulative percentage of students
35 but < 40	2	2	2.50	2.50
40 but < 45	3	5	3.75	6.25
45 but < 50	11	16	13.75	20.00
50 but < 55	21	37	26.25	46.25
55 but < 60	19	56	23.75	70.00
60 but < 65	11	67	13.75	83.75
65 but < 70	13	80	16.25	100.00
	80		**100.00**	

Oct 30-8:31 AM



Oct 30-8:32 AM

pg 475 #2, 3, 4, 7, 8, 9, 12, 16, 18, 19

Dec 17-10:30 AM