

## Class Information

**Instructor:** Chong Gu

**Office Hours:** 12:30–1:30 MWF, MATH 202, or by appointment.

**Course Work:** *Homeworks* (10%), *quizzes* (10%), *midterms* ( $2 \times 25\%$ ), and the *final* (30%).

**Textbook:** *Probability and Statistics for Engineering and the Sciences (8 Ed)*, by Devore

**Software:** R (<http://cran.r-project.org>)

The **calendar**, **assignments**, and **lecture slides** are to be found at the class homepage:

<http://www.stat.purdue.edu/~chong/stat511>

## Open-Source Software: R

R is an programming environment for data analysis and graphics not unlike S/Splus. R is packaged nicely for Linux, Windows, and Mac systems, easy to install.

R is open-source, so it is free to all. R is easily extensible and there exist hundreds of add-on packages; some of you may one day write your own R packages.

Among the most useful R commands are `help.start()`, which invokes the on-line documentation, and `q()`, which quits R.

Links to some tutorial materials are listed in the class page. These are among the many R resources found at the master CRAN site

<http://cran.r-project.org> and its mirrors.

## Introductory Example – I

A psychologist once observed 287 mothers within 4 days and noticed that 237, or 83%, held babies on the left.

Would handedness explain?

No, as 83% right-handed did this, so did 78% left-handed.

*“Newborns may be soothed by the adult heartbeat.”*

How to prove or disprove this?

In a nursery at a NYC hospital, continuous sound of human heartbeat were played for 4 days. Weight gains of babies were measured.

With a new group of babies in the same nursery, weight gains of babies were measured after 4 days without the sound of heartbeat.

*The first group gained more weight, spent much less time crying, but didn't eat more.*

- This is a **designed experiment**.

## Introductory Example – II

A survey was conducted on how often a person wore seat belt. The results are given below.

|         | M  | F  |
|---------|----|----|
| Always  | 37 | 39 |
| Usually | 60 | 58 |
| Often   | 54 | 49 |
| Never   | 64 | 39 |

Is there a difference between the genders?

*“Two tablespoons daily of lecithin may improve your memory.”* An experiment involving 61 people of ages 50 to 80, none suffering from Alzheimer, gave the following results.

|        | Lecithin | Placebo |
|--------|----------|---------|
| Better | 37       | 8       |
| Same   | 4        | 12      |

Do the data support the claim?

- These are examples of **contingency tables**.

## Introductory Example – III

To study the effect of ozone ( $O_3$ ) and sulfur dioxide ( $SO_2$ ) on plant growth, Blue Lake snap beans were grown in open-top chambers. 12 chambers were allocated to 4 treatment combinations, 3 each, with  $O_3$  present/absent and  $SO_2$  present/absent. The total yield of bean pods was measured for each chamber after one month of treatment.

| Ozone   | Sulfur Dioxide |         |
|---------|----------------|---------|
|         | Absent         | Present |
| Absent  | 1.52           | 1.49    |
|         | 1.85           | 1.55    |
|         | 1.39           | 1.21    |
| Present | 1.15           | 0.65    |
|         | 1.30           | 0.76    |
|         | 1.57           | 0.69    |

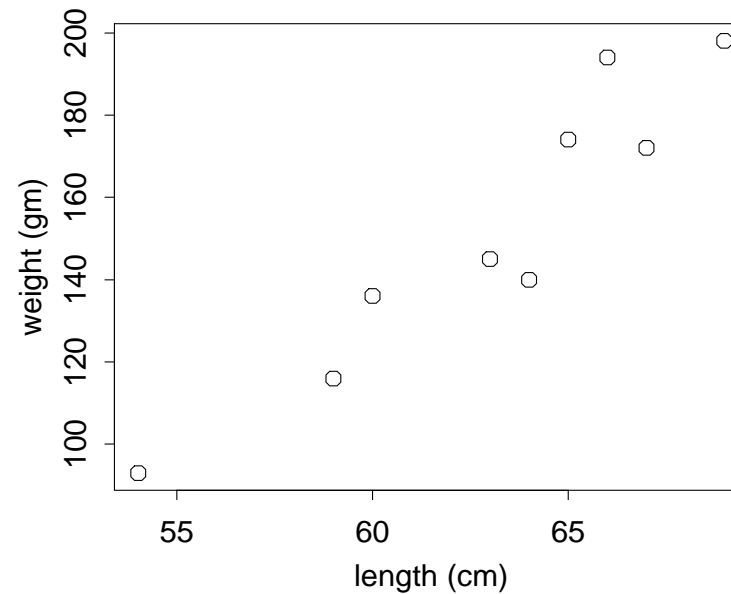
Do the pollutants affect growth?

- This can be analyzed using **analysis of variance** techniques.

## Introductory Example – IV

| Length | Weight |
|--------|--------|
| 60     | 136    |
| 69     | 198    |
| 66     | 194    |
| 64     | 140    |
| 54     | 93     |
| 67     | 172    |
| 59     | 116    |
| 65     | 174    |
| 63     | 145    |

Nine adult females of the snake *Vipera berus* were caught and measured. The lengths and weights are listed on the left and plotted below.



Can one predict the weight from the length?

- This can be analyzed using **regression** techniques.

## Introductory Example – V

A patient with a terminal disease, Mr. X, went to see Doctor Y.

**Y:** Mr. X, how can I help you today?

**X:** Well Doctor Y, you see, I happen to have this deadly disease, and I wonder if you can do something to it.

**Y:** Don't worry, you'll be fine in just a few months.

**X:** Are you sure?

**Y:** Sure! I have been practicing this new therapy to treat patients with your disease, which is 10% successful according to the literature. Now you are the lucky number ten walking into my door, as the previous nine all became statistics.

- The past is at best **independent** of the future.

## Introductory Example – VI

The admission process at a certain academic program in an university is under scrutiny, as the data over the past few years show the following.

|   | Acc. | Rej. | A. Rate |
|---|------|------|---------|
| F | 61   | 462  | .1166   |
| M | 138  | 631  | .1795   |

Is there any evidence of sex discrimination?

It can't be just by chance, but ...

A closer look at the program structure reveals that the data can be segregated into two separate specialties.

|   | Acc. | Rej. | A. Rate |
|---|------|------|---------|
| F | 9    | 19   | .3214   |
| M | 89   | 187  | .3225   |

|   | Acc. | Rej. | A. Rate |
|---|------|------|---------|
| F | 52   | 443  | .1051   |
| M | 49   | 444  | .0994   |

- This phenomenon is known as the **Simpson's paradox**.