STATISTICS170

LECTURE 5 | SAMPLING DISTRIBUTIONS | WEEK 5

Distributions of means and proportions

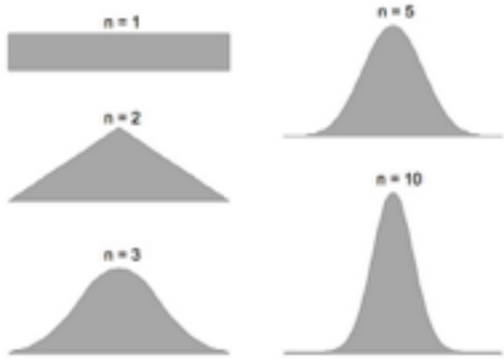
- show how more precisely normal distribution arises as the distribution of sample means
- sample means (use of averages) are used to obtain more accurate measurements and summarise of population parameters of interest
- the larger the sample, the easier it is to reflect the population and calculate a more accurate mean
- standard error provides a measure of this - important to know size and determinants

Averages of uniform random numbers

- averages are more likely to occur in the middle of the range
- distribution of these averages appear to be approximately triangular in shape

Central limit theorem

- states that the distribution of the average of a set of independent random numbers that is not normally distributed will follow a normal distribution if the sample size is large enough
- this approximation improves as a number of terms in the average increases
- tendency to become more concentrated in the centre as n increases
    - the closer the original population is to a normal distribution, the smaller the sample size required

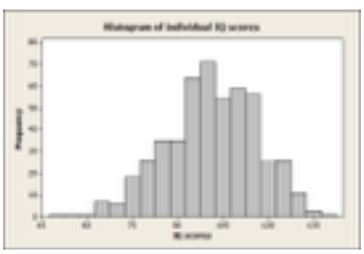


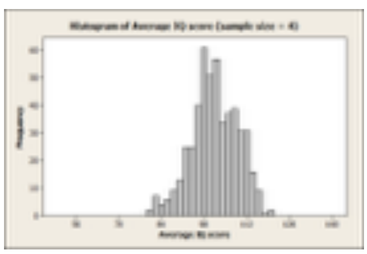
Sample means behaving in repeated sampling:

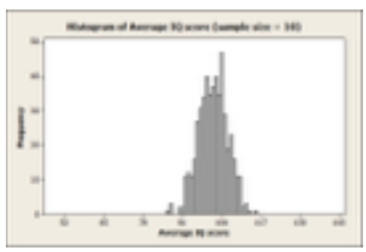
- IQ scores are normally distributed with a mean of 100 and an SD of 15

EXAMPLE:

- take random samples from population and calculate the mean **(sample sizes being n=4)**
- comparing the difference in a histogram between a sample size of 4 and 10







Simulations: descriptive statistics

Variable	# samples	Mean	SD	Min	Max
individuals	500	101.13	15.08	50	141
means (4)	500	101.36	7.61	80	121
means (10)	500	101.07	4.75	87	115

- show that the overall means of the individual scores and of the sample means are all close to 100
- SDs begin to decrease as the distributions are becoming more compact
- shape of each histogram however appears normal
- as n increases, the distribution of means becomes more compressed

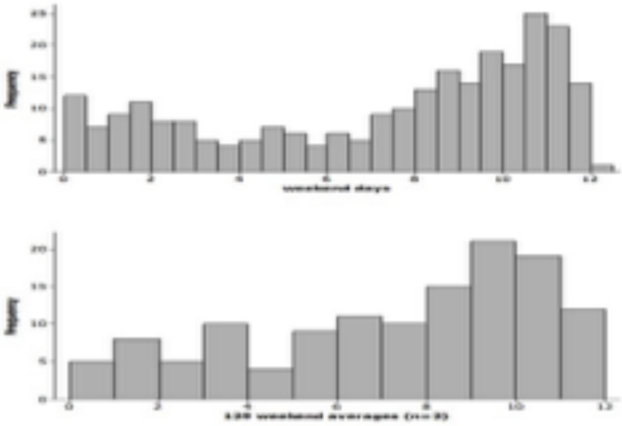
Standard deviation of an average: Accuracy

- standard deviation can be used to find the accuracy of an average
- standard deviation of averages obtained from random samples is given by:
- to increase accuracy of an estimate you would need to reduce the standard error and increase the sample size

$$\frac{\sigma}{\sqrt{n}} = \text{standard deviation of } \bar{y} = \text{standard error of } \bar{y} = \sigma_{\bar{y}}$$

- **standard error**: used for the standard deviation of an estimate of a population parameter based on a sample
- formula is for the standard error of a mean
  - applies to means of samples from any distribution having a finite standard deviation
  - based on the assumption that the components in the average are completely independent of each other
  - with correlated data there is less variation and so the SD is smaller

EXAMPLE



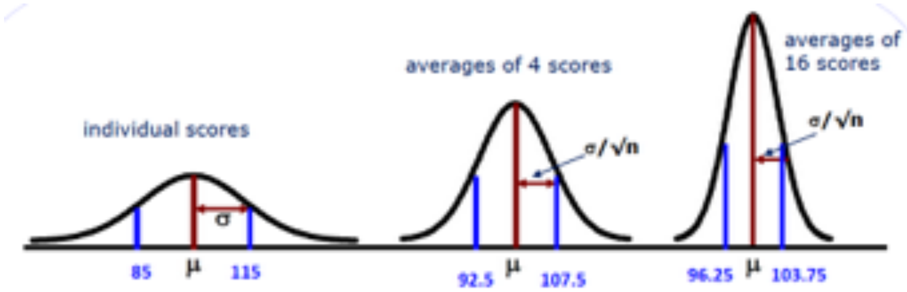
- second histogram has more concentrated averages in the middle of the range

Sampling distributions: IQ scores

- quadrupling the sample size doubles the accuracy e.g. variation among sample means from samples of size 16 is half the variation among sample means from samples of size 4

number of IQ scores in each sample	standard deviation of average (SD/ Square root of sample
1	15.00
2	10.61
4	7.50
8	5.30
16	3.75

Normal population distributions: IQ scores



- 68% people have IQs between 85 and 115
- 68% samples of size 4 have IQs between 92.5 and 107.5
- 68% of samples of size 16 have IQs between 96.25 and 103.75

Other population distributions

- uniform or rectangular distribution, each outcome has the same chance of occurring

Probability calculations involving averages

- needing to calculate the probability that an average of a fixed number of components is greater than some specified value
- dealing with an average **based on n components** instead of a single measurement uses this formula:

$$z = \frac{\bar{y} - \mu}{\sigma / \sqrt{n}}$$