BIOSTATISTICS PRACTICAL

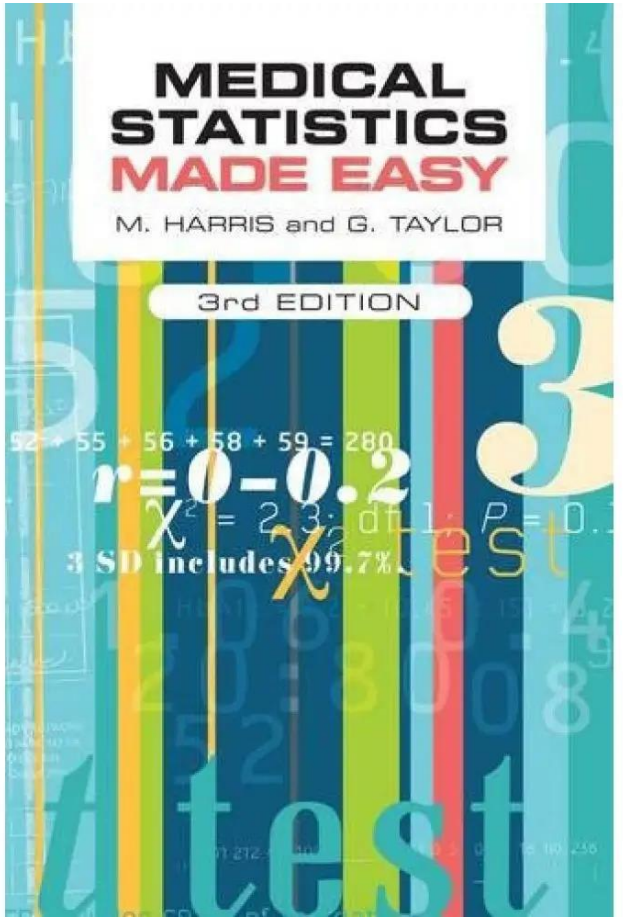
Statistics is the grammar of science.
Karl Pearson
(1857 – 1936, influential English mathematician)

Learning outcome

- 1) Describe, discuss and compare the difference level of measurements.
- 2) Describe and compare statistics & parameter.
- 3) Describe and discuss probability distribution.
- 4) Create and enter data using statistical software.
- 5) Describe, discuss, analyze & report descriptive statistics.
- 6) Describe, compare, run & report different type of univariate/bivariable analyses

JARGON

- Error type I and II
- Power
- Level of significance
- One sided vs two-sided test
- Effect size
- Sample size
- Clinical significance vs statistical significance
- Parameter vs statistics
- Variables

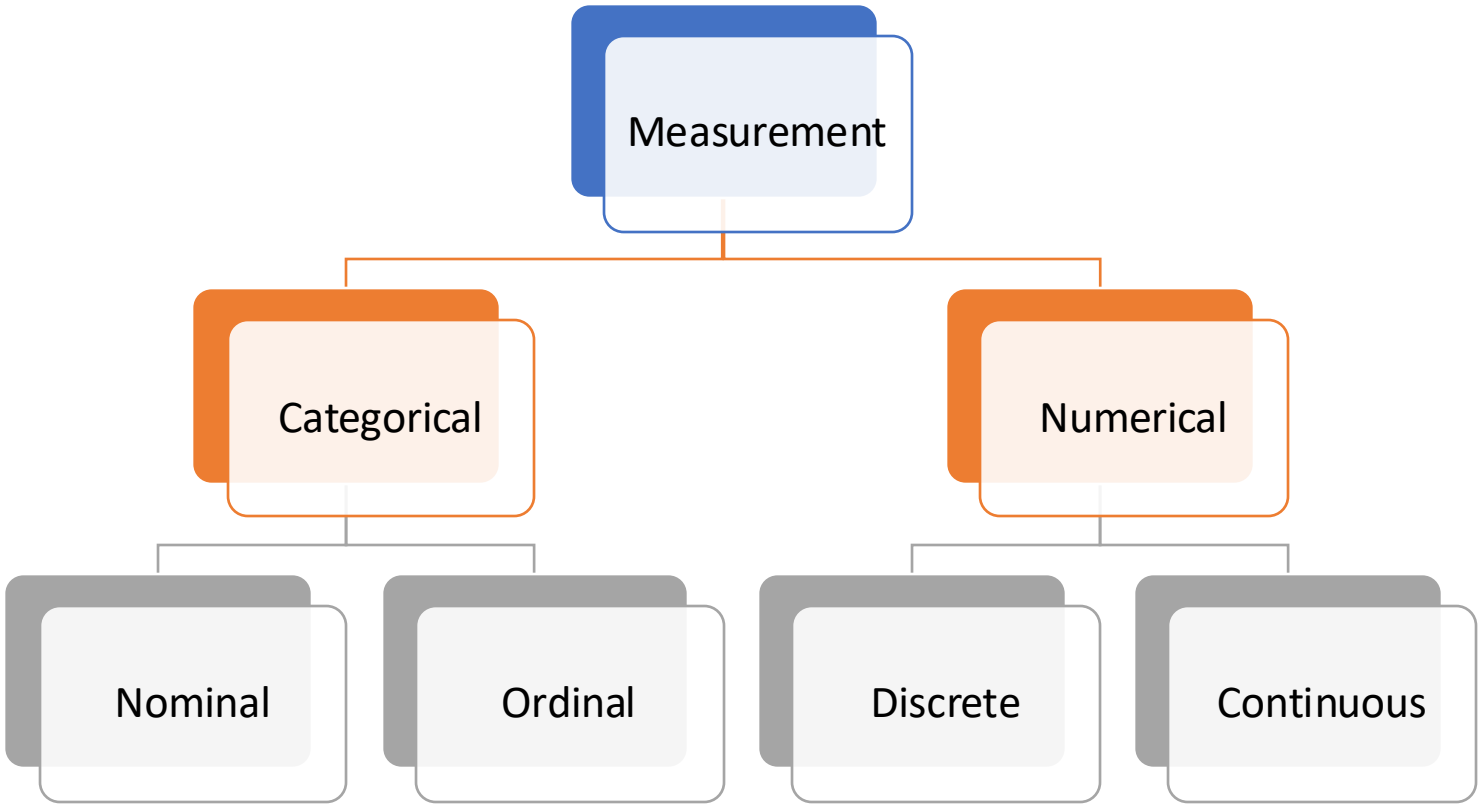


Biostatistics

"Application of statistical principles to questions ad problems in medicine, public health or biology"

Boston University School of Public Health. Basic Concepts for Biostatistics

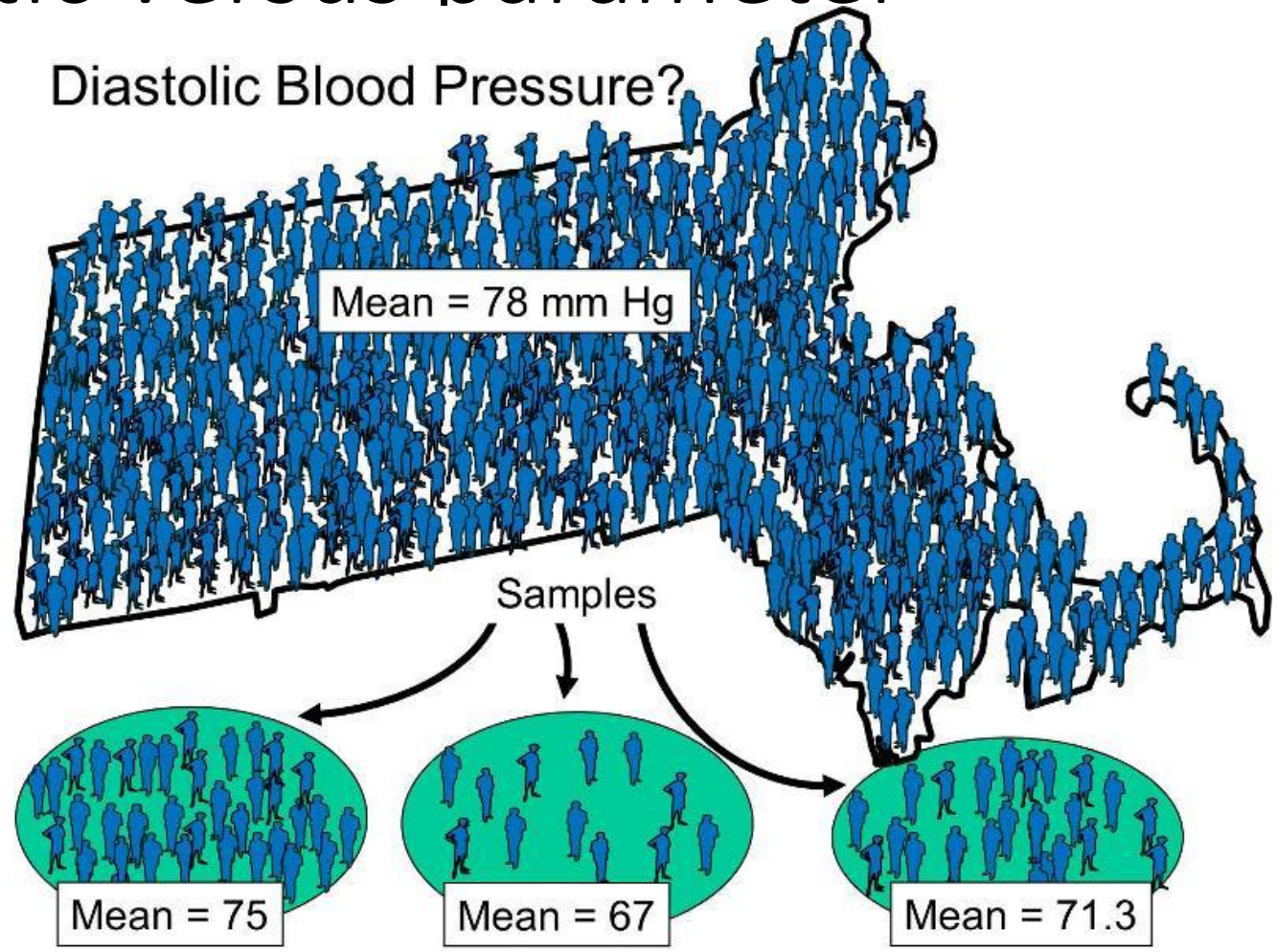
Levels of measurement



- *Gender*
- *Staging of cancer*
- *Number of household members*
- *Temperature*

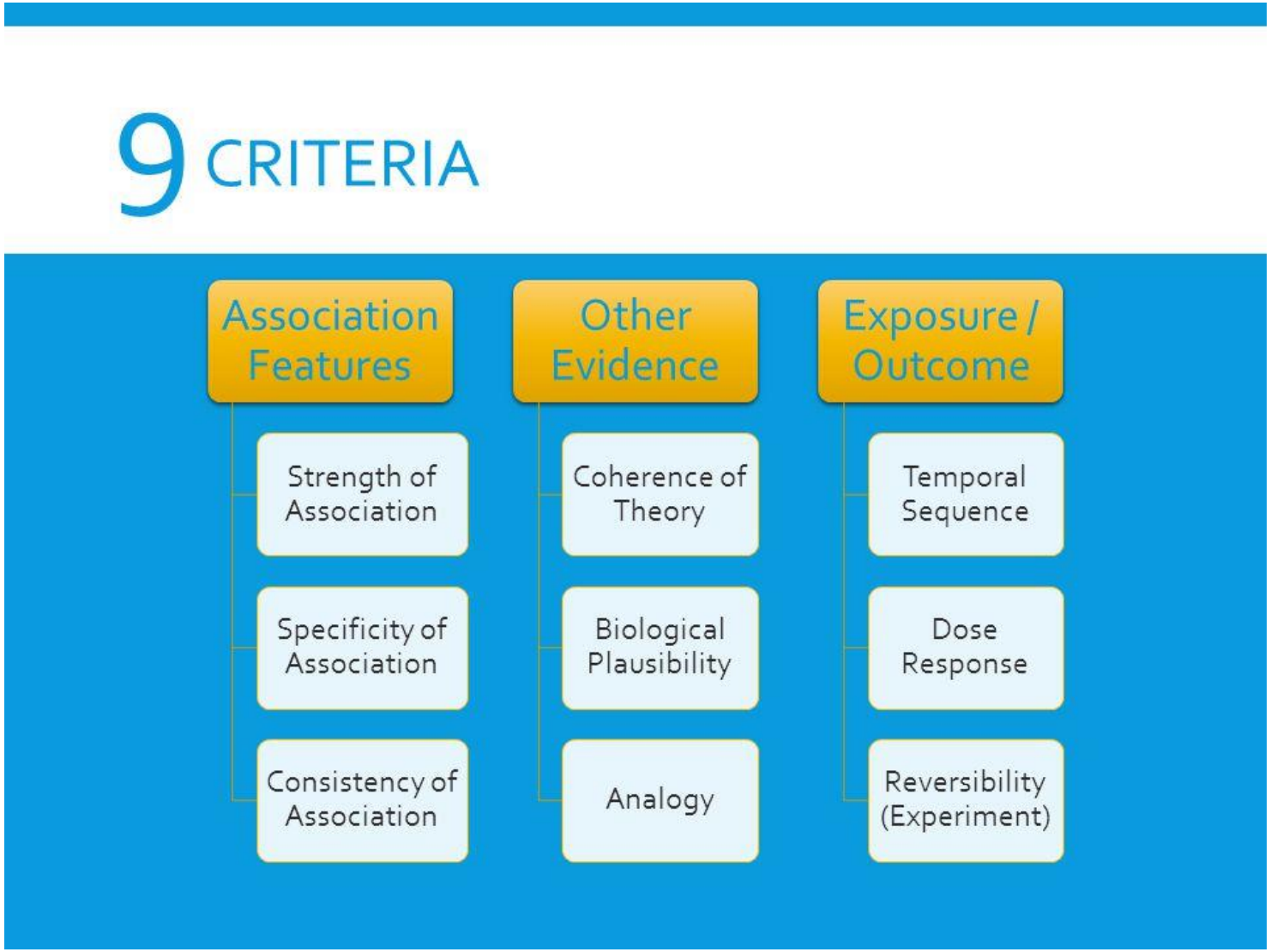
Iceberg phenomenon

Statistic versus parameter

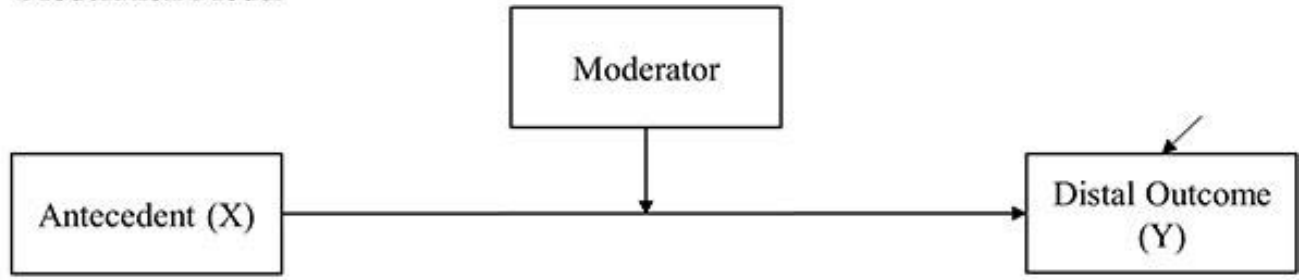


Adapted from SPH Boston University website

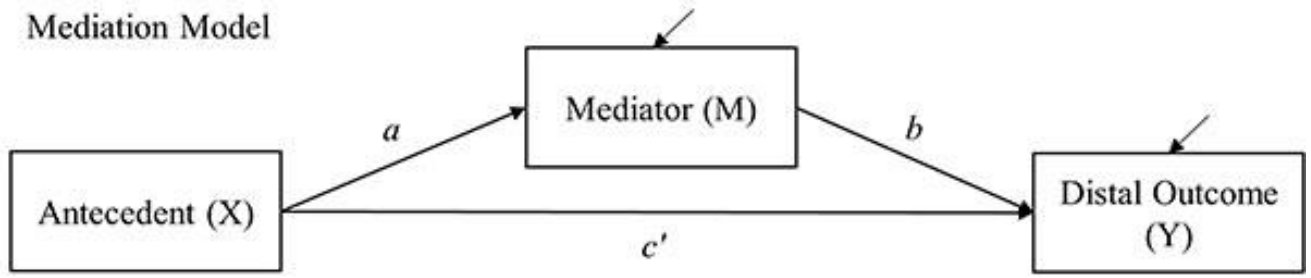
Bradford Hill Criteria



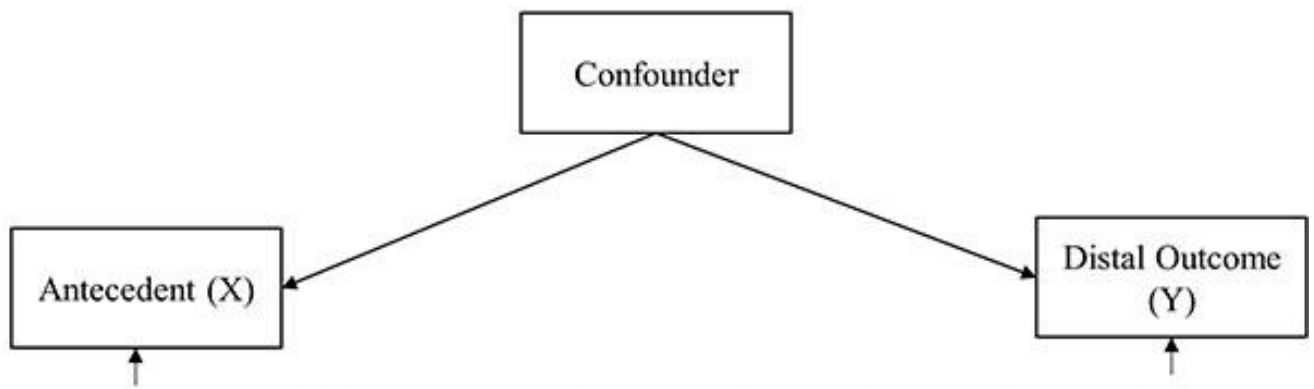
Moderation Model



Mediation Model

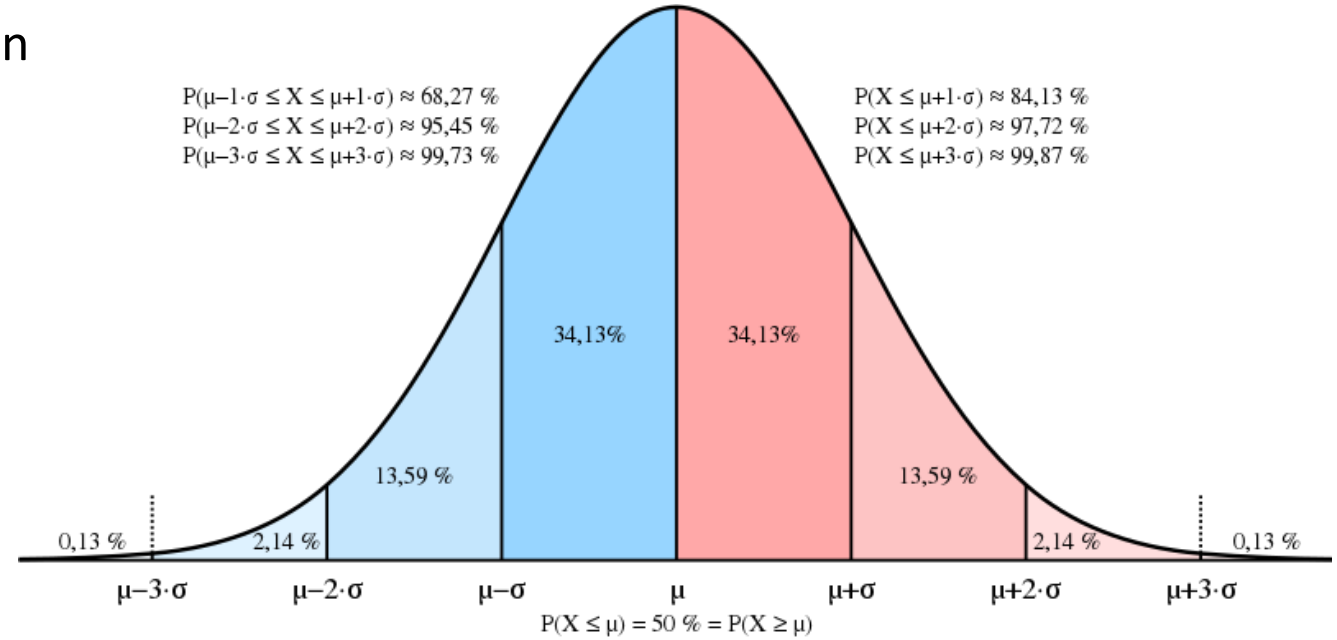


Confounding Model



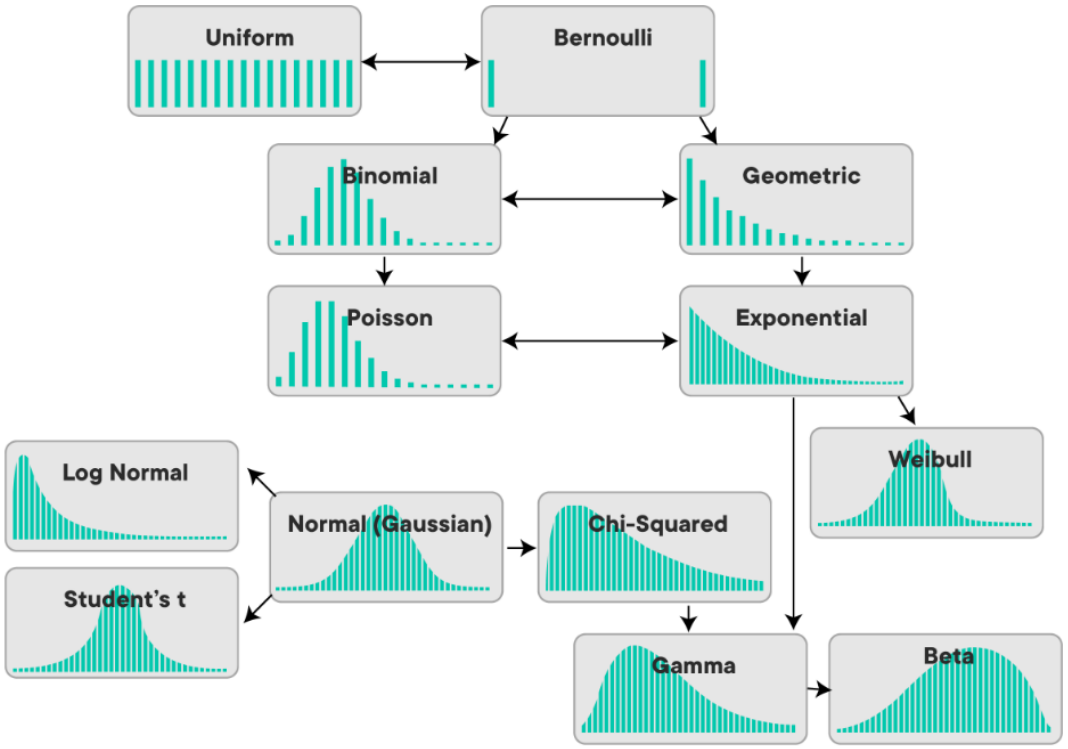
PRINCIPLE OF STATISTICS

- Sampling technique to represent population
  - Random sampling
  - Sample size
- Distribution of data
  - Normal distribution
  - Central limit theorem
  - Other distribution
  - Assumption for most statistical inference
  - Assumption for many statistical test



It is a capital mistake to theorize before one has data.
Sir Arthur Ignatius Conan Doyle
(1859 – 1930, Scottish physician and writer, most noted for his stories about Sherlock Holmes)

Distribution of data



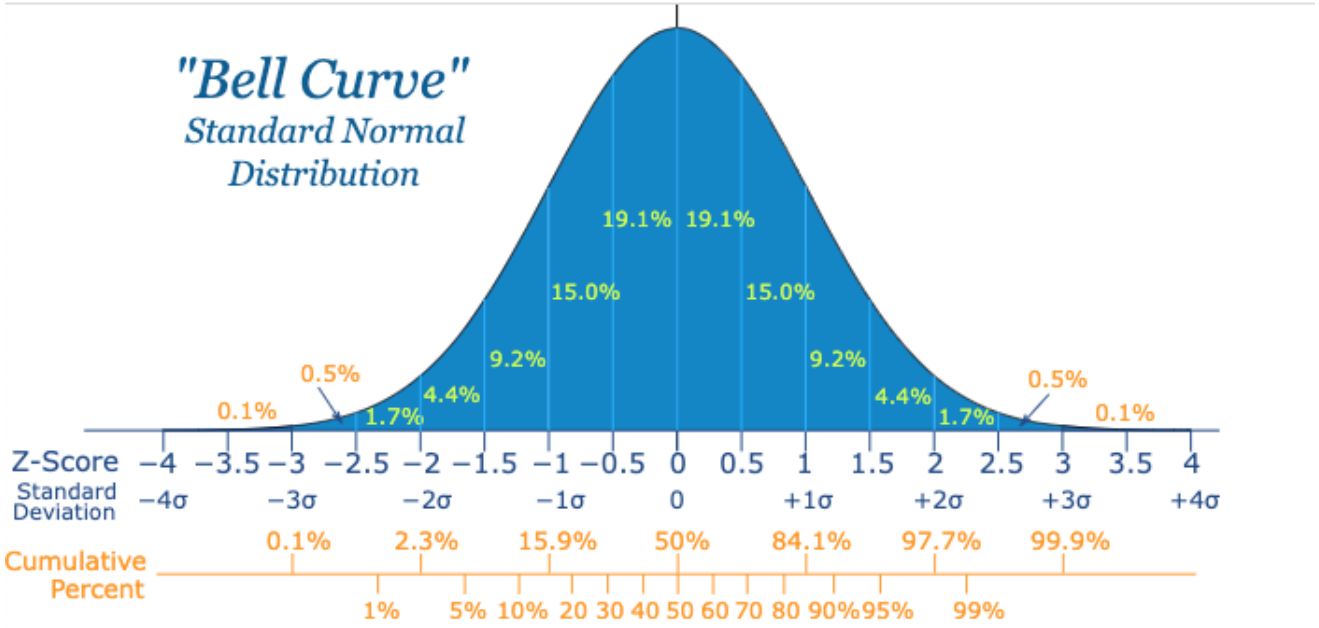
Applicable to numerical value

Discrete or Continuous data

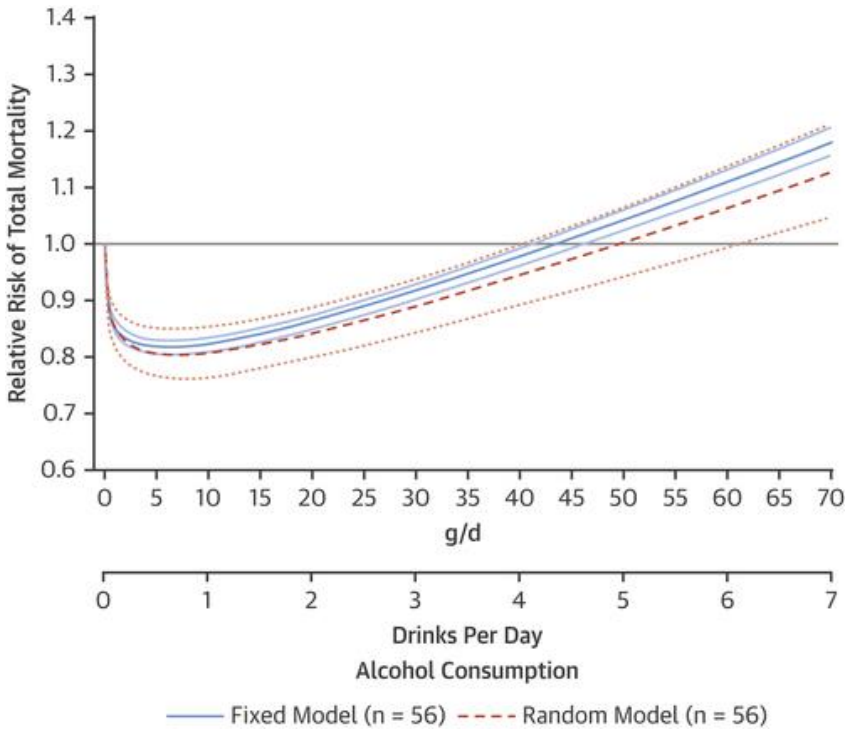
Discrete ~ *binomial, Poisson, Negative binomial, Multinomial etc*

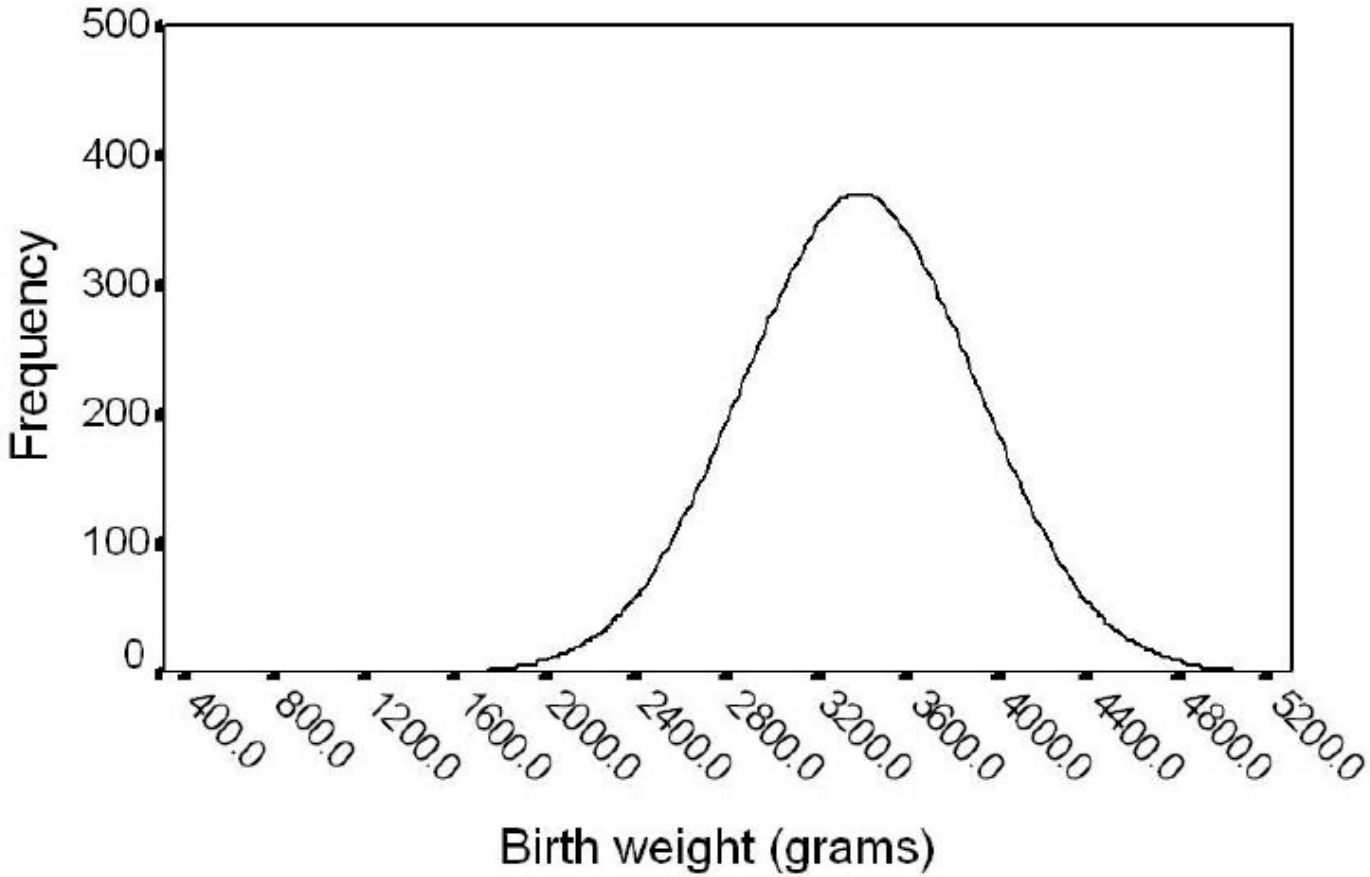
Continuous ~ *Normal, t, F, Chi square*

Normal distribution



- Most natural phenomena observe this normal distribution
- Characteristics of normal distribution
  - It is “Bell-Shaped” and has a single peak at the center of the distribution
  - The arithmetic Mean, Median and Mode are equal.
  - The total area under the curve is 1.00
  - Half the area under the normal curve is to the right of this centre point and the other half to the left of it
  - It is Symmetrical about the mean
  - It is Asymptotic: The curve gets closer and closer to the X – axis but never actually touches it.
  - To put it another way, the tails of the curve extend indefinitely in both directions.
  - The location of a normal distribution is determined by the Mean, μ
  - The Dispersion or spread of the distribution is determined by the Standard Deviation, σ.


Relative Risk of Total Mortality
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70
g/d
0 1 2 3 4 5 6 7
Drinks Per Day
Alcohol Consumption
Fixed Model (n = 56)
Random Model (n = 56)


Frequency
500
400
300
200
100
0
400.0 800.0 1200.0 1600.0 2000.0 2400.0 2800.0 3200.0 3600.0 4000.0 4400.0 4800.0 5200.0
Birth weight (grams)

Descriptive statistics

- Numerical summaries of data sets
- Exploratory data analysis (EDA) to generate inferential
- In any study , before we answer our research questions, we explore the characteristics of the variables (e.g the age, gender, ethnicity).
- Use descriptive statistics
  - Categorical data
  - Continuous data

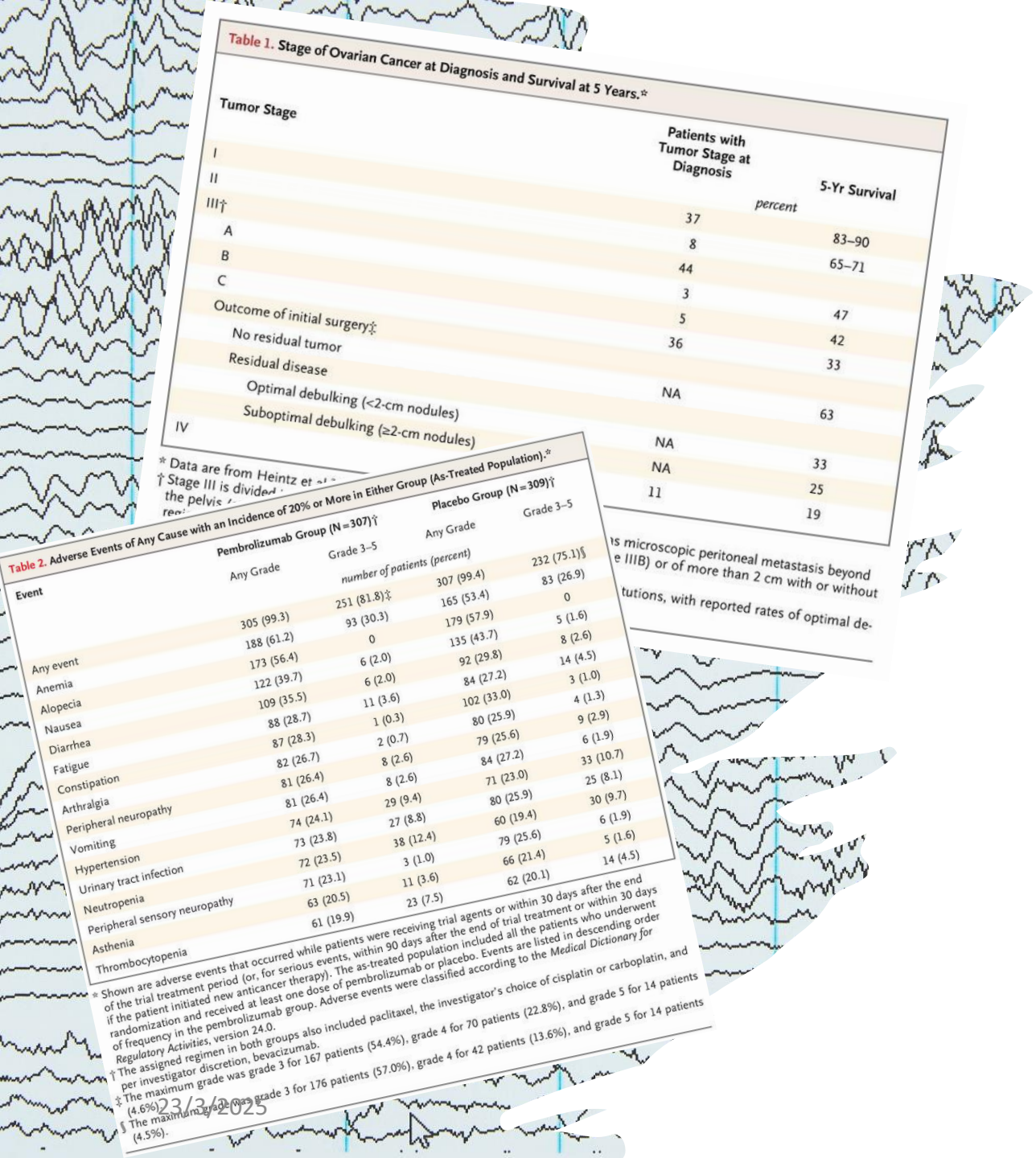


Table 1. Stage of Ovarian Cancer at Diagnosis and Survival at 5 Years.*

Tumor Stage	Patients with Tumor Stage at Diagnosis	5-Yr Survival
	percent	
I	37	83–90
II	8	65–71
III†	44	
A	3	47
B	5	42
C	36	33
Outcome of initial surgery‡		
No residual tumor	NA	63
Residual disease		
Optimal debulking (<2-cm nodules)	NA	33
Suboptimal debulking (≥2-cm nodules)	11	25
IV		19

* Data are from Heintz et al.
† Stage III is divided ... the pelvis ... regi...

...s microscopic peritoneal metastasis beyond ...e IIIB) or of more than 2 cm with or without ...

...tutions, with reported rates of optimal de-

Table 2. Adverse Events of Any Cause with an Incidence of 20% or More in Either Group (As-Treated Population).*

Event	Pembrolizumab Group (N=307)†		Placebo Group (N=309)†	
	Any Grade	Grade 3–5	Any Grade	Grade 3–5
	number of patients (percent)			
Any event	305 (99.3)	251 (81.8)‡	307 (99.4)	232 (75.1)§
Anemia	188 (61.2)	93 (30.3)	165 (53.4)	83 (26.9)
Alopecia	173 (56.4)	0	179 (57.9)	0
Nausea	122 (39.7)	6 (2.0)	135 (43.7)	5 (1.6)
Diarrhea	109 (35.5)	6 (2.0)	92 (29.8)	8 (2.6)
Fatigue	88 (28.7)	11 (3.6)	84 (27.2)	14 (4.5)
Constipation	87 (28.3)	1 (0.3)	102 (33.0)	3 (1.0)
Arthralgia	82 (26.7)	2 (0.7)	80 (25.9)	4 (1.3)
Peripheral neuropathy	81 (26.4)	8 (2.6)	79 (25.6)	9 (2.9)
Vomiting	81 (26.4)	8 (2.6)	84 (27.2)	6 (1.9)
Hypertension	74 (24.1)	29 (9.4)	71 (23.0)	33 (10.7)
Urinary tract infection	73 (23.8)	27 (8.8)	80 (25.9)	25 (8.1)
Neutropenia	72 (23.5)	38 (12.4)	60 (19.4)	30 (9.7)
Peripheral sensory neuropathy	71 (23.1)	3 (1.0)	79 (25.6)	6 (1.9)
Asthenia	63 (20.5)	11 (3.6)	66 (21.4)	5 (1.6)
Thrombocytopenia	61 (19.9)	23 (7.5)	62 (20.1)	14 (4.5)

* Shown are adverse events that occurred while patients were receiving trial agents or within 30 days after the end of the trial treatment period (or, for serious events, within 90 days after the end of trial treatment or within 30 days if the patient initiated new anticancer therapy). The as-treated population included all the patients who underwent randomization and received at least one dose of pembrolizumab or placebo. Events are listed in descending order of frequency in the pembrolizumab group. Adverse events were classified according to the *Medical Dictionary for Regulatory Activities*, version 24.0.
† The assigned regimen in both groups also included paclitaxel, the investigator's choice of cisplatin or carboplatin, and per investigator discretion, bevacizumab.
‡ The maximum grade was grade 3 for 167 patients (54.4%), grade 4 for 70 patients (22.8%), and grade 5 for 14 patients (4.6%).
§ The maximum grade was grade 3 for 176 patients (57.0%), grade 4 for 42 patients (13.6%), and grade 5 for 14 patients (4.5%).

Descriptive statistics

- Categorical data
  - Frequencies
  - Percentages
  - Fractions and/or relative frequencies
  - Cross tabulations
- Continuous data
  - Measure of central tendency
    - Mean, median, mode
    - Arithmetic mean, geometric mean
  - Measure of dispersion
    - Standard deviations, variance, interquartile range, min, max

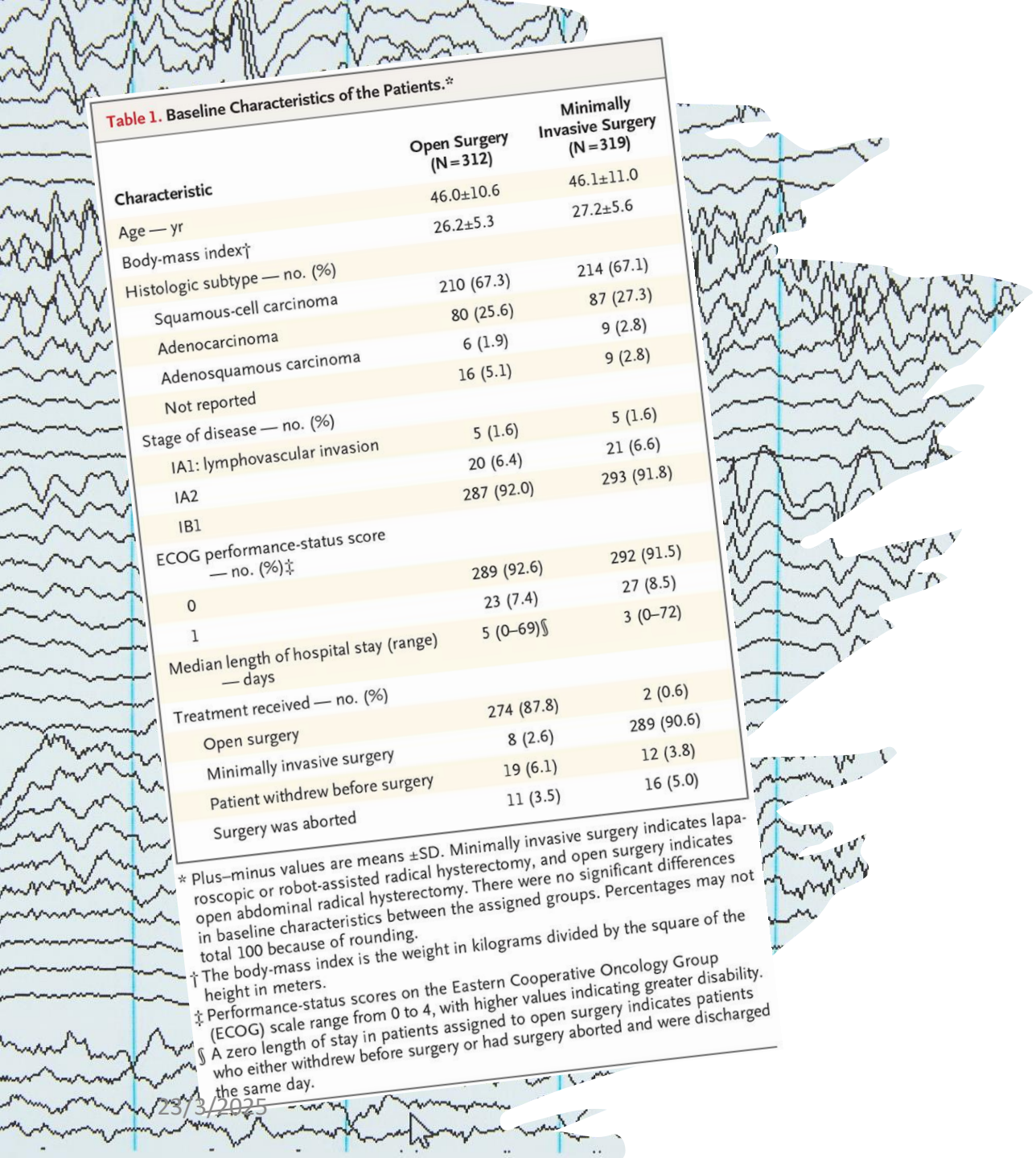


Table 1. Baseline Characteristics of the Patients.*

Characteristic	Open Surgery (N=312)	Minimally Invasive Surgery (N=319)
Age — yr	46.0±10.6	46.1±11.0
Body-mass index†	26.2±5.3	27.2±5.6
Histologic subtype — no. (%)		
Squamous-cell carcinoma	210 (67.3)	214 (67.1)
Adenocarcinoma	80 (25.6)	87 (27.3)
Adenosquamous carcinoma	6 (1.9)	9 (2.8)
Not reported	16 (5.1)	9 (2.8)
Stage of disease — no. (%)		
IA1: lymphovascular invasion	5 (1.6)	5 (1.6)
IA2	20 (6.4)	21 (6.6)
IB1	287 (92.0)	293 (91.8)
ECOG performance-status score — no. (%)‡		
0	289 (92.6)	292 (91.5)
1	23 (7.4)	27 (8.5)
Median length of hospital stay (range) — days	5 (0–69)§	3 (0–72)
Treatment received — no. (%)		
Open surgery	274 (87.8)	2 (0.6)
Minimally invasive surgery	8 (2.6)	289 (90.6)
Patient withdrew before surgery	19 (6.1)	12 (3.8)
Surgery was aborted	11 (3.5)	16 (5.0)

* Plus–minus values are means ±SD. Minimally invasive surgery indicates laparoscopic or robot-assisted radical hysterectomy, and open surgery indicates open abdominal radical hysterectomy. There were no significant differences in baseline characteristics between the assigned groups. Percentages may not total 100 because of rounding.

† The body-mass index is the weight in kilograms divided by the square of the height in meters.

‡ Performance-status scores on the Eastern Cooperative Oncology Group (ECOG) scale range from 0 to 4, with higher values indicating greater disability.

§ A zero length of stay in patients assigned to open surgery indicates patients who either withdrew before surgery or had surgery aborted and were discharged the same day.

Descriptive statistics

- Categorical data
  - Frequencies
  - Percentages
  - Fractions and/or relative frequencies
  - Cross tabulations
- Continuous data
  - Measure of central tendency
    - Mean, median, mode
    - Arithmetic mean, geometric mean
  - Measure of dispersion
    - Standard deviations, variance, interquartile range, min, max

INFERENTIAL STATISTICS

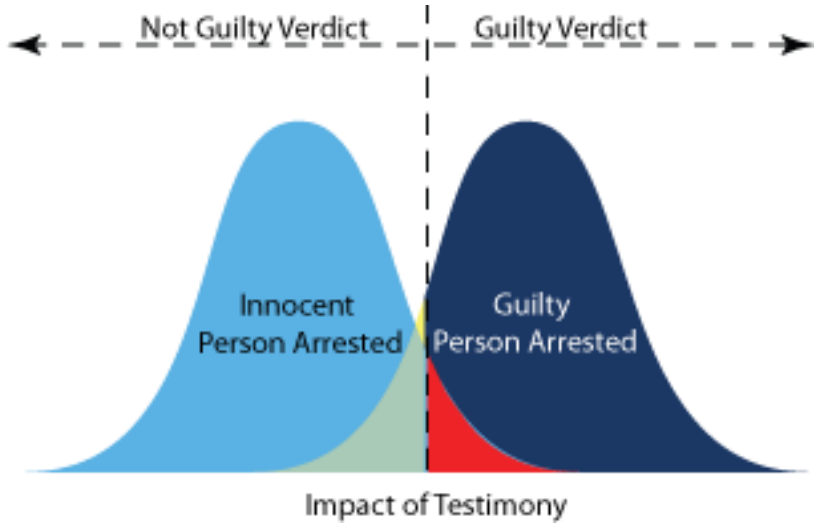
- Hypothesis testing (or statistical inference) is one of the major applications of biostatistics.
- To infer result to the population
- To test the research hypothesis in the form of statistical hypothesis

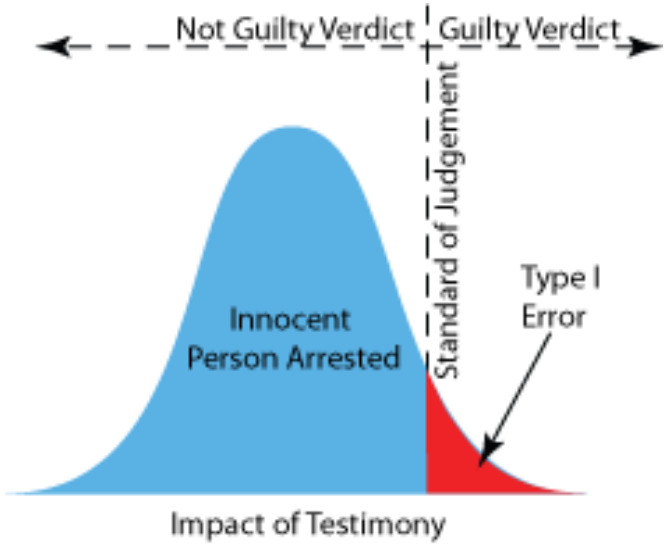
"Statistics are no substitute for judgment." Henry Clay

STATISTICAL HYPOTHESES

- **Null hypotheses, H_O**
  - is a statistical hypothesis that states there is **no differences** between a parameter and a specific value, or that there is no differences between two parameters.

- **Alternative hypotheses, H_A**
  - is a statistical hypothesis that states the **existence of a differences** between a parameter and a specific value, or states that there is a difference between two parameters.





ERROR IN DECISION MAKING

- 2 type of error
- **α is also known as Type 1 ERROR, false positive**
  - reject null when it is true, α probability 0.05
- **β is also known as Type 2 ERROR, false negative**
  - accept null when it is false, β probability chance alone
  - also known as power of study 1 - β
  - power - probability of detect a difference that exists

Types of errors

		Truth: No diff (H_0 to be not rejected)	Truth: Diff (H_0 to be rejected (H_1))
Decision based on the *p* value	H_0 not rejected, No diff	Right decision, $1-\alpha$	**β, Type II error**
	H_0 rejected (H_1), Diff	**α, Type I error**	Right decision, $1-\beta$

- **H_0 is "true" but rejected: Type I or α error**
- **H_0 is "false" but not rejected: Type II or β error**

17

SPSS Practical

1. Signup for SPSS Trial Version https://www.ibm.com/account/reg/us-en/signup?formid=urx-19774
2. Install the SPSS in your desktop or laptop
3. Download the practical dataset healthstatus.sav from Github https://github.com/adilzainal/IIUM_MBBS_Year4/blob/main/healthstatus.sav

Create and enter data

- Create a variable
- Convert from string data to categorical data.
- Create a composite variable using compute variable
- Convert from continuous data to categorical data visual binning.
- Label and give value for categorical data.

Descriptive statistics (Univariate statistics)

- Describe categorical data using frequencies.
- Describe categorical data using crosstabulation.
- Assess continuous data normality.
- Describe continuous data using explore.
- Descriptive statistics using figures.
- APA Format table.

Bivariate analysis

1. Chi square test 2. McNemar test 3. Cochran Q test 4. Independent T test 5. Paired student T test 6. One-way ANOVA 7. Repeated measure ANOVA 8. Mixed ANOVA 9. Mann-Whitney U test 10. Wilcoxon Signed Rank test 11. Kruskal-Wallis test 12. Friedman test 13. Correlation test	

Multivariate analysis (Multivariable, iv and multivariate, dv)

1. Simple vs multiple regression
2. Linear regression
3. Logistic regression
    1. Binary logistic regression
    2. Ordinal logistic regression
    3. Multinomial logistic regression
4. Other
    1. Poisson regression
    2. Cox regression

Reporting results APA format 7th Edition

- Describe in text, table or figure.
- Be concise and specifics.
- Don’t misleading
- Decimal points standardized.
- P value small, italicized, no hyphe, no zero before decimal, 2 or decimal points.
- Present exact p value or <.001
- Summary https://apastyle.apa.org/instructional-aids/numbers-statistics-guide.pdf
- Buy APA 7th Edition

STEPS FOR HYPOTHESIS TESTING

- State your null and alternate hypothesis
- Decide your critical level or alpha level
- Conduct a statistical test to derive your p-value
- Conclusion to reject or do not reject your hypothesis

THANK YOU

Muhammad Adil ZA
MBBS (IIUM) DLSHTM, MSc(Lon) MPH, DrPH(UPM) AM(Mal) FRSPH(UK)
Department of Community Medicine
Kulliyyah of Medicine
adilza@iium.edu.my

*Ab Rahman, J. (2015). *Brief guidelines for methods and statistics in medical research*. Springer Singapore.
Ahmad, W. M. A. W., Khamis, M. F., & Husein, A. (2024). *Discovering Advanced Statistics Using SPSS for Windows*. Penerbit USM.
Dodge, Y. (Ed.). (2003). *The Oxford dictionary of statistical terms*. Oxford University Press, USA.
Altman, D. G. (1990). Practical Statistics for Medical Research (1st. ed.): Chapman & Hall/CRC.
Altman, D. G., Machin, D., Bryant, T. N., & Gardner, M. J. (2000). Statistics with Confidence (Second ed.). London: BMJ Books.
*Bland, M. (2000). An Introduction to Medical Statistics (3rd. ed.): Oxford University Press.
Campbell, M., Machin, D., & Walters, S. (2007). Medical statistics: a textbook for the health sciences: Wiley-Interscience.
*Field, A. (2005). Discovering Statistics Using SPSS (Introducing Statistical Methods S.) (2nd Edition ed.): Sage Publications Ltd.
Gupta, V. (1999). SPSS for Beginners V. Gupta (Ed.)
Sabine Landau, B. S. E. (2003). A Handbook of Statistical Analyses Using SPSS (1st edition ed.): Chapman and Hall/CRC.