Chapter 8: Sampling Distributions

Section 8.2: Distribution of the Sample Proportion

p = population proportion $= \frac{X}{N}$ $\quad \hat{p}$ = sample proportion $= \frac{x}{n}$

proportions are #s b/w 0 & 1.

SAMPLE PROPORTION

We want to know the answer to the question:

Do you prefer taking your courses exclusively online instead of in person?

The answers are Yes or No, which are qualitative variables. We can make these quantitative by using **proportions**.

Poll: $n = 10$ Yes $= 1/10$ No $= 9/10$

Ex1: According to a recent survey by GCC, 20% of students prefer taking courses exclusively online.

(a) What is the proportion of students who prefer taking courses online? $p = 0.2$

(b) What is the proportion of students who prefer taking courses in-person? $1 - p = 0.8 = q$

(c) Take a poll in class:

how many prefer taking courses exclusively online: 1

how many prefer taking courses in-person: 9

proportion of people in today's class who prefer taking courses exclusively online: $1/10 = 0.1$

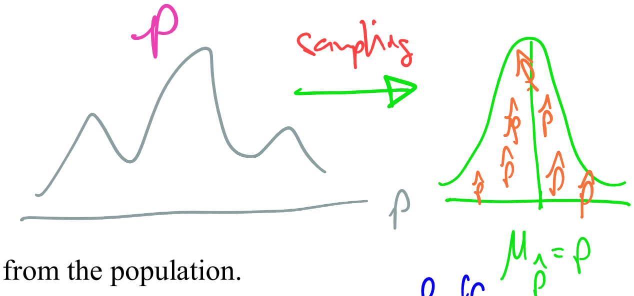
proportion of people in today's class who prefer taking courses in-person: $9/10 = 0.9$

Def The **sample proportion** is the proportion of individuals who share a characteristic obtained from a sample of the population. The **population proportion** is the proportion of all individuals who share a characteristic.

The sample proportion estimates the population proportion.

Notation:	Sample $\hat{p}$	Population
sample size:	n	N
# of people in sample who share characteristic:	x	X
sample proportion:	$\hat{p}$	p

SAMPLING DISTRIBUTION OF THE PROPORTION



Given

1. The random variable x has a distribution with proportion p.
   (the distribution may or may not be normal)
2. Simple random samples of size n are independent and selected from the population. SRS

Conclusions

1. The shape of the sampling distribution of all sample proportions $\hat{p}$ is approx. normal (use normalcdf) provided that:

$$\boxed{np(1-p) \geq 10}$$ need to check

2. The mean of all sample proportions is the population proportion: $\mu_{\hat{p}} =$ p

3. The standard deviation of all sample proportions is given by $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

$$\boxed{\mu_{\hat{p}} = p \qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}}$$

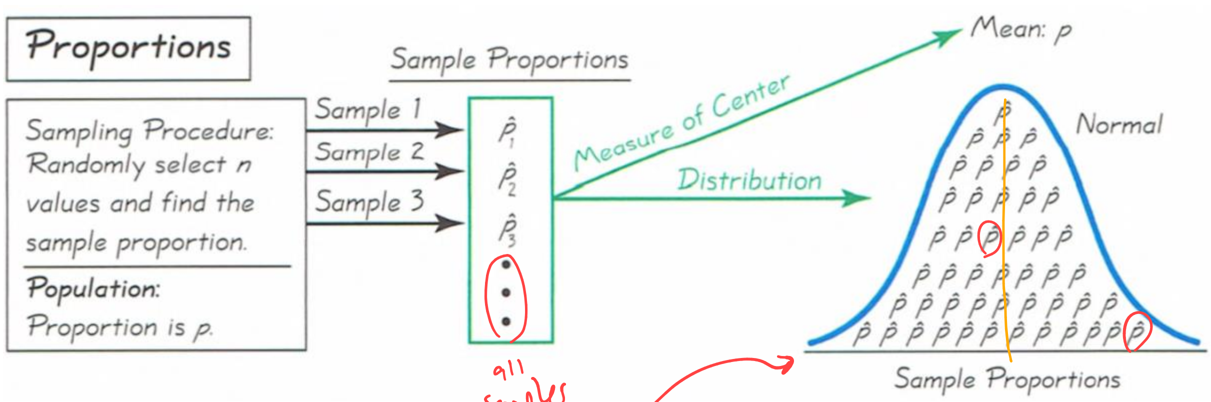
$(q = 1 - p)$

Application of Result

4. The probability of a indep SRS of size n having a (sample) proportion between a and b is $P(a < \hat{p} < b)$.

sampling

We call $\sigma_{\hat{p}}$ the **standard error** of the **proportion**.



all samples of size n

Notes About Distributions

One needs to check independence or use the *5% Guideline for independence* (i.e. the sample size n is less than 5% of the population size N, i.e. $n \le 0.05\,N$)

STEPS to Use Sampling Distribution for $\hat{p}$
Step 0: Are you selecting n people with n>1? If yes, continue:
1st: Check if SRS are independent AND $np(1-p) \ge 10$
2nd: Find $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$
3rd: Use normalcdf for the rest…just don't forget to use information from *step 2!!*

IMPORTANT DISTINCTION: WHEN DO YOU USE A SAMPLING DISTRIBUTION?

We use the sampling distribution to compute probabilities of random samples of size $n > 1$.

!

Ex2: According to a recent article, 32% of drivers had driven drowsy in the past month[1].
Law enforcement officials are planning a survey of 1000 drivers to determine what proportion are driving drowsy.
(a) Is whether a driver has driven drowsy in the past month qualitative or quantitative?

QL QN

This is qualitative ("Yes" or "No") → convert to QN using a proportion.

(b) Describe the sampling distribution of $\hat{p}$, the sample proportion of drivers who exceed the speed limit.

↳ n = 1000 > 1 ✓

- Sampling distribution will be approximately normal if:
  - * SRS? yes, as long as survey is done right.
  - * Independent? Yes, use 5% Guideline
  - * $npq \ge 10$?

    $n = 1000$
    $p = 0.32$
    $q = 1 - 0.32 = 0.68$

    $1000 * 0.32 * 0.68 = 217.6 \ge 10$ ✓
- mean of sampling dist is $\mu_{\hat{p}} = p = 0.32$

  st dev. " is $\sigma_{\hat{p}} = \sqrt{\dfrac{pq}{n}} = \sqrt{\dfrac{0.32 * 0.68}{1000}} = 0.0148$

---

[1] Source: thenationshealth.aphapublications.org/content/41/10/E52.full

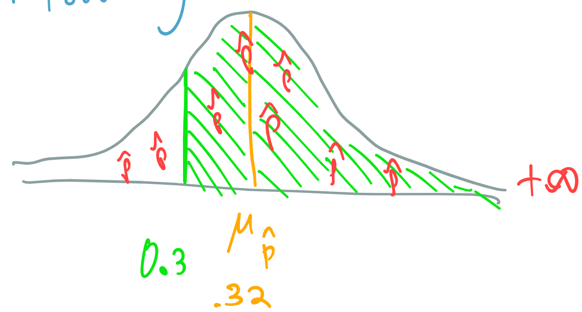
sampling

(c) In a random sample of 1000 drivers, what is the probability that more than 30% of drivers will have driven drowsy in the last month?

$n=1000$

$$P(\hat{p} > 0.3) = \text{normalcdf}(\underset{a}{0.3}, \underset{b}{1E99}, \underset{\mu_{\hat{p}}}{0.32}, \underset{\sigma_{\hat{p}}}{\sqrt{\tfrac{0.32*0.68}{1000}}})$$

$$= \boxed{0.912}$$



(d) Would it be unusual if in a random sample of 1000 drivers, 35% or more were driving drowsy in the last month?

prob ≤ 0.05

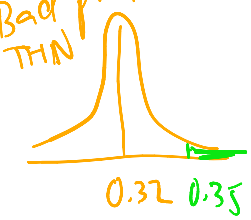
sampling

$$P(\hat{p} > 0.35) = \text{normalcdf}(\underset{a}{0.35}, \underset{b}{1E99}, \underset{\mu_{\hat{p}}}{0.32}, \underset{\sigma_{\hat{p}}}{\sqrt{\tfrac{0.32*0.68}{1000}}})$$

$$= \boxed{0.0210} \le 0.5 \text{ yes!}$$

This is unusual!

Bad picture!



(e) What is the minimum number of drivers that must be sampled to be sure that $\hat{p}$ is approximately normal?

trickier Q

sample size n

need: $n \cdot p \cdot q \geq 10$

$$\frac{n*0.32*0.68}{0.32*0.68} \geq \frac{10}{0.32*0.68}$$

$$n \geq \frac{10}{0.32*0.68} = 45.955\ldots$$ always round ↑

$\boxed{n \geq 46}$ "The minimum number of drivers that must be sampled to be sure that $\hat{p}$ is approximately normal."

Answers to Ex2:

(a) Qualitative

(b) Approximately normal with mean 0.32 and standard deviation 0.0148

(c) 0.912

(d) Yes, since the probability that 35% or more would have driven drowsy in the last month is 0.0209, or 2.09%.

(e) 46