Introductory Stats Notes

Terry Yu

2021

Contents

§1 Basics

Definition — Population: complete collection of objects, persons, or things under study.

Definition — .Parameter: Some measurable characterisc with refercne to population

Definition — Sample: is a subset of a population

Definition — Statistic: some measurable characteristic with reference to sample

Example — A survey of 2104 households ini the United States found that 65 percent subscribe to cable television. What is the population, sample, parameter, and statistic?

Solution. Teh population is all the households in the United States, whereas the sample is just the 2104 households sampled. The Parameter is □

Definition — Types of Data:

1. Attributive (Qualitative) Data: data that is not quantifaible (not represented wiht a numnber)
2. Quantitative Data: Countable data with whole numbers, such as # of thngs
3. Continuous Dta: Measurable quantities, such as time, height

§2 Descriptive Statistics

imean median mode

§2.1 Normal Distribution

Standard Deviation — Standard deviation for a sample is

$$s = \sqrt{\frac{\sum (x - \overline{x})^2}{n-1}}$$

whereas for a population is

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{n}}$$

Z Score — Standard score or Z score makes it possible to compare sores from different distributsions, with different means and different deviations, measuring in sd units

$$z = \frac{x - \mu}{\sigma}$$

§2.2 Binomial Distribution

Notation: $X\tilde{B}(n,p)$ where n is the number of times the same action is repeated, and the p is fixed for which its the probability of success and $q = 1 - p$ is its complement.

Binomial Distribution — Mean:

$$\mu = np$$

sd:

$$\sigma = \sqrt{npq}$$

Probability for exactly x chances of success:

$$p = \binom{n}{x} p^x q^{n-x}$$

§2.3 Geometric Distribution

A random variable that counts the number of trials until the first success: $X\tilde{G}(p)$.

Geometric Distribution — Probability of occurint x times before succeeding:

$$p = q^{x-1}p$$

Mean:

$$\mu = \frac{1}{p}$$

§2.4 Normal Approximation for Binomial Distribution

Continuity Correction Factor (CCF) — To convert between binomial and normal, have to approxiamte with ccf

Once approximated, use normal z score and z calculation with the binomial distribution mean and standard deviation.

Theorem (Central Limit Theorem)
The Central limit theorem compares the means in sample and in the population, saying that

1. The distribution of $\overline{x}$ is normal
2. mean of samples is population mean
$$\mu_{\overline{x}} = \mu$$
3. standard deviation of sample means is
$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

Definition — Point Estimate: a single number used as a best guess, like the sample mean $\overline{x}$

Definition — Error: distance bewteen teh point estimate and population mean

Confidence Interval — Confidence intervals are to estimate a population thing from the sample, with error. Error is computed with
$$E = z\frac{\sigma}{\sqrt{n}}$$
with the confidence interval being
$$\overline{x} \pm E$$

Example —