# Welcome to PSYC4050:

## Psychological Research Methodology IV


Michelle
Philippe
Brenda

# Course Coordinator

| | |
|---|---|
| Dr Jason Tangen | jtangen@psy.uq.edu.au |

# Tutors

| | |
|---|---|
| Philippe Lacherez | lacher@psy.uq.edu.au |
| Brenda Ocampo | b.ocampo@psy.uq.edu.au |
| Michelle Engels | mengels@psy.uq.edu.au |

Office hours by appointment, before, during, and after lectures and tutorials.

- Announcements
- Course Documents
- Discussion Board
- My Grades
- Course Profile

- Course Map
- Control Panel
- Refresh
- Detail View

[PSYC4050] PSYCHOLOGICAL RESEARCH METHODOLOGY IV (TAUGHT AT THE ST LUCIA CAMPUS), SEMESTER 1, 2008 (PSYC4050S_5820STX) > ANNOUNCEMENTS

# Announcements

Use this link regularly to access important course announcements such as changes in the syllabus, corrections/clarifications of materials, exam schedules, etc.

VIEW TODAY | VIEW LAST 7 DAYS | VIEW LAST 30 DAYS | VIEW ALL

## February 17, 2008 - February 24, 2008

**Tue, Feb 19, 2008 -- *UQ PSYC4050 2008 facebook group***

Posted by: Jason Tangen

If you prefer to use Facebook rather than Blackboard for discussion, then join the UQ PSYC4050 2008 group: http://uqedu.facebook.com/group.php?gid=8254622939

**Tue, Feb 19, 2008 -- *Tutorial Sign-on***

Posted by: Jason Tangen

You will be able to use Sign-on through mySI-net <http://www.sinet.uq.edu.au> on Wednesday, 27 February at 7am to sign up for your tutorial session.

[PSYC4050] PSYCHOLOGICAL RESEARCH METHODOLOGY IV (TAUGHT AT THE ST LUCIA CAMPUS). SEMESTER 1, 2008 (PSYC4050S_5820STX) > DISCUSSION BOARD > COURSE DISCUSSIONS: GENERAL DISCUSSION FOR PSYC4050

# Course Discussions: General Discussion for PSYC4050

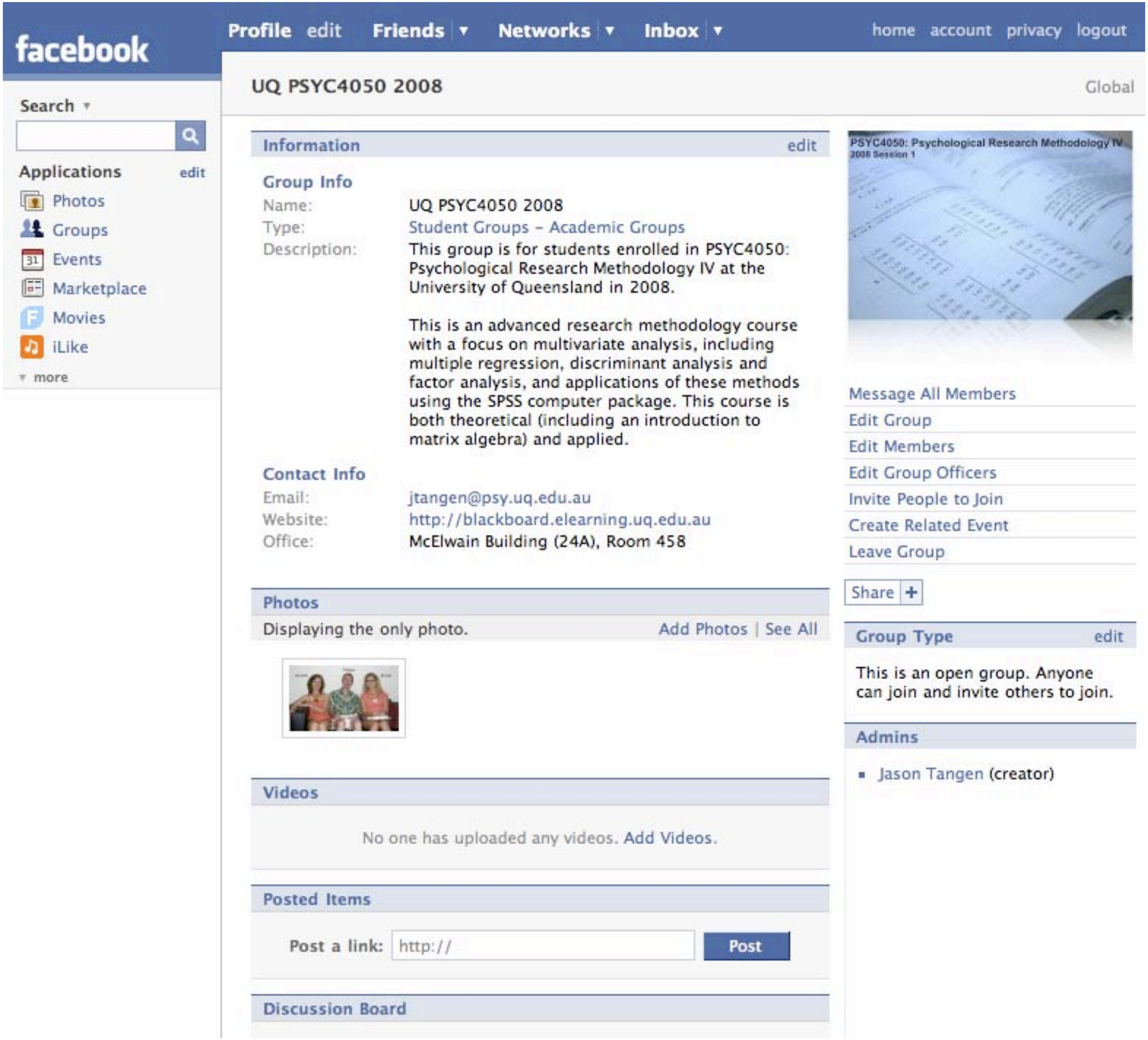
Tree View | List View

Thread | Remove | Collect | Mark Read | Mark Unread | **Change Status to:** Published | Go | Search

| | | Date | Thread | Author | Status | Tags | Unread Posts | Total Posts |
|---|---|---|---|---|---|---|---|---|
| ☐ | | 2/19/08 1:58 PM | UQ PSYC4050 2008 facebook group | Jason Tangen | Published | | 0 | 1 |

Select All | Go | Items Per Page 25 | Go | **Display** Show All | **Tags** Show All Tags | Go

OK


facebook
Profile edit Friends Networks Inbox
home account privacy logout
Search
Applications edit
Photos
Groups
Events
Marketplace
Movies
iLike
more
UQ PSYC4050 2008
Global
Information edit
Group Info
Name: UQ PSYC4050 2008
Type: Student Groups - Academic Groups
Description: This group is for students enrolled in PSYC4050: Psychological Research Methodology IV at the University of Queensland in 2008.
This is an advanced research methodology course with a focus on multivariate analysis, including multiple regression, discriminant analysis and factor analysis, and applications of these methods using the SPSS computer package. This course is both theoretical (including an introduction to matrix algebra) and applied.
Contact Info
Email: jtangen@psy.uq.edu.au
Website: http://blackboard.elearning.uq.edu.au
Office: McElwain Building (24A), Room 458
Photos
Displaying the only photo.
Add Photos | See All
Videos
No one has uploaded any videos. Add Videos.
Posted Items
Post a link: http://
Post
Discussion Board
PSYC4050: Psychological Research Methodology IV
2008 Session 1
Message All Members
Edit Group
Edit Members
Edit Group Officers
Invite People to Join
Create Related Event
Leave Group
Share
Group Type edit
This is an open group. Anyone can join and invite others to join.
Admins
Jason Tangen (creator)

# Sign-on

Wednesday, 27 February
(tomorrow) at 7am


| Class Group | Building / Room | Day / Time | |
|---|---|---|---|
| L | 24-S304 | TUE 10:00 AM - 11:50 AM | |
| P1 | 39A-226 | TUE 12:00 PM - 1:50 PM | Brenda |
| P2 | 31A-205 | TUE 12:00 PM - 1:50 PM | Philippe |
| P3 | 31A-205 | TUE 2:00 PM - 3:50 PM | Michelle |
| ~~P4~~ | ~~-~~ | ~~TUE 4:00 PM - 5:50 PM~~ | Philippe |
| P5 | 35-116 | TUE 4:00 PM - 5:50 PM | Brenda |
| P6 | 39A-227 | WED 10:00 AM - 11:50 AM | Philippe |
| P7 | 31A-205 | WED 10:00 AM - 11:50 AM | Michelle |
| P8 | 39A-227 | WED 12:00 PM - 1:50 PM | Philippe |


BLAIR
DRIVE
BASKET BALL
39A
39
37
28
26B
26
26A
35
14
CAMPBELL ROAD
UNION ROAD
24A
31B
24
25
32
38
P
E
E
TAXI
P1
39A, 226
Tue 12-2
P8
39A, 227
Wed 12-2
P6
39A, 227
Wed 10-12
P5
35, 116
Tue 4-6
P7
31A, 205
Wed 10-12
P2
31A, 205
Tue 12-2
P3
31A, 205
Tue 2-4
Lecture
24, S304
Tue 10-12

# Handouts

1. “Electronic” Course Profile
2. Matrix Booklet
3. Quiz 1 (Take home; due next week; 2%)

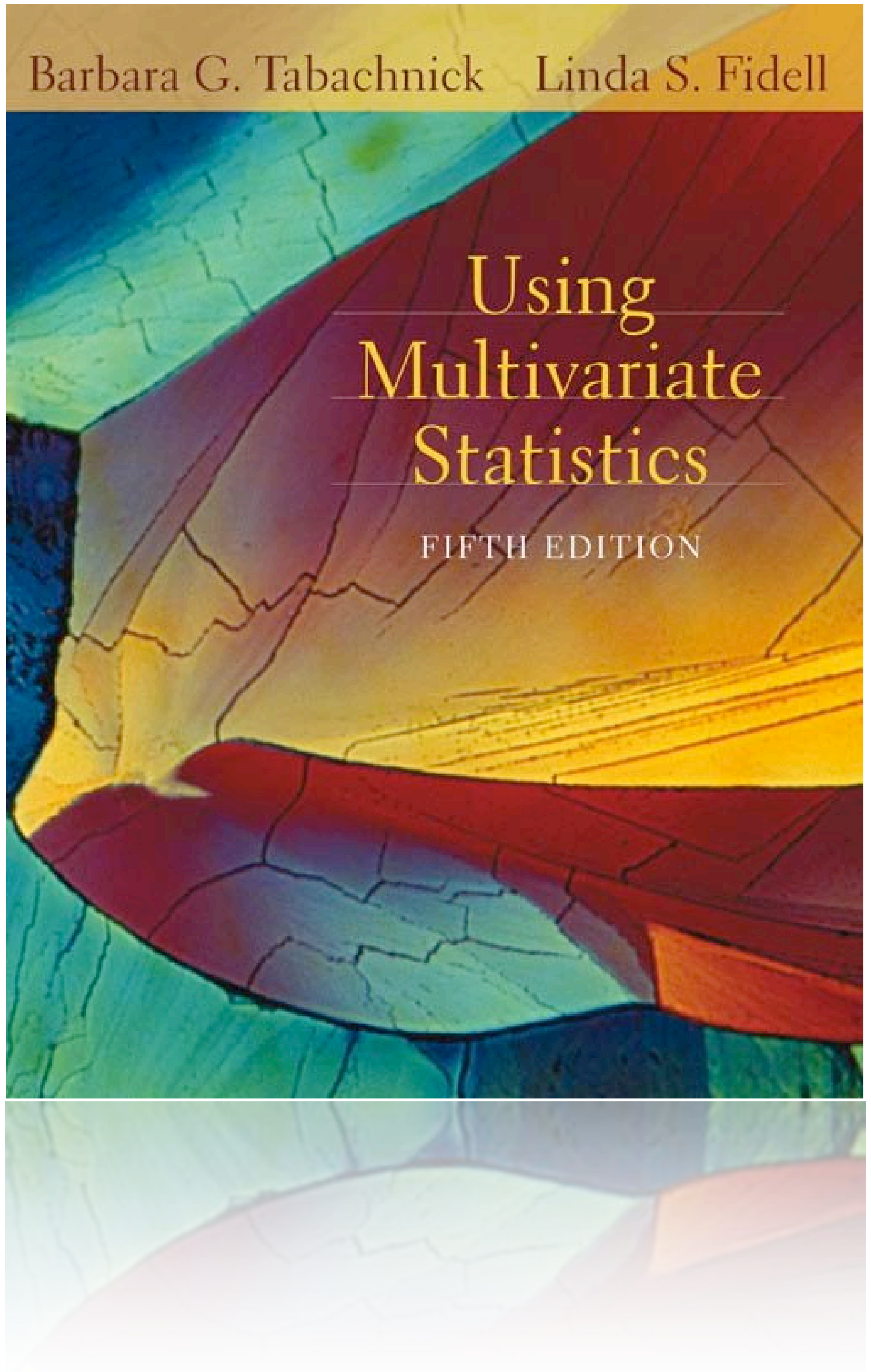
# Administrative Details

## 1. Assessment

| Assessment Task | Due Date | Weighting | Learning Objectives |
|---|---|---|---|
| *Take Home Exam*<br>Review Quiz 1 | 26 Feb 08 10:00 - 4 Mar 08 10:00 | 2% | 1, 2, 3, 4, 5 |
| *Small group exercise*<br>Small group exercise | 26 Feb 08 10:00 - 18 Mar 08 10:00 | 5% | 1, 2, 3, 4, 5 |
| *Essay*<br>The Multiple Regression assignment | 11 Mar 08 10:00 - 22 Apr 08 15:00 | 25% | 1, 2, 3, 4, 5 |
| *Essay*<br>The Discriminant Analysis assignment | 11 Mar 08 10:00 - 13 May 08 10:00 | 20% | 1, 2, 3, 4, 5 |
| *In Class Quiz*<br>Matrices (Quiz 2) | 22 Apr 08 10:00 - 22 Apr 08 11:00 | 3% | 1, 2, 3, 4, 5 |
| *Exam - during Exam Period (Central)*<br>Final Examination | Examination Period | 45% | 1, 2, 3, 4, 5 |

# Administrative Details

2. Text Book




Tabachnick, B. G. & Fidell, L. S. (2007).
*Using multivariate statistics* (5th ed.).
Boston: Pearson/Allyn & Bacon.

# Aims of the course

- to provide skills in choosing, performing and interpreting appropriate multivariate analyses
- to provide an understanding of how different multivariate methods work, and how they interrelate
- to provide the ability to critically evaluate analyses reported in the literature

# A few terms...

*Independent variables.*

Independent variables are the conditions that your subjects are exposed to... The bits of the world that you control and manipulate while holding everything else constant.

e.g., treatment vs placebo

In non-experimental scenarios, these may represent characteristics (which you can't wiggle).

e.g., tall or short

# A few terms...

*Independent variables.*

Independent variables are often thought of as predictor variables. This is most evident in experiments where we think of independent variables as causes.


*Dependent variables.*

Dependent variables are then often regarded as the effect of our manipulation.

# A few terms...

Predictor

Stimulus

Task

Input


Criterion

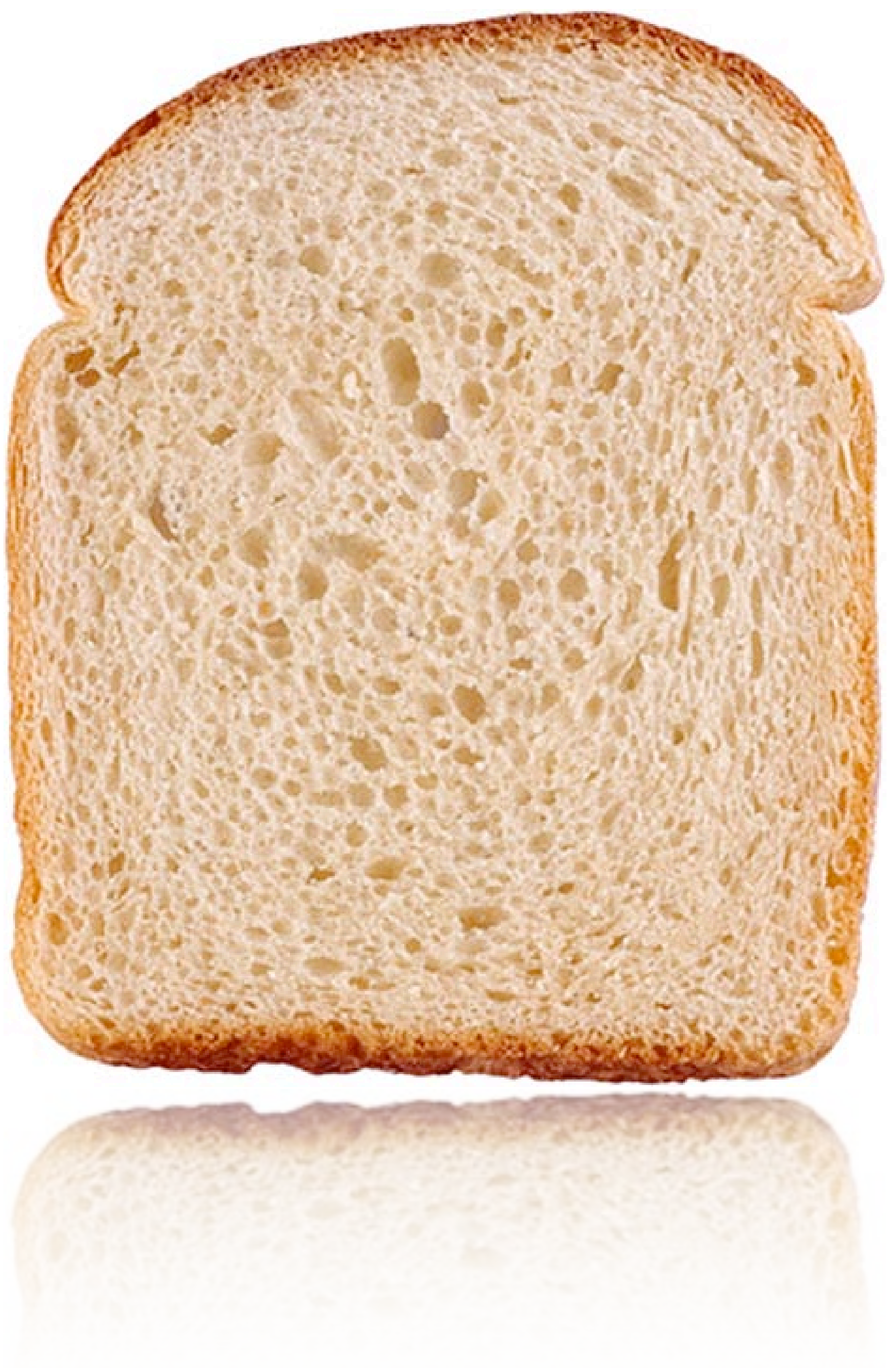
Response

Performance

Output

# A few terms...

Therefore, in many of the examples and exercises in this course, we will use these seemingly causal words for (in)dependent variables as a matter of convenience.



=

# A few terms...

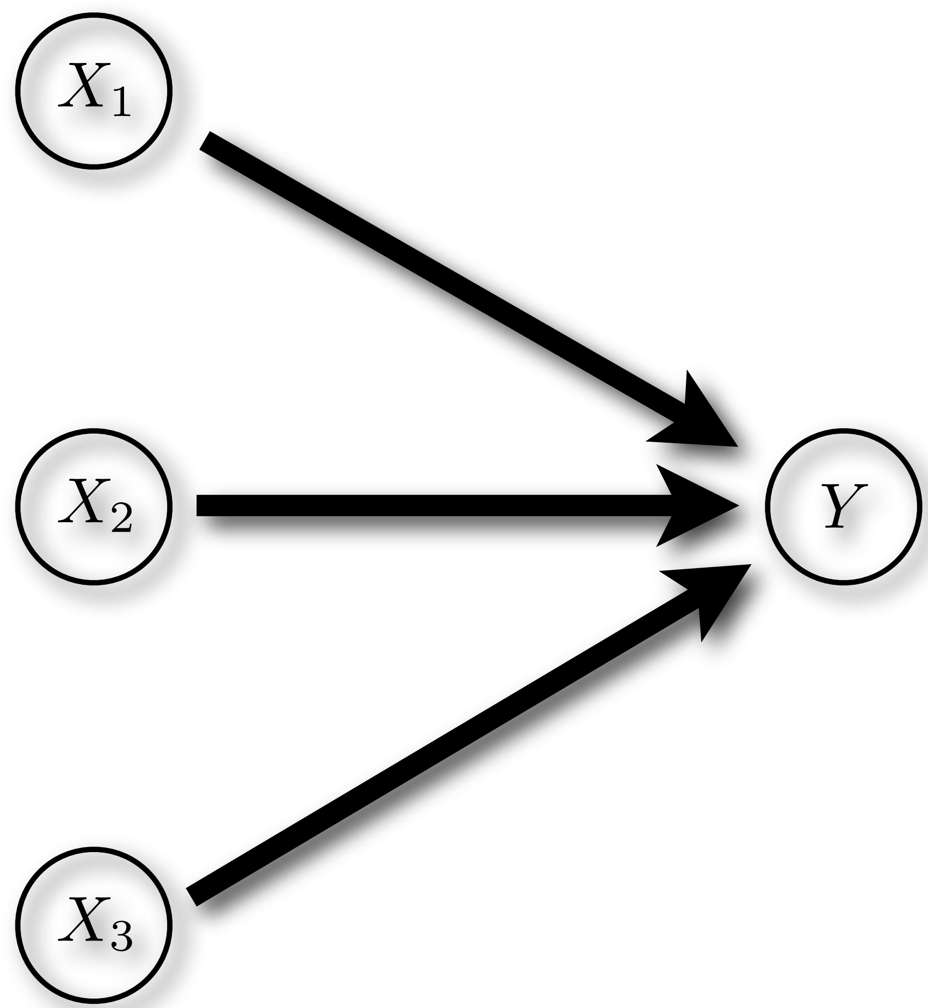
Univariate Statistics


Single
dependent variable

# A few terms...

## Univariate Statistics

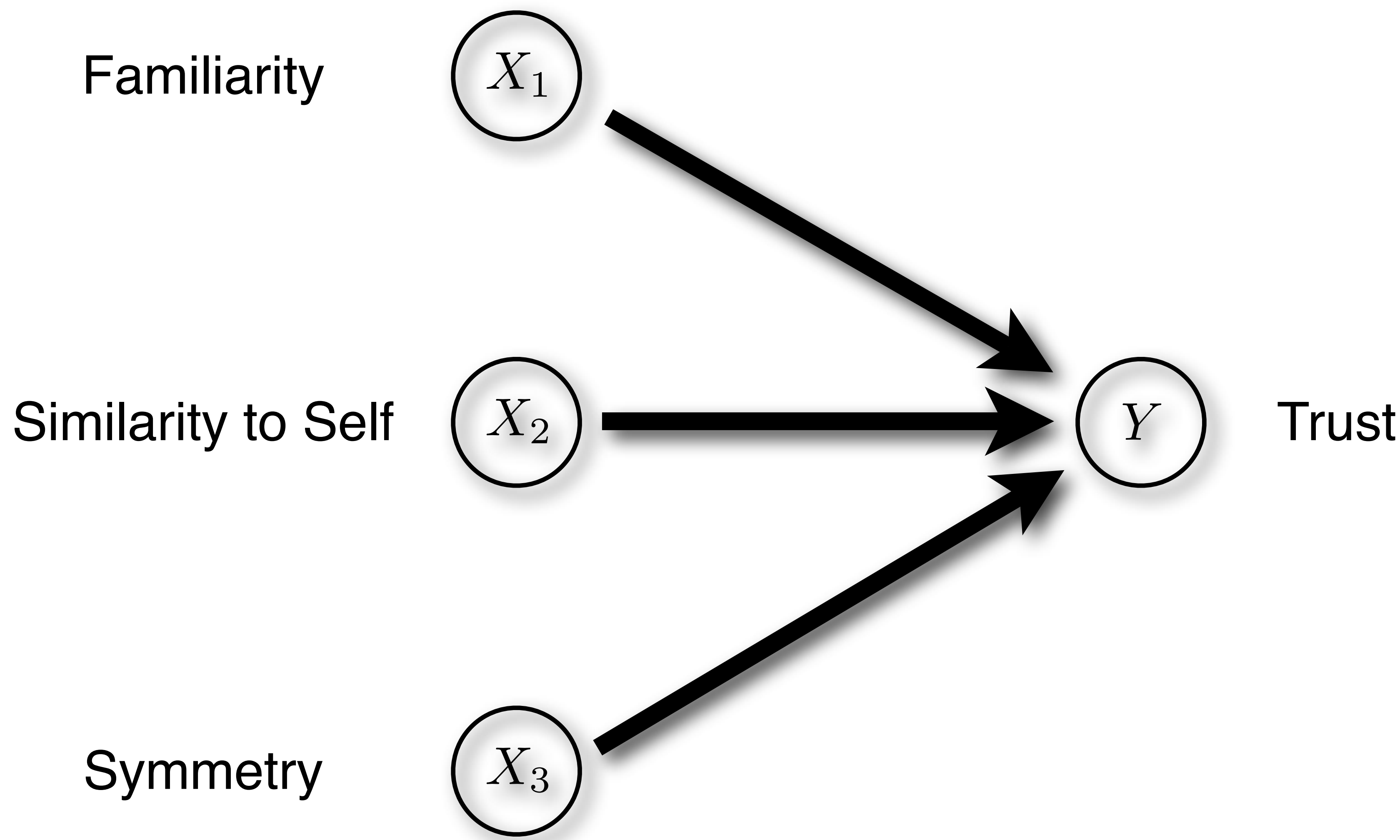
There may be more than one independent variable



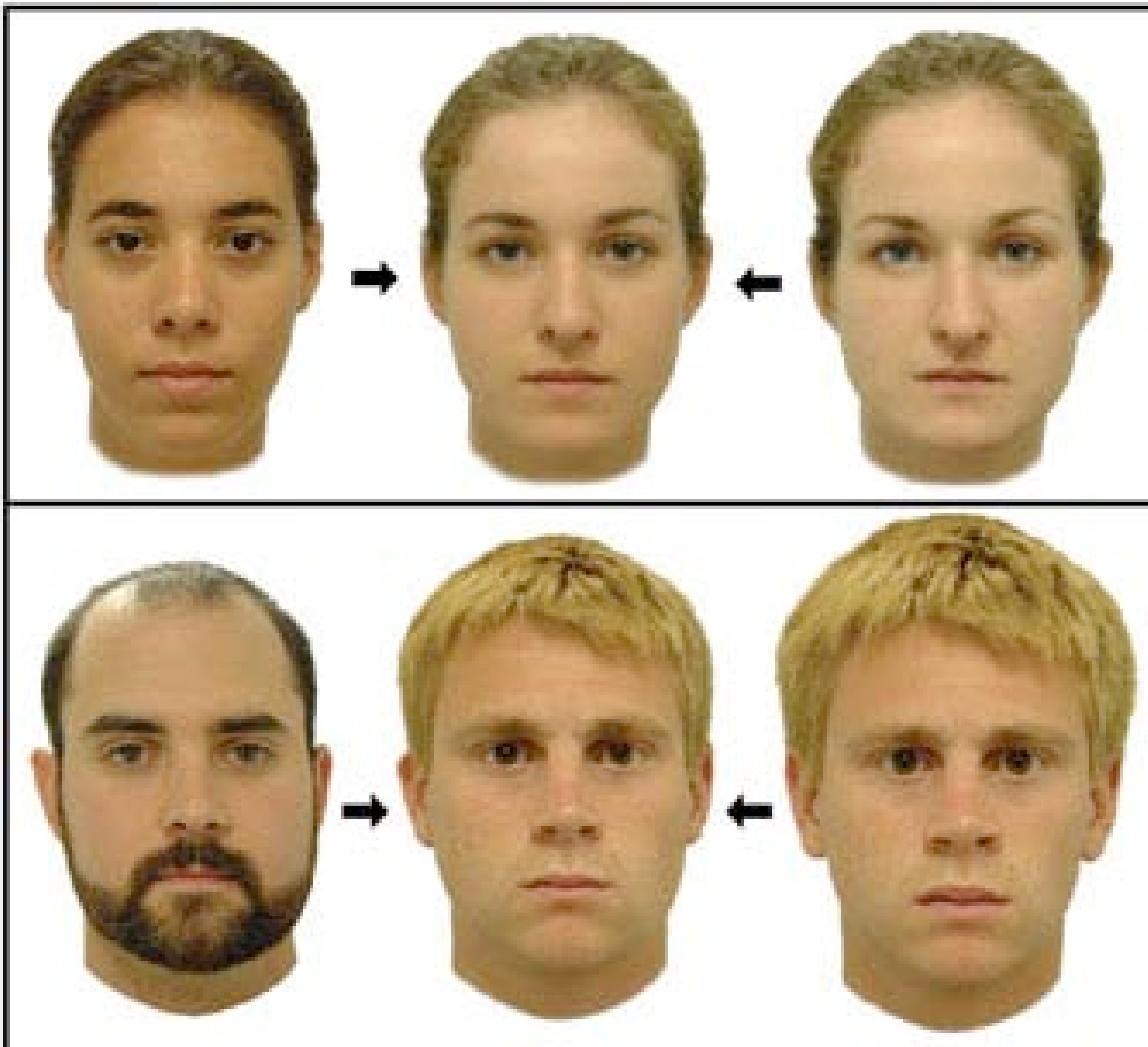

Single dependent variable

# A few terms...

## Univariate Statistics

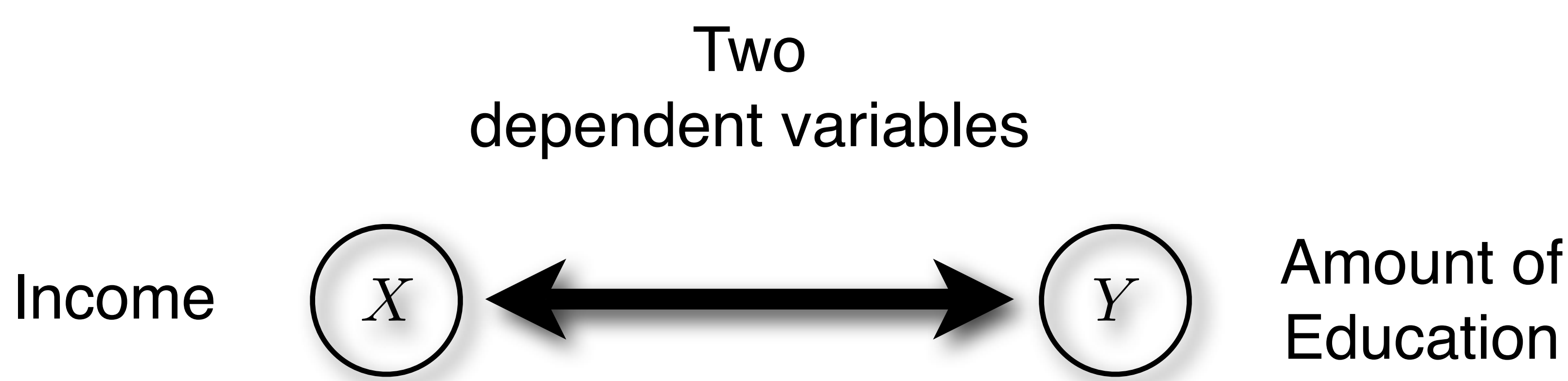



e.g., ANOVA



# A few terms...

## Bivariate Statistics

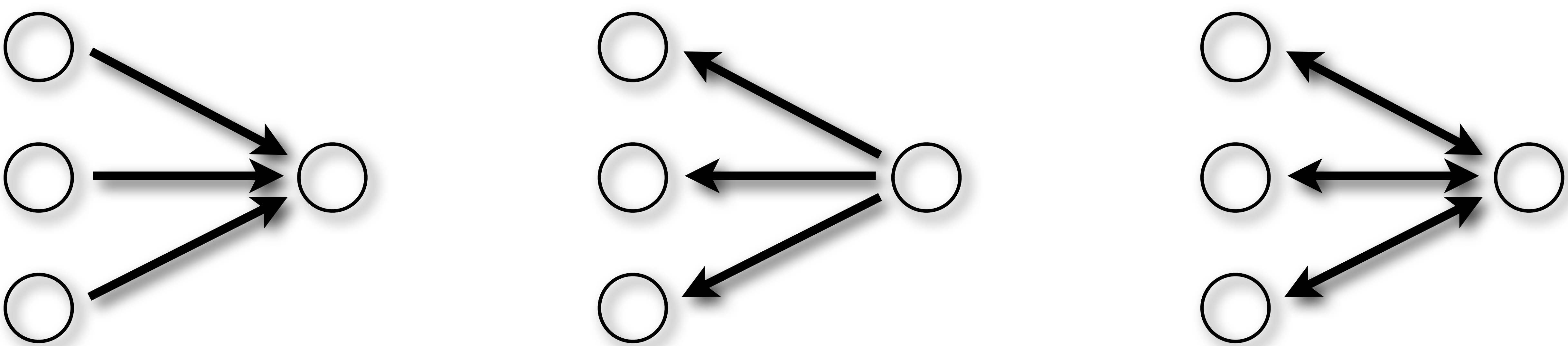



e.g., Pearson correlation coefficient

# A few terms...

## Multivariate Statistics

Multiple independent
and
dependent variables



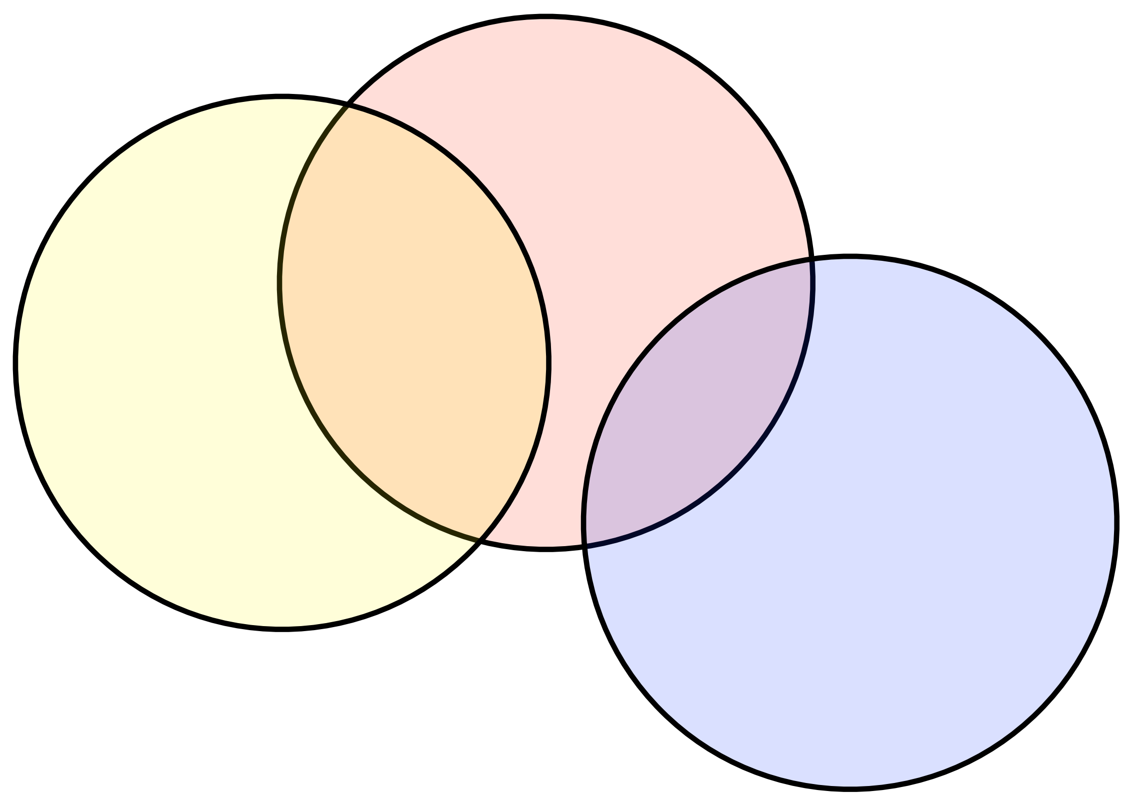
e.g., Multiple regression

# A few terms...

## Orthogonality

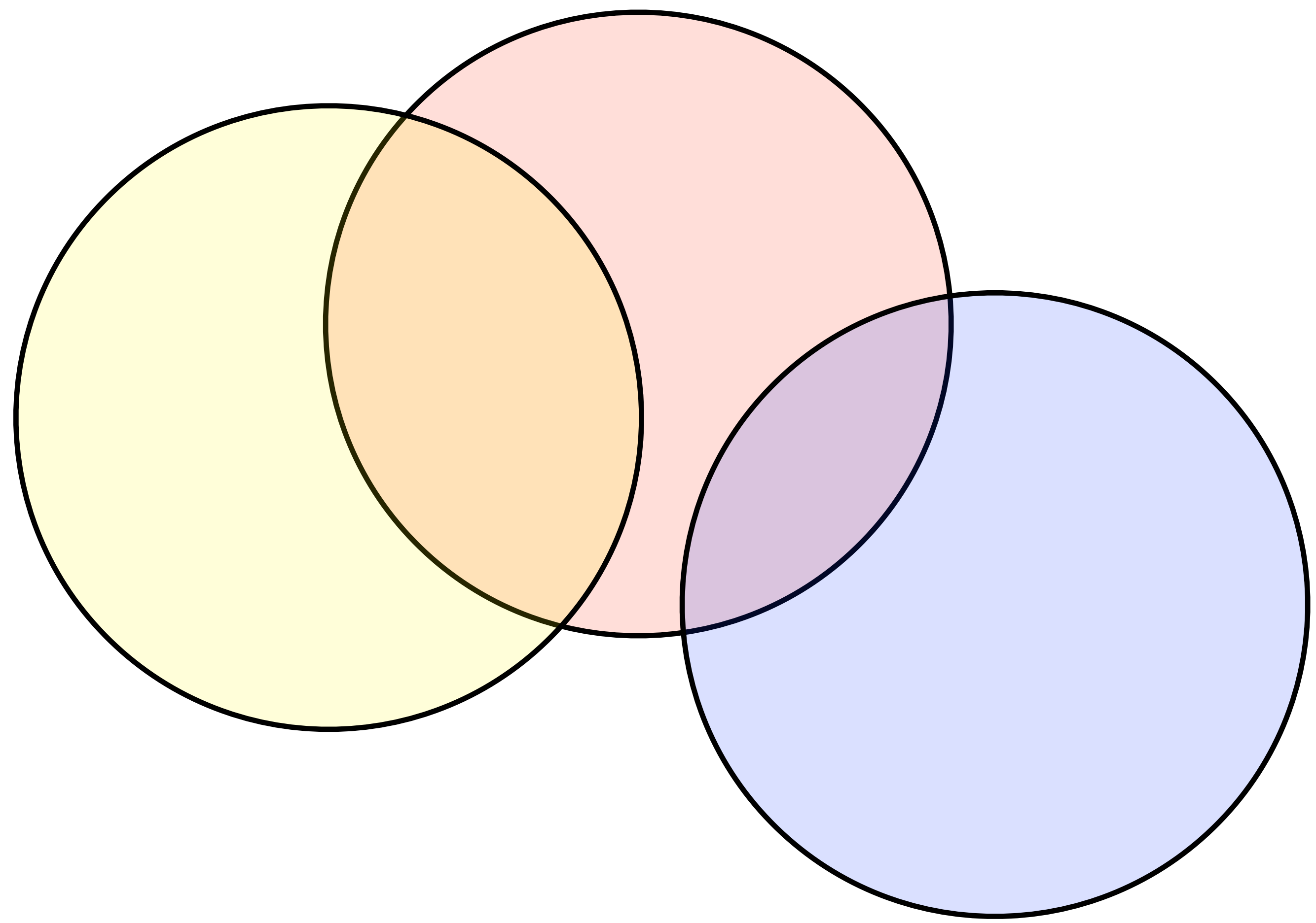
The perfect non-association between variables

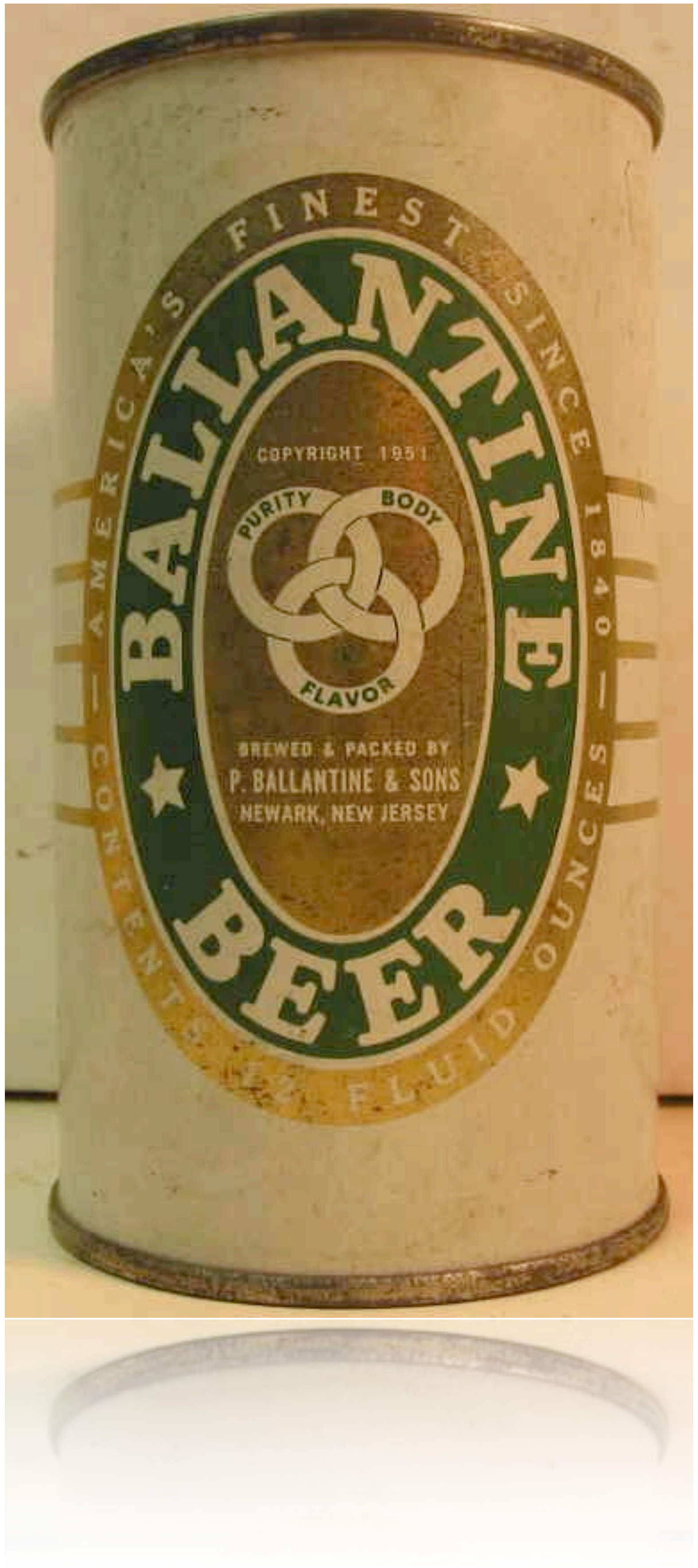
...knowing the value of one gives you no indication as to the value of the other.



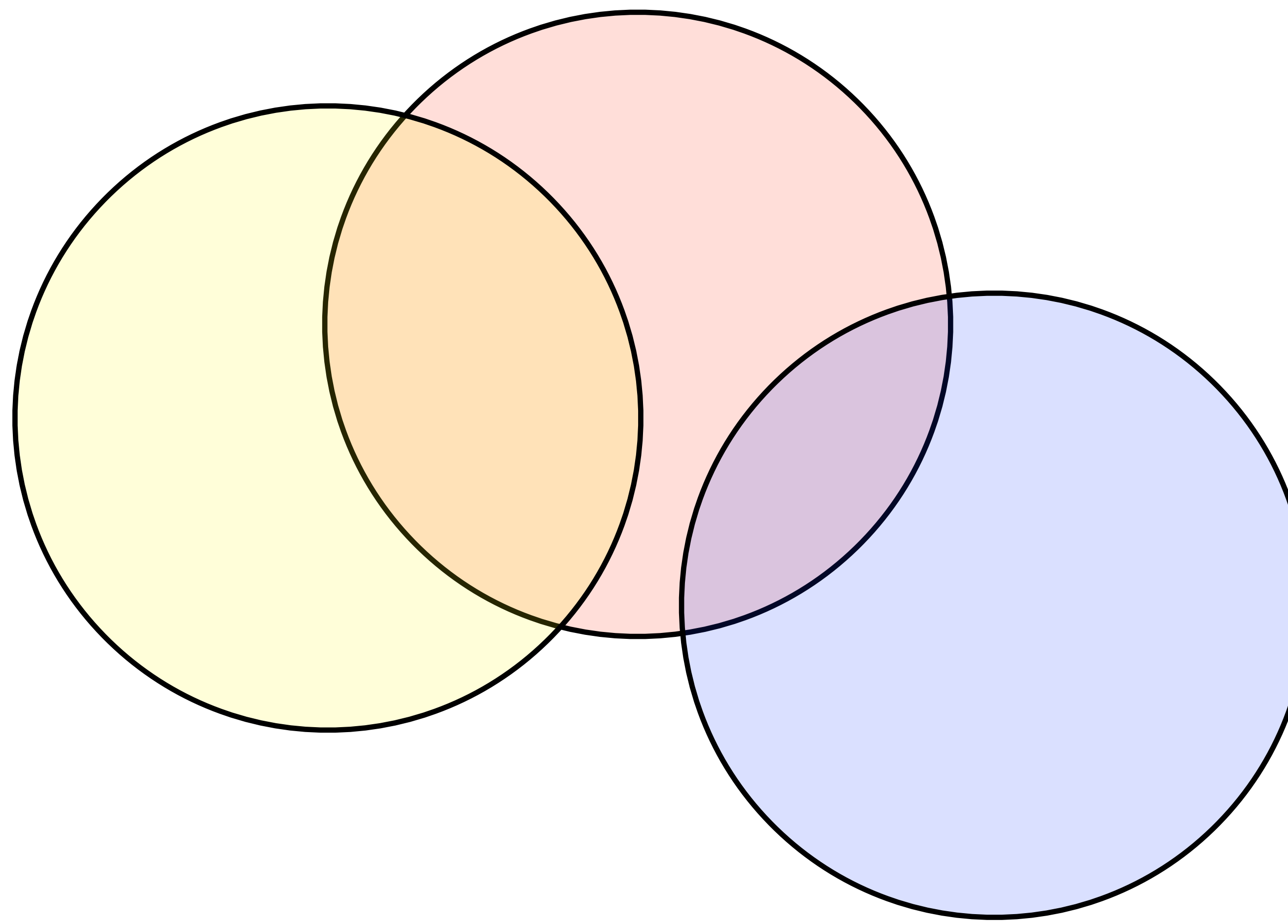
Leonhard Euler
1707-1783

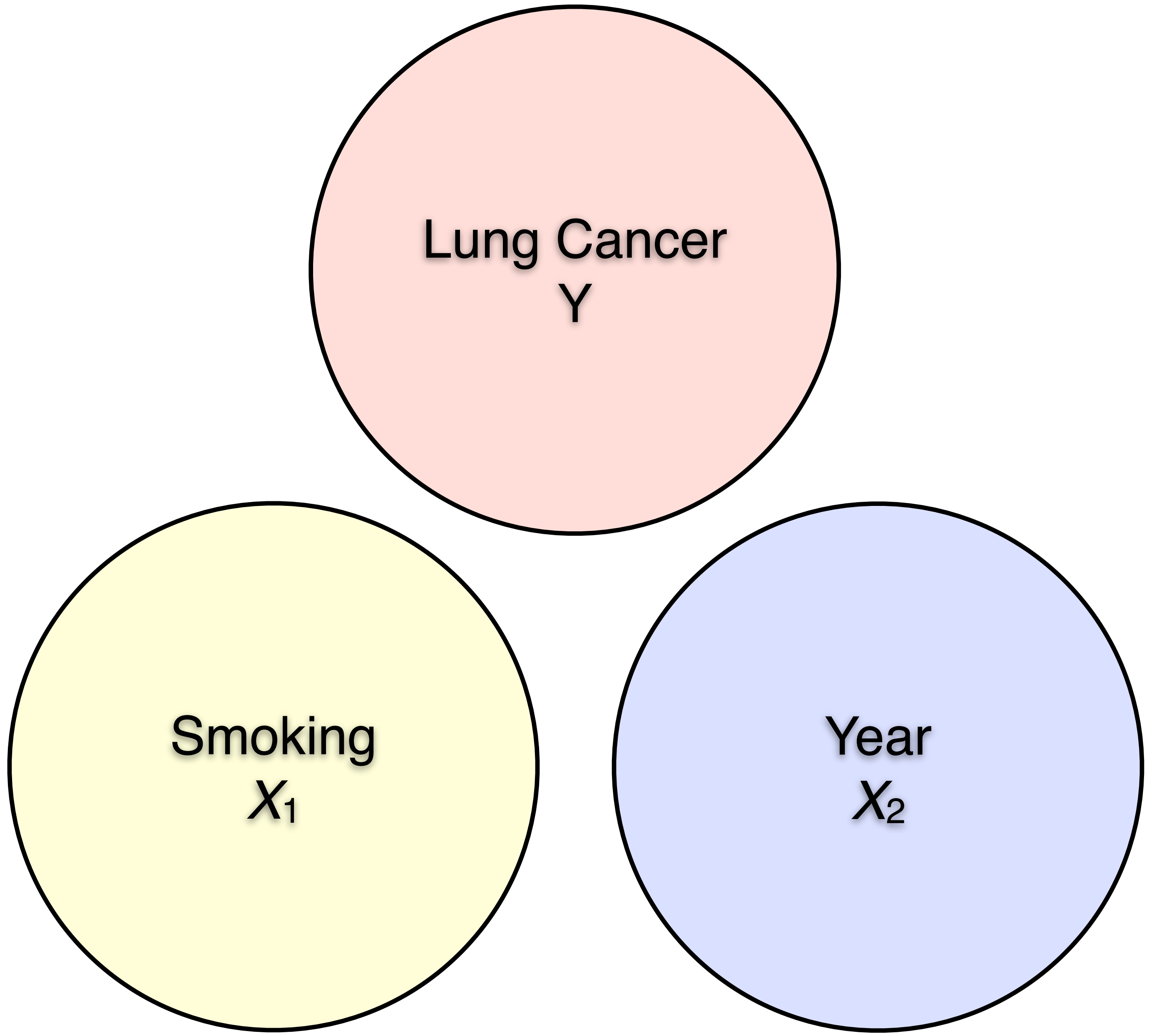
# Euler diagrams

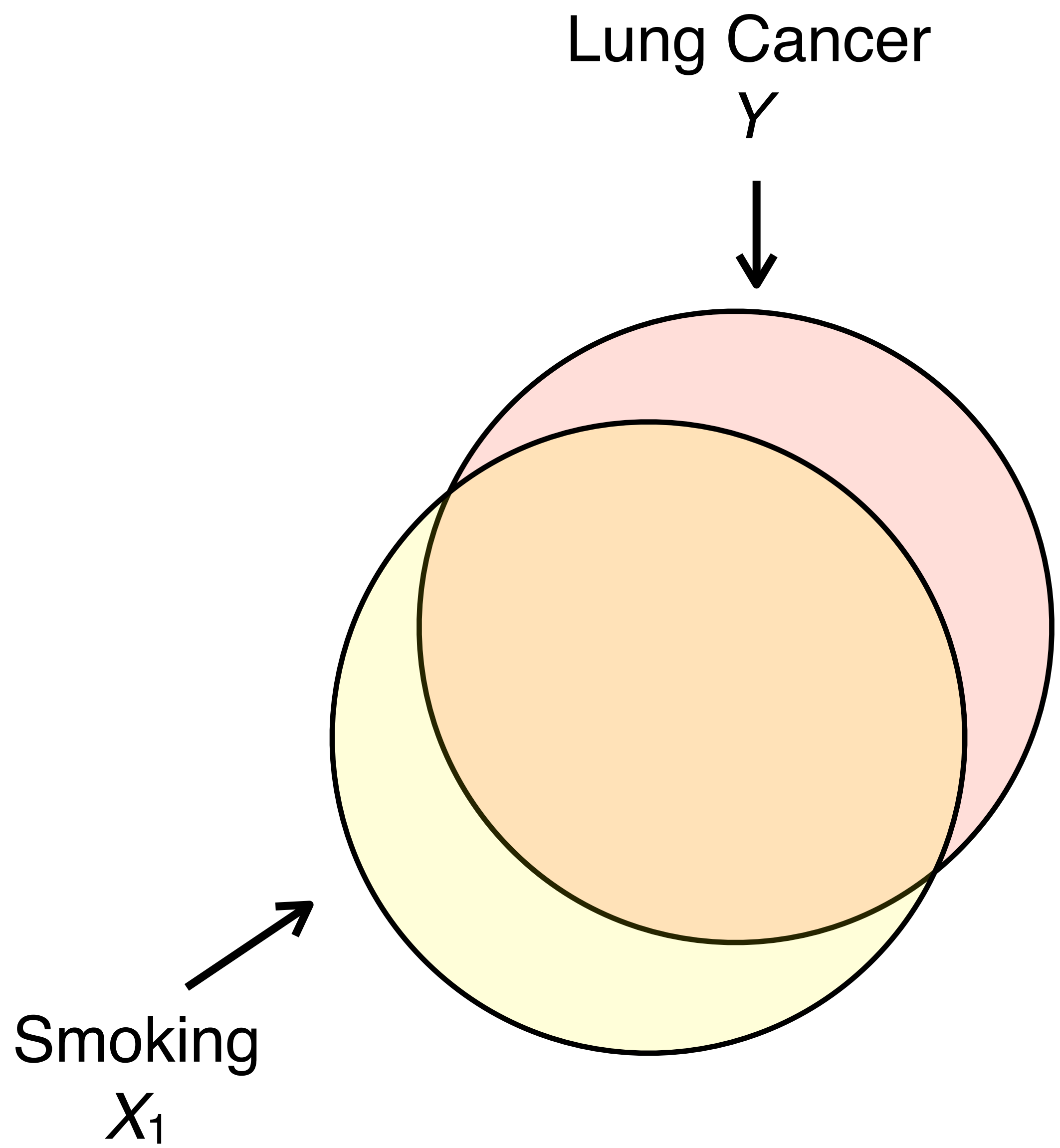




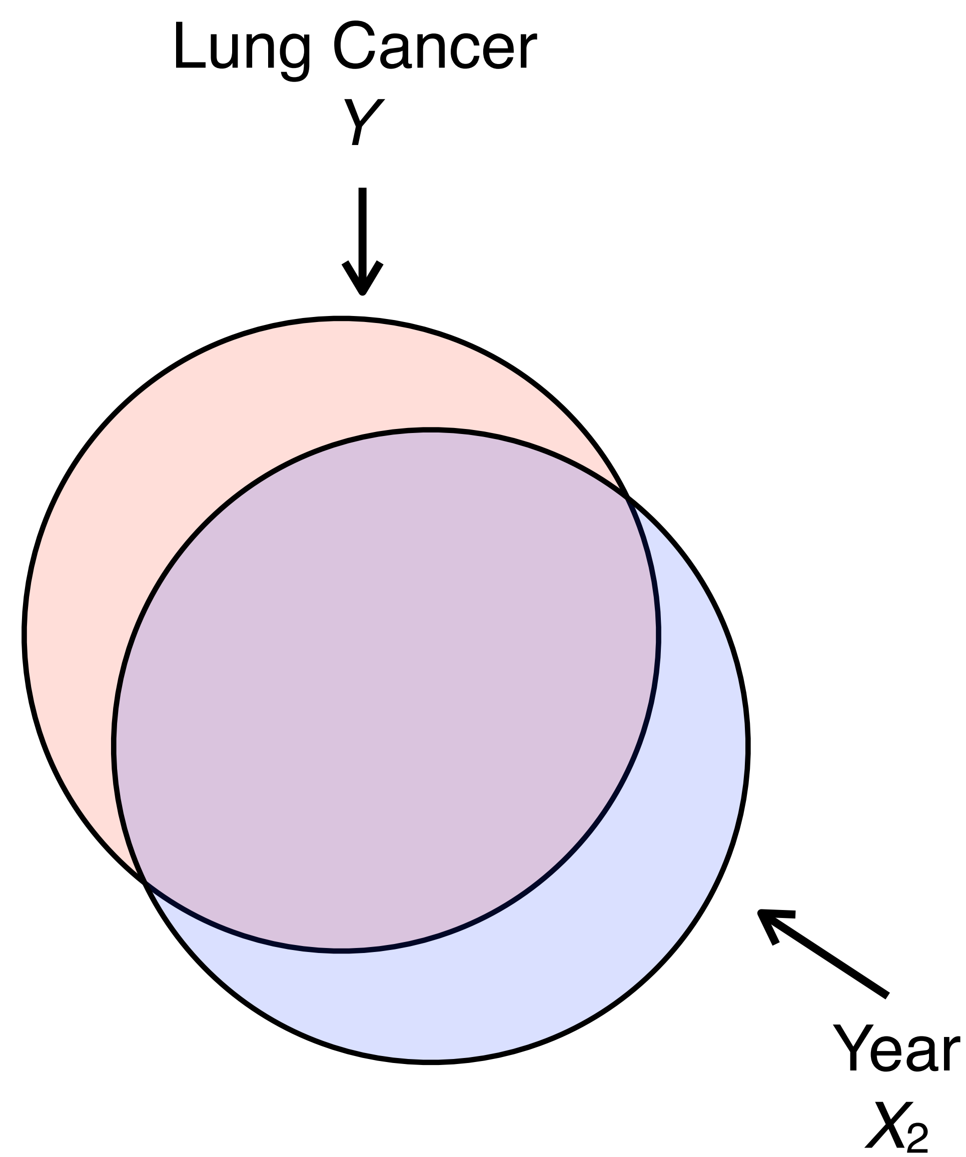

aka
“Ballantine’s”

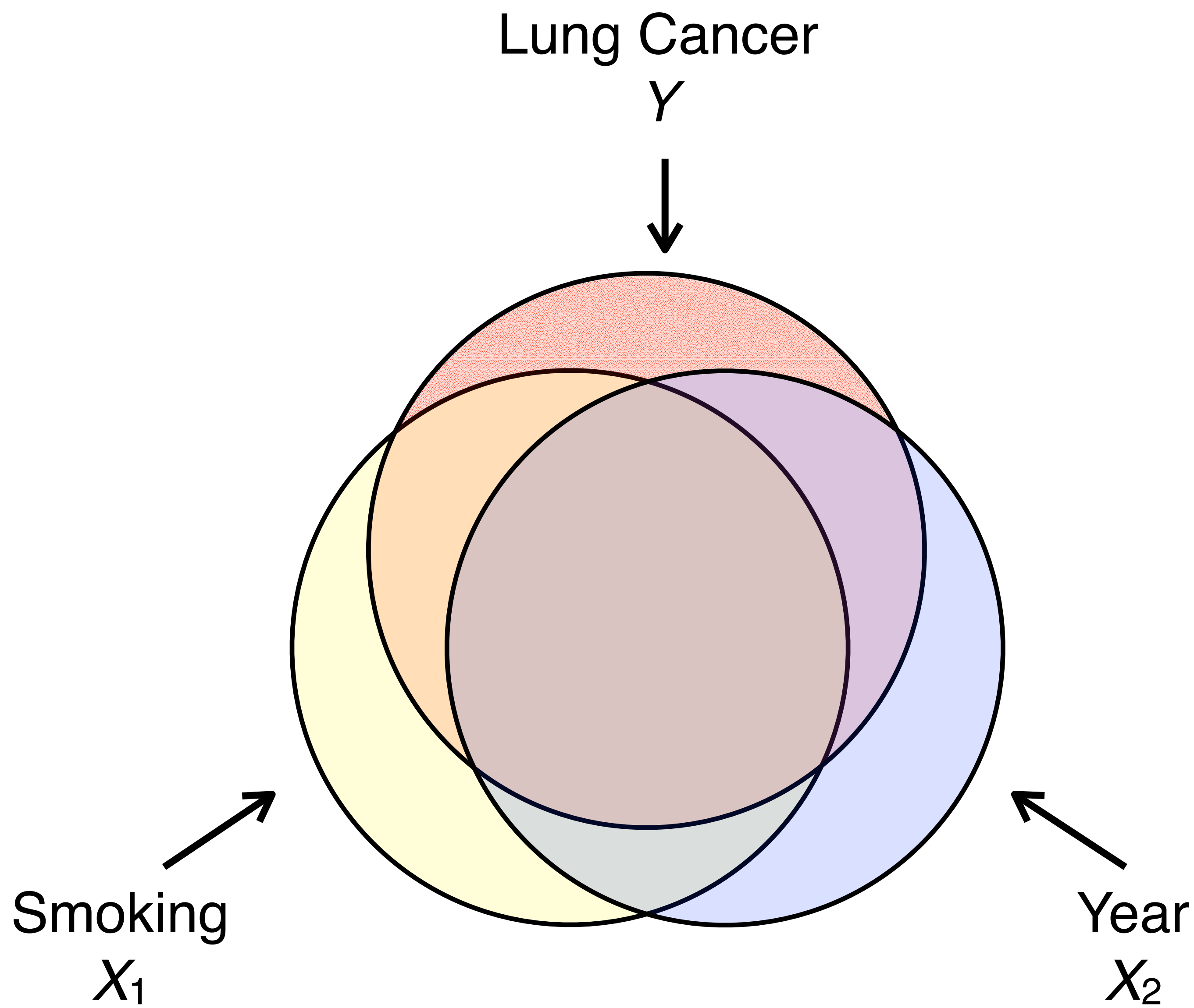


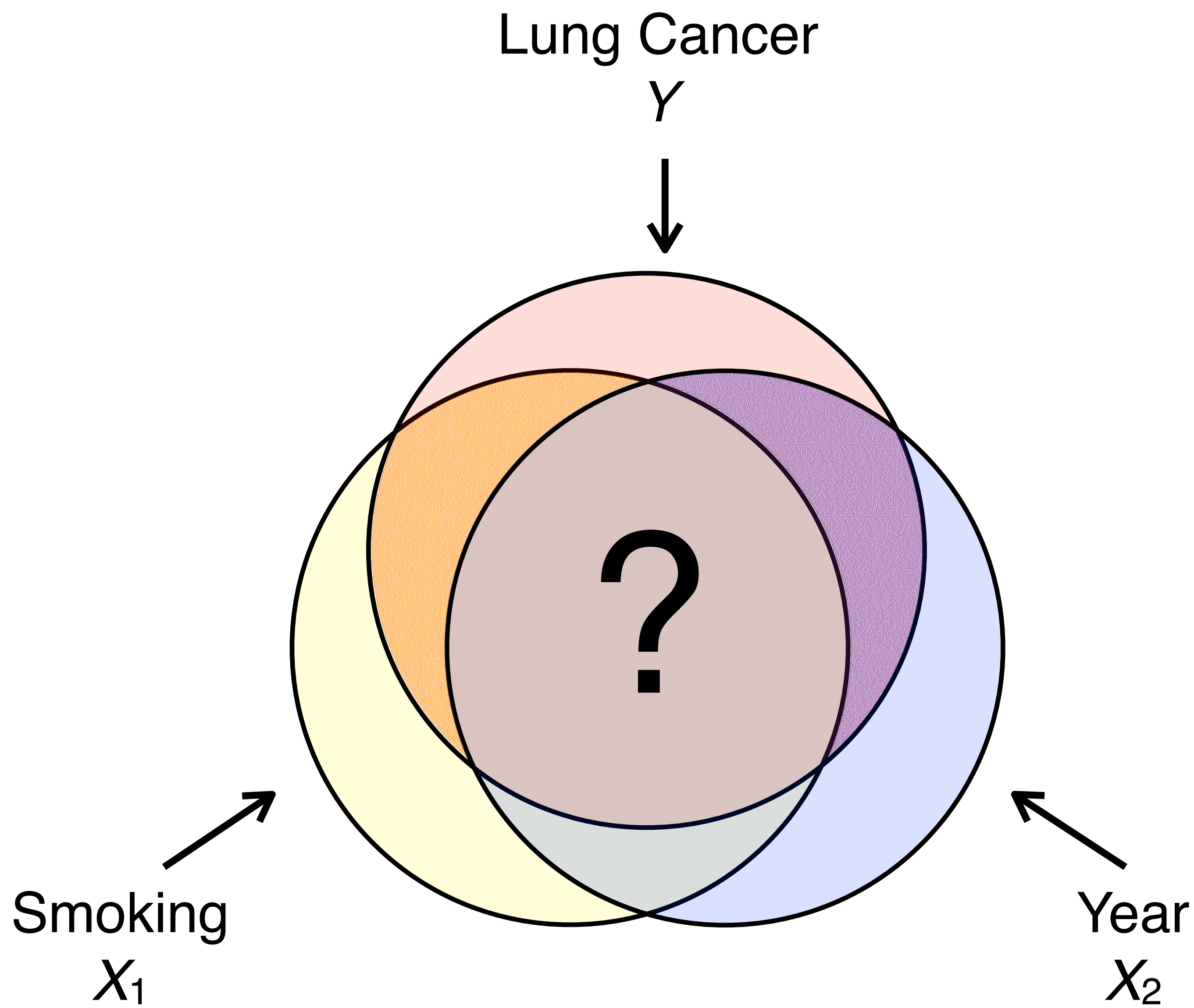

Lung Cancer
Y
Smoking
$X_1$
Year
$X_2$



Lung Cancer
$Y$
Smoking
$X_1$

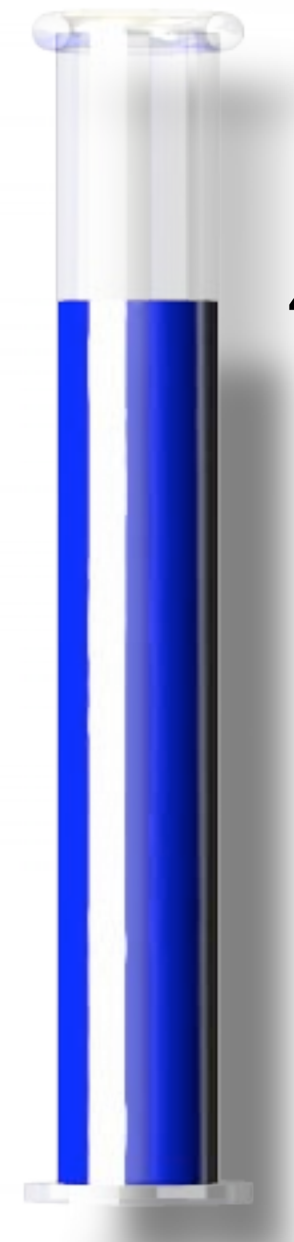

Lung Cancer
$Y$
Year
$X_2$


Lung Cancer
Y
Smoking
X1
Year
X2


Lung Cancer
Y
?
Smoking
X1
Year
X2


79 cm


1.38 mL


4.74 mL

# Definition of Multivariate Analysis

- an inquiry into the structure of interrelationships amongst multiple measures
- Three main multivariate methods covered:
    1. Multiple regression.
    2. Discriminant analysis.
    3. Factor analysis.

# Embedding data analysis in the research process

- research questions and research design help determine what analyses to use

- results of analyses are interpreted with reference to the research area

- an important aim is to *minimise information loss* between collecting data and drawing conclusions

# Aphorisms


Data do not know where they came from

Results from data analyses do not know where they came from

# Matrices

Multivariate stats implies the existence of matrix data.

So it’s important to become familiar with the manipulation of matrices and with the translation of formulas into and out of matrix notation.

A data matrix may be defined simply as a *rectangular table of numbers* on which it's legitimate to perform matrix algebra.

If a table of numbers is to be considered a matrix, it must be arranged in an orderly fashion. A necessary characteristic is that any number that is part of a matrix has a tag that specifies which row and column of the matrix it belongs to.

$$x_{ij}$$

is the value that belongs in row *i* and column *j* of matrix X.

$j$ often refers to which *variable* is involved

$i$ tells which person or other experimental unit is referred to

| Student | Variable A | Variable B | Variable C | Variable D |
|---|---|---|---|---|
| 1 | 1 | 500 | 3.2 | 1 |
| 2 | 1 | 420 | 2.5 | 2 |
| 3 | 2 | 650 | 3.9 | 1 |
| 4 | 2 | 550 | 3.5 | 2 |
| 5 | 3 | 480 | 3.3 | 1 |
| 6 | 3 | 600 | 3.2 | 2 |

In a matrix, all rows have the same number of entries, and all columns have the same number of entries. The entries may be zero, but the row and column designations can't be empty. So - to be a matrix, each entry must occupy a definite row and column, and all the entries must be filled.

**Note**: The requirement that a matrix is complete is the reason that missing data is a concern in multivariate stats. If a matrix has missing entries, then it's not really a matrix, and the matrix mathematics applies only approximately at best!

In a matrix, all rows have the same number of entries, and all columns have the same number of entries. The entries may be zero, but the row and column designations can't be empty. So - to be a matrix, each entry must occupy a definite row and column, and all the entries must be filled.

- The set of scores of each of 271 people on 43 tests.
- The number of responses of a subject observed under all combinations of four stimulus intensities and three durations of food deprivation.
- The per capita income, percentage of owner-occupied homes, and average number of years of education of people living in each of the cities having populations of more than 10,000.
- The number of messages sent from individual *i* to individual *j*.

In each of these examples, the data numbers belong in a particular row and column designation (cell) and, given this designation, we know the information (the value of the variable) that belongs in it.

In the terminology of ANOVA: the row variable and the column variable must be *crossed* for the table to be a matrix.

In order to be able to refer to single numbers as well as matrices, appropriate terms are needed.

A single number, or variable whose value is a single number, is called a *scalar*.

- The number 2 is a scalar, as is the number pi or the gross national product of Australia.

A matrix having a single row or column, or any other one-dimensional list of numbers, is called a *vector*.

- The numbers (3, 5, 7) are a vector – so is the list of scores on a test, or all the scores of a person on several tests.

We can think of a *matrix* as a two-dimensional array, a *vector* as a one-dimensional array, and a *scalar* as a zero-dimensional array.

Matrices come in all sizes and degrees of rectangularity. They may have one to an infinite number of rows and columns. They may have equal numbers of rows and columns, in which case they're referred to as *square matrices*.

The size of a matrix is referred to as its *order*, and is given as a pair of numbers, the first being the number of rows: two by three, *m* by *n*, 1 x 2, *r* x 4, 47 x 243, and so on.

A matrix presented as a table is usually enclosed in large brackets, as if there were danger of it escaping.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \qquad m \times n$$

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1,243} \\ z_{21} & z_{22} & \cdots & z_{2,243} \\ \vdots & \vdots & \ddots & \vdots \\ z_{47,1} & z_{47,2} & \cdots & z_{47,243} \end{bmatrix} \qquad 47 \times 243$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \qquad 2 \times 3$$

# Data Matrix

| Student | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|
| 1 | 1 | 500 | 3.2 | 1 |
| 2 | 1 | 420 | 2.5 | 2 |
| 3 | 2 | 650 | 3.9 | 1 |
| 4 | 2 | 550 | 3.5 | 2 |
| 5 | 3 | 480 | 3.3 | 1 |
| 6 | 3 | 600 | 3.25 | 2 |

# Correlation Matrix

| | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| $X_2$ | 1.00 | 0.85 | -0.13 |
| $X_3$ | 0.85 | 1.00 | -0.46 |
| $X_4$ | -0.13 | -0.46 | 1.00 |

# Variance-Covariance Matrix

| | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| $X_2$ | 7026.66 | 32.80 | -6.00 |
| $X_3$ | 32.80 | 0.21 | -0.12 |
| $X_4$ | -6.00 | -0.12 | 0.30 |

# Sums of Squares and Cross Products Matrix

| | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|
| $X_2$ | 35133.33 | 164.00 | -30.00 |
| $X_3$ | 164.00 | 1.05 | -0.59 |
| $X_4$ | -30.00 | -0.59 | 1.50 |

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 500 & 3.2 & 1 \\ 2 & 1 & 420 & 2.5 & 2 \\ 3 & 2 & 650 & 3.9 & 1 \\ 4 & 2 & 550 & 3.5 & 2 \\ 5 & 3 & 480 & 3.3 & 1 \\ 6 & 3 & 600 & 3.25 & 2 \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 1.00 & 0.85 & -0.13 \\ 0.85 & 1.00 & -0.46 \\ -0.13 & -0.46 & 1.00 \end{bmatrix}$$

$$\mathbf{\Sigma} = \begin{bmatrix} 7026.66 & 32.80 & -6.00 \\ 32.80 & 0.21 & -0.12 \\ -6.00 & -0.12 & 0.30 \end{bmatrix}$$

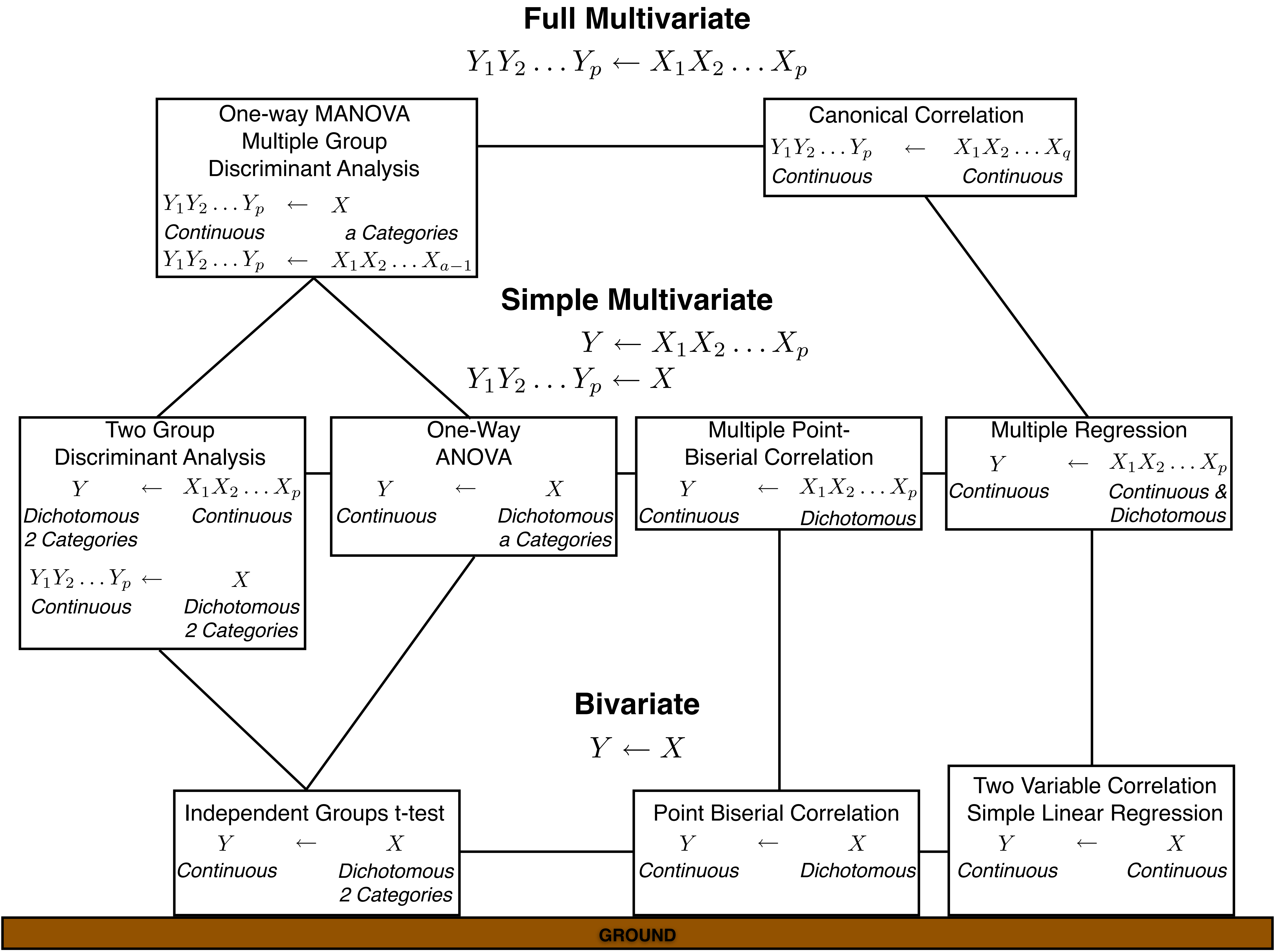
$$\mathbf{S} = \begin{bmatrix} 35133.33 & 164.00 & -30.00 \\ 164.00 & 1.05 & -0.59 \\ -30.00 & -0.59 & 1.50 \end{bmatrix}$$

# Multivariate Family Tree

The family tree is useful because it shows the links among many of the data analytic methods.

In the family tree, there are two basic ‘dimensions’:

1. The first has three levels and concerns the types of relationships between sets of variables, i.e. what variables and how many are related.
2. The second concerns the level of measurement of the variables.


Full Multivariate
$Y_1Y_2\ldots Y_p \leftarrow X_1X_2\ldots X_p$
One-way MANOVA
Multiple Group
Discriminant Analysis
$Y_1Y_2\ldots Y_p \leftarrow X$
Continuous
a Categories
$Y_1Y_2\ldots Y_p \leftarrow X_1X_2\ldots X_{a-1}$
Canonical Correlation
$Y_1Y_2\ldots Y_p \leftarrow X_1X_2\ldots X_q$
Continuous
Continuous
Simple Multivariate
$Y \leftarrow X_1X_2\ldots X_p$
$Y_1Y_2\ldots Y_p \leftarrow X$
Two Group
Discriminant Analysis
$Y \leftarrow X_1X_2\ldots X_p$
Dichotomous
2 Categories
Continuous
$Y_1Y_2\ldots Y_p \leftarrow X$
Continuous
Dichotomous
2 Categories
One-Way
ANOVA
$Y \leftarrow X$
Continuous
Dichotomous
a Categories
Multiple Point-
Biserial Correlation
$Y \leftarrow X_1X_2\ldots X_p$
Continuous
Dichotomous
Multiple Regression
$Y \leftarrow X_1X_2\ldots X_p$
Continuous
Continuous &
Dichotomous
Bivariate
$Y \leftarrow X$
Independent Groups t-test
$Y \leftarrow X$
Continuous
Dichotomous
2 Categories
Point Biserial Correlation
$Y \leftarrow X$
Continuous
Dichotomous
Two Variable Correlation
Simple Linear Regression
$Y \leftarrow X$
Continuous
Continuous
GROUND

# The Ground

## The family tree is embedded in the ground:

- Research Questions - Purposes of Research
- Raw Data Layout
- Data Summaries
- Statistical testing strategies.