Statistical Inference

Dr. M.M. Mweu,

Level II MBChB Biostatistics,

26 March, 2019

Statistical inference

- Is the procedure whereby conclusions about a pop are made based on findings from a sample obtained from the pop
- Since it's often difficult to measure every individual in the pop, samples are taken and inferences drawn from them about the pop
- Two measures of statistical inference:
    - **Confidence intervals** – give an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a set of sample data
    - **Hypothesis tests** – test whether there's sufficient evidence in a sample of data to infer that a certain condition is true for the entire population
- These measures are linked to the concept of *sampling distribution*

Statistical inference

Confidence intervals

- A confidence interval is a pair of numerical values defining an interval, which with a specified *degree of confidence* includes the parameter being estimated
- If we construct a CI for the pop mean μ with a value for the lower confidence limit (LCL) and a value for the upper confidence limit (UCL) at the 95% degree of confidence, we can say that **we are 95% certain that this CI encloses the true value of the pop mean**

Hypothesis testing

- In hypothesis testing we state that we will reject a certain hypothesis only *if there is a 5% or less chance that it is true*

Statistical inference

Hypothesis testing

(a) Null hypothesis

- Frequently, there's an *expected/natural* value for a parameter – called the **null value**
- In hypothesis testing we assess whether the statistic computed from a sample is *consistent* with the null value
- If there's consistency then the statistic will be *considered equal* to the null value except for *sampling & measurement errors*
- The argument that there's consistency betwn the statistic and the null value is the **null hypothesis** – denoted by $\boldsymbol{H_0}$
- The H_0 can be written as: $\boldsymbol{H_0}: \boldsymbol{\mu} = \boldsymbol{\mu_0}$

Statistical inference

Hypothesis testing

(b) Alternative hypothesis

- Is the opposite of the H_0 - the assertion that the null value is inconsistent with the statistic – is denoted by H_a
- H_a states that the parameter is *not equal to*, is *greater than*, or *less than* the null value
- Can be expressed as:

$$H_a: \mu \neq \mu_0 \ (two\ sided)$$

$$H_a: \mu > \mu_0 \ (one\ sided)$$

$$H_a: \mu < \mu_0 \ (one\ sided)$$

- The choice of the H_a will affect the way we conduct the test
- We choose an H_a based on the *prior knowledge* we have about possible values:

Statistical inference

Hypothesis testing

- **Two-sided test:** We are testing whether μ is, or is not equal to a specified value μ_0. We have *no strong opinion* whether μ is greater/less than μ_0 and we state:

$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

- **One-sided test:** We are testing μ to be greater/less than a given value μ_0. We need *prior knowledge* that μ is on a particular side of μ_0. We restate the hypothesis as:

$$H_0: \mu \geq \mu_0$$

$$H_a: \mu < \mu_0 \ \boldsymbol{OR}$$

$$H_0: \mu \leq \mu_0$$

$$H_a: \mu > \mu_0$$

- **Significance level:** Refers to the null hypothesis
  - When the probability (P) that the statistic is consistent with the null value becomes too small, we say that the statistic is significantly different from the null value and hence we reject $H_0: \mu = \mu_0$
  - How small must P be for us to reject the H_0? – usually **0.05** is used

Statistical inference

Errors in hypothesis testing

Type I Error

- This is denoted by $\boldsymbol{\alpha}$ – which is the probability that we will reject the $H_0: \mu = \mu_0$ when the H_0 was actually correct
- The probability that the H_0 is true is the ***P*-value**
  - More correctly, the P-value is the **probability that we would observe a statistic equal to, or more extreme, than the value we have observed if the H_0 is true**

Type II Error

- This is denoted by $\boldsymbol{\beta}$ – occurs when the H_0 is accepted when the H_a is true (NB: β is often set at **0.20**)
- This allows us to calculate the **power** of a test $\mathbf{1-\beta}$ which is the **probability of rejecting the H_0 if H_a is true**
  - Also, the power of a test is the **ability of the test to detect a real difference when that difference exists and is of a certain magnitude**
- For a given sample size n, lowering α (say below 0.05) will increase β
- Probability of type II error (β) decreases with increase in n

Statistical inference

Errors in hypothesis testing

H_0		
	True	False
Accept	$1-\alpha$ (confidence level)	Type II error (β)
Reject	Type I error (α)	$1-\beta$ (power)

Sampling distribution

- *Confidence intervals* and *hypothesis tests* are linked to the concept of *sampling distribution*
- When different samples of equal size are repeatedly taken from the same pop & we repeatedly calculate the *statistics* (e.g. estimates of μ, σ and π) for each sample we get populations of statistics with *known probability distributions*
- The pop originally sampled is called the parent *population/parent distribution* while that of the computed statistic is the *sampling distribution*
- **NB:** The idea behind estimation is that N, no. of individuals in the pop is very large compared to n the no. of individuals in the sample, so that sampling doesn't affect the probability of choosing a particular sample – means that although we are not sampling with replacement, in terms of probability, it as though we were sampling with replacement

Sampling distribution of a mean

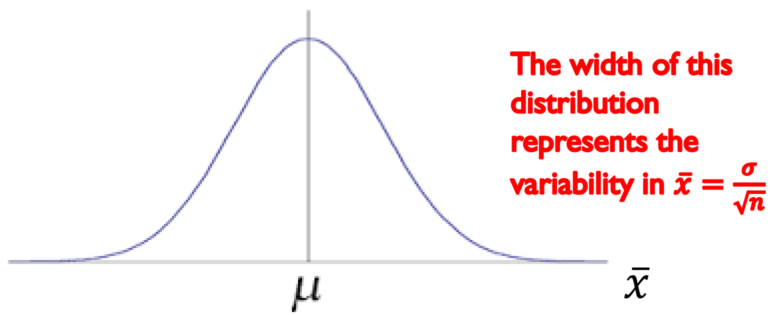
- If all possible samples of a given size, n, were picked and a $\bar{x}$ (sample mean) calculated for each, the population of $\bar{x}$'s would have a normal distribution with a mean equal to the mean of the parent distribution and a variance that is $\frac{1}{n}$ times smaller than that of the parent distribution i.e. the sampling distribution of $\bar{x}$ is **normal** with a mean= μ and a variance = $\frac{\sigma^2}{n}$

- $\sqrt{\frac{\sigma^2}{n}}$ is called the **standard error of the mean** which measures the variability of the $\bar{x}$'s obtained when taking repeated samples of size n (recall: σ measures the variability of the individual x's in the population)

- As the sample size (n) increases, the $\sqrt{\frac{\sigma^2}{n}}$ of the sample mean decreases meaning that $\bar{x}$'s become clustered more closely to the mean μ – we get more precise estimates as n increases

Sampling distribution of a mean

- If n (sample size) is large ($n \geq 20$), the sampling distribution of $\bar{x}$ will be normal with mean = μ & variance = $\frac{\sigma^2}{n}$ even if X (parent variable) is not normally distributed – called the central limit theorem



Confidence interval for a mean

- To make inferences about the true mean μ we construct a CI
- We accept that the observed mean $\bar{x}$ is generally within 1.96 (recall: $Z_{0.025} = 1.96$) standard errors of the true mean μ so that the interval: $\bar{x} \pm 1.96 \times SE(\bar{x})$ will usually include the true value
- This means that on repeated sampling, 95% of sample means would fall within 1.96 standard errors of the μ so that the interval: $\bar{x} \pm 1.96 \times SE(\bar{x})$ **includes μ approx. 95% of the time (called the 95% CI)**

Sampling distribution of a mean

Confidence interval for a mean

- A 99% CI is given by: $\bar{x} \pm 2.58 \times SE(\bar{x})$

Example

- The packed cell volume (PCV) was measured in 25 children sampled randomly from children aged 4 yrs living in a large West African village, with the following results:

$$\bar{x} = 34.1 \qquad s = 4.3$$

Using the s as an unbiased estimator of σ we obtain the 95% CI of:

$$34.1 \pm 1.96 \times \frac{4.3}{\sqrt{25}} = \mathbf{32.4\ \mathit{to}\ 35.8}$$

Use of the *t* distribution

- As the value of σ is generally unknown (recall: $95\%CI = \bar{x} \pm 1.96 \times \sigma/n$), we have to use s as an estimate of σ – introduces *sampling error* in calculation
- Due to the this error, the interval: $\bar{x} \pm 1.96 \times s/n$ includes μ *less than* 95% of the time i.e. the calculated interval is too narrow

Sampling distribution of a mean

Confidence interval for a mean

Use of the *t* distribution

- To correct for this we use a multiplying factor larger than 1.96 – makes interval wider and restores confidence level to 95%
- The multiplying factor is contained in the *t distribution*
- The factor depends on the *degrees of freedom* (v) used to calculate the sample SD s $\left(df \text{ are one less than sample size i.e.} \boldsymbol{v = n - 1}\right)$
- As n increases the factor approaches $Z_{0.025} = 1.96$ – hence t distribution only needs to be used for $n < 20$

Example

- In the PCV example, $v = 25 - 1 = 24$. Using the t distribution with 24 df, the 95% CI is: $\bar{x} \pm t_{(n-1),\alpha/2} \times \frac{s}{\sqrt{n}} = 34.1 \pm \left(2.064 \times \frac{4.3}{\sqrt{25}}\right) = \mathbf{32.3 \text{ to } 35.9}$

Approx. same as previous CI since $n > 20$

Sampling distribution of a mean

Significance test for a mean

- We may wish to test a specific hypothesis about the pop mean μ
  e.g. if data on 4yr children in USA indicate a mean PCV of 37.1 we may test whether our sample data (West African) are consistent with the H_0:

$$H_0: \mu = \mu_0 = 37.1$$

$$H_a: \mu \neq \mu_0$$

- One approach is to see whether the 95% CI includes the hypothesised value (37.1) – doesn't (some evidence against the H_0)
- More objectively we use a *significance test* and examine the P-value:

$$Z = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = \frac{34.1 - 37.1}{\frac{4.3}{\sqrt{25}}} = \mathbf{-3.49}$$

- From the *Z tables* we get the P-value: $2 \times 0.00024 = \mathbf{0.00048}$ (NB: P-value is normally one-tailed so multiply the resulting probability by 2)
- <u>Interpretation:</u> **The data provide strong evidence against H_0 hence the mean PCV in 4yr old children in the West African village is less than that of children of the same age in the USA**

Sampling distribution of a mean

Significance test for a mean

- If $n < 20$ then *t distribution* is more appropriate:

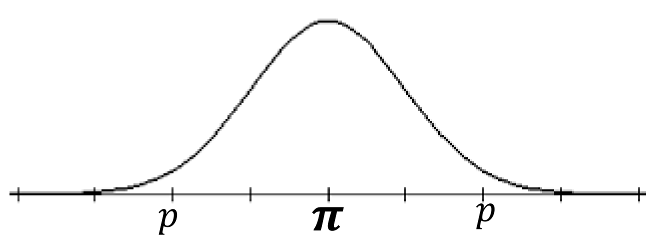
$$t = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = \mathbf{-3.49}$$

- From *t tables* 3.49 falls betwn 3.467 & 3.745 hence $\mathbf{0.001} < \boldsymbol{P} < \mathbf{0.002}$

Sampling distribution of a proportion

Example: In a survey of 335 men attending a health centre in Guilford (UK), 127 (37.9%) men said they were current smokers

- How we can we use the result from the above sample to say something about the population which it represents? – we use the concept of *sampling distribution*
- Suppose we repeatedly took a *new* sample of 335 men from this health centre (assuming a large no. of men are registered there) and calculated the proportion who smoked & then created a histogram of these values – histogram would represent the *sampling distribution of the proportion*:



The distribution is centred over the value of the proportion of smokers in the pop π. Some p's will be smaller/larger than π but many will be closer to π

Sampling distribution of a proportion

- In practice we only conduct one survey from which to estimate p
- Is p close to π or is it very different from π?
- In any random sample there's some *sampling variation* in p so that the larger the n the smaller the sampling variation
- The sampling variation of a proportion is described by its *standard error*:

$$\sqrt{\frac{p \times (1 - p)}{n}}$$

- The SE of a proportion is a measure of how far our observed proportion p differs from the true pop proportion π
- In the previous UK example, the estimated SE of the proportion of smokers is: $\sqrt{\frac{0.379 \times (1 - 0.379)}{335}} = \mathbf{0.0265}$ ***or*** $\mathbf{2.65\%}$

Sampling distribution of a proportion

- To estimate the interval of possible values within which the true pop proportion π lies we compute the CI:

$$p \pm Z_{\alpha/2} \times \sqrt{\frac{p \times (1-p)}{n}} = 0.379 \pm 1.96 \times \sqrt{\frac{0.379 \times 0.621}{335}}$$

$$= 0.327\ to\ 0.431\ or\ \mathbf{32.7}\%(\boldsymbol{LCL})\ \boldsymbol{to}\ \mathbf{43.1}\%(\boldsymbol{UCL})$$

Interpretation: We are 95% confident that the true percentage of smokers in Guilford, UK lies between 32.7% and 43.1%