

Review of basic statistics (Chapter 14 and 15 part I)

Bold and Italic typed is my prediction that it will appear on the exam

A. Terminology

Population: entire set of all elements that share common characteristics

Sample: subset of population

Parameter: summary that describe the characteristics of a population (e.g. μ , σ)

Statistics: summary that describe characteristics of a sample

Application question testing the difference between parameter and statistics

Goal of marketing research: learn about the population through collecting samples which can help describe and make inference about the population

How do we learn about the population?

B. Types and definitions of statistics

- Descriptive statistics: summarize data (e.g. s , \bar{x})
 - Expression: graphical (e.g. histogram) and numerical (mode, median, mean)
- Inferential statistics: drawing conclusions about populations based upon statistics
- Population distribution: frequency distribution of all possible observation in the population (e.g. μ , σ)

It is unknown

- Sample distribution: frequency distribution of observations in the sample (e.g. s , \bar{x})

It is known

Used to get info about the population

What are the differences between population and sample distribution

C. Normal Distribution

1) Properties

→ area under the curve equals to probability of occurrence

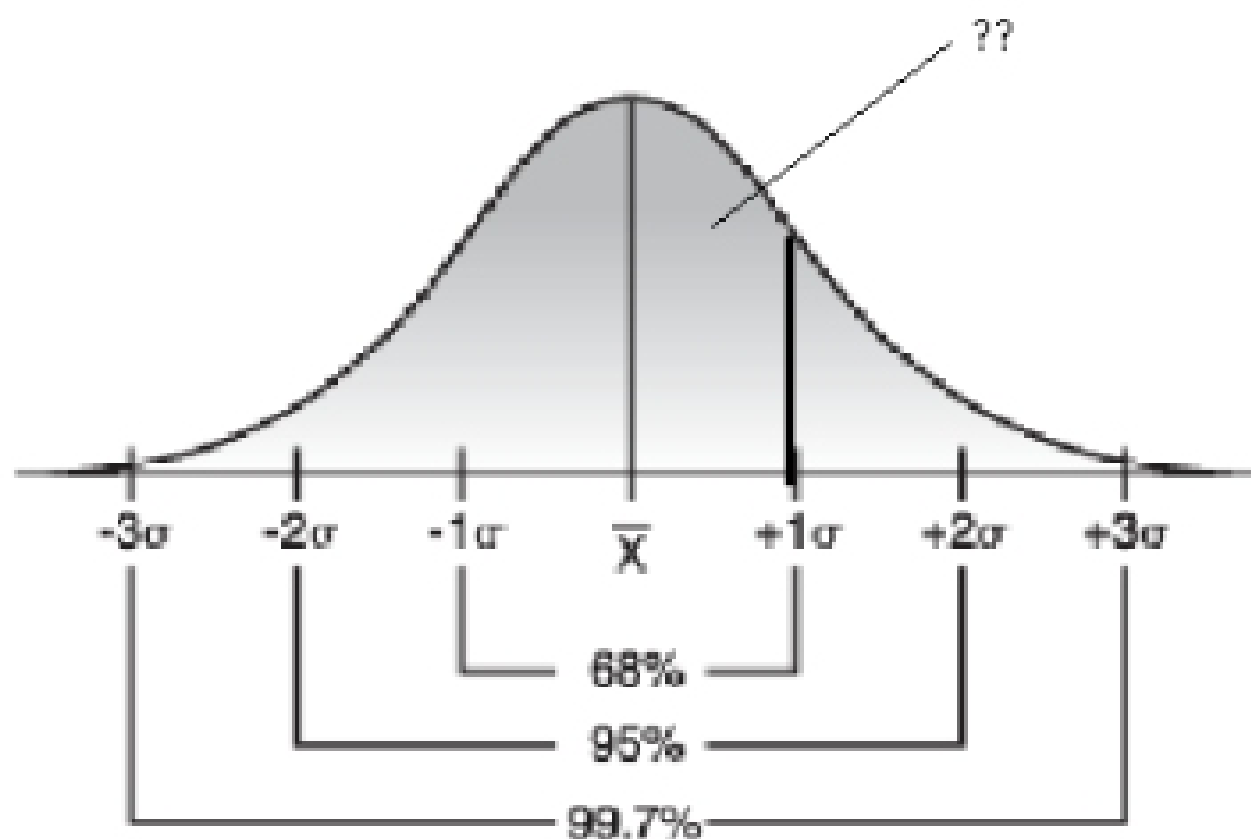
→ symmetric across the mean; each side contains the probability of 0.5

(mean=median=mode)

→ total area under the curve = 1

→ bell-shaped

2) Empirical rule



Area of ?? = $0.68/2 = 0.34$

3) Standard normal distribution

- has a mean of 0 and a standard deviation of 1

Review of Basic Statistics Part II (Distribution) Chapter 14

Bold and Italic typed is my prediction that it will appear on the exam

A. Variables and statistics

- Variable: characteristics of interest of population or sample
- Variable of interests of population is very large (infinite), while it is the opposite for sample

Comparison between sample and population

B. Graphical summaries of descriptive statistics

- Open-ended question: word cloud
- Nominal and ordinal data: pie chart, bar chart, and frequency table
- Interval and ratio: histogram

Test whether nominal and ordinal data use bar chart or histogram

Why can't we use pie chart to represent interval or ratio data?

➔ Hard to summarize and interpret

C. Graphical summaries of descriptive statistics

- Measure of central tendency and calculations (median, mean, mode)

Remember median requires arranging data from small to large

When to use mean or median? Median is suitable for skewed data as skewedness will affect the mean

- Numerical summaries: measures of dispersion
 - i) Variance, standard deviation and range
 - ii) *Cannot report means and standard deviation for nominal and ordinal data*

D. Application of statistics to marketing research

- first use descriptive statistics

Review of Inferential Statistics (Sampling distributions and CLM)

Bold and Italic typed is my prediction that it will appear on the exam

A. Sampling distribution of sample mean

- *Understand the population, variable of interest, and truth of population*
- 1) Sampling distribution of sample mean
 - *sampling distribution of sample mean* is all possible values of \bar{x} for across all possible samples of the same size
- 2) Central Limit Theorem
 - for any variable, regardless of its distribution, the distribution of variable will approach a normal distribution as n increases
 - implications:
 - mean of \bar{X} = mean of X
 - standard error = σ divided by square root of n
 - the result holds for any distribution
- 3) Distributions of interest
 - i) population distribution
 - smooth line μ, σ
 - *again, it is unknown*
 - ii) sample distribution
 - frequency distribution of observations of variable in the sample
 - \bar{x} and s
 - *again, it is known*
 - iii) Sampling distribution
 - THEORETICAL frequency distribution of ALL possible sample means
 - mean of \bar{X} and standard error

B. Applying the CLT

- 1) Analysis of ordinal/nominal data
 - want to know the population proportion (because it is not ratio/interval data)