Lecture 3

Sample statistic; describes a sample eg; sample mean (x bar)

Population parameter; describes a population eg: μ (population mean)

Sampling distribution; distribution of the sample statistic (take more than one sample and compare the means of the sample).

Error; unexplained variation

- Bias; difference b/w the true value and the measured value.
    - Concerns the centre of the sampling distribution
    - Reduced using random sampling
- Precision; variability of the measurements
    - Concerns the spread of the sampling distribution
    - Reduced by increasing the sample size

Sampling error; difference b/w sample statistics and population parameters. Reduced by increasing the sample size eg: the difference b/w μ and x bar.

Non-sampling error; any other error eg; sample is not reflective of the entire population, biased language which imposes a response (selection bias), measurement bias, measurement error. Reduce by random sampling and carful examining.

Observational studies; just observe without controlling or treating.

Experiment; intervention eg; treatments, comparing the effects.

Lurking variables; anything you haven't observed which could have an impact on the relationship/result of the study eg; gender. Aim to remove when collecting data.

Randomisation; using chance in a study

- Random sampling; any difference between the sample and the population is due to sampling variation
- Can carry out after blocking if a variable may have an impact on the response variable.

Replication; repeating data collection to reduce variation.

Blocking; randomisation within different subsets – if believed to impact the results.

Designing a study;

- Objective
- Population of interest
- Sampling procedure
- Variables which need measuring
- How much data is needed
- Data analysis

Probability; how likely something is to occur.

P = no. of times the event occurs/sample size.

P(A) = 0; never occurs.

P(A) = 1; certain to occur.

The sum of the probabilities must be equal to 1.

Determining probabilities;

- Subjective; informed by prior experience
- Empirical; perform experiment to count the proportion of successful outcomes.
- Model-based; imply particular probabilities.

Mutually exclusive events; cannot occur at the same time- therefore no overlap of events. $P(A \cup B) = P(A) + P(B)$

Collectively exhaustive events; all outcomes (one outcome must occur) eg; coin toss, one outcome must occur, each outcome makes up the fixed number of outcomes.

Union; event A or event b or both [$A \cup B$]

Intersection; event A and B occur at the same time – there is overlap [$A \cap B$]

For events that aren't mutually exclusive; both may occur at the same time and therefore we must remove the double count (overlap) when calculating probability.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Contingency tables; total must equal 100% or 1.

- Marginal probabilities in the margins
- Join probabilities in the intersections.

Conditional probability; $P(A|B) = P(A \text{ and } B)/P(B)$ OR $P(A \text{ and } B) = P(A|B)P(B)$. Calculates the probability of one variable given that another has already occurred.

Tree diagrams;

- 1^{st} branch; marginal probabilities.
- 2^{nd} branch; conditional probabilities.

Multiply the two to get the joint probabilities.

Sensitivity; true positive result

P(+ve result | disease present)/P(disease present)

Specificity; true negative result

P(-ve result | no disease present)/P(no disease present)

Independence; if A and B are independent than whether A happens or not has no effect on the probability of B occurring. Conditions for independence;

$$P(A \mid B) = P(A)$$
$$P(B \mid A) = P(B)$$
$$P(A \text{ and } B) = P(A)\,P(B)$$