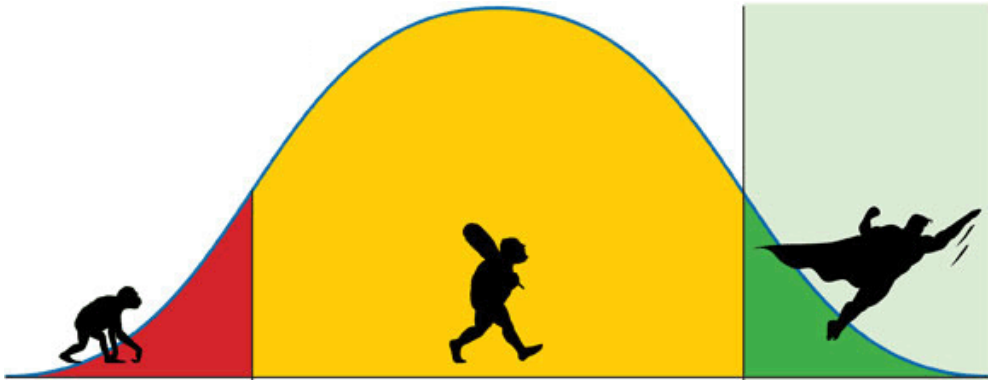


Measurement, variables and distributions

2015 Quantitative Psychological Research Methods – Session 2

Aims for this session

- To introduce various types of measurement scales
- To demonstrate how psychological data can be distributed across a population
- To discuss notions of centre and spread of a distribution

Recommended readings: Gravetter & Wallnau,

Chapters 1 - 3

SLEEP QUESTIONNAIRE

Scales of Measurement

- Four scales of measurement are commonly used in psychological research:
  - Categorical/Nominal
  - Ordinal
  - Ratio
  - Interval

- Each of these scales may be used with a particular kind of research data. Let's explore each one now…

Scales of measurement (G& W, chap 1.4)

Nominal (sometimes called Categorical)

- numbers are assigned to indicate a different category with no intrinsic ordering

Example:
Mental health study
Gender
0 = Female
1 = Male

16th participant
(ID no 00165)
is female

*dart.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

8 : id 00104

	id	gender	age	alcohol	cigs	marij	amphet	employ	swl	pa	na	kessler
1	00001	0	21.02	1	1	1	1	1	29.00	26.00	32.00	39.00
2	00010	1	49.15	1	2	1	1	7	28.00	43.00	13.00	14.00
3	00034	0	53.41	1	4	1	1	3	14.00	42.00	11.00	13.00
4	00038	1	22.67	1	4	1	1	1	26.00	42.00	13.00	19.00
5	00041	0	39.62	2	1	1	1	6	19.00	36.00	27.00	28.00
6	00059	0	51.69	1	1	1	1	5	29.00	50.00	14.00	10.00
7	00085	0	50.76	1	2	1	1	5	27.00	18.00	11.00	15.00
8	00104	0	56.60	2	1	1	1	6	26.00	21.00	10.00	11.00
9	00109	0	23.10	1	3	1	1	6	17.00	38.00	25.00	13.00
10	00110	1	67.18	2	1	1	1	4	7.00	36.00	26.00	24.00
11	00124	1	62.70	2	1	1	1	7	30.00	31.11	20.00	11.00
12	00129	1	49.58	1	1	1	1	7	20.00	41.00	10.00	12.00
13	00133	0	28.35	1	1	1	1	7	30.00	34.00	20.00	15.00
14	00141	0	34.45	1	1	1	1	1	17.00	23.00	29.00	44.00
15	00155	1	53.95	2	1	1	1	7	15.00	34.00	15.00	24.00
16	00165	0	51.90	1	1	1	1	7	35.00	50.00	10.00	10.00
17	00195	0	31.97	2	1	1	1	5	31.00	36.00	14.00	16.00
18	00226	0	30.14	2	5	1	1	6	.	20.00	13.00	17.00
19	00234	0	41.46	1	4	1	1	5	30.00	22.00	20.00	31.00
20	00235	1	29.98	1	1	1	1	7	32.00	42.00	15.00	10.00
21	00252	1	53.95	1	1	1	1	6	19.00	28.89	19.00	20.00
22	00255	0	37.10	1	1	1	1	6	16.00	33.00	24.00	24.00
23	00261	0	29.52	1	1	1	1	6	28.00	28.00	21.00	22.00
24	00276	0	42.13	1	2	1	1	7	30.00	35.00	16.00	20.00
25	00282	1	57.52	1	1	1	1	3	12.00	12.00	31.11	35.00
26	00294	1	31.17	2	1	1	1	7	29.00	38.00	15.00	17.00
27	00295	0	53.47	2	1	1	1	6	30.00	33.00	10.00	11.00
28	00304	0	24.52	1	5	1	2	6	22.00	36.00	10.00	12.00
29	00305	0	45.32	1	3	1	1	6	9.00	17.00	17.00	22.00

Data View | Variable View

SPSS Processor is ready

Scales of measurement (G& W, chap 1.4)

Ordinal: numbers indicate an order (e.g. one response is greater than another) but give no idea of the size of the difference

Example:
Mental health study
Amphetamine use
1 = Never use
2 = Irregular use
3 = Daily use

28th participant (ID no 00304) uses irregularly

*dart.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

8 : id 00104

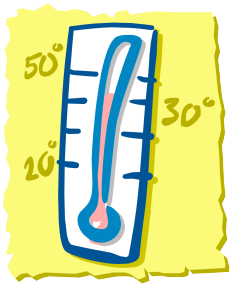
	id	gender	age	alcohol	cigs	marij	amphet	employ	swl	pa	na	kessler
1	00001	0	21.02	1	1	1	1	1	29.00	26.00	32.00	39.00
2	00010	1	49.15	1	2	1	1	7	28.00	43.00	13.00	14.00
3	00034	0	53.41	1	4	1	1	3	14.00	42.00	11.00	13.00
4	00038	1	22.67	1	4	1	1	1	26.00	42.00	13.00	19.00
5	00041	0	39.62	2	1	1	1	6	19.00	36.00	27.00	28.00
6	00059	0	51.69	1	1	1	1	5	29.00	50.00	14.00	10.00
7	00085	0	50.76	1	2	1	1	5	27.00	18.00	11.00	15.00
8	00104	0	56.60	2	1	1	1	6	26.00	21.00	10.00	11.00
9	00109	0	23.10	1	3	1	1	6	17.00	38.00	25.00	13.00
10	00110	1	67.18	2	1	1	1	4	7.00	36.00	26.00	24.00
11	00124	1	62.70	2	1	1	1	7	30.00	31.11	20.00	11.00
12	00129	1	49.58	1	1	1	1	7	20.00	41.00	10.00	12.00
13	00133	0	28.35	1	1	1	1	7	30.00	34.00	20.00	15.00
14	00141	0	34.45	1	1	1	1	1	17.00	23.00	29.00	44.00
15	00155	1	53.95	2	1	1	1	7	15.00	34.00	15.00	24.00
16	00165	0	51.90	1	1	1	1	7	35.00	50.00	10.00	10.00
17	00195	0	31.97	2	1	1	1	5	31.00	36.00	14.00	16.00
18	00226	0	30.14	2	5	1	1	6	.	20.00	13.00	17.00
19	00234	0	41.46	1	4	1	1	5	30.00	22.00	20.00	31.00
20	00235	1	29.98	1	1	1	1	7	32.00	42.00	15.00	10.00
21	00252	1	53.95	1	1	1	1	6	19.00	28.89	19.00	20.00
22	00255	0	37.10	1	1	1	1	6	16.00	33.00	24.00	24.00
23	00261	0	29.52	1	1	1	1	6	28.00	28.00	21.00	22.00
24	00276	0	42.13	1	2	1	1	7	30.00	35.00	16.00	20.00
25	00282	1	57.52	1	1	1	1	3	12.00	12.00	31.11	35.00
26	00294	1	31.17	2	1	1	1	7	29.00	38.00	15.00	17.00
27	00295	0	53.47	2	1	1	1	6	30.00	33.00	10.00	11.00
28	00304	0	24.52	1	5	1	2	6	22.00	36.00	10.00	12.00
29	00305	0	45.32	1	3	1	1	6	9.00	17.00	17.00	22.00

Data View | Variable View

SPSS Processor is ready

Scales of measurement (G& W, chap 1.4)

Interval: numbers indicate an actual amount, with equal unit of measurements separating two scores but not a real zero



Example:
Mental health study
Satisfaction with life scale
Measured by questionnaire on a scale from 5 – 35
Zero is meaningless

21st participant (ID no 00252) has SWL score of 19

*dart.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

8 : id 00104

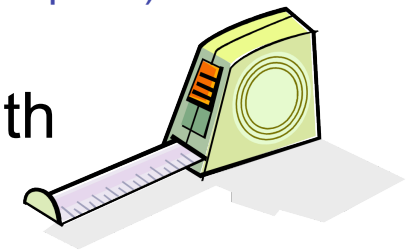
	id	gender	age	alcohol	cigs	marij	amphet	employ	swl	pa	na	kessler
1	00001	0	21.02	1	1	1	1	1	29.00	26.00	32.00	39.00
2	00010	1	49.15	1	2	1	1	7	28.00	43.00	13.00	14.00
3	00034	0	53.41	1	4	1	1	3	14.00	42.00	11.00	13.00
4	00038	1	22.67	1	4	1	1	1	26.00	42.00	13.00	19.00
5	00041	0	39.62	2	1	1	1	6	19.00	36.00	27.00	28.00
6	00059	0	51.69	1	1	1	1	5	29.00	50.00	14.00	10.00
7	00085	0	50.76	1	2	1	1	5	27.00	18.00	11.00	15.00
8	00104	0	56.60	2	1	1	1	6	26.00	21.00	10.00	11.00
9	00109	0	23.10	1	3	1	1	6	17.00	38.00	25.00	13.00
10	00110	1	67.18	2	1	1	1	4	7.00	36.00	26.00	24.00
11	00124	1	62.70	2	1	1	1	7	30.00	31.11	20.00	11.00
12	00129	1	49.58	1	1	1	1	7	20.00	41.00	10.00	12.00
13	00133	0	28.35	1	1	1	1	7	30.00	34.00	20.00	15.00
14	00141	0	34.45	1	1	1	1	1	17.00	23.00	29.00	44.00
15	00155	1	53.95	2	1	1	1	7	15.00	34.00	15.00	24.00
16	00165	0	51.90	1	1	1	1	7	35.00	50.00	10.00	10.00
17	00195	0	31.97	2	1	1	1	5	31.00	36.00	14.00	16.00
18	00226	0	30.14	2	5	1	1	6	.	20.00	13.00	17.00
19	00234	0	41.46	1	4	1	1	5	30.00	22.00	20.00	31.00
20	00235	1	29.98	1	1	1	1	7	[illegible]	42.00	15.00	10.00
21	00252	1	53.95	1	1	1	1	6	19.00	28.89	19.00	20.00
22	00255	0	37.10	1	1	1	1	6	[illegible]	33.00	24.00	24.00
23	00261	0	29.52	1	1	1	1	6	28.00	28.00	21.00	22.00
24	00276	0	42.13	1	2	1	1	7	30.00	35.00	16.00	20.00
25	00282	1	57.52	1	1	1	1	3	12.00	12.00	31.11	35.00
26	00294	1	31.17	2	1	1	1	7	29.00	38.00	15.00	17.00
27	00295	0	53.47	2	1	1	1	6	30.00	33.00	10.00	11.00
28	00304	0	24.52	1	5	1	2	6	22.00	36.00	10.00	12.00
29	00305	0	45.32	1	3	1	1	6	9.00	17.00	17.00	22.00

Data View | Variable View

SPSS Processor is ready

Scales of measurement (G& W, chap 1.4)

Ratio: numbers indicate an actual amount, with a real zero.



Example:
Mental health study
Age in years
0 means just born

17th participant
(ID no 00195)
is 32 years old

*dart.sav [DataSet1] - SPSS Data Edi

File Edit View Data Transform Analyze ... Utilities Window Help

8 : id 00104

	id	gender	age	alcohol	cigs	marij	amphet	employ	swl	pa	na	kessler
1	00001	0	21.02	1	1	1	1	1	29.00	26.00	32.00	39.00
2	00010	1	49.15	1	2	1	1	7	28.00	43.00	13.00	14.00
3	00034	0	53.41	1	4	1	1	3	14.00	42.00	11.00	13.00
4	00038	1	22.67	1	4	1	1	1	26.00	42.00	13.00	19.00
5	00041	0	39.62	2	1	1	1	6	19.00	36.00	27.00	28.00
6	00059	0	51.69	1	1	1	1	5	29.00	50.00	14.00	10.00
7	00085	0	50.76	1	2	1	1	5	27.00	18.00	11.00	15.00
8	00104	0	56.60	2	1	1	1	6	26.00	21.00	10.00	11.00
9	00109	0	23.10	1	3	1	1	6	17.00	38.00	25.00	13.00
10	00110	1	67.18	2	1	1	1	4	7.00	36.00	26.00	24.00
11	00124	1	62.70	2	1	1	1	7	30.00	31.11	20.00	11.00
12	00129	1	49.58	1	1	1	1	7	20.00	41.00	10.00	12.00
13	00133	0	28.35	1	1	1	1	7	30.00	34.00	20.00	15.00
14	00141	0	34.45	1	1	1	1	1	17.00	23.00	29.00	44.00
15	00155	1	53.95	2	1	1	1	7	15.00	34.00	15.00	24.00
16	00165	0	51.96	1	1	1	1	7	35.00	50.00	10.00	10.00
17	00195	0	31.97	2	1	1	1	5	31.00	36.00	14.00	16.00
18	00226	0	30.14	2	5	1	1	6	.	20.00	13.00	17.00
19	00234	0	41.46	1	4	1	1	5	30.00	22.00	20.00	31.00
20	00235	1	29.98	1	1	1	1	7	32.00	42.00	15.00	10.00
21	00252	1	53.95	1	1	1	1	6	19.00	28.89	19.00	20.00
22	00255	0	37.10	1	1	1	1	6	16.00	33.00	24.00	24.00
23	00261	0	29.52	1	1	1	1	6	28.00	28.00	21.00	22.00
24	00276	0	42.13	1	2	1	1	7	30.00	35.00	16.00	20.00
25	00282	1	57.52	1	1	1	1	3	12.00	12.00	31.11	35.00
26	00294	1	31.17	2	1	1	1	7	29.00	38.00	15.00	17.00
27	00295	0	53.47	2	1	1	1	6	30.00	33.00	10.00	11.00
28	00304	0	24.52	1	5	1	2	6	22.00	36.00	10.00	12.00
29	00305	0	45.32	1	3	1	1	6	9.00	17.00	17.00	22.00

Data View | Variable View

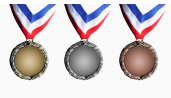
SPSS Processor is ready

Discrete and Continuous Variables

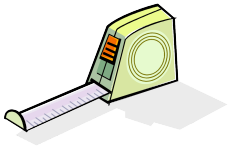
- Scores that comprise data from a research study come from observation and measuring variables
- Variables in a study can be categorized by the types of values assigned to them
- Today, we will consider two types of variables:
  1. Discrete variables
  2. Continuous variables

Discrete & Continuous Variables

Discrete — only whole number amounts



Continuous — allows for numbers to be fractions or decimals (at least in principle)



Amphetamine use:
Discrete

Satisfaction with life:
Continuous

*dart.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

8 : id 00104

	id	gender	age	alcohol	cigs	marij	amphet	employ	swl	pa	na	kessler
1	00001	0	21.02	1	1	1	1	1	29.00	26.00	32.00	39.00
2	00010	1	49.15	1	2	1	1	7	28.00	43.00	13.00	14.00
3	00034	0	53.41	1	4	1	1	3	14.00	42.00	11.00	13.00
4	00038	1	22.67	1	4	1	1	1	26.00	42.00	13.00	19.00
5	00041	0	39.62	2	1	1	1	6	19.00	36.00	27.00	28.00
6	00059	0	51.69	1	1	1	1	5	29.00	50.00	14.00	10.00
7	00085	0	50.76	1	2	1	1	5	27.00	18.00	11.00	15.00
8	00104	0	56.60	2	1	1	1	6	26.00	21.00	10.00	11.00
9	00109	0	23.10	1	3	1	1	6	17.00	38.00	25.00	13.00
10	00110	1	67.18	2	1	1	1	4	7.00	36.00	26.00	24.00
11	00124	1	62.70	2	1	1	1	7	30.00	31.11	20.00	11.00
12	00129	1	49.58	1	1	1	1	7	20.00	41.00	10.00	12.00
13	00133	0	28.35	1	1	1	1	7	30.00	34.00	20.00	15.00
14	00141	0	34.45	1	1	1	1	1	17.00	23.00	29.00	44.00
15	00155	1	53.95	2	1	1	1	7	15.00	34.00	15.00	24.00
16	00165	0	51.90	1	1	1	1	7	35.00	50.00	10.00	10.00
17	00195	0	31.97	2	1	1	1	5	31.00	36.00	14.00	16.00
18	00226	0	30.14	2	5	1	1	6	.	20.00	13.00	17.00
19	00234	0	41.46	1	4	1	1	5	30.00	22.00	20.00	31.00
20	00235	1	29.98	1	1	1	1	7	32.00	42.00	15.00	10.00
21	00252	1	53.95	1	1	1	1	6	19.00	28.89	19.00	20.00
22	00255	0	37.10	1	1	1	1	6	16.00	33.00	24.00	24.00
23	00261	0	29.52	1	1	1	1	6	28.00	28.00	21.00	22.00
24	00276	0	42.13	1	2	1	1	7	30.00	35.00	16.00	20.00
25	00282	1	57.52	1	1	1	1	3	12.00	12.00	31.11	35.00
26	00294	1	31.17	2	1	1	1	7	29.00	38.00	15.00	17.00
27	00295	0	53.47	2	1	1	1	6	30.00	33.00	10.00	11.00
28	00304	0	24.52	1	5	1	2	6	22.00	36.00	10.00	12.00
29	00305	0	45.32	1	3	1	1	6	9.00	17.00	17.00	22.00

Data View / Variable View

SPSS Processor is ready

Sleep Questionnaire

- Let's revisit the sleep questionnaire
- What scales of measurement are used in each part?
- Are the variables being measured in each part discrete or continuous?

Distributions of data

Distributions of discrete variables

Frequency distribution for gender

A frequency distribution table
(G&W, chap 2.2)

Missing: number of respondents who did not complete this item on the questionnaire

gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	4286	56.3	56.8	56.8
	male	3256	42.8	43.2	100.0
	Total	7542	99.0	100.0	
Missing	999	73	1.0		
Total		7615	100.0		

Total sample of 7615

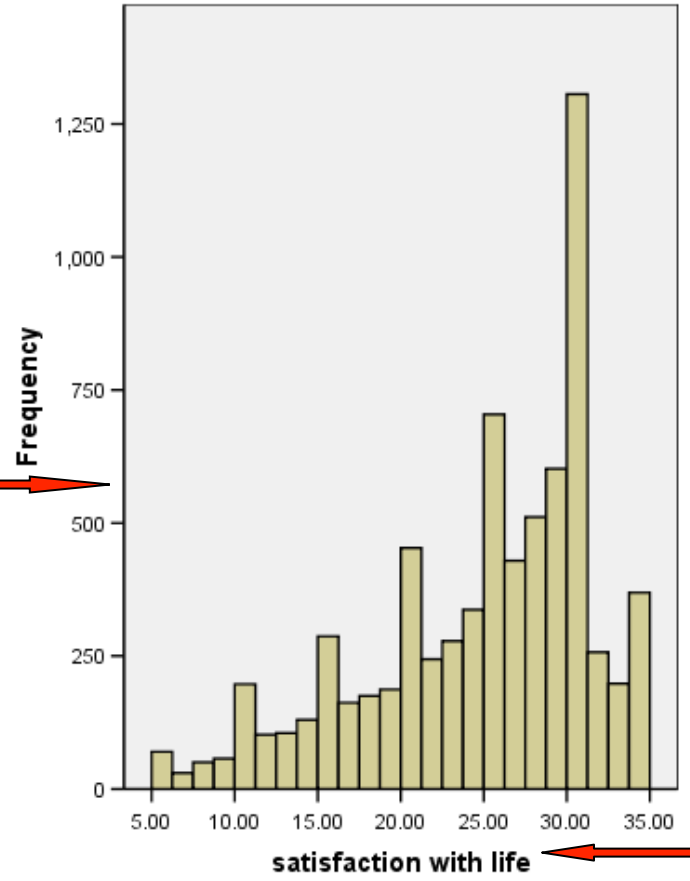
Of non-missing data, 57% and 43% females and males, respectively

Distributions of continuous variables

Distribution of scores on the satisfaction with life scale

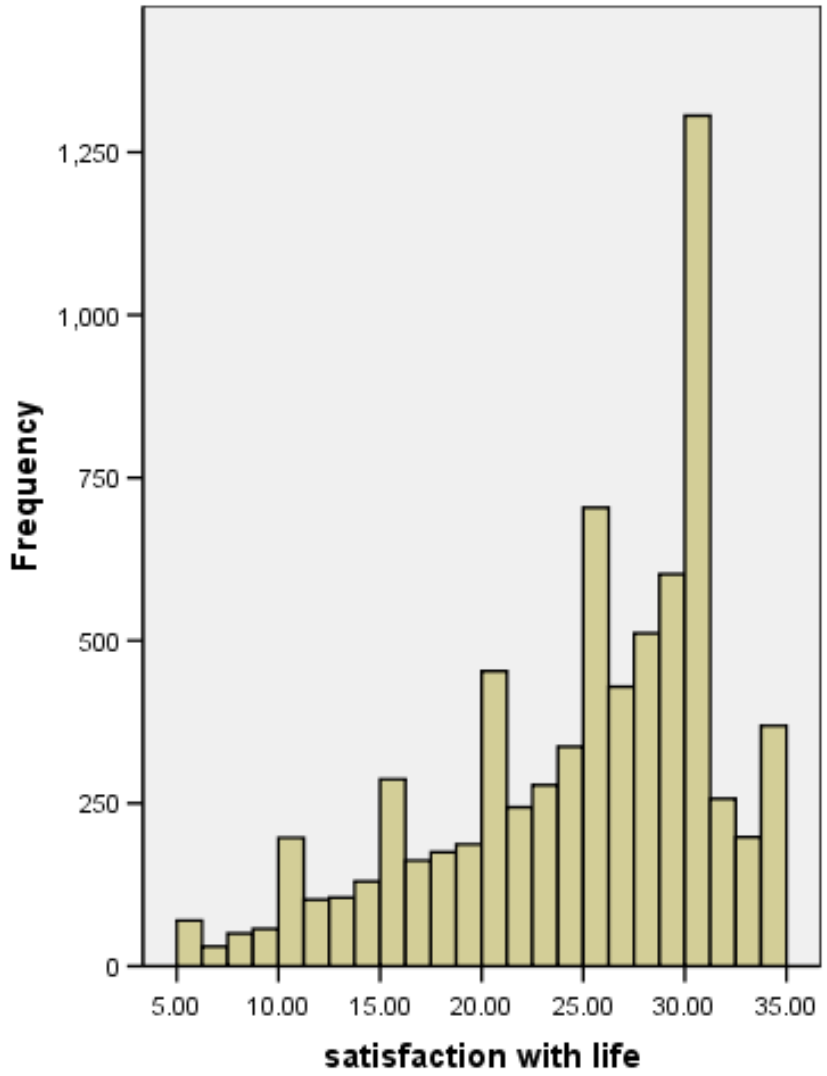
A histogram
(G&W, chap 2.3)

Frequency: number of respondents with a given score



Score on SWL scale

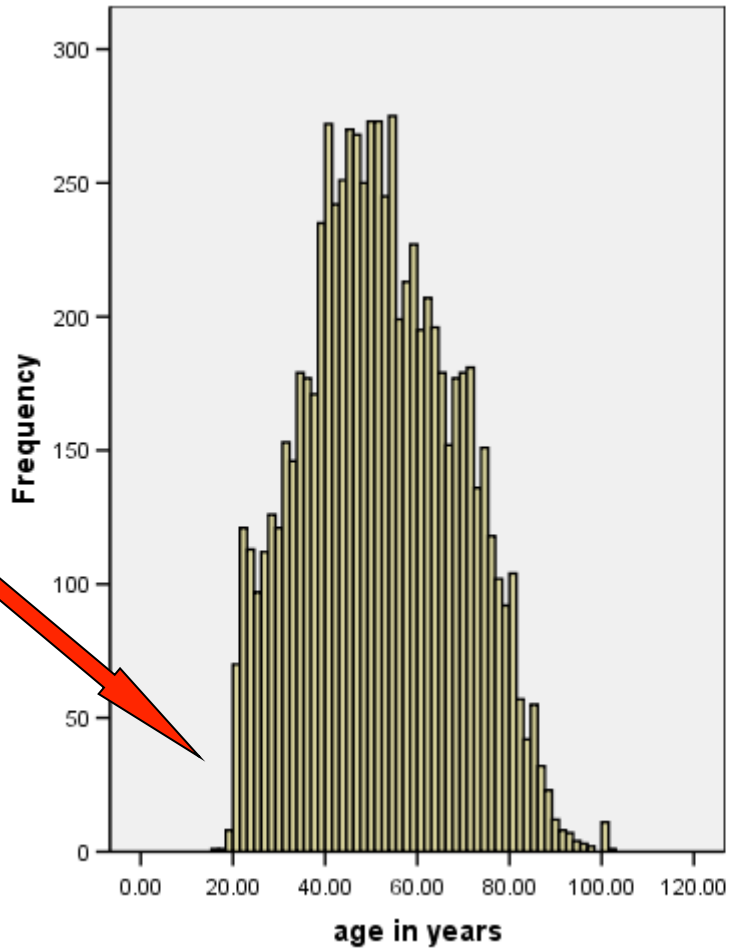
Satisfaction with life scale



Distributions of continuous variables

Distribution of age

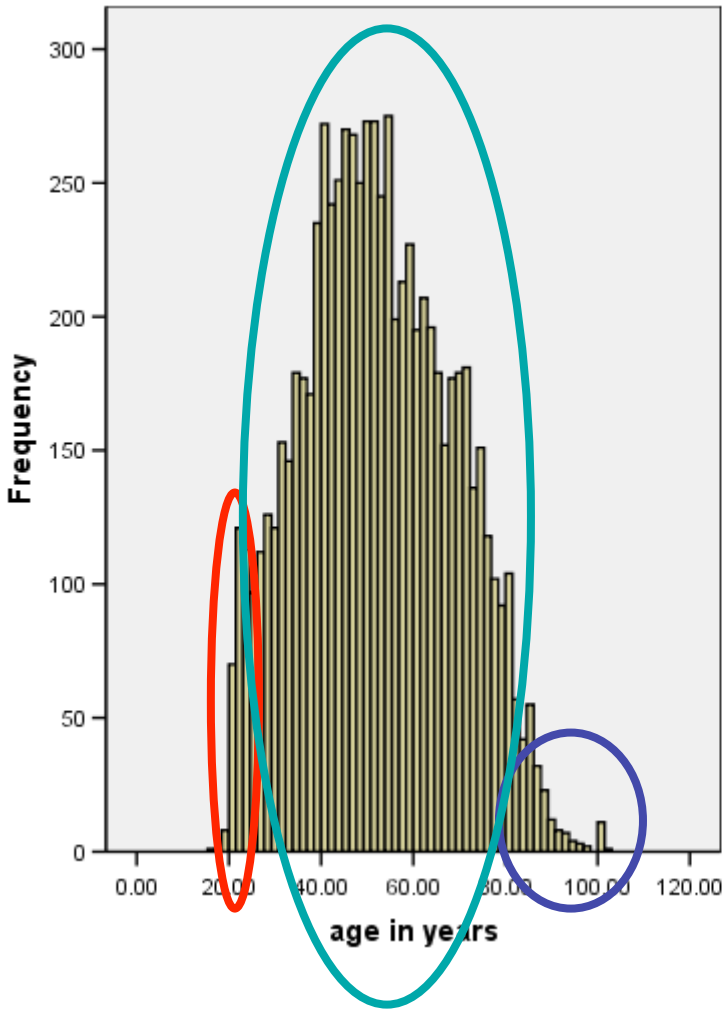
Note cutoff: participants less than 20 were not involved



Distributions of continuous variables

Distribution of age

Some participants are young adults



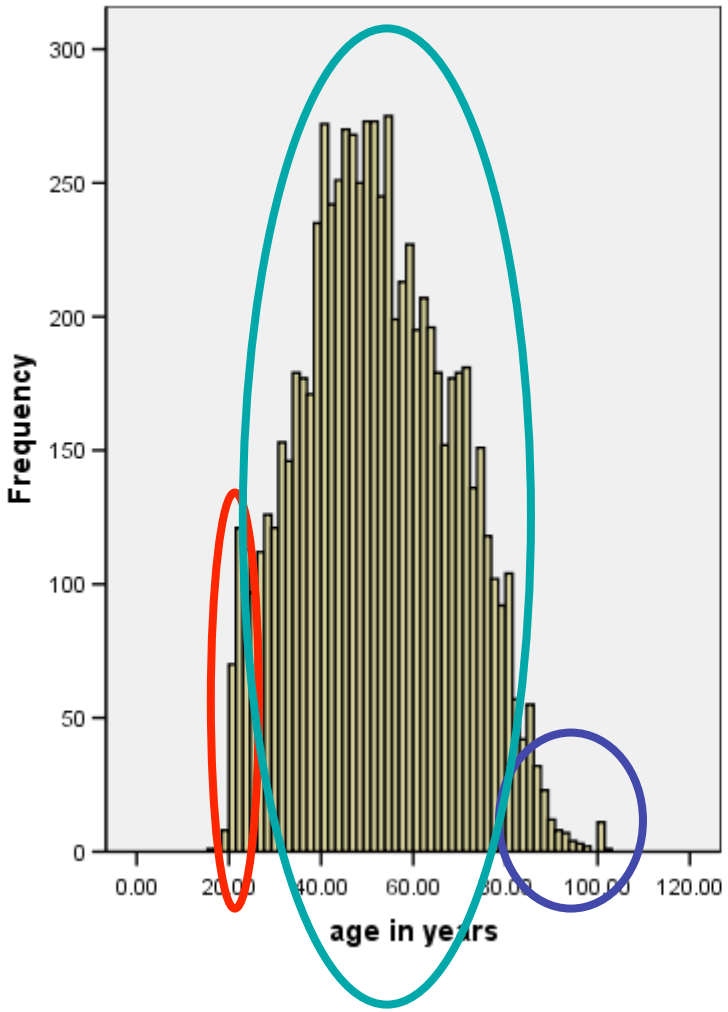
Some participants are very old adults

Most participants are neither young nor very old

Distributions of continuous variables

Distribution of age

In this sample it is unusual to have young adults (low probability)



It is unusual to have very old adults (low probability)

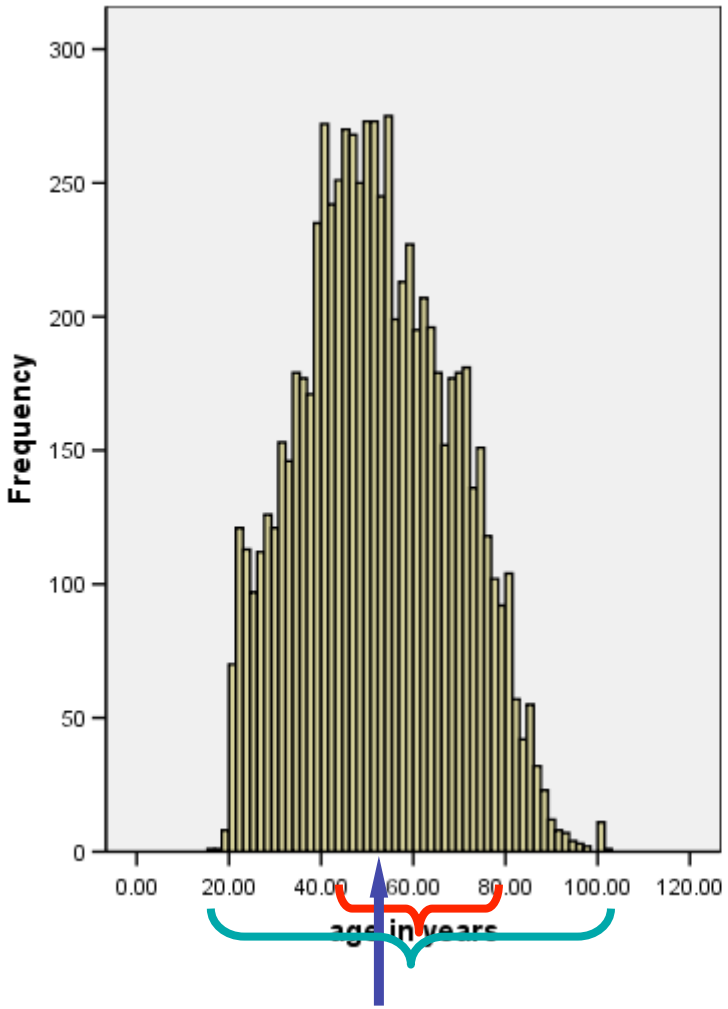
It is not unusual for participants to be neither young nor very old (high probability)

Distributions of continuous variables

Distribution of age

A typical age is in the centre of the distribution

(Scores in the centre of this distribution are not unusual and occur with high probability)



The average age is 52

But the ages range from 16 to 101 (the spread of a distribution)

Important features of distributions

- Central tendency: A score that represents the "typical" or "expected" score in a distribution.
  - Calculated as either:
    - the mean: the simple average
      - typically use with continuous data
    - the median: 50% of scores below and 50% above
      - typically use with discrete data
    - (rarely) the mode: the most common score
      - typically use with categorical data

  Gravetter & Wallnau, Chapter 3.

Some statistical notation for the mean

An individual score usually indicated by X

Mean of sample indicated by M

n … the number of people in a sample

Σ… "the sum of" (Greek capital letter Sigma

$$\sum X$$ the sum of all scores

$$M = \frac{\sum X}{n}$$ Mean equals the sum of all scores divided by n

Mean (Average) Sleep

- What was the mean number of hours that people in the class slept last night?

- To work this out we can use the mean formula:

$$M = \frac{\sum X}{n}$$

- Put differently, let's sum (Σ) everyone's sleep scores (x) and divide that total by the number of people in the class (n).

Important features of distributions

- Spread or variation: Humans differ!
- It is important to know how much the scores can vary.
  - Typical measures of variation:
    - the standard deviation:
      - use with the mean
    - the interquartile range:
      - use with the median
    - (also the range: highest score minus lowest score)

More later on how these measures are calculated: details are in Gravetter & Wallnau, Chapter 4.

Why are distributions important for psychology?

Take an example:

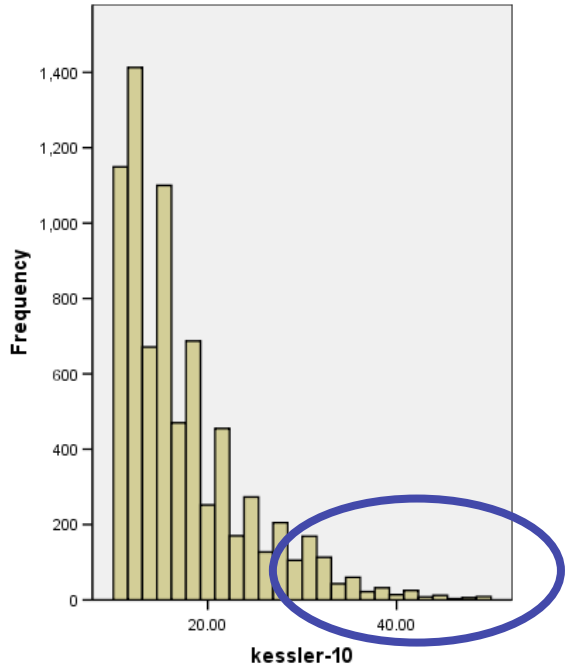
K10 in the mental health study:

The Kessler 10 is a measure of psychological distress and a high score (above 30) may indicate more serious problems.

Why are distributions important for psychology?

Distribution of K10 in the mental health study:

Mean = 17.4
Std.dev.= 6.7



Very high scores are unusual (not many of them).

What counts as very high? They are a long way from the mean (more than 2 standard deviations).

17.4+2×6.7 = 30.8

Very high (or extreme) scores indicate psychological distress – they are scores that are very different from "typical"or "expected" scores

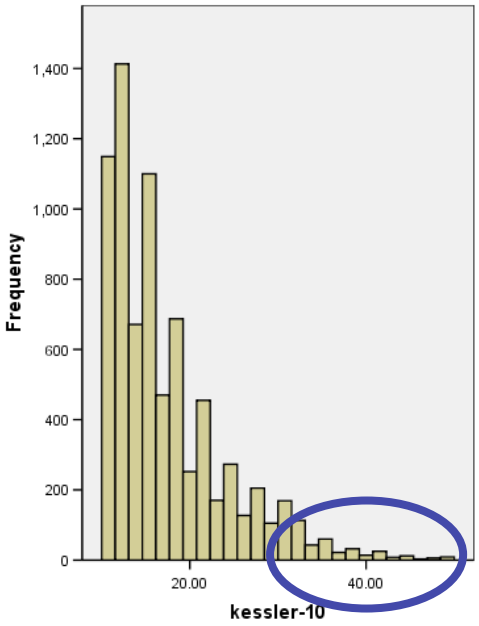
Why are distributions important for psychology?

In dealing with mental health (for instance), we need to know:

1. what is typical?
2. how do human beings vary?

We can use this information to help us identify people in need of assistance.

We need to know the distribution of scores, the central tendency and the variation.



Mean = 17.4
Std.dev.= 6.7

Distribution of arrival time in class

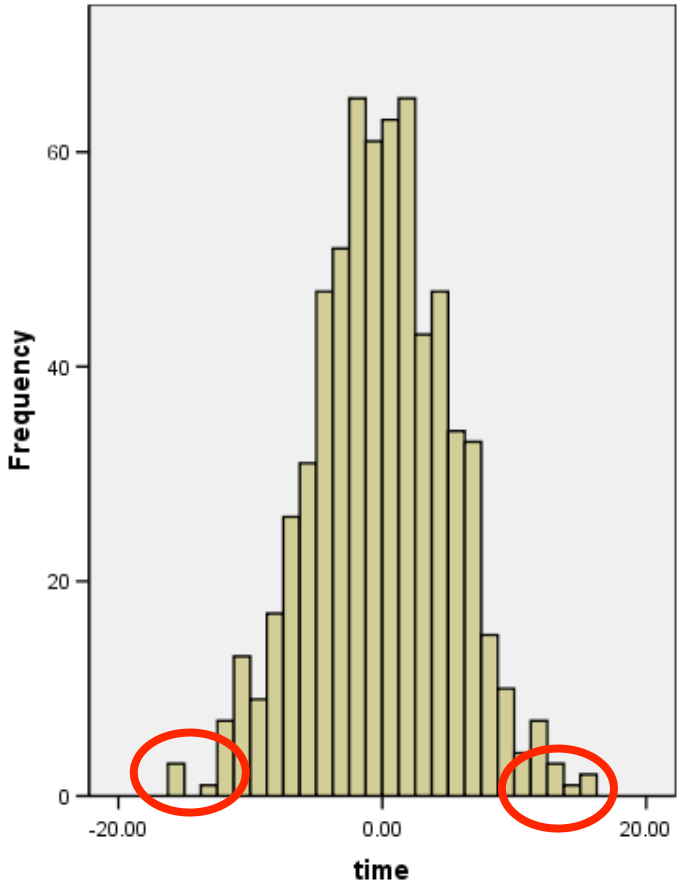
658 1st year students recorded their arrival time in a given lecture in terms of minutes arrived before the lecture start (negative) and minutes after lecture start (positive).

So those who arrived exactly on time had a score of 0.

Distribution of arrival time in class

Mean = - 0.1
Std. dev. = 5.2

Range is from
-16 to +16



Very early and very late attendees are rare.

What counts as "very early" or "very late"?

As with K10, let's check within 2 s.d.'s of the mean.

Mean ± 2×s.d
= -0.1 ± 2×5.2
i.e. from -10.5 to +10.3

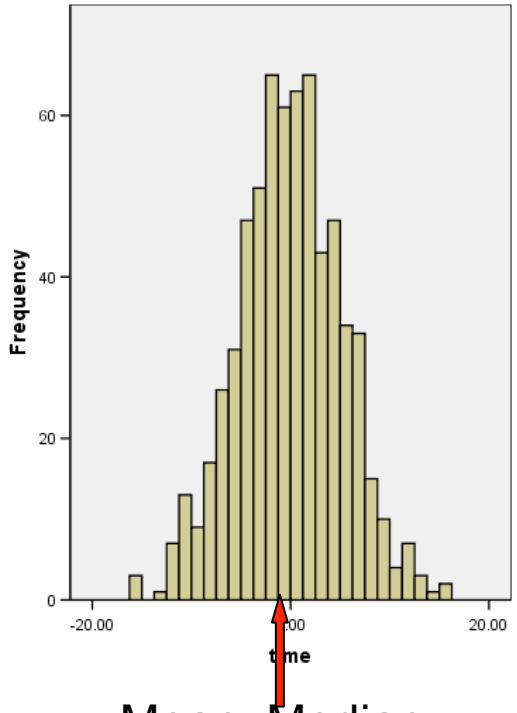
2.9% are very early (less than -10.5)
2.3% are very late (greater than +10.3).

Extreme cases are rare (low probability)
Extreme cases are different
Extreme cases are worthy of attention

Comparing the mean and the median

For symmetric distributions (eg arrival times)

Mean = Median

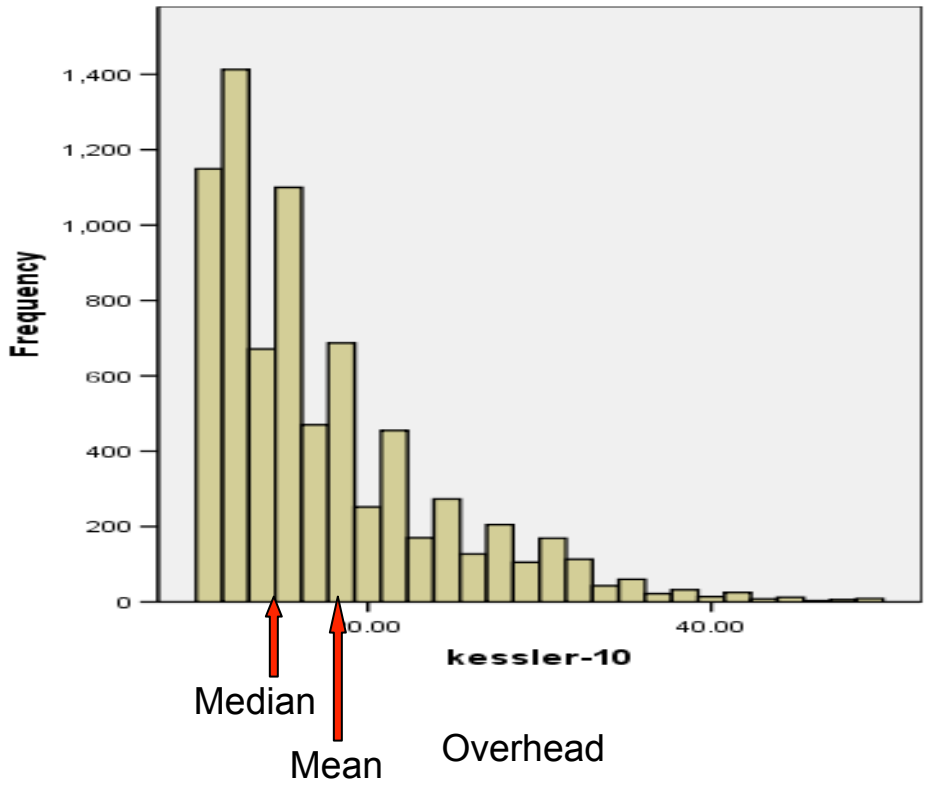


Comparing the mean and the median

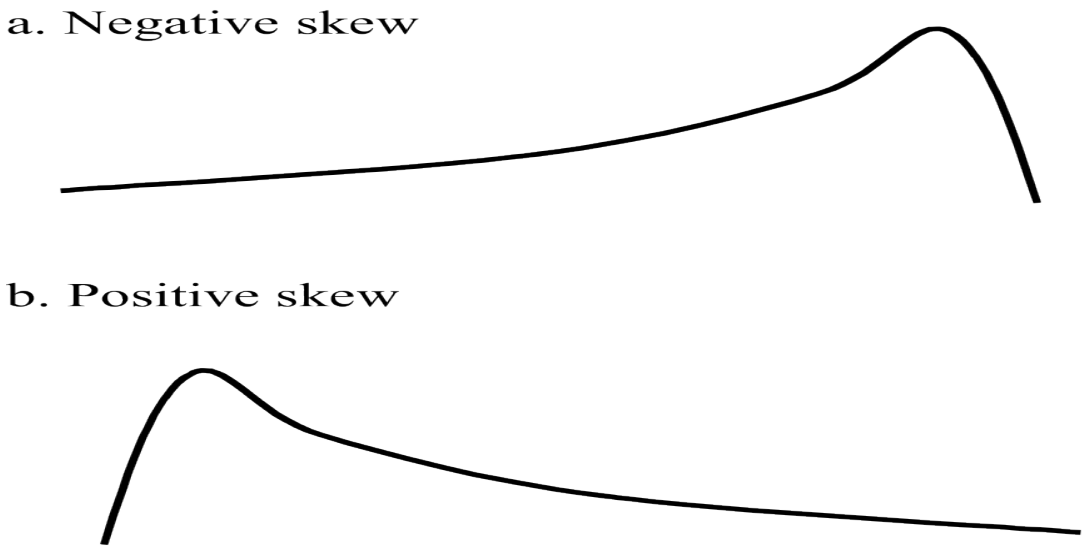
For non-symmetric distributions: (skewed such as K10)

Mean is sensitive to skewness and outliers

(see, e.g., Figure 3.14, G & W, chapter 3)



Skewed distributions



K10 is positively skewed

Outliers

- Observations that are an extremely long way from the rest of the data

Example: Seven people in a small business
(6 workers and boss)

20,000 20,000 20,000 20,000
20,000 20,000 100,000

"100,000" is an outlier because it is a long way from most other data values

Median = 20,000
Mean = 31,429

Other types of distributions

(G&W, Fig 3-13)

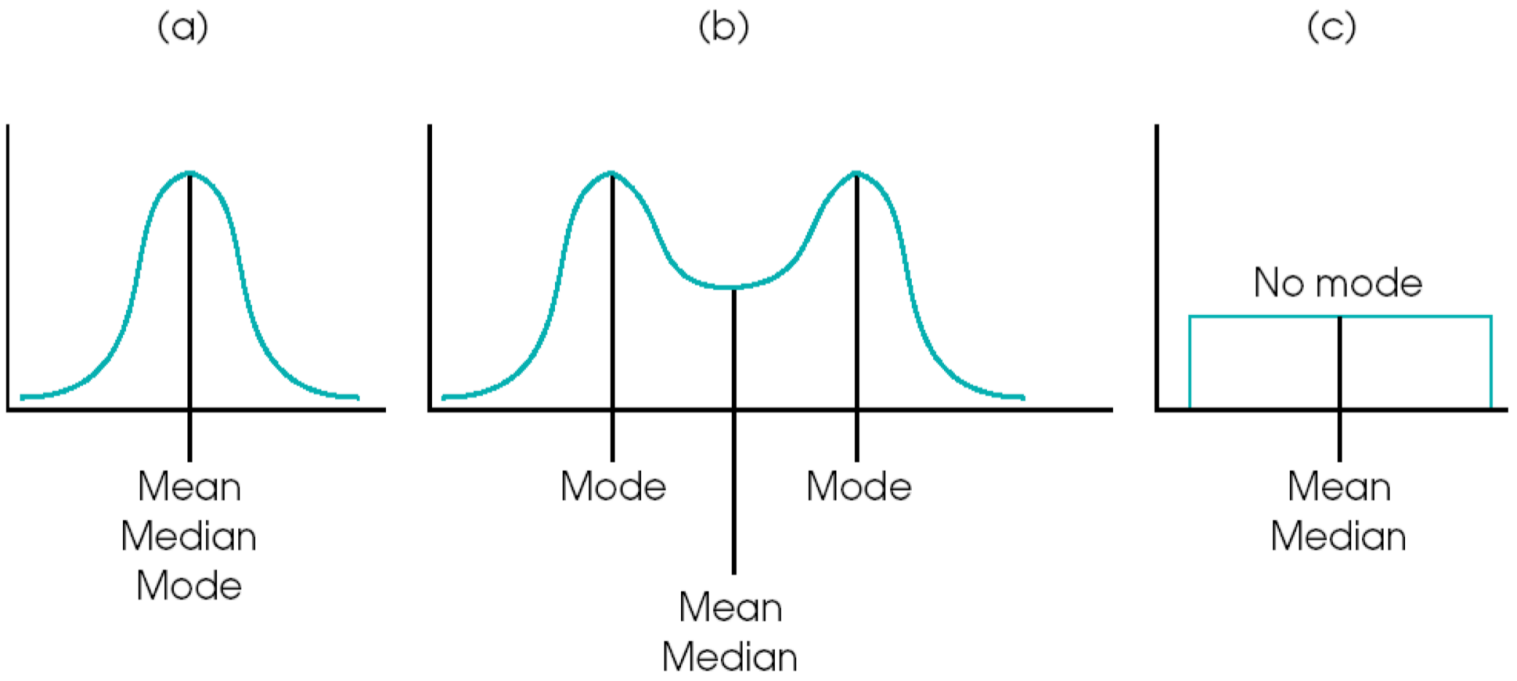


Figure 3-13 (p. 95)

Measures of central tendency for three symmetrical distributions: normal, bimodal, and rectangular.

Overhead

Problem set

Overhead

Linking summary

First session:

Research questions lead to hypotheses

Design the research (including issues of validity) in order to get data

Decide how to take a sample from the population

This session:

Measurement

Leads to collecting data

Beginning to understand what the data say through distributions and means

Overhead