Statistics 3 Notes

Will Bevington   Alex Pace

September 26, 2015

Contents

1 Combining Random Variables

X and Y are two independent random variables with means $E(X)$ and $E(Y)$, and variables $Var(X)$ and $Var(Y)$

Our aim is to find: $E(X \pm Y)$ and $Var(X \pm Y)$

$$E(X \pm Y) = E(X) \pm E(Y)$$
$$Var(X \pm Y) = Var(X) + Var(Y)$$

$E(aX + b) = aE(X) + b$
$Var(aX + b) = a^2 Var(X) + b$

These last two formulas should be recalled from S1. We can combine these last four formulas to acquire:

$$E(aX \pm bY) = aE(X) \pm bE(y)$$
$$Var(aX \pm bY) = a^2 Var(X) + b^2 E(Y)$$

The combination of two independent Normal distributions is also a Normal distribution, and can be expressed thus:

$X \ N(\mu x, \sigma^2 x)$
$Y \ N(\mu y, \sigma^2 y)$

$$aX \pm bY \ N(a\mu x \pm b\mu y, a^2 \sigma x^2 + b^2 \sigma y^2)$$

2 Sampling Frames

2.1 Key Words

Population The whole set of items that are of interest.

Census Observes or measures every memeber of a population.

Sample Survey A selection of observations taken from a subset of the population which is used to find out information about the population as a whole.

Random Sample A sample in which every possible sample of size n has an equal chance of being selected.

Sampling Frame A list identifying every single sampling unit that could be included in the sample

2.2 Random Sample

Random Number Sample
Give each sampling unit in the sampling frame a number and use a random number generator or random number tables to select required number of samplng units.

2

Lottery Sample

Put the sampling units from the sampling frame into a "hat" and select randomly without replacing.

Positives

- Random and free from bias
- Easy to carry out

Negatives

- Not suitable for large sample sizes

2.3   Systematic Sampling

Pick at required intervals from an ordered list.

e.g. I want a sample of 15 from 60: $\frac{60}{15} = 4$ Therefore, choose a starting point randomly from one of the first four sampling unit from the ordered list, then choose every fourth sampling unit after until you have selected 15.

Positives

- Suitable for large samples
- Is easy to carry out

Negatives

- Sample is not random unless the ordered list is random
- Can introduce bias

2.4   Stratified Sampling

A form of random sampling: The population is split into mutually exclusive groups (strata). Random samples are taken from each strata, the relative size of each corresponds to the same ratio as each strata's representation in the total population.

Positives

- Works well with large samples that can be split into mutually exclusive groups
- Reflects a populations structure

Negatives

- Takes longer than random sampling
- Within each strata the problems are the same as with any random sample.
- Ill defined strata can overlap (meaning they are no longer mutually exclusive)
- Can't provide accurate data when strata overlap/

2.5 Quota Sampling

When no sampling frame is available, quota sampling may be used. The population is divided nto groups (as with sratified sampling). Quotas for each group are created that corrrespond with the groups representation in the total population. The interviewer then selects sampling units until each quota is reached.

Positives

- Administering the test is easy
- Test is low cost
- Test is quick if the sample is small

Negatives

- Introduces interviewer bias
- Can't estimate sampling errors geamssymb geometry [round]natbib
  Statistics 3 Notes Will Bevington Alex Pace

3 Types of Data

3.1 Primary Data

When you collect data, or someone collects data on your behalf.

Positives

- You have control over the type and method of collection
- The exact data needed is collected
- The Accuracy is known

Negatives

- Expensive (money and time)

3.2 Seconday Data

Second ahnd data, collected by another person or organisation.

Positives

- Cheaper than gathering primary data (time and money)
- Large amounts of data are easily available on the internet
- Access to data over time (trends)

Negatives

- Bias is not always acknowledged
- Accuracy is not known
- Certain data can be in a form that is difficult to deal with