Lecture 8

Sampling Distributions

agenda

Announcements	HW 3 due ~~Friday~~ Sun (does this work??)
M&M Activity	The best activity ever!
Simulation Activity	Simulate M&M's in R
Mini-Lecture	Sampling distributions

Quick Review

Parameters

A summary characteristic (e.g., mean, proportion, etc.) of a population

A parameter is a fixed, unchanging value.

Statistics

A summary characteristic (e.g. sample mean, sample proportion, etc.) of a sample

A statistic is a random variable, which describes a sample.

Binomial Distribution

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Bern}(p)$$

$$X_i = \{0, 1\}$$

$$E[X_i] = p$$

$$\text{var}[X_i] = p(1-p)$$

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Binom}(n, p)$$

$$X_i = \{0, 1, 2, \ldots, n\}$$

$$E[X_i] = np$$

$$\text{var}[X_i] = np(1-p)$$

M&M activity directions

01

Group up!

Groups of 5 students

02

Questions

Answer the preliminary thought questions

03

Get M&M packs

Send ONE representative to pick up one pack of M&M per group member

04

Count your M&Ms

Compute the proportion of **blue** M&M's in your bag (round to 2 decimal places)

05

Make a dot plot

mark a dot on the board corresponding to your individual proportion

06

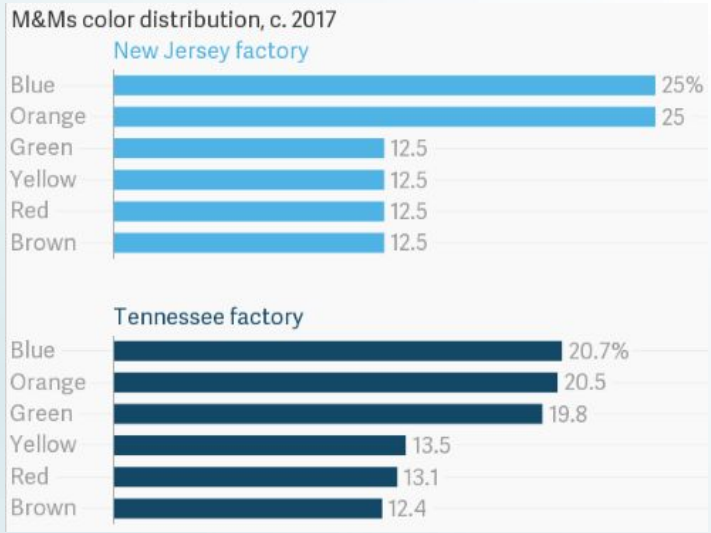
Reflect

answer reflection questions

Types of distributions

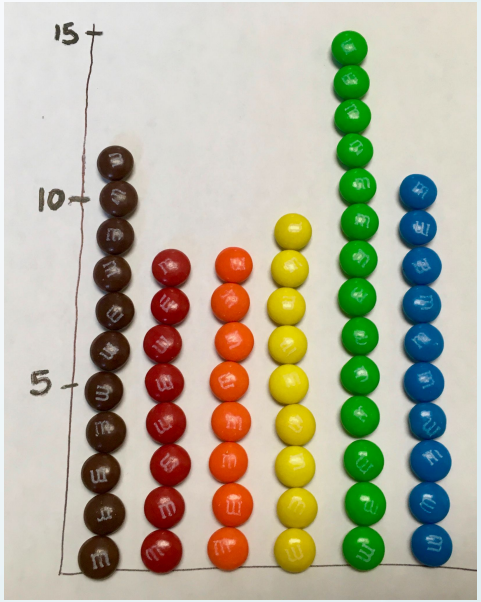
Population distribution

describes the variation of values from a population



Sample distribution

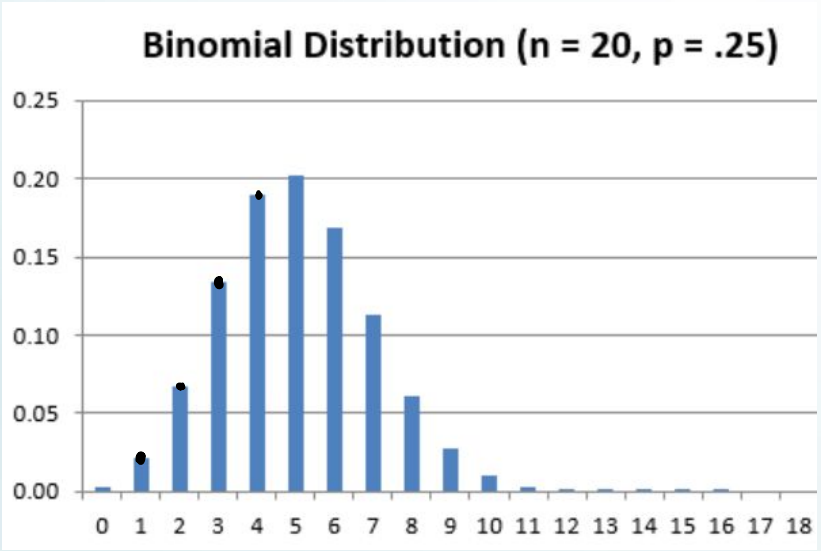
describes the variation of values from a sample



Types of distributions

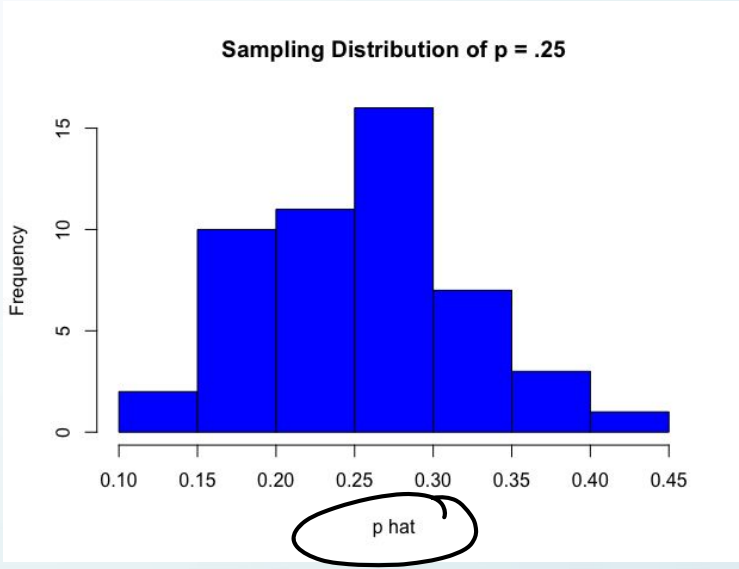
Probability distribution

describes the variation of all possible values of a random variables by it's probability



Sampling distribution

describes the variability of all possible values of a sample statistics across multiple samples



Properties of sampling distributions

The general format of the **sampling distribution** of a statistic is the same for all of the statistics:

- The mean of the sampling distribution is the population parameter that the statistic is estimating.
  - Ex: $E[\hat{p}] = p$
- The standard deviation of the sampling distribution measures how the values of the sample statistic might vary across different samples from the same population
  - A function of the population parameter and the sample size n.
  - Ex: $E[\hat{p}] = \sqrt{\frac{p(1-p)}{n}}$
- As long as certain conditions are met (i.e., sample size is large), the sampling distribution is *approximately normal*.

Central Limit Theorem

CLT (for proportions)

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} Bern(p)$. Define $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then,

$$\hat{p} \dot{\sim} N\left(\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$$