Statistics
Chapter Four

- **Bivariate Descriptive Statistics**: used to describe the relationship between two variables
  - → Eg. height and weight, smoking status and lung cancer
  - → Appropriate statistic depends on the variables level of measurement

- **Crosstabulation**: Researchers crosstabulate the frequencies of all categories of two variables in a 2D frequency distribution (contingency table)
  - → Crosstabulated variables should be nominal level (or ordinal level with a small number of categories)

- **Reading a crosstab table**
  - → Cell Count
  - → Row Percentage
  - → Column Percentage
  - → Overall Percentage

- **Risk Indexes**: Developed to describe risk outcomes and facilitate clinical decision making
  - → For situations w/ two nominal variables
    - → One is a risk factor (or intervention status) eg. wore a mask, didn't wear a mask
    - → The other is outcome eg. got COVID, didn't get it

Chapter Four

- Risk Index Scenarios
  - → Prospective (cohort) design
    - → Some people have exposure to the risk factor, others don't
    - → Both groups are followed to assess outcome
  - → Retrospective (case : control) design
    - → Some people have bad outcomes (cases) others don't (control)
    - → Groups are compared regarding prior exposure to the risk factors
  - → Experimental Design (clinical trials)
    - → Some people are asigned to a control group in which they have ongoing or "baseline" exposure to risk, while others are assigned to an experimental group in which they receive an intervention hypothesized to reduce risk
    - → Both groups are followed to assess outcome

- Risk Index Types
  - → Risk indexes capture two aspects of the effects of risk exposure.
    - → Absolute Risk: Indexes quantify the actual amount of risk related to different exposures
    - → Relative Risk: Indexes compare risks in the two risk exposure groups

- When Setting up a risk index table
  - → Outcomes go in columns and risk behaviours go in rows

a	b
c	d

  - → Always put bad outcome in 1st column and risk in 1st row

Chapter Four

- **Absolute Risk**: the proportion of people w a negative outcome
  - → You can calculate absolute risk for the exposed group and for the non-exposed group
  - → $AR_E = (a \div (a+b))$
  - → $AR_{NE} = (c \div (c+d))$

- **Absolute Risk Reduction**: the absolute difference between the two risk groups
  - → $ARR = AR_E - AR_{NE}$

- **Relative Risk**: the ratio of absolute risks (adverse outcomes) in the two groups
  - → $RR = AR_E \div AR_{NE}$

- **Relative Risk Reduction**: the proportion of baseline risk that is reduced through nonexposure
  - → $RRR = ARR \div AR_{NE}$

- **Odds**: the proportion of people in each risk group who have the adverse outcome, relative to the proportion who do not
  - → $Odd_E = A \div B$
  - → $Odd_{NE} = C \div D$

- **Odds Ratio**: the ratio of two odds
  - → $OR = Odd_E \div odd_{NE}$

- **Number needed to treat**: Estimate of how many people would need to avoid exposure (or get treatment) to prevent one negative outcome
  - → $NNT = 1 \div ARR$

Chapter Four

- **Correlation**: a bond or connection btwn high level variables
  - → Variation in one variable is systematically related to variation in another.
  - → Correlation btwn two quantitative variables can be graphed in a scatterplot

- **Scatterplot**: graphs the values of one variable on the x axis and the values of the second variable on the y axis
  - → Indicates whether the variables have a linear relationship
    - → A linear relationship occurs when there is a constant rate of change btwn the two variables
  - → Scatterplots indicate direction and magnitude of the relationship

- **Positive Relationship**: line slopes from lower left to upper right
  - → low values of one variable correspond to low values of the other and high values also correspond

- **Negative Relationship**: line slopes from upper left to lower right
  - → low values of one variable correspond to high values of the other, and vice versa

- If the data points are tightly packed along the diagonal it indicates a **strong relationship**

- If data points are loosely spaced but suggest a diagonal it indicates a **weak relationship**

- A correlation coefficient (**Pearson's R**) is a statistic that summarizes the magnitude and direction of relationships btwn two variables

Chapter Four

- Correlation Coefficient Values
  - → Range from -1.00 and 1.00
    - → -1.00 is a perfect negative correlation
    - → 1.00 is a perfect positive correlation
    - → 0 is no correlation
  - → The absolute value of the coefficient indicates strength