Class 19

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Copyright by D.B. Rowe

Agenda:

Recap Chapter 9.2

Lecture 9.2-9.3

Recap Chapter 9.1-9.2

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

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

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

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

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

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 $\sqrt{\frac{p(1-p)}{n}}$
3. An approximately normal distribution if n is sufficiently "large."

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

$q^* = 1 - p^*$

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

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

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

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

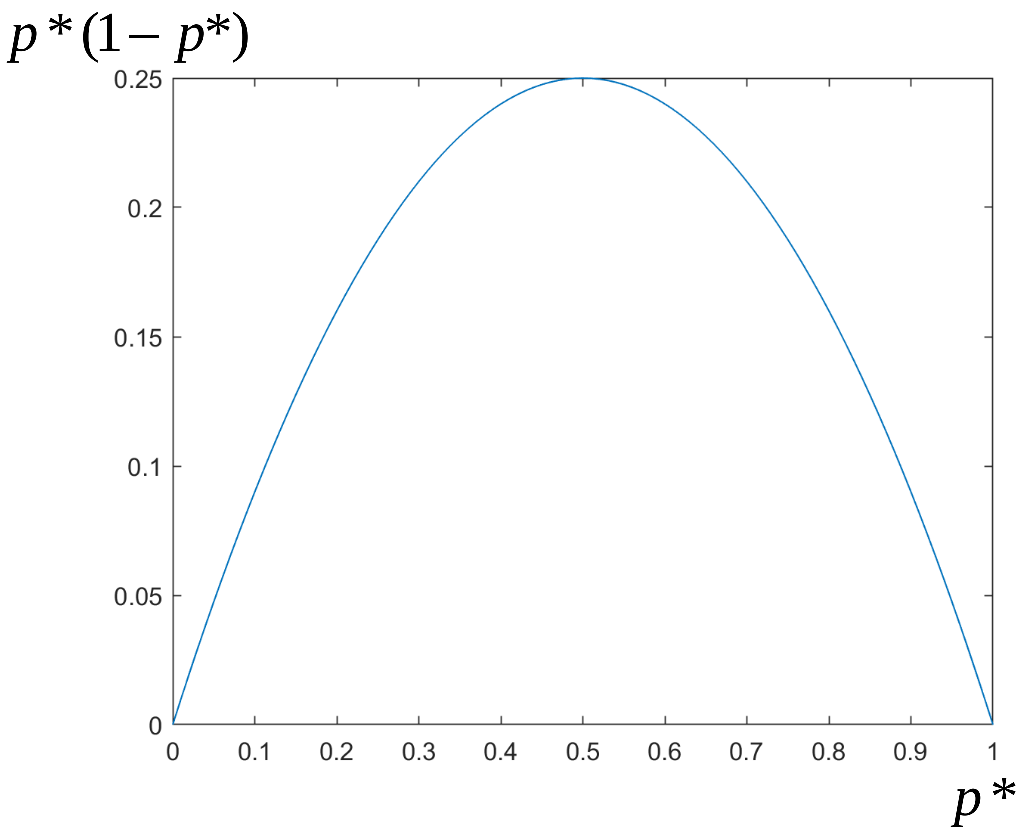
Determining the Sample Size

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, 93, 95, 97

Chapter 9: Inferences Involving One population (continued)

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Hypothesis Testing Procedure

We can perform hypothesis tests on the proportion

H_0: $p \geq p_0$ vs. H_a: $p < p_0$

H_0: $p \leq p_0$ vs. H_a: $p > p_0$

H_0: $p = p_0$ vs. H_a: $p \neq p_0$

Test Statistic for a Proportion p

$$z^* = \frac{p' - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad \text{with} \quad p' = \frac{x}{n} \qquad (9.9)$$

Assume n large for CLT and z.

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

There are three possible hypothesis pairs for the proportion.

H_0: $p \geq p_0$ vs. H_a: $p < p_0$

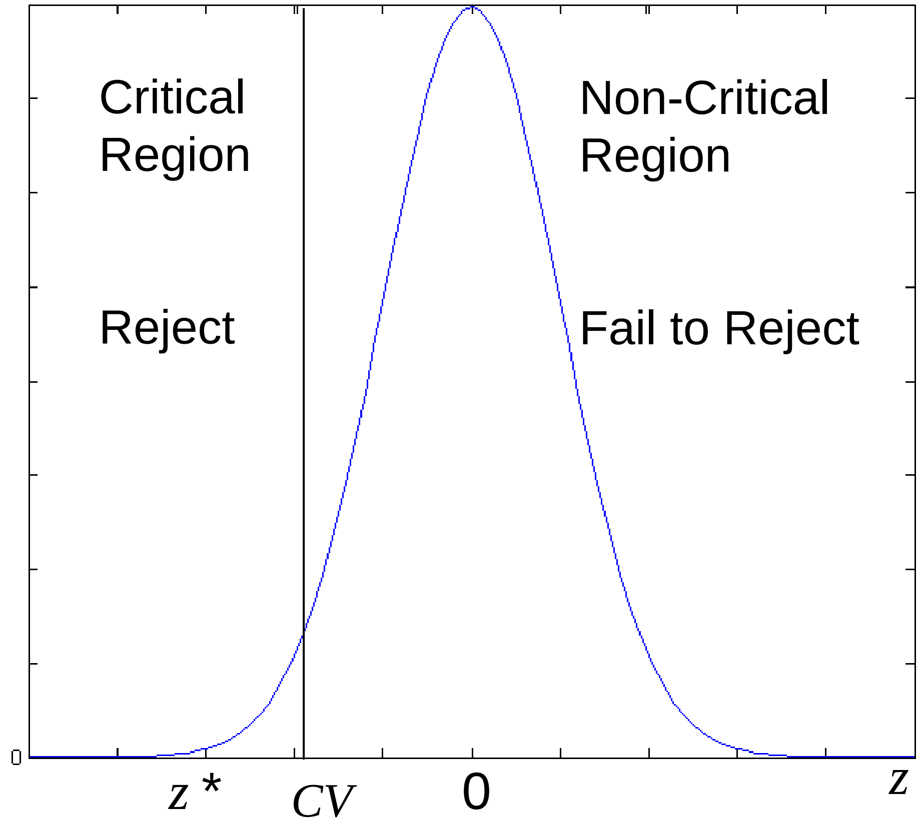
Reject H_0 if

$$z^* = \frac{p' - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

less than

$$-z(\alpha)$$

data indicates $p < p_0$ because p' is "a lot" smaller than p_0



9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

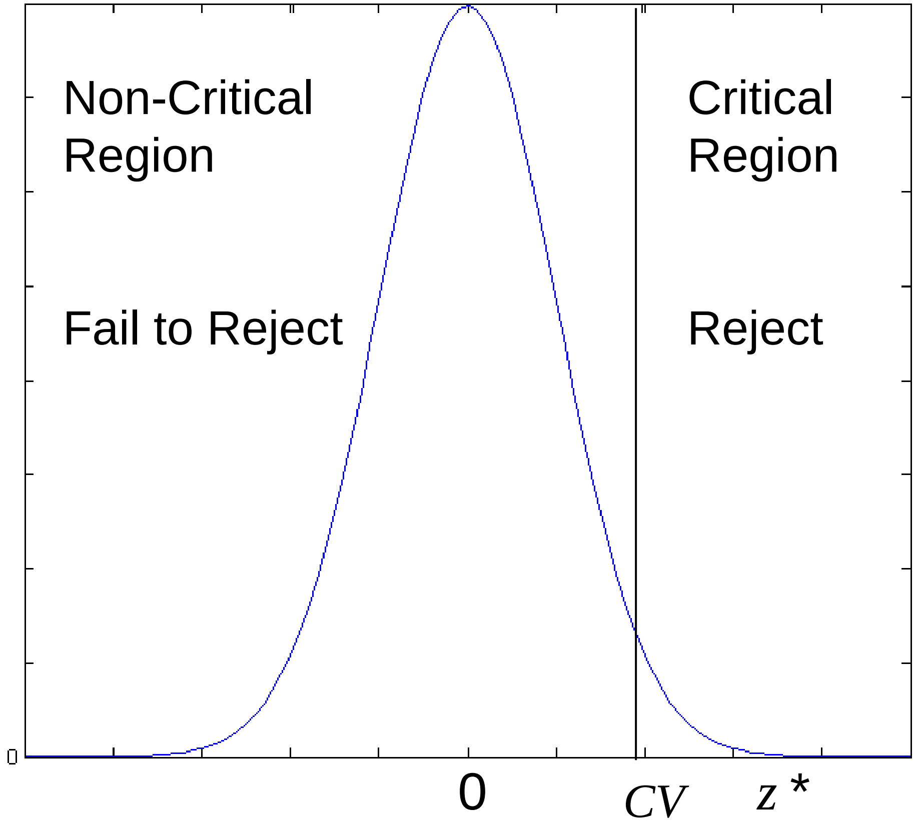
There are three possible hypothesis pairs for the proportion.

H_0: $p \leq p_0$ vs. H_a: $p > p_0$

Reject H_0 if $\qquad$ greater then

$$z^* = \frac{p' - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad z(\alpha)$$

data indicates $p > p_0$ because p' is “a lot” smaller than p_0



9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

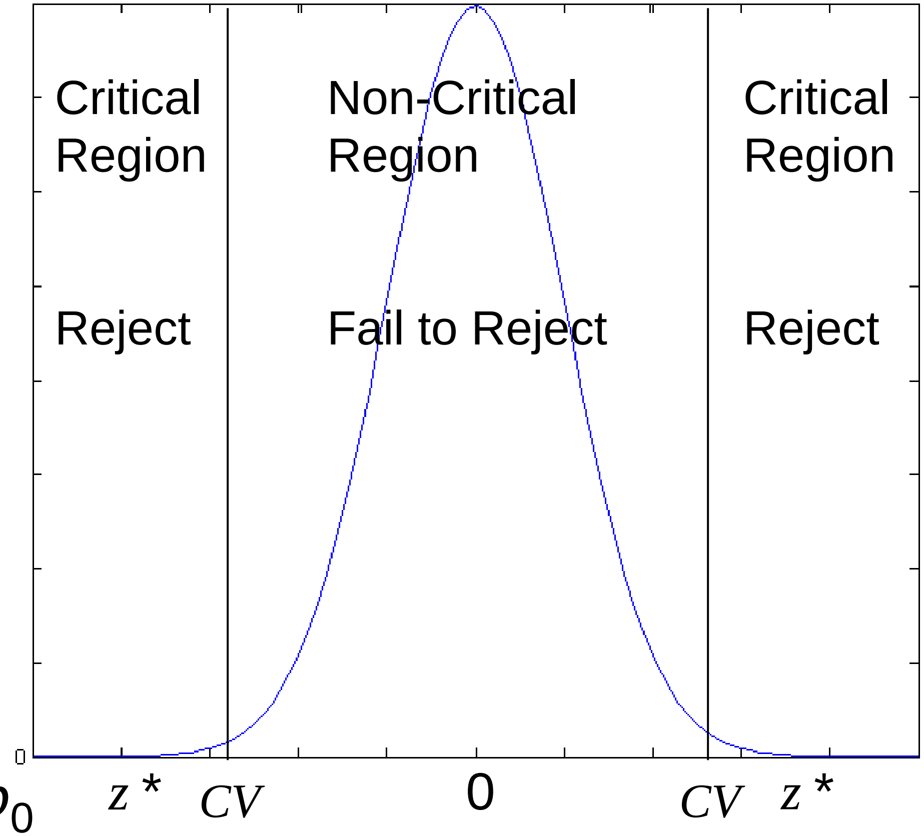
There are three possible hypothesis pairs for the proportion.

H_0: $p = p_0$ vs. H_a: $p \neq p_0$

Reject H_0 if $z^* = \dfrac{p' - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ less than $-z(\alpha/2)$

or if $z^* = \dfrac{p' - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ is greater than $z(\alpha/2)$

data indicates $p \neq p_0$, p' far from p_0



9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Example:

It is reported that 61% get more than 7 hrs of sleep per night on the weekend.

A sample n=350 found that x=235 had more than 7 hours sleep.

With α=.05, does the evidence show that more than 61% sleep More than 7 hrs on the weekend?

$p_0 = .61$

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Fill in the steps.

Step 1:

Step 2:

Step 3:

Step 4:

Step 5:

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Step 1: H_0: p=.61 (≤) vs. H_a: p > .61

Step 2:

Step 3:

Step 4:

Step 5:

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Step 1: H_0: p=.61 (≤) vs. H_a: p > .61

Step 2: $z^* = \dfrac{p' - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ $\qquad p' = \dfrac{x}{n}$

Step 3:

Step 4:

Step 5:

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Step 1: H_0: p=.61 (≤) vs. H_a: p > .61

Step 2: $z^* = \dfrac{p' - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ $\qquad p' = \dfrac{x}{n}$

$$p' = \frac{235}{350} = 0.671$$

Step 3: $z^* = \dfrac{0.671 - 0.61}{\sqrt{\dfrac{0.61(1-0.61)}{350}}} = 2.34$

Step 4:

Step 5:

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

Step 1: H_0: p=.61 (≤) vs. H_a: p > .61

Step 2: $z^* = \dfrac{p' - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ $\quad p' = \dfrac{x}{n}$

$$p' = \frac{235}{350} = 0.671$$

Step 3: $z^* = \dfrac{0.671 - 0.61}{\sqrt{\dfrac{0.61(1-0.61)}{350}}} = 2.34$

Step 4: $z(0.05) = 1.65$

Step 5:

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

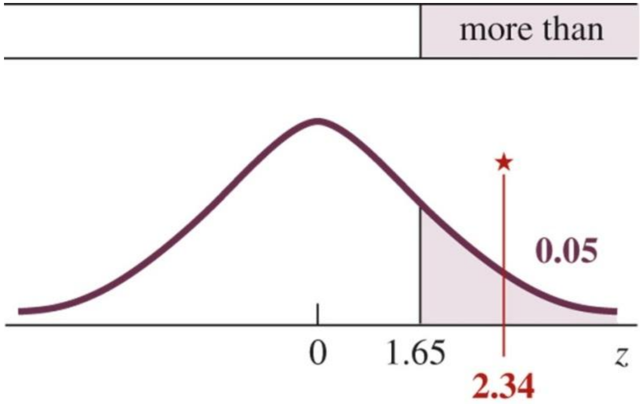
Step 1: H_0: p=.61 (≤) vs. H_a: p > .61

Step 2: $$z^* = \frac{p' - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \qquad p' = \frac{x}{n}$$

$$p' = \frac{235}{350} = 0.671$$

Step 3: $$z^* = \frac{0.671 - 0.61}{\sqrt{\frac{0.61(1-0.61)}{350}}} = 2.34$$

Step 4: $$z(0.05) = 1.65$$



Step 5: Since 2.34>1.65, Reject H_0.

Figure from Johnson & Kuby, 2012.

9: Inferences Involving One Population

9.3 Inference about the Variance and Standard Deviation

We can perform hypothesis tests on the variance.

H_0: $\sigma^2 \geq \sigma_0^2$ vs. H_a: $\sigma^2 < \sigma_0^2$

H_0: $\sigma^2 \leq \sigma_0^2$ vs. H_a: $\sigma^2 > \sigma_0^2$

H_0: $\sigma^2 = \sigma_0^2$ vs. H_a: $\sigma^2 \neq \sigma_0^2$

The assumptions for inferences about the variance σ^2 or standard deviation σ :
The sampled population is normally distributed.

9: Inferences Involving One Population

9.1 Inference about the Mean μ (σ Unknown)

What is the Student t-distribution and how do we get it?

Background Information

If the data comes from normally distributed population, then

$$x \sim N(\mu, \sigma^2)$$

mean variance

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

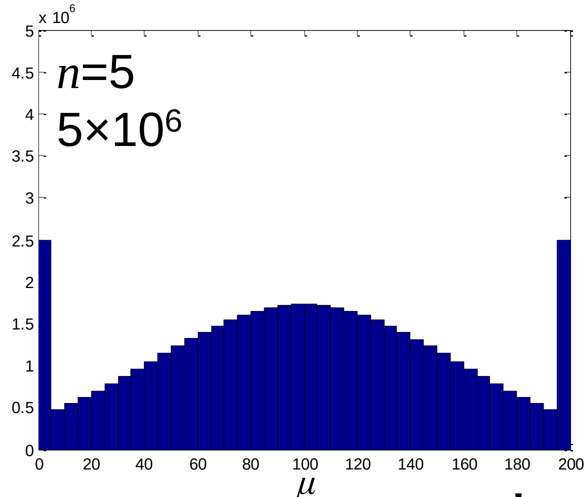
mean variance

generate 5×10^6 random values

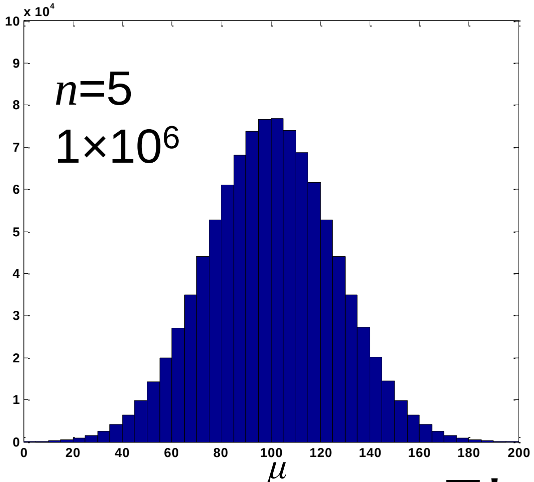
$\mu = 100$

$\sigma = 57.7$

$n = 5$



x's



sample means from each 5 → $\bar{x}$'s

Recall

9: Inferences Involving One Population

9.1 Inference about the Mean μ (σ Unknown)

$\mu = 100$

$\sigma = 57.7$

$n = 5$

It turns out that with the variance σ^2 known, the distribution of

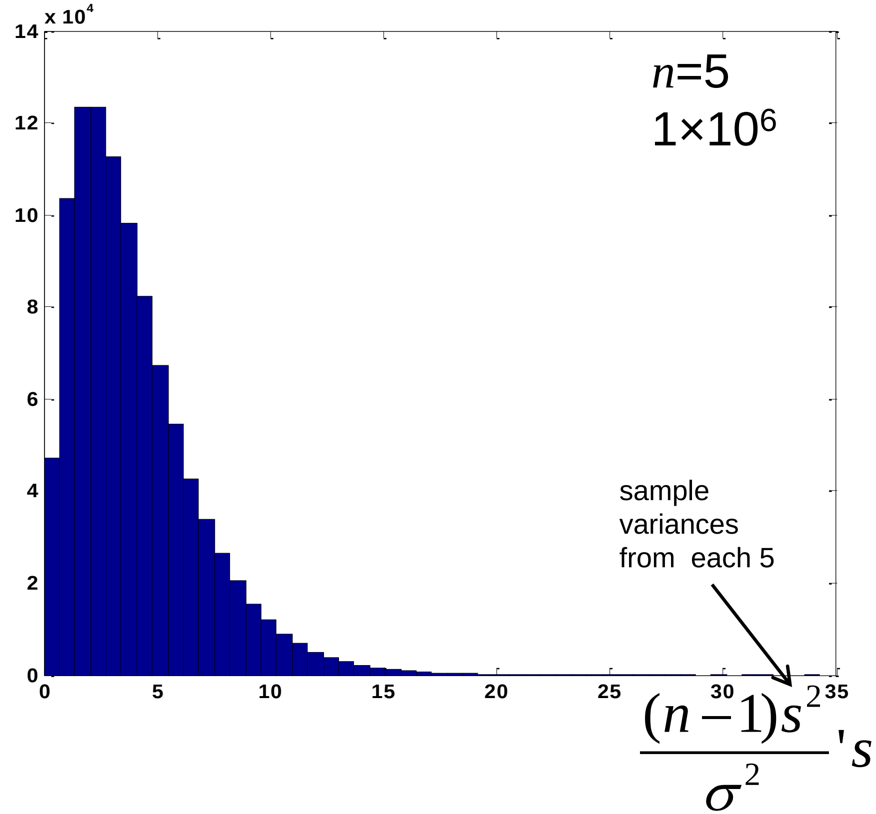
$$\frac{(n-1)s^2}{\sigma^2}$$

sample variance

population variance

has a chi-square distribution with n-1 degrees of freedom.

(χ^2 distribution on Pages 453-454)



Rowe, D.B.

9: Inferences Involving One Population

9.3 Inference about the Variance and Standard Deviation

Properties of the chi-square distribution

1. χ^2 is nonnegative
2. χ^2 is not symmetric, skewed to right mode<median<mean
3. χ^2 is distributed to form a family each determined by $df=\nu=n-1$.

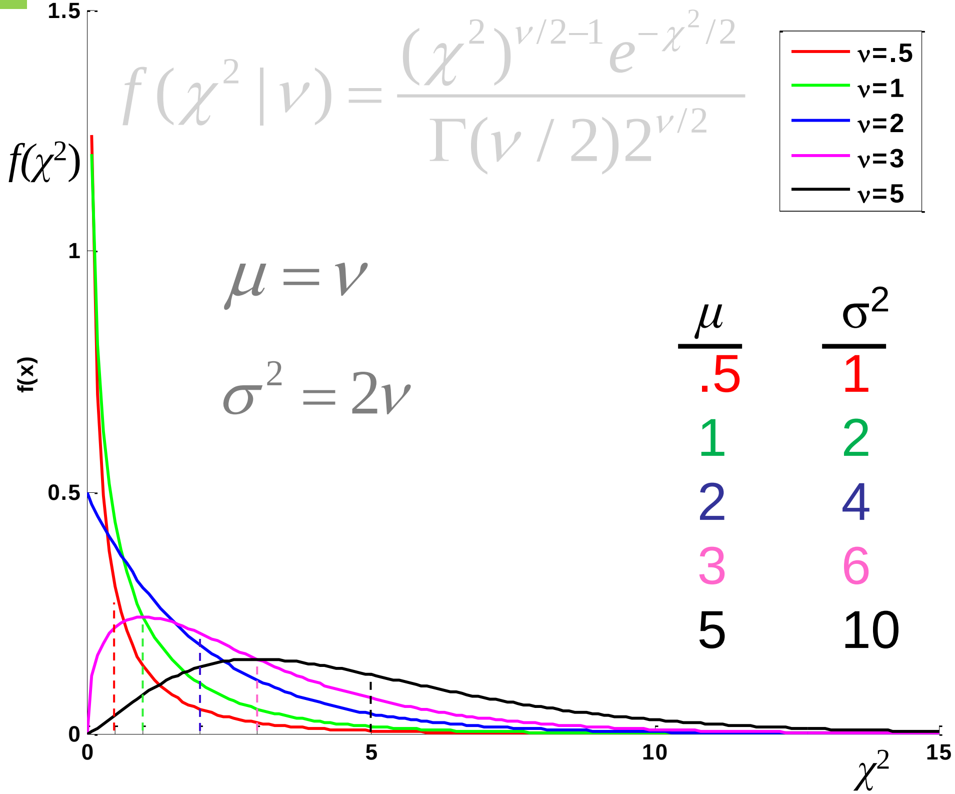
ignore

$$f(\chi^2|\nu)=\frac{(\chi^2)^{\nu/2-1}e^{-\chi^2/2}}{\Gamma(\nu/2)2^{\nu/2}}$$

$$\mu=\nu$$

$$\sigma^2=2\nu$$

μ	σ^2
.5	1
1	2
2	4
3	6
5	10



$$\text{median}\approx\nu-\frac{2}{3}+\frac{4}{27\nu}-\frac{8}{729\nu^2}$$ ignore

$$\text{mode}=\nu-2,\ \nu>2$$

9: Inferences Involving One Population

9.3 Inference about the Variance and Standard Deviation

Test Statistic for Variance (and Standard Deviation)

$$\chi^{2}* = \frac{(n-1)s^{2}}{\sigma_{0}^{2}}, \quad \text{with } df = n-1. \qquad (9.10)$$

sample variance

hypothesized population variance

Will also need critical values.

$$P\left(\chi^{2} > \chi^{2}(df,\alpha)\right) = \alpha$$

Table 8
Appendix B
Page 721

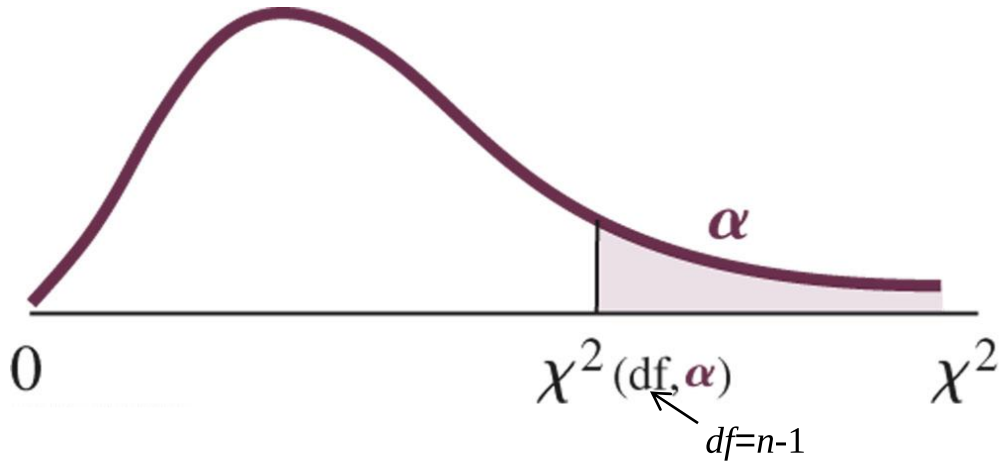
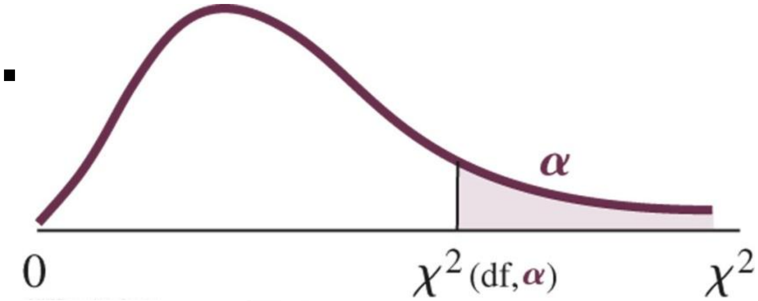


Figure from Johnson & Kuby, 2012.

9: Inferences Involving One Pop.

Example: Find $\chi^2(20,0.05)$.

Table 8, Appendix B, Page 721.



a) Area to the Right

b) Area to the Left (the Cumulative Area)

	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
							Median						
df	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975	0.99	0.995
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.34	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.34	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.34	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.34	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.34	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.34	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.34	23.8	28.4	31.4	34.2	37.6	40.0

Figures from Johnson & Kuby, 2012.

9: Inferences Involving One Population

Example:

A soft-drink bottling company wants to control variability by not allowing the variance to exceed 0.0004.

A sample is taken, n=28, s^2=0.0007 and α=0.05.

Fill in the steps.

Step 1

Step 2

Step 3

Step 4

Step 5

9: Inferences Involving One Population

Example:

A soft-drink bottling company wants to control variability by not allowing the variance to exceed 0.0004.

A sample is taken, n=28, s^2=0.0007 and α=0.05.

Step 1

H_0: $\sigma^2 \leq 0.0004$ vs. H_a: $\sigma^2 > 0.0004$

Step 2

Step 3

Step 4

Step 5

9: Inferences Involving One Population

Example:

A soft-drink bottling company wants to control variability by not allowing the variance to exceed 0.0004.
A sample is taken, n=28, s^2=0.0007 and α=0.05.

Step 1

H_0: $\sigma^2 \leq 0.0004$ vs. H_a: $\sigma^2 > 0.0004$

Step 2

$$\chi^{2}* = \frac{(n-1)s^2}{\sigma_0^2}$$

$df=n$-1

Step 3

Step 4

Step 5

9: Inferences Involving One Population

Example:

A soft-drink bottling company wants to control variability by not allowing the variance to exceed 0.0004.
A sample is taken, n=28, s^2=0.0007 and α=0.05.

Step 1

H_0: $\sigma^2 \leq 0.0004$ vs. H_a: $\sigma^2 > 0.0004$

$\sigma_0^2=0.0004$

Step 2

$$\chi^2{*}=\frac{(n-1)s^2}{\sigma_0^2} \longrightarrow$$

$df=n$-1

Step 3

$$\chi^2{*}=\frac{(28-1)(0.0007)}{0.0004}=47.25$$

Step 4

Step 5

9: Inferences Involving One Population

Example:

A soft-drink bottling company wants to control variability by not allowing the variance to exceed 0.0004.

A sample is taken, n=28, s^2=0.0007 and α=0.05.

Step 1

H_0: $\sigma^2 \leq 0.0004$ vs. H_a: $\sigma^2 > 0.0004$ $\sigma_0^2=0.0004$

Step 2

$$\chi^2{*} = \frac{(n-1)s^2}{\sigma_0^2} \longrightarrow$$

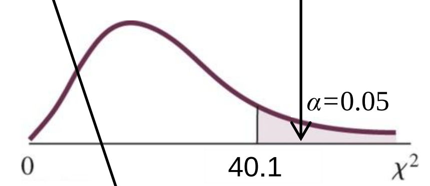
$df=n$-1

Step 3

$$\chi^2{*} = \frac{(28-1)(0.0007)}{0.0004} = 47.25$$

Step 4

$0.005<p$-value<0.01 and $\chi^2(27,.05) = 40.1$



Step 5

	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
27	11.8	12.9	14.6	16.2	18.1	21.7	26.34	31.5	36.7	40.1	43.2	47.0	49.6

9: Inferences Involving One Population

Example:

A soft-drink bottling company wants to control variability by not allowing the variance to exceed 0.0004.
A sample is taken, n=28, s^2=0.0007 and α=0.05.

Step 1

H_0: $\sigma^2 \leq 0.0004$ vs. H_a: $\sigma^2 > 0.0004$ $\qquad \sigma_0^2=0.0004$

Step 2

$$\chi^2{*} = \frac{(n-1)s^2}{\sigma_0^2} \qquad df=n\text{-}1$$

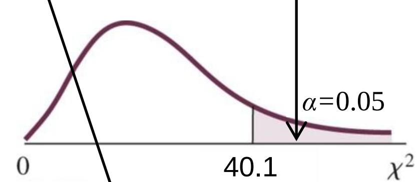
$\longrightarrow$

Step 3

$$\chi^2{*} = \frac{(28-1)(0.0007)}{0.0004} = 47.25$$

Step 4

$0.005 < p$-value< 0.01 and $\chi^2(27,.05) = 40.1$



Step 5

Reject H_0 since p-value<.05 or because 47.25>40.1.

	0.995	0.99	0.975	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005
27	11.8	12.9	14.6	16.2	18.1	21.7	26.34	31.5	36.7	40.1	43.2	47.0	49.6

Chapter 9: Inferences Involving One Population

Questions?

Homework: Read Chapter 9.2-9.3

WebAssign

Chapter 9 # 93, 95, 97, 119, 121, 129, 131, 135