$$\sigma=( (1-p)/(p^2) )^{1/2} = \left( \frac{1-p}{p^2} \right)^{1/2}$$

1. **Course Logistics**

2. **Guest: Professor D. Maynard (Stats Tutorial)**

3. **Prologue: A Basic mathematical review**

4. **Chapter 1:**

The role of statistics in the research process
Statistical applications
Types of variables

# Instructor

- **Prof**. Don Kerr
- **Office: Labatt Hall Building: LH 208**
- **Office Hours:**

  Mondays (2:30-3:30 p.m.)
- Please note this hour is drop-in – no need to make an appointment …

# Instructor

- **Phone:** 519-433-3491
  - not at all a good method of communication

- **E-mail:** dkerr@uwo.ca
  - Use this to contact me for any 'official' reason
    - **Ex. You are about to miss a test due to illness**
  - to make an appointment to see me if you can't drop into my listed hours
  - For extremely short questions
    - Otherwise just drop in

# Required Readings

- The following text is a required resource for this course.
- Healey/Prus/Lieflander. Statistics: A Tool for Social Research, **4th Canadian edition**

# MY COURSE WEBSITE (I don't use OWL very much):

- ***Includes***:
  - Professor/class info
  - Course outline
  - Lectures
  - Assignments
  - Grades
  - Extra documents you need like appendices and formula sheets

- http://dkerr.kingsfaculty.ca/

# Other bits ....

Monday 11:30 am class … please watch lateness …

- Copying other students on assignments, cheating on tests etc.
  - Obviously don't do these things
  - See course outline

- Electronics use
  - No hand-held devices in the classroom please
  - Laptop use should be appropriate (no social media, videogames etc.)
  - Do not video/audio record this class or your fellow students without direct consent

  - **MY SUGGESTION: PRINT UP THE COURSE LECTURE AHEAD OF TIME EACH WEEK, AND BRING IT TO CLASS (FOR NOTE TAKING)**

Course Requirement:

Problem solving assignments (50% of course grade)

#1 Sept 30th
#2 Oct 14th
#3 Oct 21st
#4 Nov 18th
#5 Dec 2nd

Late assignments (5% daily, not including Saturdays & Sundays)

Midterm exam (25% of course grade) – Oct 28th

Final exam (25% of course grade) – Final examination period

Note: King's Academic Policy

If you require any form of accommodation for a test and/or assignment (e.g. due to a sickness or similar event), please let me know, as soon as possible (ASAP).

If you make a "formal request" to the university -you must be in touch with me ASAP (via email: dkerr@uwo.ca)

- **AVAILABLE THIS YEAR:**

- **Sociology Statistics Review/Help sessions for students registered in SOC2205**

- **NOTE: Completely optional; non-mandatory – starting next week, i.e. the 2nd week of the term**

- **TIME (To be announced)**

- **Hour of tutoring with Statistics professor: Prof. Donna Maynard**
- **Weekly review of course material and 'questions of the week'.**

- **THIS is where you can get help with your assignments!!!**

# Chapters 1 & 2: In this presentation you will learn about:

- **Prologue: A Basic mathematical review**
- A word on rounding variables

- **CHAPTER 1:**
- The role of statistics in the research process
- Statistical applications
- Types of variables

# Some basic math

$$\sigma=( (1-p)/(p^2) )^{1/2} = \left( \frac{1-p}{p^2} \right)^{1/2}$$

**1.** Multiplication

All of the below imply the exact same thing (MULTIPLICATION):

**a * b**

**a X b**

**ab** **10 X 5 = 50**

**a . b**

**(a)b**

**a(b)**

**(a)(b)**

2. Division

All of the below imply the exact same thing (DIVISION)

**a/b** **a ÷ b** $\frac{a}{b}$

# Some basic math

$$\sigma = ((1-p)/(p^2))^{1/2} = \left(\frac{1-p}{p^2}\right)^{1/2}$$

3. Squares and square root:

$X^2 = (X)(X)$ Eg. If X is 4, then $4^2 = (4)(4) = 16$

$\sqrt{X}$ square root of X Eg. If X is 4, then $\sqrt{4} = 2$

Note:
$X^{1/2}$ is another way of denoting the square root of X

# Some basic math

$$\sigma=( (1-p)/(p^2) )^{1/2} = \left( \frac{1-p}{p^2} \right)^{1/2}$$

3. Squares and square root:

$X^2 = (X)(X)$ Eg. If X is 4, then $4^2 = (4)(4) = 16$

$\sqrt{X}$ square root of X Eg. If X is 4, then $\sqrt{4} = 2$

Note:
$X^{1/2}$ is another way of denoting the square root of X

4. Summation: $\sum X_i$

Assume we had 5 cases, with values 21, 20, 19, 18, 18

$$\sum X_i = 21 + 20 + 19 + 18 + 18$$

# Some basic math

$$\sigma = ((1-p)/(p^2))^{1/2} = \left(\frac{1-p}{p^2}\right)^{1/2}$$

Summation continued, with 5 cases with the values 21, 20, 19, 18, 18

$$\sum X^2_i = 21^2 + 20^2 + 19^2 + 18^2 + 18^2$$

$$= 441 + 400 + 361 + 324 + 324 = 1850$$

# Some basic math

$$\sigma=( (1-p)/(p^2) )^{1/2} = \left( \frac{1-p}{p^2} \right)^{1/2}$$

Summation continued, with 5 cases with the values 21, 20, 19, 18, 18

$$\sum X^2_i \quad = 21^2 + 20^2 + 19^2 + 18^2 + 18^2$$

$$= 441 + 400 + 361 + 324 + 324 = 1850$$

Note: this is different from the below:

$$\left(\sum X_i\right)^2 \quad = ( 21 + 20 + 19 + 18 + 18 )^2 = 96^2 = 9216$$

$\sum X_i$ Add up each value/score

$\sum X^2{}_i$ Square each value/score and then add them up

$\left(\sum X_i\right)^2$ Add up each value/score and then square them up..

# Some basic math

$$\sigma=( (1-p)/(p^2) )^{1/2} = \left( \frac{1-p}{p^2} \right)^{1/2}$$

5 + (-2) = 5 – 2 = 3

5 - (-2) = 5 + 2 = 7

(3)(-3) = -9

(-3)(3) = -9

(-3)(-3) = +9

(9)/(-3) = -3

(-9)/(3) = -3

(-9)/(-3) = 3

***Order of operations:*** the rules of which calculation comes first in an expression.


They are:

- Do everything inside parentheses first: ()
- Exponents
- then do multiplies and divides
- lastly do the adds and subtracts

2 + 5 X 3=
2 + 15 = 17….  NOT: 7 X 3 !!!!!

(8 + 2) X 4 + 1 =
(10) X 4 + 1 =  NOT (10) X 5 !!!!
40 + 1 = 41

How we handle parentheses matters!

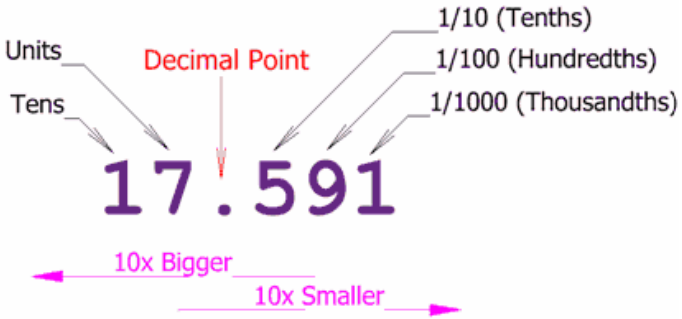
$(8 + 2) – 4(3)^2 / (8 - 6) = ?$
$(10) – 4(3)^2 / (2) =$
$10 – 4(9) / 2 =$
$10 - 36/2 =$
$10 – 18 = -8$

Example with the same equation but with no parentheses

$8 + 2 – 4 \times 3^2 / 8 - 6 = ?$
$8 + 2 – 4 \times 9 / 8 – 6 =$
$8 + 2 – 36 / 8 – 6 =$
$10 – 4.5 – 6 = 0.5$

A word on “rounding”

If possible, do all your calculations without rounding…




In providing your final answer on any question:
ROUND TO THE SECOND DECIMAL POINT

17.59122 is rounded to 17.59
17.238111 is rounded to 17.24

Always round 5 upward in the context of this class.. 15.52511 -> 15.53

If for practical purposes, you must round (time constraints? Many steps in a calculation, round to the “fourth decimal point”)

5.3939211 - > 5.3939

8.339555 -> 8. 3396
Why, because the fifth figure past the decimal point is 5 (i.e. round previous figure up)…

# The Role of Statistics

**Statistics** are mathematical tools used to organize, summarize, and manipulate ***data***.

What are **data**?

information expressed as numbers (quantitatively).

scores (or values) on "*variables*"

# EXAMPLE OF DATA

| wrkstat | marital | agewed | sibs | childs | age | birthmo | zodiac | educ | degree |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 20 | 3 | 1 | 43 | 5 | 2 | 11 | 1 |
| 1 | 5 | 0 | 2 | 0 | 44 | 8 | 6 | 16 | 3 |
| 1 | 3 | 25 | 2 | 0 | 43 | 2 | 11 | 16 | 3 |
| 2 | 5 | 0 | 4 | 0 | 45 | 99 | 99 | 15 | 1 |
| 5 | 5 | 0 | 1 | 0 | 78 | 10 | 7 | 17 | 4 |
| 5 | 1 | 25 | 2 | 2 | 83 | 3 | 12 | 11 | 1 |
| 1 | 1 | 22 | 2 | 2 | 55 | 10 | 7 | 12 | 1 |
| 5 | 1 | 24 | 3 | 2 | 75 | 11 | 9 | 12 | 1 |
| 1 | 3 | 22 | 1 | 2 | 31 | 7 | 4 | 18 | 4 |
| 2 | 5 | 0 | 1 | 0 | 54 | 3 | 12 | 18 | 4 |
| 1 | 5 | 0 | 1 | 0 | 29 | 4 | 2 | 18 | 4 |
| 1 | 5 | 0 | 0 | 0 | 23 | 10 | 8 | 15 | 1 |
| 1 | 1 | 31 | 0 | 1 | 61 | 99 | 99 | 12 | 1 |
| 5 | 4 | 24 | 3 | 4 | 63 | 3 | 1 | 4 | 0 |
| 4 | 5 | 0 | 4 | 3 | 33 | 3 | 12 | 10 | 0 |
| 1 | 5 | 0 | 0 | 1 | 36 | 11 | 8 | 14 | 1 |
| 7 | 5 | 0 | 98 | 4 | 39 | 3 | 12 | 8 | 0 |

One would need documentation to figure out what this data means?
What are the variables, cases, etc. ?

# The Role of Statistics

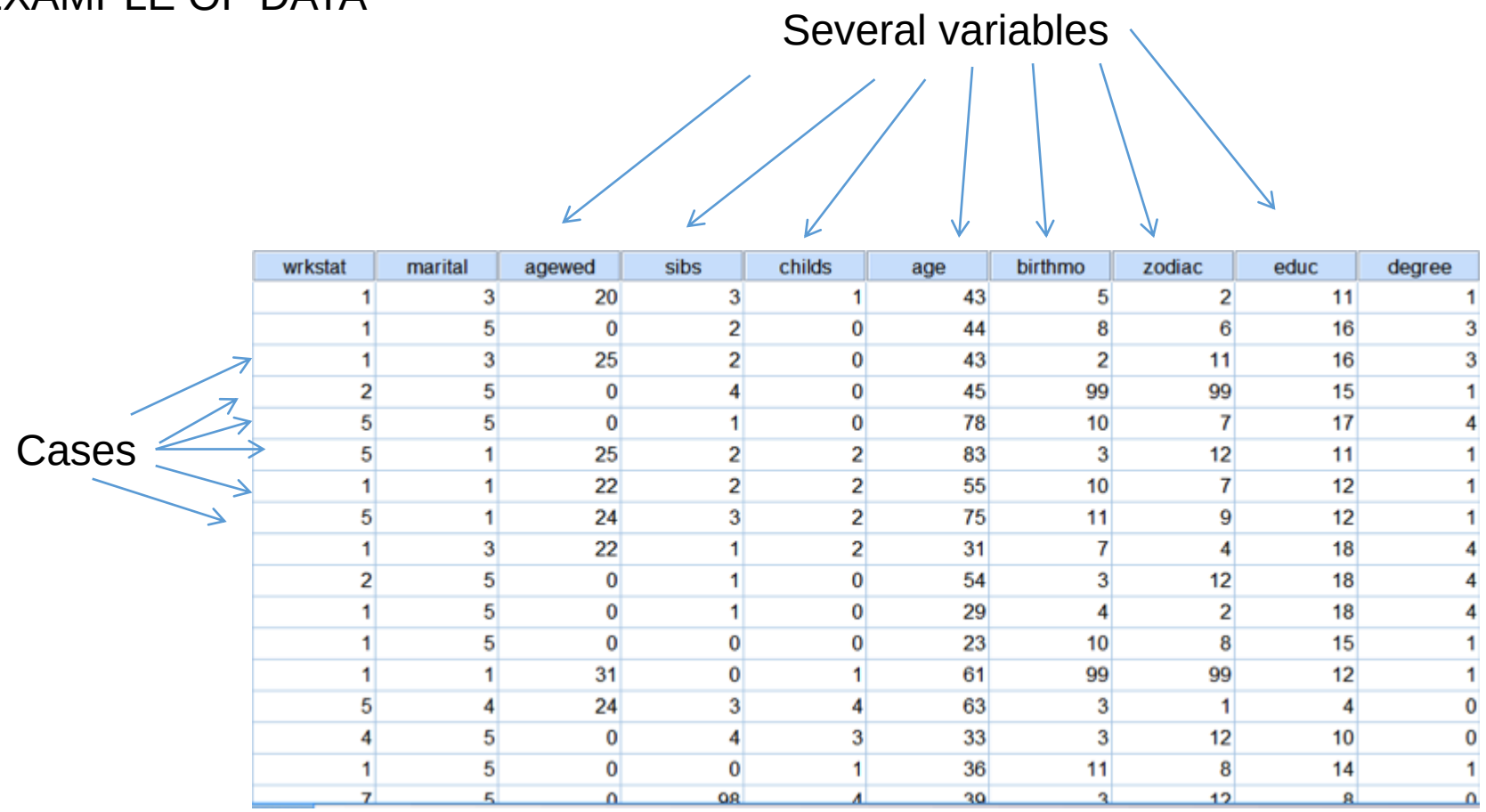
*What are "**variables**"?*

traits that can change values from *case* to case (e.g., age, gender, social class).

What are "**cases**"?

Cases are the entities from which data are gathered (e.g., people, groups, provinces, countries).

# EXAMPLE OF DATA




| wrkstat | marital | agewed | sibs | childs | age | birthmo | zodiac | educ | degree |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 20 | 3 | 1 | 43 | 5 | 2 | 11 | 1 |
| 1 | 5 | 0 | 2 | 0 | 44 | 8 | 6 | 16 | 3 |
| 1 | 3 | 25 | 2 | 0 | 43 | 2 | 11 | 16 | 3 |
| 2 | 5 | 0 | 4 | 0 | 45 | 99 | 99 | 15 | 1 |
| 5 | 5 | 0 | 1 | 0 | 78 | 10 | 7 | 17 | 4 |
| 5 | 1 | 25 | 2 | 2 | 83 | 3 | 12 | 11 | 1 |
| 1 | 1 | 22 | 2 | 2 | 55 | 10 | 7 | 12 | 1 |
| 5 | 1 | 24 | 3 | 2 | 75 | 11 | 9 | 12 | 1 |
| 1 | 3 | 22 | 1 | 2 | 31 | 7 | 4 | 18 | 4 |
| 2 | 5 | 0 | 1 | 0 | 54 | 3 | 12 | 18 | 4 |
| 1 | 5 | 0 | 1 | 0 | 29 | 4 | 2 | 18 | 4 |
| 1 | 5 | 0 | 0 | 0 | 23 | 10 | 8 | 15 | 1 |
| 1 | 1 | 31 | 0 | 1 | 61 | 99 | 99 | 12 | 1 |
| 5 | 4 | 24 | 3 | 4 | 63 | 3 | 1 | 4 | 0 |
| 4 | 5 | 0 | 4 | 3 | 33 | 3 | 12 | 10 | 0 |
| 1 | 5 | 0 | 0 | 1 | 36 | 11 | 8 | 14 | 1 |
| 7 | 5 | 0 | 98 | 4 | 39 | 3 | 12 | 8 | 0 |

- A new study by the Centre for Addiction and Mental Health (CAMH), found Canadians drink more than alcohol than the global average. On average, Canadians consume 8.2 litres of pure alcohol a person aged 15 and over

1. What is the variable?
2. What are the cases?
3. What is the statistic used?

**Variable** is **amount of alcohol consumed**.
**Cases** are the **Canadians sampled**.
**Statistic** is the **average** - average amount of alcohol consumed

(Note: that amounts to 480 bottles of beer, 91 bottles of wine or 27 bottles of spirits, the Center reports).

- In a recent survey of university students, Statistics Canada found that the average age of university students in Canada was 21.7 years. Identify the following:
  1. What is the variable?
  2. What are the cases?
  3. What is the statistic used?

Oldest and youngest university grads

94 years old
10 years old

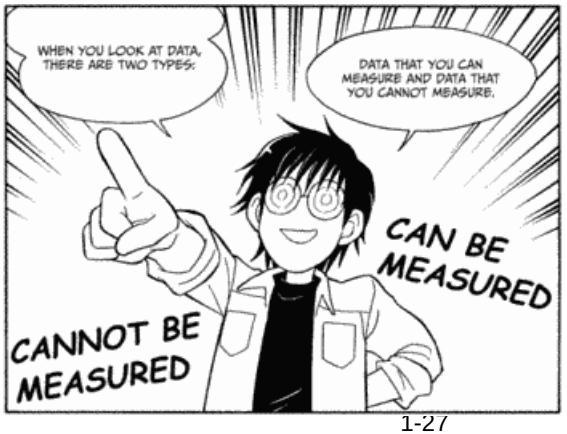

**Variable** is **age**.
**Cases** are the **university students**.
**Statistic** is the **average** - average age of university students in Canada

# Statistical Applications

- Two main statistical applications:
  - Descriptive statistics
  - Inferential statistics (sample based)

# Descriptive Statistics

- Summarize one variable **(univariate)**.
- Summarize the relationship between two variables **(bivariate)** .
- Summarize the relationship between three or more variables **(multivariate)** .



- ***Univariate*** descriptive statistics include:
  - Percentages, averages, and various charts and graphs.
  - Example: Canadians have on average 1.6 children

- ***Bivariate*** descriptive statistics describe the strength and direction of the relationship between two variables.
  - Example: University educated Canadians are more likely than others to have university educated parents.

- ***Multivariate*** descriptive statistics describe the relationships between three or more variables.
  - Example: Income tends to increase with education more so for native born Canadians then is true of Canadians born abroad.

# Inferential Statistics

- Generalize, or infer, from a sample to a population.
  - Population includes **all** cases in which the research is interested.
  - Samples include carefully chosen **subsets** of the population.

# Inferential Statistics

*(continued)*

- Voter surveys are a common application of inferential statistics.
  - A thousand or so carefully selected voters are interviewed about their voting intentions.
  - This information is used to estimate the intentions of **all** voters (millions of people).
  - Example: The Green Party of Canada will receive about 10% of the vote.

# Types of Variables

- There are many schemes used to classify variables including:
    1. Independent or dependent variables
    2. Discrete or continuous variables
    3. Nominal, ordinal, or interval-ratio variables (or levels of measurement)

# Independent or Dependent Variables

- In causal relationships:

CAUSE → EFFECT

independent variable → dependent variable

Which variable is the dependent?

Smoking behavior, lung cancer

Weekly exercise (in hours), pulse rate (resting)

Longevity, SES,

# Discrete or Continuous Variables

- **Discrete** variables are measured in units that cannot be subdivided.
  - Example: Nationality (Canadian; Polish; Irish, Syrian, etc.)

- **Continuous** variables are measured in a unit that can be subdivided to a very refined degree…
  - Example: Age

**What about the following?**

**Religious affiliation?**

**Income?**

# Nominal, Ordinal, or Interval-Ratio Level of Measurement

- The mathematical quality of the scores of a variable is measured on three different levels, called levels of measurement:
  - Nominal - Scores are **labels** only, they can not be treated like numbers in the normal sense.
  - Ordinal - Scores have some numerical quality and can be **ranked**.
  - Interval-ratio - Scores are **numbers**.

# Nominal Level Variables

- Scores are different from each other but cannot be treated as numbers.
  - Examples:
    - Gender
      - 1 = Female, 2 = Male
    - Immigrant Status
      - 1 = Canadian-born, 2 =Foreign-born

# Ordinal Level Variables

- Scores can be ranked from high to low or from more to less.
  - Survey items that measure opinions and attitudes are typically ordinal.
  - If you can distinguish between the ***scores*** of the variable using terms such as "more, less, higher, or lower" the variable is ordinal.
    - Example: Students at a university were asked "Do you (1) strongly agree, (2) agree, (3) disagree or (4) strongly disagree that smoking should be banned on campus?"

(A student that strongly agreed would be ***more*** in favour to ban smoking on campus than a student who merely "agreed" or "strongly disagreed").

# Interval-Ratio Level Variables

- Scores are actual numbers and have a true zero point and equal intervals between scores.
- Examples:
  - Age (in years)
  - Income (in dollars)
  - Number of children
    - A true zero point (0 = no children)
    - Equal intervals: each child adds one unit

- Different statistics require different mathematical operations (ranking, addition, square root, etc.), so the level of measurement of a variable tells us which statistics are permissible and appropriate.

| wrkstat | marital | sibs | childs | age | zodiac | educ | degree |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 1 | 43 | 2 | 11 | 1 |
| 1 | 5 | 2 | 0 | 44 | 6 | 16 | 3 |
| 1 | 3 | 2 | 0 | 43 | 11 | 16 | 3 |
| 2 | 5 | 4 | 0 | 45 | 99 | 15 | 1 |
| 5 | 5 | 1 | 0 | 78 | 7 | 17 | 4 |
| 5 | 1 | 2 | 2 | 83 | 12 | 11 | 1 |
| 1 | 1 | 2 | 2 | 55 | 7 | 12 | 1 |
| 5 | 1 | 3 | 2 | 75 | 9 | 12 | 1 |
| 1 | 3 | 1 | 2 | 31 | 4 | 18 | 4 |
| 2 | 5 | 1 | 0 | 54 | 12 | 18 | 4 |
| 1 | 5 | 1 | 0 | 29 | 2 | 18 | 4 |
| 1 | 5 | 0 | 0 | 23 | 8 | 15 | 1 |
| 1 | 1 | 0 | 1 | 61 | 99 | 12 | 1 |
| 5 | 4 | 3 | 4 | 63 | 1 | 4 | 0 |
| 4 | 5 | 4 | 3 | 33 | 12 | 10 | 0 |
| 1 | 5 | 0 | 1 | 36 | 8 | 14 | 1 |

What's the level of measurement?

**Wrkstat**
1- full time
2- part time
3- unemployed
4- dropped out the labour force
5. Not applicable (too young/retired, etc)

**marital**
1- married
2- cohabiting
3- divorced
4- same sex married
5. Never married

**Sibs** – number of siblings
**Childs** – number of children
**Age** – current age

**Zodiac**
1- Aries
2- Cancer
3-Sagittarius
4., etc.

**Educ**
- Years of education

**Degree**
1- Hs or less
2. College
3. University
4. Graduate/professional degree