

# Lecture 1 Introduction To BIO210

BIO210 Biostatistics

---

Xi Chen

Spring, 2022

School of Life Sciences  
Southern University of Science and Technology



南方科技大学生命科学学院  
SUSTech · SCHOOL OF  
**LIFE SCIENCES**

## Administrative

**Instructor:** Xi Chen, chenx9@sustech.edu.cn

**Office:** Research Building B, Room 414

### Teaching Assistants:

Yunlong Zhang: 12131328@mail.sustech.edu.cn

Pengfei Yin: 12133074@mail.sustech.edu.cn

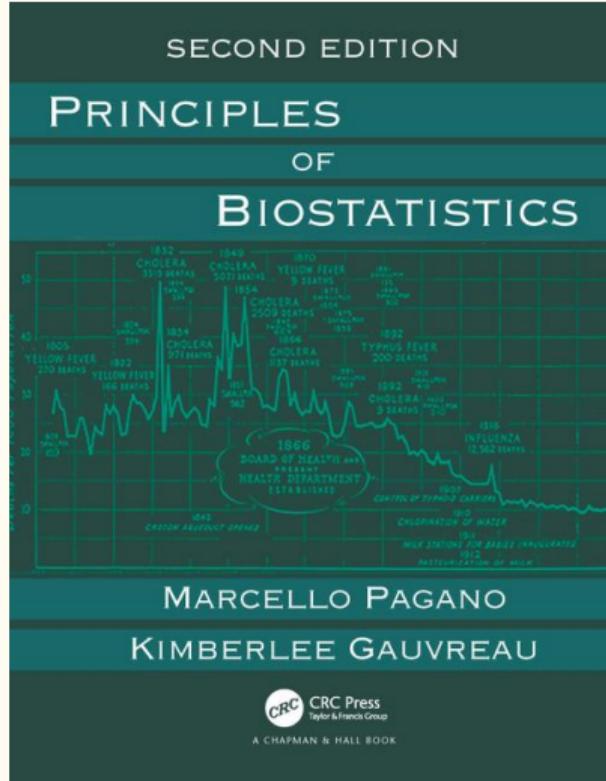
QQ group: 181040126

## Grading system

Attendance	Quizzes ??	Assignments	Mid-term exam	Final exam
10%	10% ??	10%	30%	40%

**Quizzes and exams: open notes, calculators can be used !**

# Textbook



## Principles of Biostatistics, 2nd Edition, by Pagano and Gauvreau

- Pages: 525
- Available from our library
- Not required for the course

## Goals of BIO210

- Introduce basic concepts of statistics to students with no prior knowledge.
- Help students feel justifiably confident of their ability to interpret data/information from research articles and daily lives.
- Select appropriate statistical methods for your problem.
- Help students formulate a statistical problem from real-life situation and use the numerical techniques to solve and extract information from it.

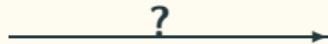
## Difference to MA212

- Focused on data from basic biology and medicine.
- Focused on application.
- Focused on statistics.

# Introduction to biostatistics

- What is statistics?
  - Statistics is the science of getting generalisable knowledge out of a set of data.
  - Statistics is the science whereby inferences are made about specific random phenomena on the basis of relatively limited sample material.
  
- Why should biologists care about it?
  1. Daily life
  2. Scientific research in biology

# Design of the phone buttons



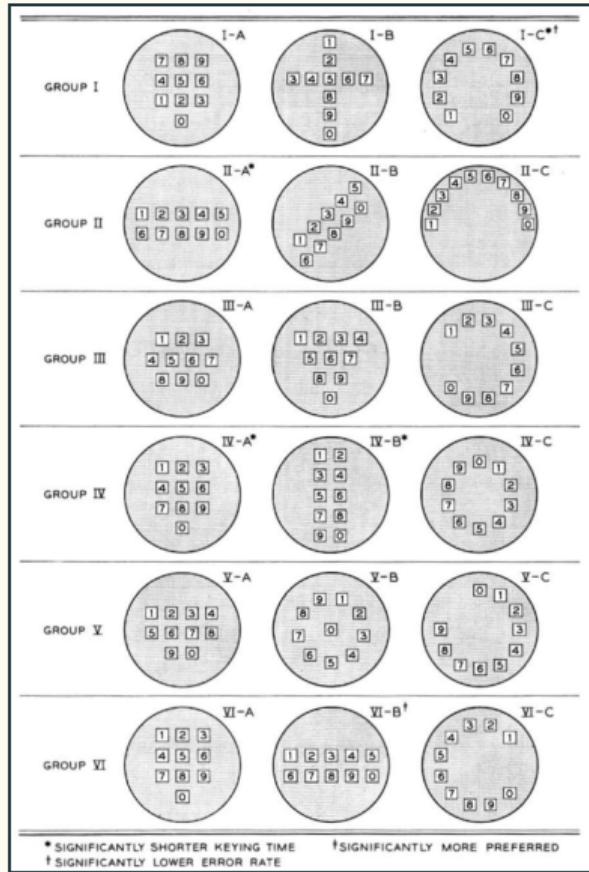
## Human Factors Engineering Studies of the Design and Use of Pushbutton Telephone Sets

By R. L. DEININGER

(Manuscript received February 16, 1960)

*From the user's point of view, what are the desirable characteristics of pushbuttons for use in 500-type telephone sets? The studies reported bear on this question and also on questions of how people process information when keying telephone numbers. Four categories of design features were studied: key arrangement, force-displacement characteristics, button-top design and central office factors. The results indicate that considerable latitude exists for key set design in terms of user performance; however, the preference judgments are more selective. The studies also showed that the manner in which the person acquired and keyed the telephone number influenced performance appreciably.*

# Design of the phone buttons



ARRANGEMENT	KEYING TIME (SECONDS)	PER CENT ERRORS	RANKING FOR	RANKING AGAINST
THREE-BY-THREE PLUS ONE	6.01	2.5	3RD	2ND
TWO HORIZONTAL ROWS	6.17	2.3	1ST (MOST)	4TH
TWO VERTICAL COLUMNS	6.12	1.3	5TH (LEAST)	1ST (MOST)
TELEPHONE	5.90	2.0	2ND	5TH (LEAST)
SPEEDOMETER	5.97	3.0	4TH	3RD

# Design of the phone buttons

The Journal of Applied Psychology  
Vol. 38, No. 3, 1953

## Expected Locations of Digits and Letters on Ten-Button Keysets<sup>1</sup>

Mary Champion Lutz

Bell Telephone Laboratories, Murray Hill, New Jersey

and Alphonse Chapanis

The Johns Hopkins University

Although keysets are used on a great variety of machine devices—computers, coding devices, and communications equipment—there appear to be few systematic studies concerned with the design factors that make keysets easy or hard to use. The study reported here deals with one aspect of keyset design, viz., the locations of numbers and letters on individual keys. In addition, we are concerned here with a particular class of keysets—ten-button sets used by long-distance telephone operators—but the results probably can be generalized to other practical situations.

In making long-distance calls, telephone operators use a set of ten keys, arranged in two vertical rows of five, with letters and numbers on the keys as shown in Fig. 1.

To complete a call, the operator usually keys a letter-number combination which looks like this:

815 RE 4-0267

The patterns of errors made by operators suggest that a different arrangement of the letters and numbers on the keys, or of the keys themselves, might help to reduce errors. As a first step in the determination of the best arrangement of the keys and of the letters and numbers on them, we decided to find out



FIG. 1. Arrangement of letters and numbers on a toll operator's keyset.

<sup>1</sup>This study was done at the Bell Telephone Laboratories

**Method**  
*Subjects.* The subjects for this experiment were classified according to (a) age, (b) sex, (c) previous experience on keysets such as appear on computing machines, typewriters, and musical instruments. Three hundred Ss were used, one hundred to answer each of the three questions, each one hundred chosen as in Table I.

*Test Materials.* The test materials consisted of bookletlets containing order arranged in each of the six configurations shown in the top row of Fig. 3. Each configuration appeared on a separate page. In Part I, a random arrangement of the digits 0 to 9 was printed on the page opposite each configuration of circles. In Parts II and III a random arrangement of the alphabet (except the letters Q and Z) was printed on the page opposite each configuration. For Part III see the bookletlets used configurations with numbers already printed in the circles (see Fig. 2). The numbering arrangements selected were

316

Mary Champion Lutz and Alphonse Chapanis

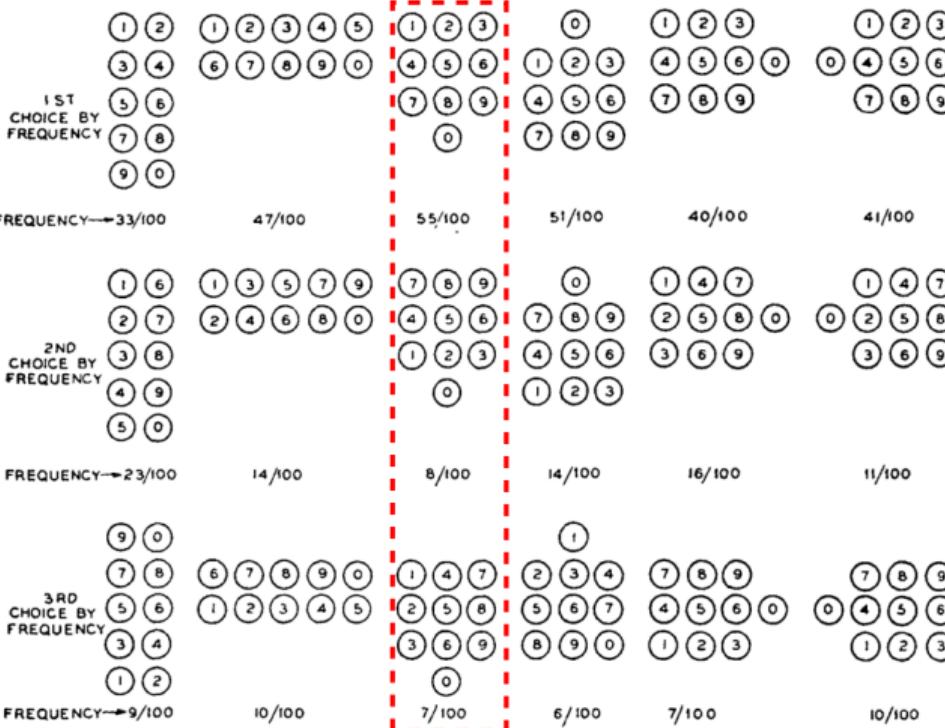
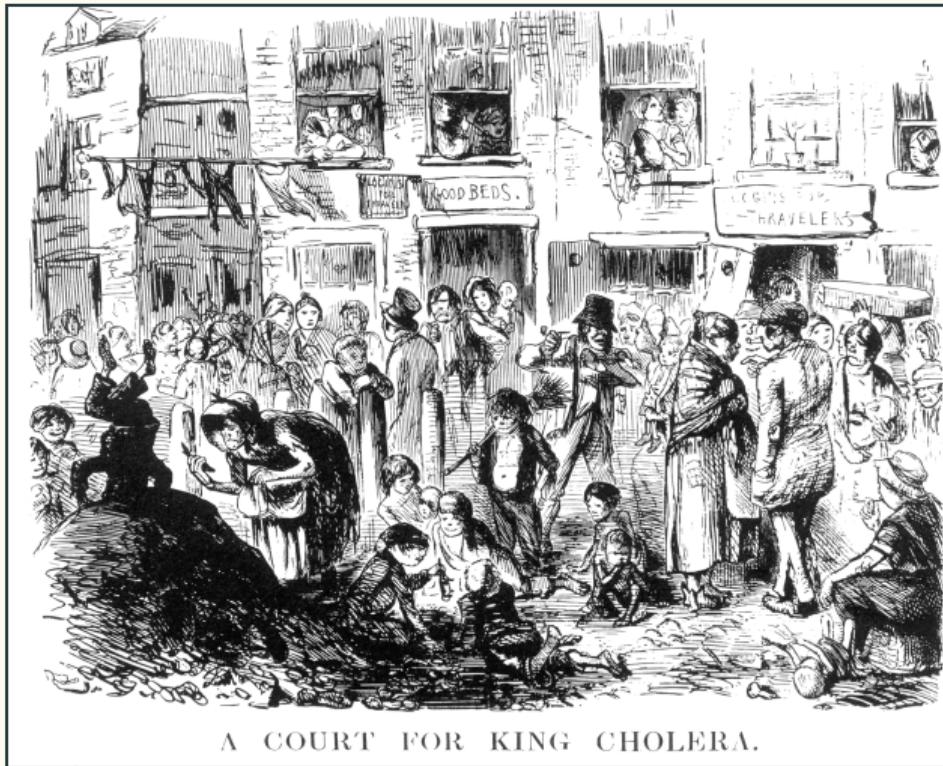


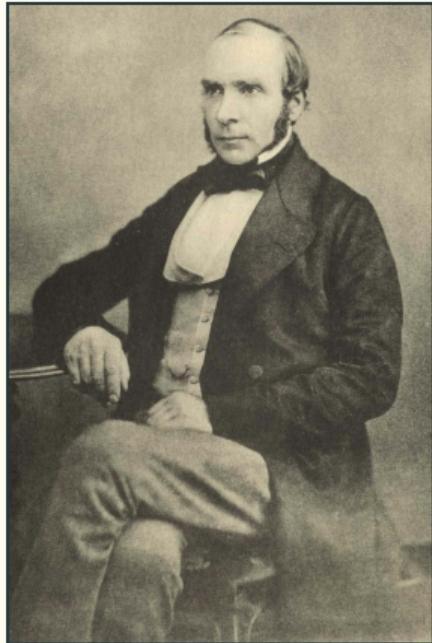
FIG. 3. First three choices by frequency for number arrangements on each of the six configurations tested in Part I

# The Cholera outbreak in London, 1854



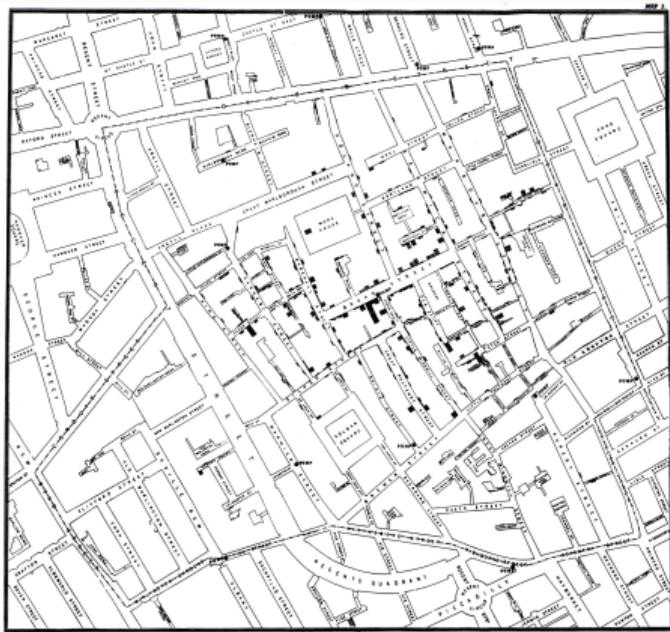
by John Leech (1852) *Punch*

# The Cholera outbreak in London, 1854



# The Cholera outbreak in London, 1854

The spot map:



<http://johnsnow.matrix.msu.edu>



<https://mjdanielson.github.io/Cholera-Map>

# The Cholera outbreak in London, 1854



39 Broadwick Street  
London

# Pfizer Vaccine

The New York Times

Covid-19 Vaccines > | Vaccine Questions Rollout by State Chinese Vaccine Setbacks How 9 Vaccines Work

## New Pfizer Results: Coronavirus Vaccine Is Safe and 95% Effective

The company said it planned to apply for emergency approval from the Food and Drug Administration “within days.”

---

REUTERS

World Business Markets Breakingviews Video More

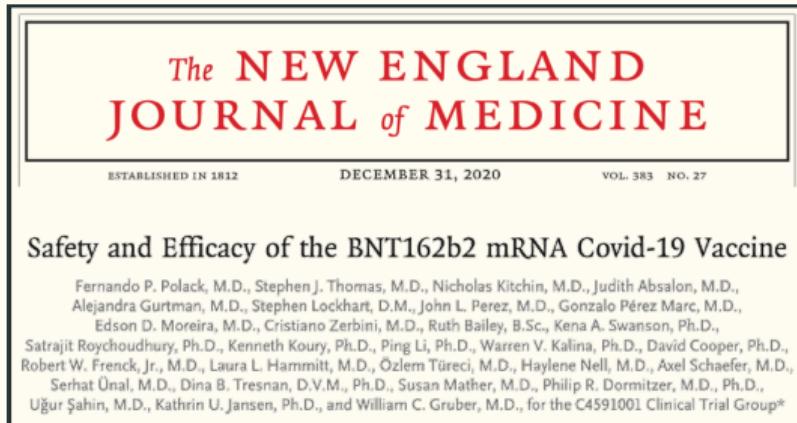
HEALTHCARE & PHARMA NOVEMBER 18, 2020 / 8:55 PM / UPDATED 2 MONTHS AGO

## Instant View: Pfizer ends COVID-19 trial with 95% efficacy

By Reuters Staff 5 MIN READ

(Reuters) - Pfizer Inc said on Wednesday that final results from the late-stage trial of its COVID-19 vaccine showed it was 95% effective, adding it had the required two-months of safety data and would apply for emergency U.S. authorization within days.

# Pfizer Vaccine



**Key results:** A total of 43,548 participants underwent randomization, of whom 43,448 received injections: 21,720 with BNT162b2 and 21,728 with placebo. There were 8 cases of Covid-19 with onset at least 7 days after the second dose among participants assigned to receive BNT162b2 and 162 cases among those assigned to placebo; BNT162b2 was 95% effective in preventing Covid-19 (95% credible interval, 90.3 to 97.6). Similar vaccine efficacy (generally 90 to 100%) was observed across subgroups defined by age, sex, race, ethnicity, baseline body-mass index, and the presence of coexisting conditions. Among 10 cases of severe Covid-19 with onset after the first dose, 9 occurred in placebo recipients and 1 in a BNT162b2 recipient. The safety profile of BNT162b2 was characterized by short-term, mild-to-moderate pain at the injection site, fatigue, and headache. The incidence of serious adverse events was low and was similar in the vaccine and placebo groups.

# Course outline

## Data Presentation

- Types of numerical data
- Tables and graphs
- Measures of central tendency
- Measures of dispersion

# Course outline

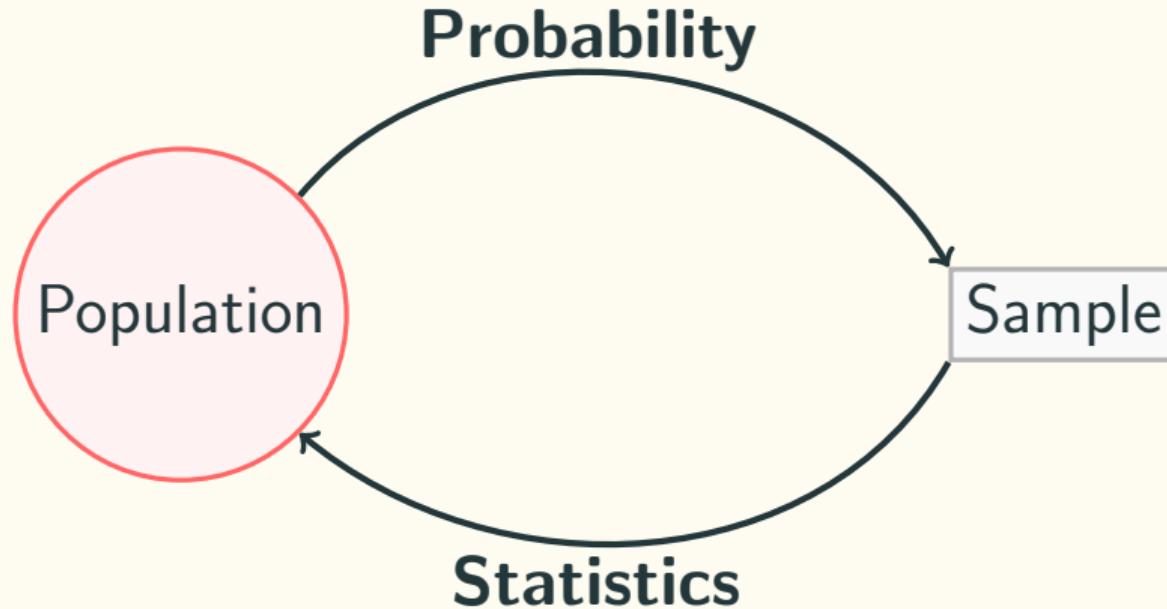
## Probability

- Notation
- Set Notation and Operations
- Interpretations of Probability
- Probability Properties
- Conditional Probability
- Bayes' Theorem
- Independent Events

# Course outline

## Probability Distributions

- Random Variables
- Discrete Probability Distributions
- Continuous Probability Distributions



### Probability vs. Statistics

**Probability:** Previous studies showed that the drug was 80% effective. Then we can anticipate that for a study on 100 patients, in average 80 will be cured and at least 65 will be cured with 99.99% chances.

**Statistics:** Observe that 78/100 patients were cured. We will be able to conclude that we are 95% confident that for other studies the drug will be effective on between 69.88% and 86.11% of patients.

# Course outline

## Estimation

- The Relationship Between Population and Sample
- Sampling Distribution
- Point Estimation
- Maximum Likelihood Estimation

# Course outline

## Confidence Intervals

- Two-sided confidence intervals
- One-sided confidence intervals
- Student's t distribution

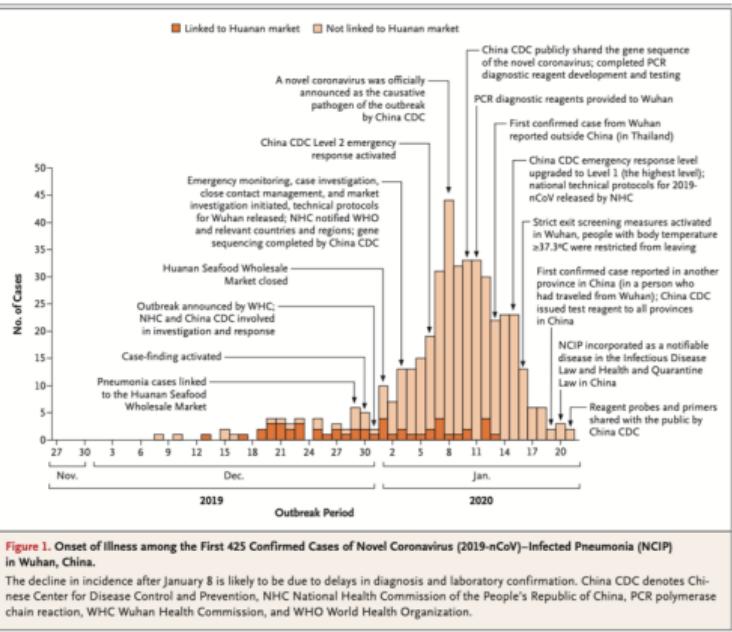
# Course outline

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

## Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia

Qun Li, M.Med., Xuhua Guan, Ph.D., Peng Wu, Ph.D., Xiaoye Wang, M.P.H.,  
Lei Zhou, M.Med., Yeqing Tong, Ph.D., Ruiqi Ren, M.Med.,  
Kathy S.M. Leung, Ph.D., Eric H.Y. Lau, Ph.D., Jessica Y. Wong, Ph.D.,  
Xuesen Xing, Ph.D., Nijuan Xiang, M.Med., Yang Wu, M.Sc., Chao Li, M.P.H.,  
Qi Chen, M.Sc., Dan Li, M.P.H., Tian Liu, B.Med., Jing Zhao, M.Sc.,  
Man Li, M.Sc., Wenxiao Tu, M.Med., Chuding Chen, M.Sc., Lianmei Jin, M.Med.,  
Rui Yang, M.Med., Qi Wang, M.P.H., Suhua Zhou, M.Med., Rui Wang, M.D.,  
Hui Liu, M.Med., Yingbo Luo, M.Sc., Yuan Liu, M.Med., Ge Shao, B.Med.,  
Huan Li, M.P.H., Zhongfa Tao, M.P.H., Yang Yang, M.Med.,  
Zhiqiang Deng, M.Med., Boxi Liu, M.P.H., Zhitao Ma, M.Med.,  
Yanping Zhang, M.Med., Guoqing Shi, M.P.H., Tommy T.Y. Lam, Ph.D.,  
Joseph T.K. Wu, Ph.D., George F. Gao, D.Phil., Benjamin J. Cowling, Ph.D.,  
Bo Yang, M.Sc., Gabriel M. Leung, M.D., and Zijian Feng, M.Med.



"... The mean **incubation period** was 5.2 days (95% confidence interval [CI], 4.1 to 7.0), with the 95th percentile of the distribution at 12.5 days. In its early stages, the epidemic doubled in size every 7.4 days. With a mean **serial interval** of 7.5 days (95% CI, 5.3 to 19), the basic reproductive number was estimated to be 2.2 (95% CI, 1.4 to 3.9). . . ."

# Course outline

## Hypothesis Testing

- Introduction to Hypothesis Testing
- Two-Sided Tests of Hypotheses
- One-sided Tests of Hypotheses
- Types of Error
- Power
- Sample Size Estimation

# Course outline

## Comparison of Two Means and Proportions

- Paired Samples
- Independent Samples
- The behaviour of the p-value

# Course outline

## Analysis of Variance (ANOVA)

- One-Way Analysis of Variance
- Multiple Comparisons Procedures

## Contingency Tables

- The Chi-Square Test
- McNemar's Test
- The Odds Ratio

## Correlation & Linear Regression

- The Two-Way Scatter Plot
- Pearson's Correlation Coefficient
- Regression Concepts
- Linear Regression Model
- Simple Linear Regression

## Nonparametric Methods

- The Sign Test
- The Wilcoxon Signed-Rank Test
- The Wilcoxon Rank Sum Test
- Advantages and Disadvantages of Nonparametric Methods

### Other practical data analysis techniques

- Monte Carlo Simulation
- Bootstrapping methods
- Permutation tests

# Course outline

## Summary and Review

- Summary of Statistical Techniques
- Choose the Correct Statistical Technique

## Course outline

### What is this course **NOT** about

- Bayesian statistics
- Mathematical proof
- Implementation
- How and where to find data