Stat 202 - 2015S - W9 - Friday (Pg 1)

~~Yesterday~~ Last Wednesday - Sampling Distributions for means

Today - Sampling Distribution for Counts and Proportions

Review from ~~Yesterday~~ Yesterday

A statistic computed from observations made based on a random sample is a random variable.

Its probability distribution is called a <u>sampling distribution</u>

Examples Pick 30 AU students at random.

Wed - Compute their average ~~G~~ GPA, (across the 30 in the sample)

today - Compute the proportion (of the 30 students) who prefer snow days to class days

- OR count the number of students (out of 30) who prefer snow days to class days.

Main idea - if you do this experiment over and over again (with different samples of 30 students each time) you will see a regular distribution of values that is predictable because the sample is random.

Wed - mean of population μ
This is what you are trying to estimate by computing the sample mean $\overline{x}$.

Facts about sample means

1. Sample means are less variable than individual observations
2. Sample means are more normal than individual observations

The case $n=1$ (one person in sample) is special.

Sampling Distribution = Population Distribution when $n=1$

Rules $\begin{cases} \text{For} \\ \text{Population mean: } \mu \\ \text{Population standard deviation: } \sigma \end{cases}$

$\mu_{\bar{x}} = \mu$ (always an unbiased estimat if sample is random)

that's why we like random samples

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$ | $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

IF Population Distribution $\sim N(\mu, \sigma)$

Sampling Distribution for mean $\sim$

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Even if population distribution is not normal

Sampling Distribution For mean

$$\overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

For large n

(Approximation gets better with statistical certainty as n get increasingly large).

The mean was for a quantitative variable: an individual's GPA or height

Today we are going to do categorical variables
- Does an individual prefer snow days to class days

OR - Does an individual prefer red over blue.

OR - Is an individual Male or Female,

Let's say we want to find out ~~how many~~ what percentage of AV students prefer snow days, what percentage prefer red over blue and what percentage are female.

We chose 30 students at random and give them a survey.

Our class is not a random sample. Nonetheless let's do it

How many prefer snow days?

⋮

What are the estimates for the proportions of all AU students

$\boxed{\text{Same numbers}}$

Because our class is not a random sample of the population of AU students our estimates are biased.

If they were from a random sample we would have

$\mu_{\hat{p}} = p \leftarrow$ true population proportion

$\hat{p} \leftarrow$ symbol for estimate of proportion

$\mu \leftarrow$ mean (of estimates of proportion from samples of size n)

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

If our estimates of proportion were right, what would the standard deviations of our estimates be from n = number of students in our class

Where do these formulas come from?

If you pick n people at random from a population, ~~[illegible]~~ with a proportion p of having a certain characteristic, the <u>count</u> not proportion of people with that characteristic in the sample is a Binomial Random Variable (approximately)

To Review when binomial applies

1. There is a fixed number of obs (n)
2. The n obs are all independent
3. Each obs fall into two categories:
   have characteristic (success)
   don't have characteristic (Failure),
4. The probability of success is the same for all obs. (p)

X = Count of successes
$X \sim B(n, p)$

↳ the way we denote a binomial distrib. with paraments n (obs) p (prob of success for each)

I said that the count of people in a random sample that possess a certain characteristic is approximately but is not exactly a binomial random variable.

Let's say we choose 5 people at random from our class.

yes 1) Is there a fixed number of obs?

no 2) Are they independent (Knowing one tells you nothing about the rest)
Knowing we pick Mark first means we know we don't pick Mark second

yes 3) Two categories? yes

no 4) Probability p the same for all
no because proportion of men changes when we pick one man first

So I derived the other day (tuesday)

$$\left.\begin{aligned} \mu_X &= np \\ \sigma_X &= \sqrt{np(1-p)} \end{aligned}\right\}$$ For a binomial distribution X

~~binomial distribution (approximate counts of people in random sample that possess characteristic)~~

If X is the count of the number of people in the random sample with a given characteristic, then X is approximatly Binomial

The rule of thumb says the population size has to be at least 20 times as large as the sample size

$\frac{X}{n}$ is then an estimate of the proportion p

$$\mu_X = np \qquad \mu_{\frac{X}{n}} = \frac{np}{n} = p$$

$$\sigma_X^2 = np(1-p) \qquad \sigma_{\frac{X}{n}}^2 = \frac{1}{n^2}(n)(p)(1-p)$$

$$\sigma_X = \sqrt{np(1-p)} \qquad = \frac{p(1-p)}{n}$$

$$\sigma_{\frac{X}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\frac{X}{n} = \hat{p}$$

$$\left.\begin{aligned} \mu_{\hat{p}} &= p \\ \sigma_{\hat{p}} &= \sqrt{\frac{p(1-p)}{n}} \end{aligned}\right\} \text{what I said earlier.}$$

Normal approximation for counts
and proportions – in certain situations a binomial random variable is approximately Normal

X is approximately $N(np, \sqrt{np(1-p)})$

$\hat{p}$ is approximately $N(p, \sqrt{\frac{p(1-p)}{n}}$

$\boxed{\text{Show StatCrunch}}$