Foundations for Inference I

Dajiang Liu

@PHS525

Feb-09-2016

Statistical Inference

- Statistical inference is usually performed using a randomly selected sample from the population
  - The estimates obtained from the sample may not actually reflect properties of the population

- Understanding the quality of the parameter
  - How close is the estimated mean value to the true (population) mean value
  - Normally, inference is done by using a sub-sample to infer the properties of the population

Point Estimates – Sample Mean

For a sample of size n

- Estimate population mean by sample mean

$$\bar{X} = (X_1 + X_2 + \cdots + X_n)/n$$

- Estimate population standard deviation by sample standard deviation

$$\sigma = \sqrt{[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2]/n}$$

Variations in the Point Estimate of $\bar{X}$

- Since samples are randomly chosen from a population, sample means are usually different from population mean

- Variations in the sample mean can be quantified using standard errors of the sample mean estimate (point estimate)

$$SE_{\bar{X}} = \sigma/\sqrt{n}$$

Summary of What We learnt So Far

- Samples means (standard deviations) can be used to estimate population means (standard deviations)

- Sample means are not accurate

- Uncertainties in the sample means can be quantified by standard deviations

Exercises – Calculate Moving Averages for the Sample

- Load data *run10*
  - *run10=read.table('run10.txt',* stringsAsFactors=TRUE, *header=TRUE,sep='\t')*
- Compute the population mean and standard deviation
  - *mean.age=mean(run10$age,na.rm=T)*
  - *sd.age=sqrt(var(run10$age,na.rm=T))*
  - *Or*

    sd.age=sd(run10$age,na.rm=T);
- *moving.average=0;*

for(ii in 1:length(run10$age))
moving.average[ii]=mean(run10$age[ii:(ii+100)],na.rm=T);

Exercises

- Plot moving averages
- Plot histograms
- Can you summarize properties of moving averages

Confidence Intervals

- Point estimates are not perfect

- They contain errors in the estimates

- Instead of providing a single estimate, it is often necessary to provide a range of possible values for the population parameters of interest, e.g. the sample mean point estimate $\bar{X}$

How to Interpret Confidence Intervals

- Confidence interval is always associated with a size, say 95%, 90% or 99%

- What is a 95% confidence interval: An interval of values that contains the true parameter value with probability of 95%

- Approximate 95% confidence intervals

$$\text{Point Estimate} \pm 1.96 \times \text{SE} \quad (1)$$

- Another interpretation:

Assume that we draw 100 samples, for each sample, we calculate confidence interval according to (1),

    - ~95 of those intervals would contain the true parameter value

How to Rescale Confidence Interval

- How about you are interested in more precise/broader confidence intervals
  - Replace 1.96 by some other numbers
    - 2.58 for 99% confidence interval
    - 1.64 for 90% confidence interval

- Question asked: which confidence interval is wider, 90% or 99%
  - Answer: 99% CI is wider

Example 4.10

- In run10Samp, the sample mean is 95.61 and the standard error is 1.58,
- What is the 95%/90%/99%-confidence interval for the time?
- How to interpret the results:

- Which confidence interval is more precise/broad?

Example

- In a clinical trial of Lipitor, a common drug used to lower cholesterol, 863 patients were given a treatment of 10mg tablets. That group consists of 19 patients who experienced flu symptom. The probability of an average person getting a flu is 1.9%.
- What is the mean value of the number of people that have flu symptoms
- What is the confidence interval
- Do you think it is usual to see 19 patients to develop flu after taking Lipitor?

Example

- In a clinical trial of Lipitor, a common drug used to lower cholesterol, 863 patients were given a treatment of 10mg tablets. That group consists of 19 patients who experienced flu symptom. The probability of an average person getting a flu is 1.9%.
- What is the mean value of the number of people that have flu symptoms
  - $\bar{X} = \frac{19}{863}$
- What is the 95% confidence interval
  - The CI is $\bar{X} \pm \sqrt{(\bar{X}(1 - \bar{X})/863}$
- Do you think it is usual to see 19 patients to develop flu after taking Lipitor?
  - Check if the CI overlaps 1.9%

Homework

- Page 204: 4.3, 4.4, 4.7, 4.8 for version 3 of the text book