Class 18

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

Recap Chapter 9.1

Lecture Chapter 9.2

Recap Chapter 9.1

9: Inferences Involving One Population

9.1 Inference about the Mean μ (σ Unknown)

In Chapter 8, we performed hypothesis tests on the mean by

1) assuming that $\bar{x}$ was normally distributed (n "large"),

2) assuming the hypothesized mean μ_0 were true,

3) assuming that σ was known, so that we could form

$z^* = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$ which with 1) – 3) has standard normal dist.

9: Inferences Involving One Population

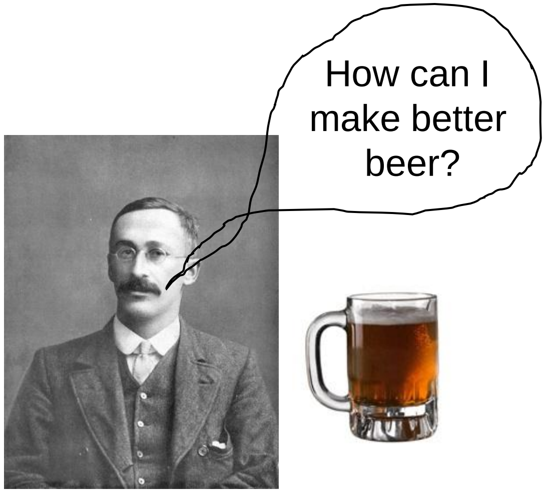
9.1 Inference about the Mean μ (σ Unknown)

However, in real life, we never know σ for

$$z^* = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

so we would like to estimate σ by s, then use

$$t^* = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \quad .$$



But t^* does not have a standard normal distribution.

It has what is called a Student t-distribution. ⟵ Gosset Guinness Brewery

9: Inferences Involving One Population

9.1 Inference about the Mean μ (σ Unknown)

Using the t-Distribution Table

Finding critical value from a Student t-distribution, df=n-1

$t(df,\alpha)$, t value with α area larger than it

with df degrees of freedom

Table 6
Appendix B
Page 719.

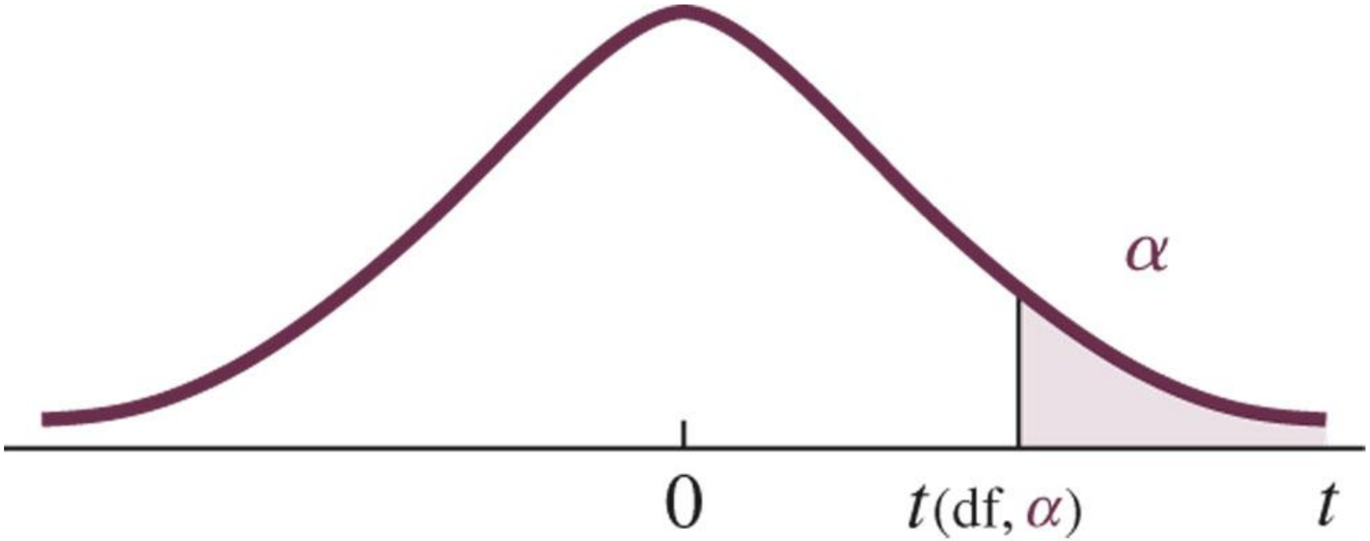
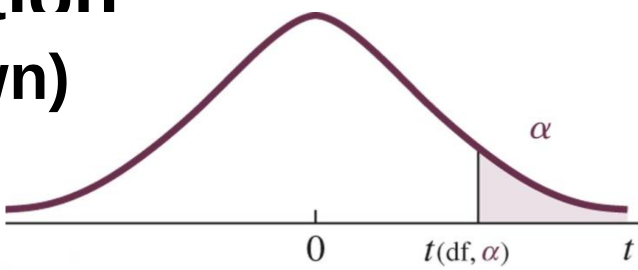


Figure from Johnson & Kuby, 2012.

9: Inferences Involving One Population

9.1 Inference about the Mean μ (σ Unknown)

Example: Find the value of $t(10,0.05)$,
$df=10$, $\alpha=0.05$.



Area in One Tail	0.25	0.10	0.05	0.025	0.01	0.005
Area in Two Tails / df	**0.50**	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**
3	0.765	1.64	2.35	3.18	4.54	5.84
4	0.741	1.53	2.13	2.78	3.75	4.60
5	0.727	1.48	2.02	2.57	3.36	4.03
6	0.718	1.44	1.94	2.45	3.14	3.71
7	0.711	1.41	1.89	2.36	3.00	3.50
8	0.706	1.40	1.86	2.31	2.90	3.36
9	0.703	1.38	1.83	2.26	2.82	3.25
10	0.700	1.37	1.81	2.23	2.76	3.17
⋮						
35	0.682	1.31	1.69	2.03	2.44	2.72
40	0.681	1.30	1.68	2.02	2.42	2.70
50	0.679	1.30	1.68	2.01	2.40	2.68
70	0.678	1.29	1.67	1.99	2.38	2.65
100	0.677	1.29	1.66	1.98	2.36	2.63
df > 100	0.675	1.28	1.65	1.96	2.33	2.58

Table 6
Appendix B
Page 719.

Go to 0.05 One Tail column and down to 10 *df* row.

Figures from
Johnson & Kuby, 2012.

9: Inferences Involving One Population

9.1 Inference about the Mean μ (σ Unknown)

Confidence Interval Procedure

Discussed a confidence interval for the μ when σ was known,

Confidence Interval for Mean:

$$\bar{x} - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \quad \textbf{to} \quad \bar{x} + z(\alpha/2)\frac{\sigma}{\sqrt{n}} \qquad \textbf{(8.1)}$$

now, with sigma unknown, the CI for the mean is

Confidence Interval for Mean:

$$\bar{x} - t(df, \alpha/2)\frac{s}{\sqrt{n}} \quad \textbf{to} \quad \bar{x} + t(df, \alpha/2)\frac{s}{\sqrt{n}} \qquad \textbf{(9.1)}$$

9: Inferences Involving One Population

9.1 Inference about the Mean μ (σ Unknown)

Recap 9.1:

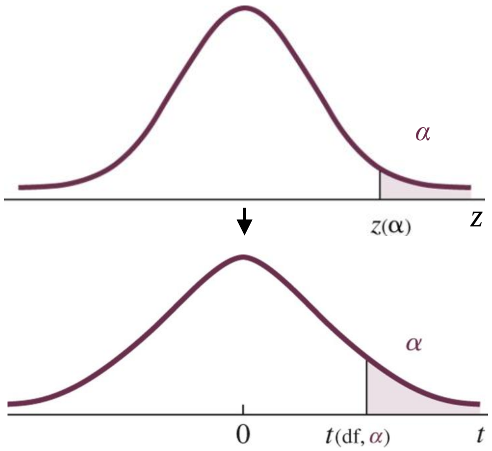
Essentially have new critical value, $t(df,\alpha)$ to look up

in a table when σ is unknown. Used same way as before.

<u>σ assumed known</u> → <u>σ assumed unknown</u>

$$\bar{x} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}} \longrightarrow \bar{x} \pm t(df,\alpha/2)\frac{s}{\sqrt{n}}$$

$$z^* = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \qquad t^* = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$$



Lecture Chapter 9.2

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Chapter 5

We talked about a Binomial experiment with two outcomes.
Each performance of the experiment is called a trial.
Each trial is independent.

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$n = 1, 2, 3, \ldots$$
$$0 \le p \le 1$$
$$x = 0, 1 \ldots, n$$

n = number of trials or times we repeat the experiment.
x = the number of successes out of n trials.
p = the probability of success on an individual trial.

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

When we perform a binomial experiment we can estimate the probability of heads as

Sample Binomial Probability i.e. number of H out of n flips

$$p' = \frac{x}{n} \qquad (9.3)$$

where x is the number of successes in n trials.

This is a point estimate. Recall the rule for a CI is

point estimate $\pm$ some amount

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Background

In Statistics, if we have a random variable x with

$$\text{mean}(x) = \mu \quad \text{and} \quad \text{variance}(x) = \sigma^2$$

then the mean and variance of cx where c is a constant is

$$\text{mean}(cx) = c\mu \quad \text{and} \quad \text{variance}(cx) = c^2\sigma^2 .$$

This is a rule.

If x has a binomial distribution then

$$\text{mean}(cx) = c\underbrace{np} \quad \text{and} \quad \text{variance}(cx) = c^2\underbrace{np(1-p)} .$$

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Background

With $p'=\dfrac{x}{n}$, the constant is $c=\dfrac{1}{n}$, and

$$\text{mean}\left(\frac{x}{n}\right)=\left(\frac{1}{n}\right)\text{mean}(x)=\left(\frac{1}{n}\right)np=p$$

so the variance of $p'=\dfrac{x}{n}$ is $\text{variance}\left(\dfrac{x}{n}\right)=\dfrac{\sigma^2}{n^2}=\dfrac{p(1-p)}{n}$

standard error of $p'=\dfrac{x}{n}$ is $\sigma_{p'}=\sqrt{\dfrac{\sigma^2}{n^2}}=\sqrt{\dfrac{p(1-p)}{n}}$.

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

That is where 1. and 2. in the green box below come from

If a random sample of size n is selected from a large population with p= P(success), then the sampling distribution of p' has:

1. A mean $\mu_{p'}$ equal to p
2. A standard error $\sigma_{p'}$ equal to $\sqrt{\dfrac{p(1-p)}{n}}$
3. An approximately normal distribution if n is sufficiently "large."

6: Normal Probability Distributions

6.5 Normal Approximation of the Binomial Distribution

If we flip the coin a large number of times

$$P(x)=\frac{n!}{x!(n-x)!}p^{x}(1-p)^{n-x} \qquad x=0,\ldots.,n$$

x = # of heads when we flip a coin n times

n=14
p=1/2

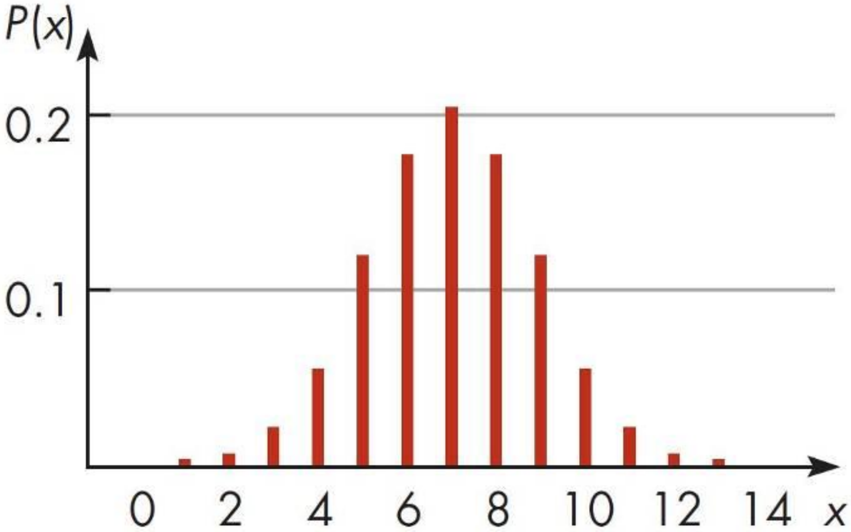


Figure from Johnson & Kuby, 2012.

It gets tedious to find the n=14 probabilities!

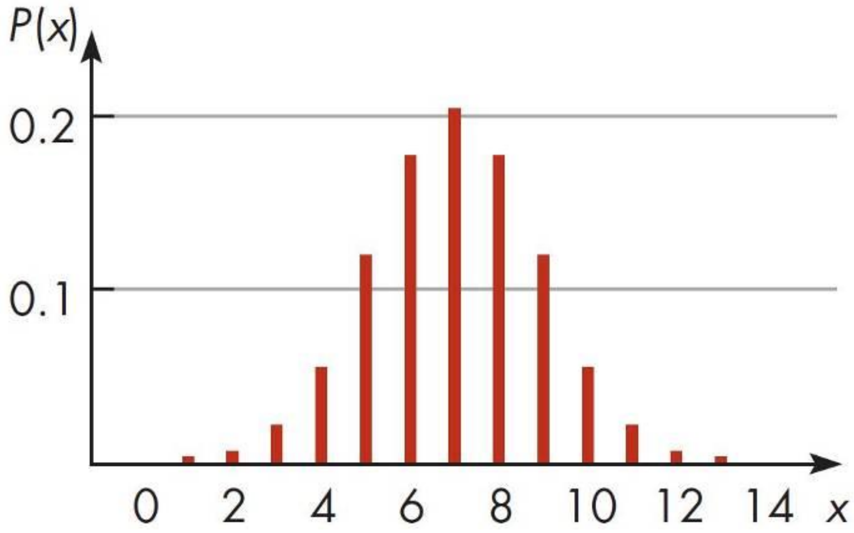
6: Normal Probability Distributions

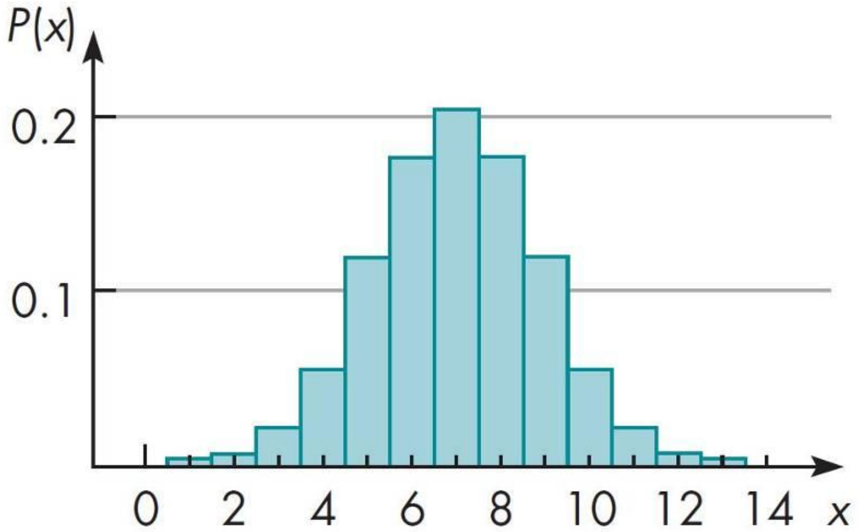
6.5 Normal Approximation of the Binomial Distribution

n=14
p=1/2

It gets tedious to find the n=14 probabilities!

So what we can do is use a histogram representation,





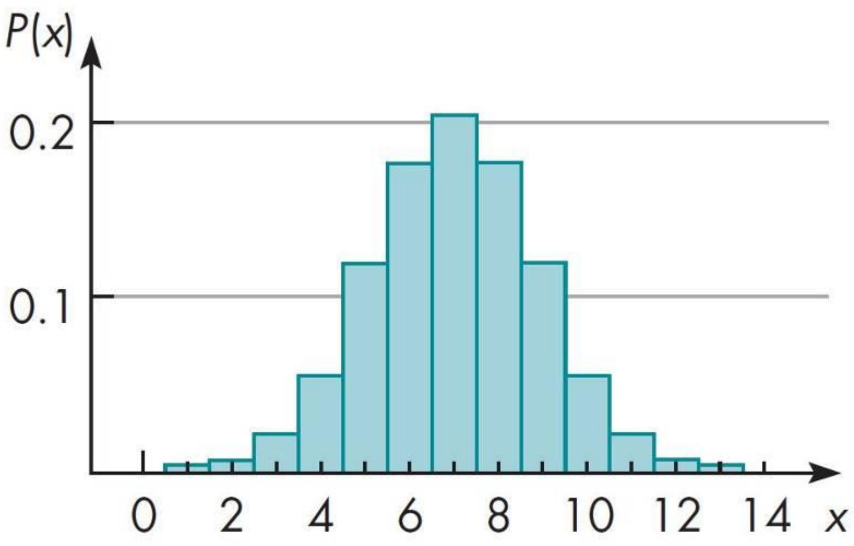
Figures from Johnson & Kuby, 2012.

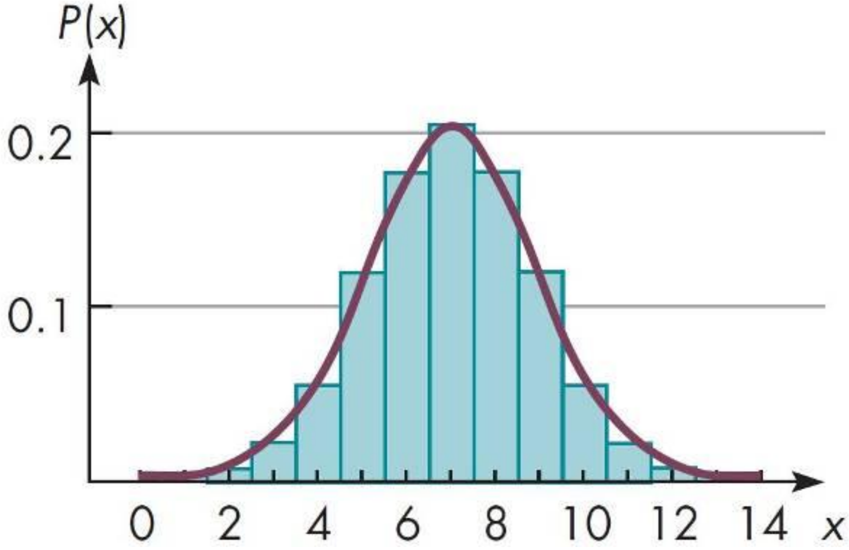
6: Normal Probability Distributions

6.5 Normal Approximation of the Binomial Distribution

So what we can do is use a histogram representation, n=14 p=1/2

Then approximate binomial probabilities with normal areas.





Figures from Johnson & Kuby, 2012.

6: Normal Probability Distributions

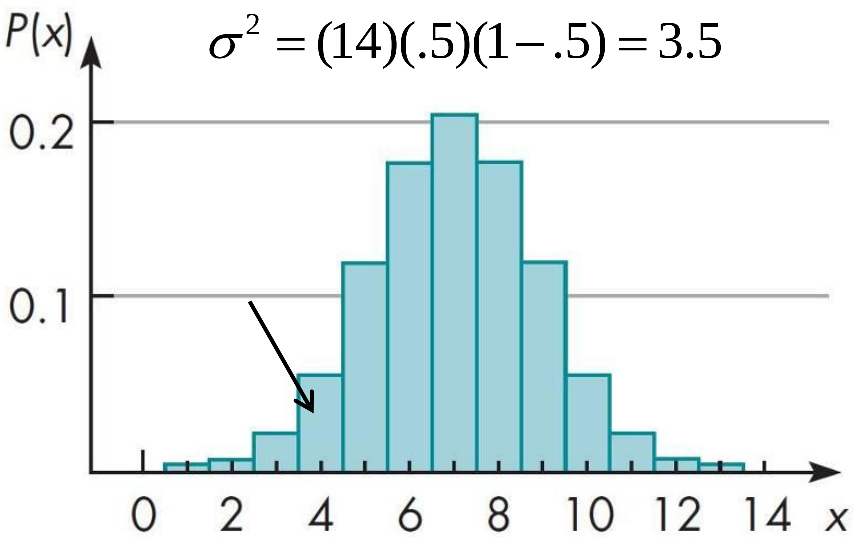
6.5 Normal Approximation of the Binomial Distribution

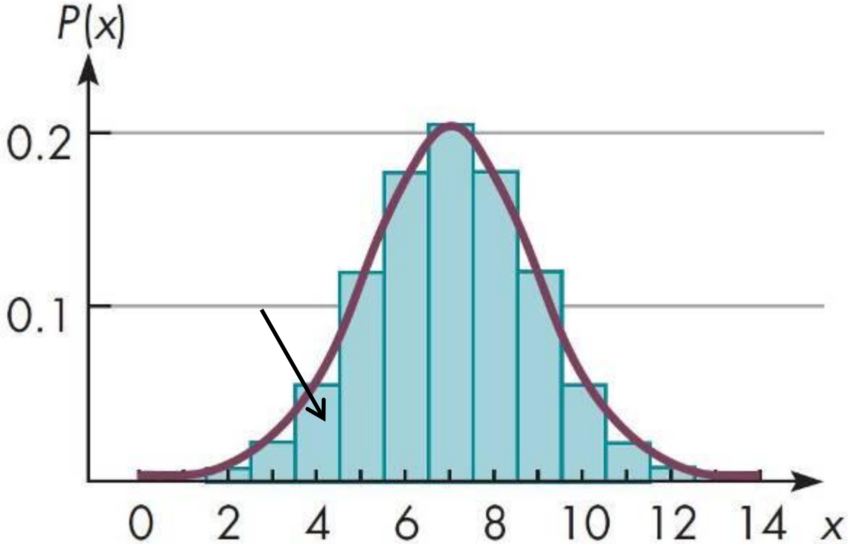
Approximate binomial probabilities with normal areas.
Use a normal with $\mu = np,\ \sigma^2 = np(1-p)$

n=14
p=1/2

$$\mu = (14)(.5) = 7$$

$$\sigma^2 = (14)(.5)(1-.5) = 3.5$$





Figures from Johnson & Kuby, 2012.

6: Normal Probability Distributions

6.5 Normal Approximation of the Binomial Distribution

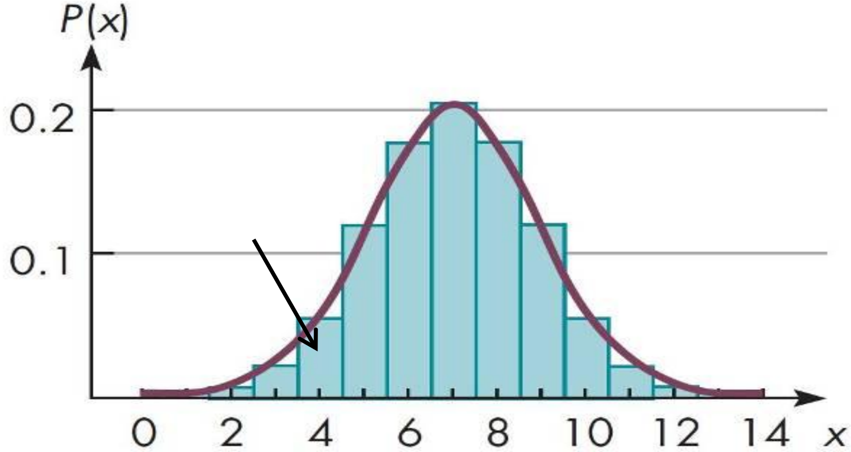
n=14, p=1/2

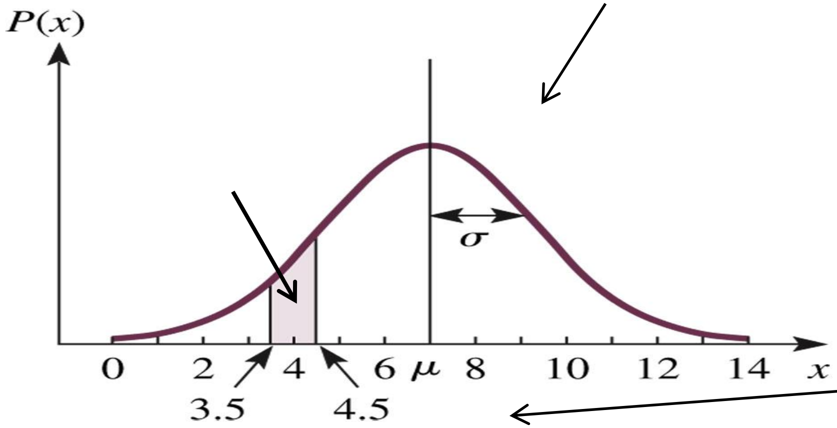
We then approximate binomial probabilities with normal areas.

$P(x=4)$ from the binomial formula

is approximately $P(3.5<x<4.5)$

from the normal with $\mu=7,\ \sigma^2=3.5$





the ±.5 is called a "continuity correction"

Figures from Johnson & Kuby, 2012.

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

In practice, using these guidelines will ensure normality of x:

1. The sample size n is greater than 20.

page 435

2. The product np and $n(1\text{-}p)$ are both greater than 5.

3. The sample consists of less than 10% of the population.

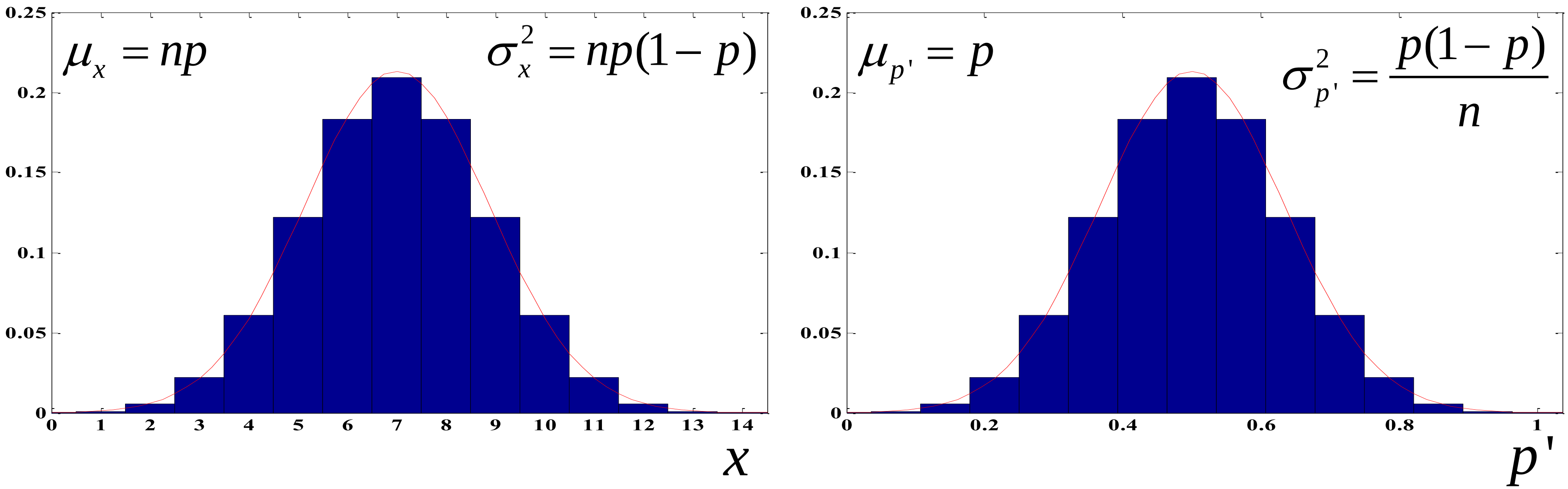
1. $n \geq 20$, 2. $np \geq 5$ and $n(1-p) \geq 5$, 3. $\frac{n}{N} < .10$.

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

But we're not using x, we're scaling it and using $p' = \frac{x}{n}$.

It turns out that $p' = \frac{x}{n}$ also has an approx. normal distribution.



9: Inferences Involving One Population

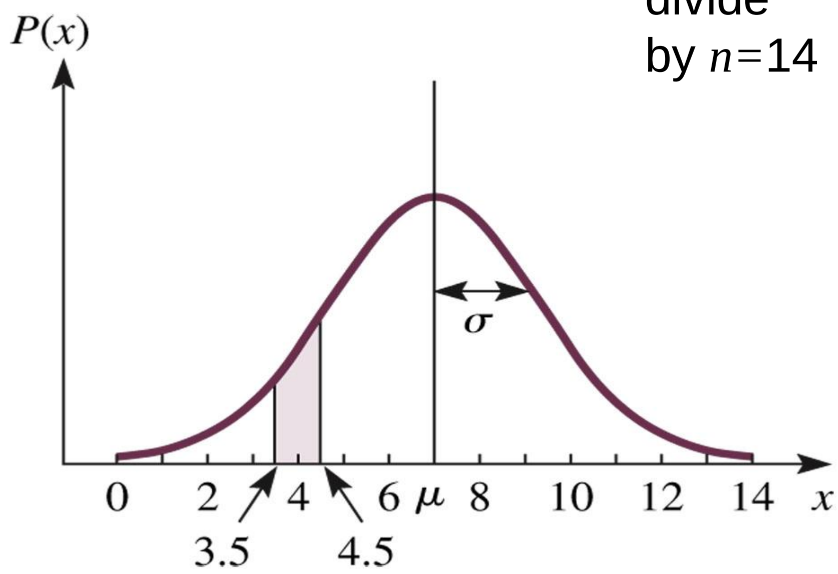
9.2 Inference about the Binomial Probability of Success

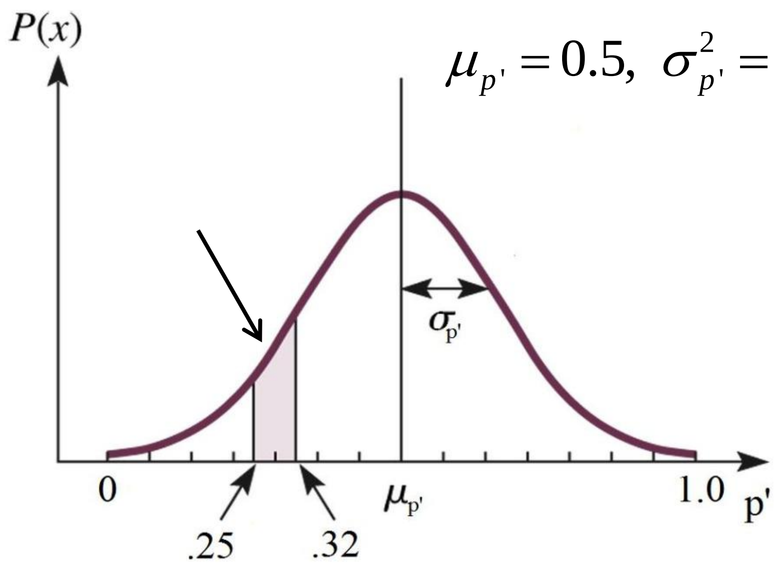
n=14, p=1/2

Now we can determine probabilities with normal areas. $p' = \frac{x}{n}$

$P(3.5 < x < 4.5) = 0.0594$ $\longrightarrow$ divide by n=14 $P(3.5/14 < p' < 4.5/14) = 0.0594$

$\mu_{p'} = 0.5,\ \sigma^2_{p'} = 0.018$





Need to convert to z's.

Figure left from and right modified Johnson & Kuby, 2012.

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

n=14, p=1/2

Now we can determine probabilities with normal areas. $p' = \frac{x}{n}$

For x

$$P(3.5 < x < 4.5)$$

$$P(3.5 - 7 < x - np < 4.5 - 7)$$

$$P\left(\frac{3.5-7}{\sqrt{3.5}} < \underbrace{\frac{x-np}{\sqrt{np(1-p)}}}_{z} < \frac{4.5-7}{\sqrt{3.5}}\right)$$

For p'

$$P(.25 < p' < .32)$$

$$P(.25 - .5 < p' - p < .32 - .5)$$

$$P\left(\frac{.25-.5}{\sqrt{\frac{.5(1-.5)}{14}}} < \underbrace{\frac{p'-p}{\sqrt{\frac{p(1-p)}{n}}}}_{z} < \frac{.32-.5}{\sqrt{\frac{.5(1-.5)}{14}}}\right)$$

Now we can look up areas.

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

For a confidence interval, we would use

Confidence Interval for a Proportion

$$p' - z(\alpha/2)\sqrt{\frac{p'q'}{n}} \quad \text{to} \quad p' + z(\alpha/2)\sqrt{\frac{p'q'}{n}} \qquad (9.6)$$

where $p' = \frac{x}{n}$ and $q' = (1-p')$.

Since we didn't know the true value for p, we estimate it by p'.

This is of the form point estimate $\pm$ some amount .

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Example:

Dana randomly selected n=200 cars and found x=17 convertibles. Find the 90% CI for the proportion of cars that are convertibles.

$$p' = \frac{x}{n} = \frac{17}{200}$$

$$\alpha = 0.1$$

$$z(\alpha/2) = z(0.1/2) = 1.65$$

$$\longrightarrow$$

$$p' \pm z(\alpha/2)\sqrt{\frac{p'q'}{n}}$$

$$\frac{17}{200} \pm 1.65\sqrt{\frac{(17/200)(1-17/200)}{200}}$$

$$0.052 \text{ to } 0.118$$

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Determining the Sample Size

Using the error part of the CI, we determine the sample size n.

Maximum Error of Estimate for a Proportion

$$E = z(\alpha/2)\sqrt{\frac{p'(1-p')}{n}} \qquad (9.7)$$

Sample Size for 1- α Confidence Interval of p

$$n = \frac{[z(\alpha/2)]^2 p^*(1-p^*)}{E^2} \qquad (9.8)$$

From prior data, experience, gut feelings, séance. Or use 1/2.

where p^* and q^* are provisional values used for planning.

over

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

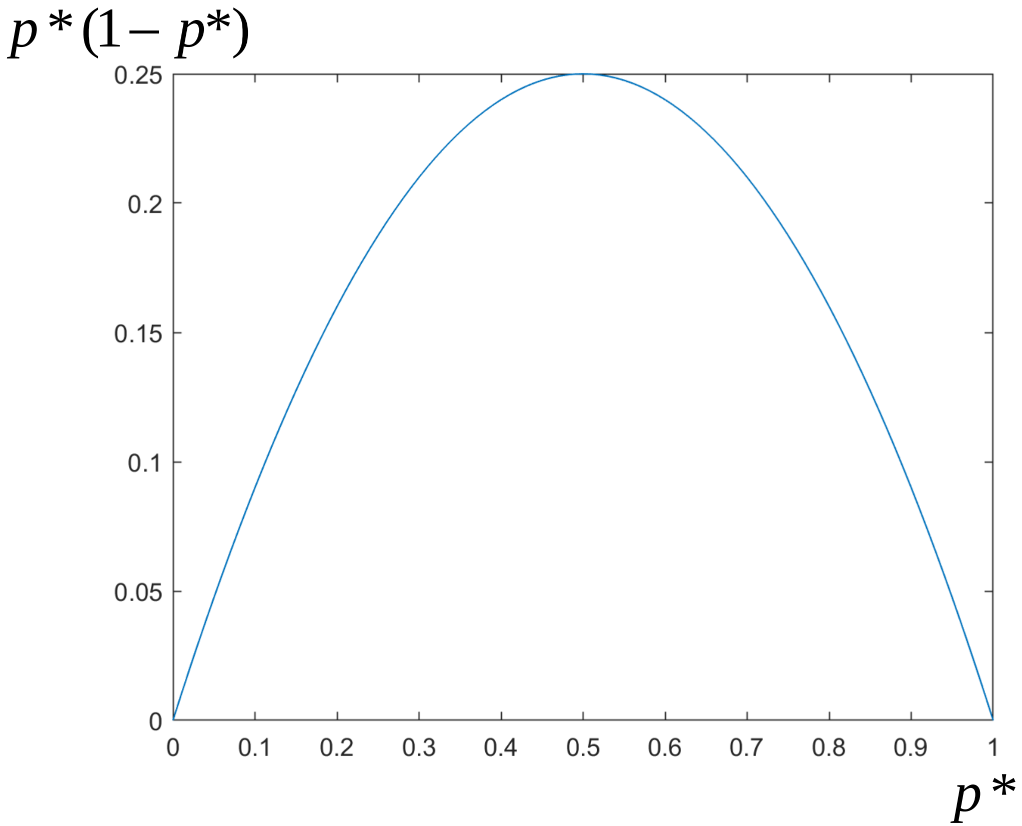
Determining the Sample Size

Why use p*=1/2?

It makes $p^*(1-p^*)$ largest and hence makes

$$n = \frac{[z(\alpha/2)]^2 p^*(1-p^*)}{E^2}$$

the largest.



9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Determining the Sample Size

Example:

A supplier claims bolts are approx. 5% defective. Determine the sample size n if we want our estimate within ±0.02 with 90% confidence.

$$n = \frac{[z(\alpha/2)]^2 p^*(1-p^*)}{E^2} \qquad 1-\alpha = 0.90 \qquad E = 0.02$$

$$z(0.1/2) = 1.65 \qquad p^* = 0.05$$

$$n = \frac{[1.65]^2(0.05)(1-0.05)}{(0.02)^2} = \frac{0.12931875}{0.0004} = 323.4 \quad \rightarrow \quad n = 324$$

Chapter 9: Inferences Involving One Population

Questions?

Homework: Read Chapter 9.2
WebAssign Homework
Chapter 9 # 75, 89