


Statistics

an Overview




What We Will Cover in This Section



- What statistics are.
- Descriptive Statistics
 - Frequency distributions
 - Graphs
 - Mean
 - Standard deviation
- Inferential Statistics
 - Z-scores
 - Correlation



Statistics, an Overview 2



Basic Terms and Concepts



Basic Terminology

STATISTICS

Numerical techniques for describing groups of people or events.

Fundamental Uses

DESCRIPTIVE STATISTICS

Techniques used to organize, summarize, and describe sets of numbers.

INFERENCE STATISTICS

Techniques that allow us to make estimates about populations based on sample data.

Using Numbers

3.14159

Levels of Measurement

NOMINAL SCALE

Numbers are used as labels.

ORDINAL SCALE

Numbers are used to indicate rank order.

Levels of Measurement

INTERVAL SCALE

Numbers are used to indicate an actual amount and there is an equal unit of measurement between adjacent numbers.

RATIO SCALE

Numbers indicate an actual amount and there is a true zero.

Variables

DISCRETE VARIABLE

A variable that can take on only whole values.

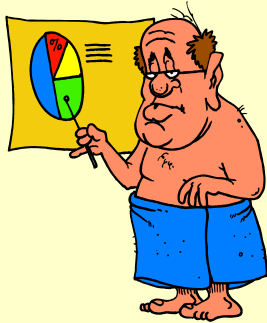
Example: Number of toes or number of cars you have.

CONTINUOUS VARIABLE

A variable that can take on fractional values.

Example: Speedometer reading, height.

Frequency Tables and Graphs

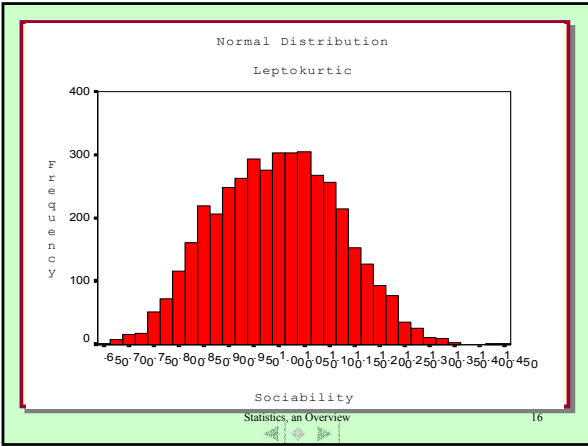


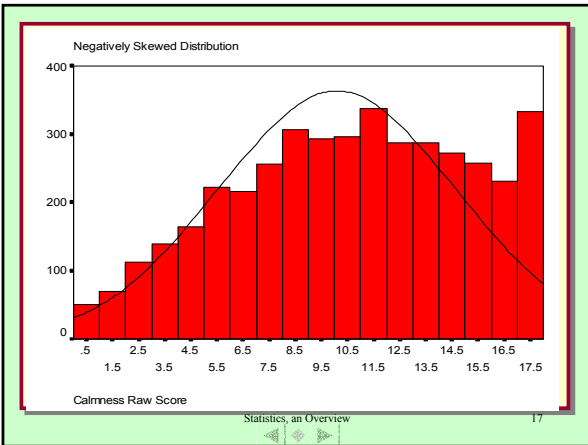
Common Statistics

- Frequency
 - The number of people who got a certain score.
 - Symbolized with f .
- Number
 - The total count of observations in a sample.
 - Symbolized with N .
- Percentile
 - The percent of people who got a score and lower.
 - Symbolized with P_n .

Simple frequency distribution (N=20)

Score	Frequency (f)	%	Cum % (Percentile)
17	1	5	100
16	0	0	95
15	4	20	95
14	5	25	75
13	4	20	50
12	3	15	30
11	2	10	15
10	1	5	5





Measures
of
the
Middle

Statistics, an Overview 18

Question

What number would you use to describe the typical height of people in this class?



Mean

- Sum the scores and divide by the number of scores.
- Symbols
 - Sample: M or \bar{X}
 - Population: μ

Median

- The score below which 50% of the scores fall.
- Represents P_{50} .
- Divides the distribution in half.
- Symbol.
 - Sample: Mdn

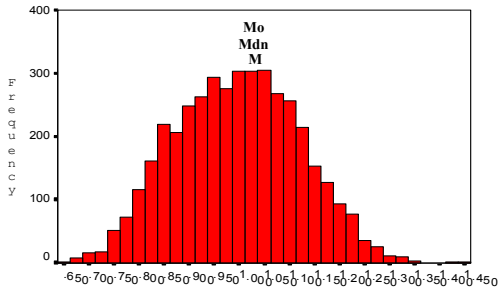
Example

8 9 10 11 12 13 14 15 16

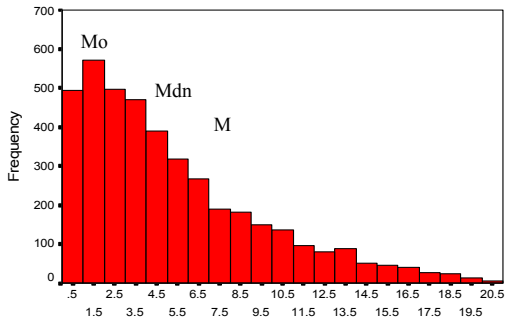
8 9 10 11 12 13 16 16 46

Normal Distribution

Leptokurtic



Positively Skewed Distribution



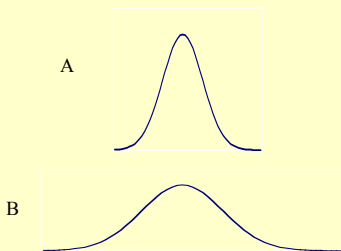
Measures of Variability



Overview

The Mean describes the 'typical' score; measures of variability give an index of how much the rest of the scores in the distribution are spread out around the mean.

Two Normal Distributions with the Same Mean



Range

- The distance between the lowest and highest score.
- Formula
Range = Highest Score – Lowest Score + 1
- Example

1 3 4 6 8 12 15 16 18 19
1 3 4 6 8 12 15 16 18 79

Deviation Score

Score	$X - M_x$	$(X - M_x)^2$
5	-2.5	6.25
6	-1.5	2.25
7	-.5	.25
8	.5	.25
9	1.5	2.25
10	2.5	6.25

Sum 45 0 **17.50**
Mean 7.5 0 **2.92**

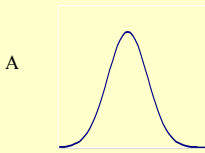
Variance

Mean squared
deviation score
around the mean.

Standard Deviation

Square root of
the variance.

Distributions with Different Variability

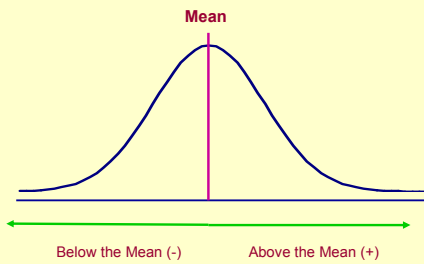


Standard Scores (Z-scores)

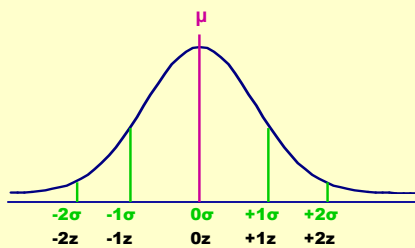
Standard Score (z-score)

The purpose of the STANDARD SCORE is to describe the location of every score in a distribution relative to the mean.

Concept



Units of Measurement



Equations

Defining
Equation

$$z = \frac{x - \mu}{\sigma}$$

Working
Equation

$$z = \frac{x - \bar{X}}{S}$$

Examples

Mean = 50 SD = 10

What is the z score for
a raw score of 65?

$$Z = (65 - 50)/10$$

$$Z = 15/10$$

$$Z = 1.5$$

What is the z score for
a raw score of 45?

$$Z = (45 - 50)/10$$

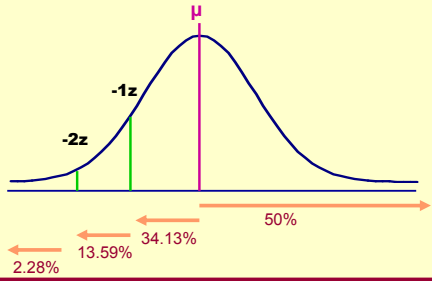
$$Z = -5/10$$

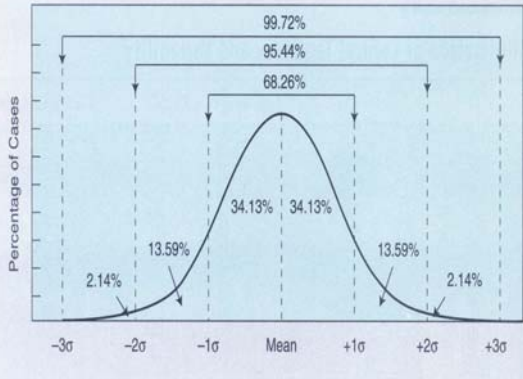
$$Z = -.5$$

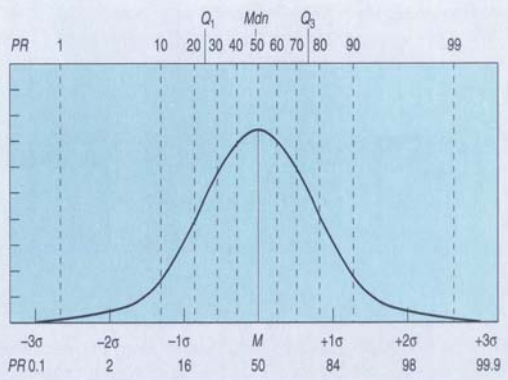
Uses of the z-score

- Comparing different people on the same test.
- Comparing same person across different measures.
- Comparing different people across different tests.

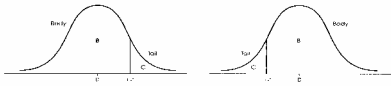
Slicing the Normal Curve







Areas Under the Normal Curve (p 690)

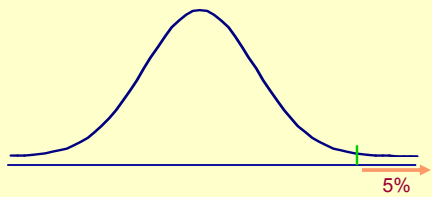


(A) z	(B) PROPORTION IN BODY	(C) PROPORTION IN TAIL	(A) z	(B) PROPORTION IN BODY	(C) PROPORTION IN TAIL	(A) z	(B) PROPORTION IN BODY	(C) PROPORTION IN TAIL
0.00	.5000	.5000	0.20	.5793	.4207	0.40	.6554	.3446
0.01	.5039	.4961	0.21	.5832	.4168	0.41	.6591	.3409
0.02	.5079	.4921	0.22	.5871	.4129	0.42	.6628	.3372
0.03	.5119	.4881	0.23	.5910	.4089	0.43	.6664	.3336
0.04	.5159	.4841	0.24	.5948	.4052	0.44	.6700	.3300
0.05	.5199	.4801	0.25	.5987	.4013	0.45	.6736	.3264
0.06	.5239	.4761	0.26	.6026	.3974	0.46	.6772	.3228
0.07	.5279	.4721	0.27	.6064	.3936	0.47	.6808	.3192
0.08	.5319	.4681	0.28	.6103	.3897	0.48	.6844	.3156
0.09	.5359	.4641	0.29	.6141	.3859	0.49	.6879	.3121
0.10	.5398	.4602	0.30	.6179	.3821	0.50	.6915	.3085
0.11	.5438	.4562	0.31	.6217	.3783	0.51	.6950	.3050
0.12	.5478	.4523	0.32	.6255	.3745	0.52	.6985	.3015
0.13	.5517	.4483	0.33	.6293	.3707	0.53	.7019	.2981
0.14	.5557	.4443	0.34	.6331	.3669	0.54	.7054	.2946
0.15	.5596	.4404	0.35	.6368	.3632	0.55	.7088	.2912
0.16	.5636	.4364	0.36	.6406	.3594	0.56	.7123	.2877
0.17	.5675	.4325	0.37	.6443	.3557	0.57	.7157	.2843
0.18	.5714	.4286	0.38	.6480	.3520	0.58	.7190	.2810
0.19	.5753	.4247	0.39	.6517	.3483	0.59	.7224	.2776

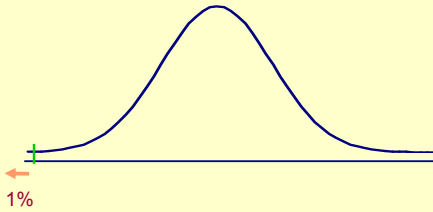
Key Probabilities

Above what z-score do 5% of the cases fall?	1.64
Below what z-score does 1% of the cases fall?	2.33
Between which two z-scores do 95% of the cases fall?	± 1.96
Between which two z-scores do 99% of the cases fall?	± 2.58

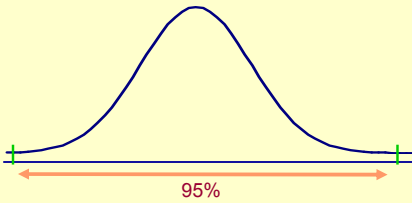
Above what z-score do 5% of the cases fall?



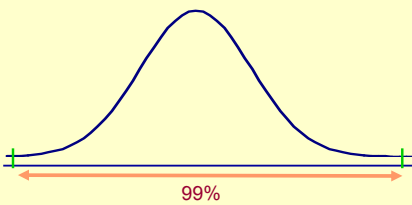
Below what z-score does 1% of the cases fall?



Between what two z-scores do 95% of the cases fall?



Between what two z-scores do 99% of the cases fall?



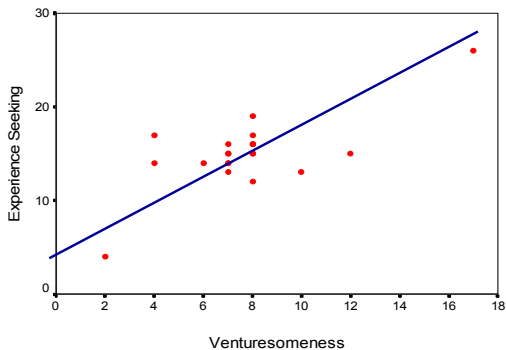
Measure of Association The Correlation



What Correlational Statistics Do

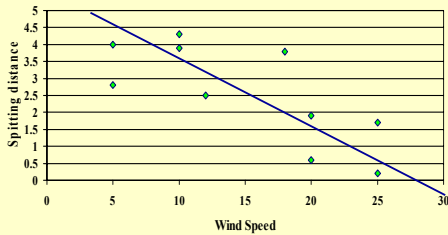
1. Assess the strength of the relationship between two or more variables.
2. Determine the direction of the relationship.
 - Positive.
 - Negative.

Positive correlation example ($r=.716$)

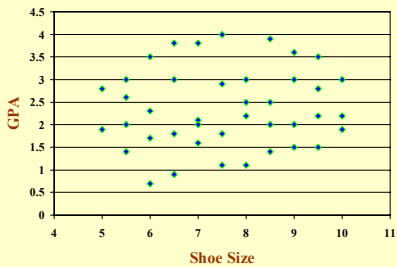


Negative Correlation

CUP Wind Spitting Olympiad



Zero Correlation Example



The Correlation Coefficient

- Indicated by r .
- Ranges from -1.00 to $+1.00$
 - The number indicates the strength of the relationship.
 - The sign indicates whether the relationship is positive or negative.
- Does NOT indicate causality.

Measuring the Correlation

Coefficient	Strength
.60 to 1.00 -.60 to -1.00	Very strong
.40 to .59 -.40 to -.59	Moderate
.20 to .39 -.20 to -.39	Weak
-.19 to +.19	Very weak

Where Simple Correlations are Used

1. Prediction.
2. Validation studies.
3. Reliability studies.
4. Theoretical studies.
5. Identification of surrogate variables.

