MATH 205: Statistical methods

Lecture 22: Samples and populations

Announcements

Next week

- Monday–Wednesday: Midterm-Simulation in the labs.
  Open book + internet.
- Wednesday's lecture: Review session before the written exam
- Friday: Midterm-Written.
  Closed book.
  Can use calculator.
  Can bring a A4-sized hand-written one-sided note to the exam.

Chapter 6: Samples and Populations

6.1 The Sample Mean

6.2 Confidence Intervals

Samples and populations

- Very often, the data we see is a small part of the data we could have seen
- The data we could have observed, if we could have seen everything, is the *population*
- The data we actually have is the sample

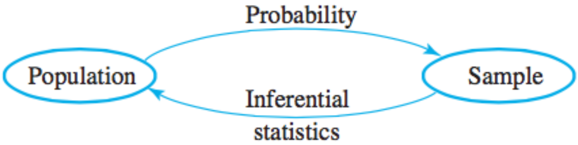
Samples and populations

This situation occurs very often

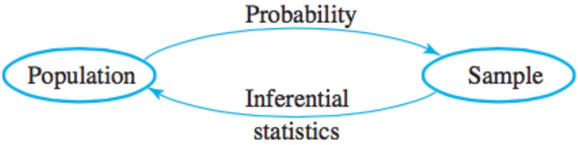
- Imagine we wish to know the average weight of a rat. This isn't random; you could weigh every rat on the planet, and then average the answers.
- Instead, we weigh a small set of rats, chosen at random but rather carefully so.
- If we have chosen sufficiently carefully, then we can say a great deal from the sample alone

Distributions are like populations

- we can think about population as a probability distribution P
- the samples are random variables X generated from P
- from the observed values of the samples, we want to infer properties about P



Random sample



Definition

The random variables $X_1, X_2, ..., X_n$ are said to form a (simple) random sample of size n if

1. the X_i's are independent random variables
2. every X_i has the same probability distribution

The sample mean is an estimate of the population mean

Definition

Let $X_1, X_2, ..., X_n$ be a random sample from a distribution. The sample mean is defined as

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- The sample mean is a random variable.
- It is random, because different samples from the population will have different values of the sample mean.

Reminder: notations

- Let $X_1, X_2, \dots, X_n$ be a random sample of size n
- The sample mean of $X_1, X_2, \dots, X_n$, defined by

$$\bar{X} = \frac{X_1 + X_2 + \dots X_n}{n},$$

  is a random variables
- When the values of $x_1, x_2, \dots, x_n$ are collected,

$$\bar{x} = \frac{x_1 + x_2 + \dots x_n}{n},$$

  is a realization of the $\bar{X}$, and is a number

The sample mean is an estimate of the population mean

Questions:

- What can we say about the distribution of $\bar{X}$?
- When can we use $\bar{X}$ to estimate the population mean with confidence?

Linear combination of random variables

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent random variables (with possibly different means and/or variances). Define

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

then the mean and the standard deviation of T can be computed by

- $E(T) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$
- $Var(T) = a_1^2Var(X_1) + a_2^2Var(X_2) + \ldots + a_n^2Var(X_n)$

Linear combination of normal random variables

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent normal random variables (with possibly different means and/or variances). Then

$$T = a_1 X_1 + a_2 X_2 + \ldots a_n X_n$$

also follows the normal distribution with

- $E(T) = a_1 E(X_1) + a_2 E(X_2) + \ldots + a_n E(X_n)$
- $Var(T) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + \ldots + a_n^2 Var(X_n)$

Recall: Example

Problem

A gas station sells three grades of gasoline: regular unleaded, extra unleaded, and super unleaded. These are priced at 2.20, 2.35, and 2.50 per gallon, respectively.
Let X_1, X_2, and X_3 denote the amounts of these grades purchased (gallons) on a particular day. Suppose the X_i's are independent with $\mu_1 = 1000$, $\mu_2 = 500$, $\mu_3 = 300$, $\sigma_1 = 100$, $\sigma_2 = 80$, $\sigma_3 = 50$. Compute the expected value and the standard deviation of the revenue from sales

$$Y = 2.2X_1 + 2.35X_2 + 2.5X_3.$$

Mean and variance of the sample mean

Problem

Given independent random samples $X_1, X_2, ..., X_n$ *from a distribution with mean* μ *and standard deviation* σ*, the mean is modeled by a random variable* $\bar{X}$*,*

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Compute $E(\bar{X})$ *and* $Var(\bar{X})$*(in terms of* μ *and* σ*.)*

Mean and variance of the sample mean

Theorem

Given independent random samples $X_1, X_2, ..., X_n$ from a distribution with mean μ and standard deviation σ, the mean is modeled by a random variable $\bar{X}$,

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Then

$$E[\bar{X}] = \mu$$

and

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

Law of large numbers

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then

$$\bar{X} \to \mu$$

as n approaches infinity