

Stat 202-2015~~S~~-W~~9~~-Friday^{Monday} (Pg 1)

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

Today - Sampling Distribution for Counts and Proportions

Review from ~~Monday~~ Yesterday

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

Its probability distribution is called a Sampling distribution

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 $\bar{x}$.

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
when $n=1$ = Population Distribution

Rules { For
 Population mean: μ
 Population standard deviation: σ
 $\mu_{\bar{x}} = \mu$ (always an unbiased estimate if sample is random)
 $\nearrow$ that's why we like random samples

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

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

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

for large n

(Approximation gets better with statistical certainty $\rightarrow n$ 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 ~~what percentage of students prefer snow days~~ 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

Same numbers

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

IF they were ~~from~~ from a random sample we would have

$\mu_{\hat{p}} = p$ true population proportion
 $\uparrow$ symbol for estimate of proportion
 $\uparrow$ 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, ~~that~~ with a
proportion p of having a certain
characteristic, the count ^{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)

~~Check~~
 $X =$ (count of successes)

$$X \sim B(n, p)$$

↳ the way we denote a binomial distrib.
with parameters 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)

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

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

~~The proportion~~

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

The rule of thumb says the population size has to be ^{at least} 120 times ^{as large as} the sample size

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

$$\mu_X = np$$

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

$$\sigma_X^2 = np(1-p)$$

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

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

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

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

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

$$\mu_{\hat{p}} = p$$

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

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

Show Statcrunch

→ Use this approximation when
 $np \geq 10$ and $n(1-p) \geq 10$