M358K: October 28th, 2020.

Normal Approximation to the Binomial [practice].

Problem. "True/False!"

An exam w/ 100 questions total; all True/False. A student knows correct answers to exactly 36 questions; the rest of the questions, he guesses @ random. What is the probab. that he gets @ least 70 on the test?

→: # of questions he still needs to guess correctly is

$$70-36=34$$

X... # of questions he guesses correctly

$X \sim \text{Binomial}(size = n = 64, prob = p = 0.5)$

$\mathbb{P}[X \geq 34] = ?$

1st Using R: $\mathbb{P}[X \geq 34] = 1 - \mathbb{P}[X < 34]$

$$= 1 - \underbrace{\mathbb{P}[X \leq 33]}_{\text{cdf of } X}$$

$= 1 - \text{pbinom}(33, \text{size} = 64, \text{prob} = 0.5)$

$= 0.35399$

2nd Using the Normal Approximation:

- **mean**: $\mathbb{E}[X] = n \cdot p = 64(0.5) = 32$

- variance: $\text{Var}[X] = n \cdot p \cdot (1-p) = 64(0.5)(0.5) = 16$

$\Rightarrow$ the std deviation: $SD[X] = 4$

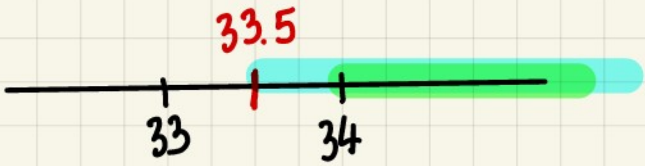
$$\mathbb{P}[X \geq 34] = \mathbb{P}\left[\underbrace{\frac{X-32}{4}}_{\text{"}\sim\text{"}\, N(0,1)} \geq \frac{34-32}{4}\right]$$

naïve

$$= \mathbb{P}[Z \geq 0.5] = 1 - \mathbb{P}[Z < 0.5]$$

$$= 1 - 0.6915 = 0.3085$$

std normal tables



$$\mathbb{P}[X \geq 34] = \mathbb{P}[X \geq 33.5] = \mathbb{P}\left[\underbrace{\frac{X-32}{4}}_{\text{"}\sim\text{"}\, N(0,1)} \geq \frac{33.5-32}{4}\right]$$

the continuity correction

$$= \mathbb{P}[Z \geq 0.375]$$

$$= 1 - \mathbb{P}[Z < 0.375]$$

$$= 1 - \frac{1}{2}(0.6443 + 0.6480)$$

std normal tables

$$= 0.35385$$

Problem. A fair coin is tossed 100 times.

Q: What's the probability of getting exactly 50 heads?

(i) "Guesstimate": 0.001, 0.01, 0.1, 0.5, 0.9, 0.99.

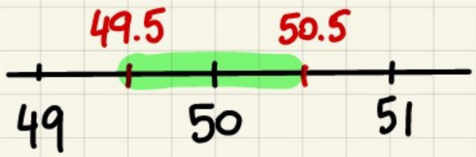
(ii) "Exactly": Using R:

dbinom(50, size = 100, prob = 0.5) = 0.0796

(iii) "Approximately": $X \sim \text{Binomial}(n=100,\ p=0.5)$

1^{st} $\mathbb{P}[X=50] = \binom{100}{50}(0.5)^{50}(0.5)^{50}$ $\rightarrow$ Sterling Formula

2^{nd} $\mathbb{P}[X=50] = \mathbb{P}[49.5 < X < 50.5]$



- $\mathbb{E}[X] = n \cdot p = 100(0.5) = 50$
- $\text{Var}[X] = n \cdot p \cdot (1-p) = 100(0.5)(0.5) = 25$

$\Rightarrow SD[X] = 5$

$$\mathbb{P}\left[\frac{49.5-50}{5} < \underbrace{\frac{X-50}{5}}_{\text{"}\sim\text{"}\ N(0,1)} < \frac{50.5-50}{5}\right]$$

$$= \mathbb{P}[-0.1 < Z < 0.1]$$

$$= \mathbb{P}[Z < 0.1] - \underbrace{\mathbb{P}[Z \leq -0.1]}_{= 1-\mathbb{P}[Z \leq 0.1]}$$

$$= 2 \cdot \mathbb{P}[Z \leq 0.1] - 1$$

$$= 2 \cdot (0.5398) - 1 = 0.07966$$

std normal tables

Statistical Inference for the Population Proportion

Let p denote the probability that a randomly chosen member of the population has a certain property, i.e., an opinion on the mayor,
vote for the pink party,
allergic to Wensleydale, ...

p is unknown and is our parameter of interest :)

Let's say that every time that we have an experimental unit in the sample it's a "success".

X ... # of successes in a sample of size n

$$\boxed{X \sim \text{Binomial}(\#\text{ of trials} = n, \text{ prob. of success} = p)}$$

$\hat{p}$ ... statistic of interest; i.e., the sample proportion

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