Class 20

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Final Exam

Thursday Dec 12, 8:00am – 10:00 pm Cudahy 001

https://bulletin.marquette.edu/undergrad/academicregulations/#examinations-midtermandfinal

“Final examinations are held in most subjects and must be held on the days/times, as published on the university calendar website. No final exam may be rescheduled for the convenience of the faculty or students.”

“Students who miss a final examination risk receiving a failing grade for the course.”

Agenda:

Recap Chapter 9.2 and 9.3

Lecture Chapter 10.1-10.2

Recap Chapter 9.2 and 9.3

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

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

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

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

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.

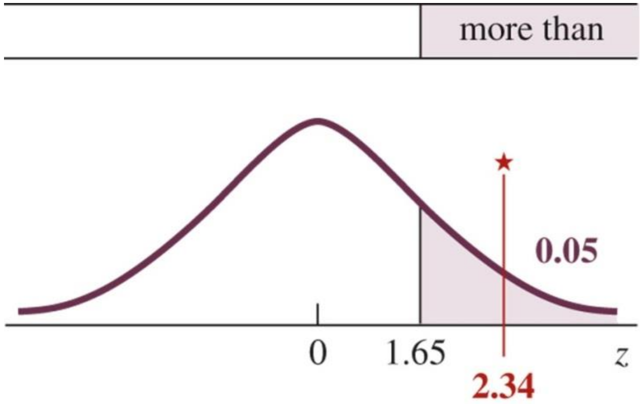


Figure from Johnson & Kuby, 2012.

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: $\sigma^2 \neq \sigma_0^2$

For this hypothesis test, use the $\chi 2$ distribution ⟶

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

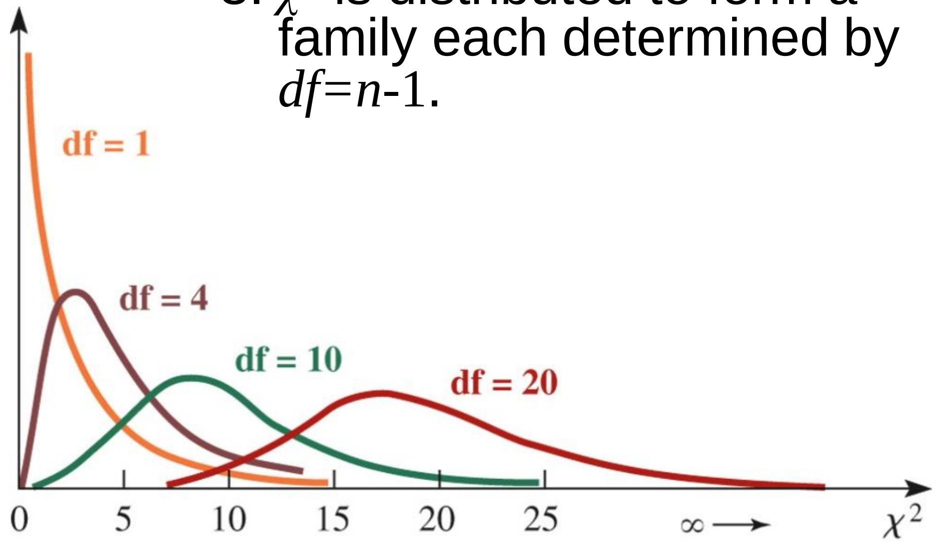


Figure from Johnson & Kuby, 2012.

9: Inferences Involving One Population

9.3 Inference about the Variance and Standard Deviation

Test Statistic for Variance (and Standard Deviation)

sample variance

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

hypothesized population variance

Will also need critical values.

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

Table 8
Appendix B
Page 721

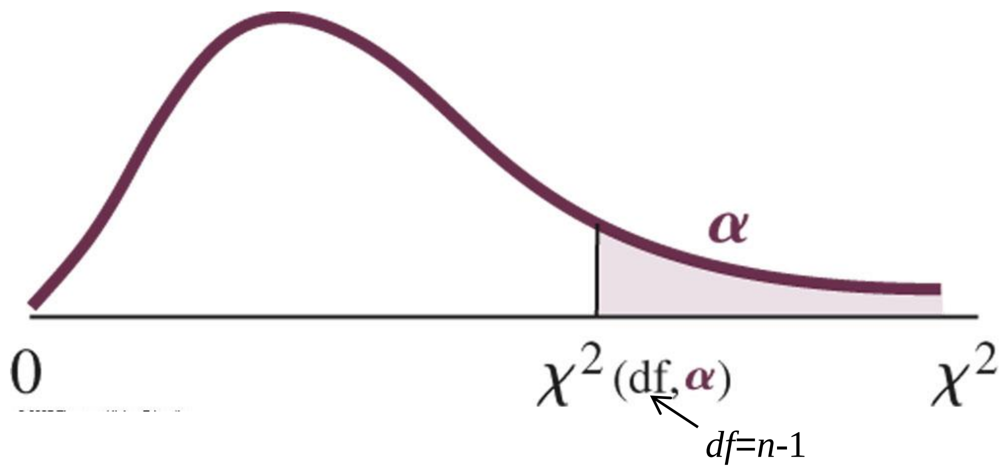
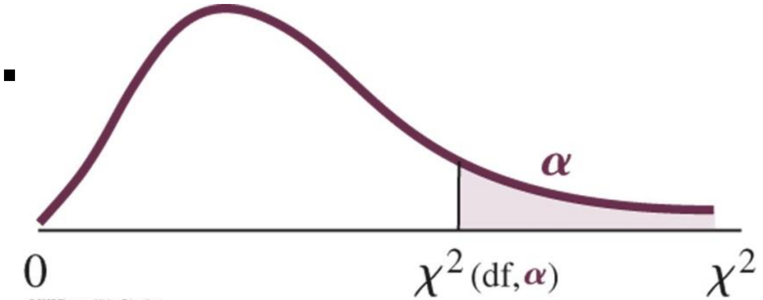


Figure from Johnson & Kuby, 2012.

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)

							Median						
a)	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
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.

Step 1

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

Step 2

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

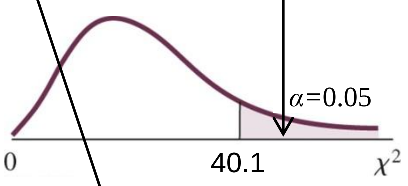
df=n-1

Step 3

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

Step 4

$0.005 < p\text{-value} < 0.01$ 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

Lecture Chapter 10

Chapter 10: Inference Involving Two Populations

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

MARQUETTE UNIVERSITY

Be The Difference.

10: Inferences Involving Two Populations
10.1 Dependent and Independent Samples

In this chapter we will have samples from two populations.

The two populations can either be dependent or independent.

Dependent Samples: If samples have related pairs.
Random sample of married couples.
Male Height vs. Female Height

not same

Independent Samples: If samples are unrelated.
Random sample of males, Random Sample of females.
Male Height vs. Female Height

10: Inferences Involving Two Populations

10.2 Inference Concerning the Mean Difference Using Two Dependent Samples

When we have dependent samples, there is a commonality between the two items in the pair. Quite often before and after.

Population 1: $\mu_1 = \mu_c + \mu_{before}$

common or baseline mean

Population 2: $\mu_2 = \mu_c + \mu_{after}$

common or baseline mean

But we're interested in the difference in means:

$$\begin{aligned} \mu_1 - \mu_2 &= (\mu_c + \mu_{before}) - (\mu_c + \mu_{after}) \\ &= \mu_{before} - \mu_{after} \end{aligned}$$

common mean subtracts out

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

We form a paired difference from the data

Paired Difference

$$d = x_1 - x_2 \tag{10.1}$$

This means that we are subtracting the sample value from

population 2 from the sample value from population 1.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Imagine that we have paired data $(x_{1,1}, x_{2,1}),\ldots,(x_{1,n}, x_{2,n})$
$x_{j,i}$, population j, observation i where j=1,2 i=1,…,n.

We form a paired difference from the data: $d_i = x_{1,i} - x_{2,i}$
$d_1 = x_{1,1} - x_{2,1},\ d_2 = x_{1,2} - x_{2,2},\ldots,\ d_n = x_{1,n} - x_{2,n}$.

When paired observations are randomly selected from normal populations, the paired difference, $d_i = x_{1,i} - x_{2,i}$ will be approximately normally distributed about a mean μ_d with a standard deviation σ_d .

Is actually exactly normally distributed if the populations are (dependent) normally distributed.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

With the d_i's being sampled from populations

with a mean of μ_d and a standard deviation of σ_d, then

$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ is approximately normally distributed (recall CLT)

with a mean $\mu_{\bar{d}} = \mu_d$, and standard deviation $\sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}}$.

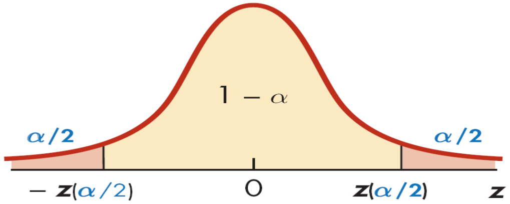
10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

This would allow us to form a z statistic for the mean of differences $\bar{d}$, $z = \dfrac{\bar{d} - \mu_d}{\sigma_d / \sqrt{n}}$ with a standard normal distribution.

assuming known

We can then look up probabilities in the table,



find critical values $z(\alpha/2)$, construct confidence intervals

$$\bar{d} \pm z(\alpha/2)\frac{\sigma_d}{\sqrt{n}}$$

and test hypotheses using $z^* = \dfrac{\bar{d} - \mu_{0d}}{\sigma_d / \sqrt{n}}$.

Figure from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

However, as in Inferences for One Population, we never know the true value of σ_d . So we estimate it with sample standard deviation s_d. This changes $z=\dfrac{\bar{d}-\mu_d}{\sigma_d/\sqrt{n}}$ to $t=\dfrac{\bar{d}-\mu_d}{s_d/\sqrt{n}}$ and the distribution from standard normal to Student t with $df=n-1$ where $s_d^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})^2$.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Confidence Interval Procedure

With σ_d unknown, a 1-α confidence interval for μ_d is:

Confidence Interval for Mean Difference (Dependent Samples)

$$\bar{d} - t(df, \alpha/2)\frac{s_d}{\sqrt{n}} \quad \text{to} \quad \bar{d} + t(df, \alpha/2)\frac{s_d}{\sqrt{n}} \quad \text{where } df = n\text{-}1 \tag{10.2}$$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{10.3}$$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 \tag{10.4}$$

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Example:

Construct a 95% CI for mean difference in Brand B – A tire wear.

d_i's: $8,\ 1,\ 9,\ -1,\ 12,\ 9$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$n = 6$ $\quad df = 5$ $\quad t(df, \alpha/2) =$

$\bar{d} =$ $\quad \alpha = 0.05$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

$s_d =$

$$\bar{d} \pm t(df, \alpha/2)\frac{s_d}{\sqrt{n}} \longrightarrow$$

Figure from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Example:

Construct a 95% CI for mean difference in Brand B – A tire wear.

d_i's: $8,\ 1,\ 9,\ -1,\ 12,\ 9$

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$n = 6$ $df = 5$ $t(df, \alpha/2) = 2.57$

$\bar{d} = 6.3$ $\alpha = 0.05$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2$$

$s_d = 5.1$

$$\bar{d} \pm t(df, \alpha/2)\frac{s_d}{\sqrt{n}} \longrightarrow (0.090, 11.7)$$

Figure from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Hypothesis Testing Procedure

We can test for differences in the population means:

H_0: $\mu_1 \geq \mu_2$ vs. H_a: $\mu_1 < \mu_2$ $\rightarrow$ H_0: $\mu_1 - \mu_2 \geq 0$ vs. H_a: $\mu_1 - \mu_2 < 0$

H_0: $\mu_1 \leq \mu_2$ vs. H_a: $\mu_1 > \mu_2$ $\rightarrow$ H_0: $\mu_1 - \mu_2 \leq 0$ vs. H_a: $\mu_1 - \mu_2 > 0$

H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$ $\rightarrow$ H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 \neq 0$

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Hypothesis Testing Procedure

We can test for differences in the population means:

H_0: $\mu_1 \geq \mu_2$ vs. H_a: $\mu_1 < \mu_2$ $\rightarrow$ H_0: $\mu_1 - \mu_2 \geq 0$ vs. H_a: $\mu_1 - \mu_2 < 0$

H_0: $\mu_1 \leq \mu_2$ vs. H_a: $\mu_1 > \mu_2$ $\rightarrow$ H_0: $\mu_1 - \mu_2 \leq 0$ vs. H_a: $\mu_1 - \mu_2 > 0$

H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$ $\rightarrow$ H_0: $\mu_1 - \mu_2 = 0$ vs. H_a: $\mu_1 - \mu_2 \neq 0$

$\mu_d = \mu_1 - \mu_2$ $\rightarrow$ H_0: $\mu_d \geq 0$ vs. H_a: $\mu_d < 0$

$(\mu_d = \mu_{before} - \mu_{after})$ H_0: $\mu_d \leq 0$ vs. H_a: $\mu_d > 0$

H_0: $\mu_d = 0$ vs. H_a: $\mu_d \neq 0$

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Hypothesis Testing Procedure

With σ_d unknown, the test statistic for μ_d is:

Test Statistic for Mean Difference (Dependent Samples)

$$t^* = \frac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$$

where $df = n - 1$

(10.5)

Go through the same five hypothesis testing steps.

9: Inferences Involving One Population

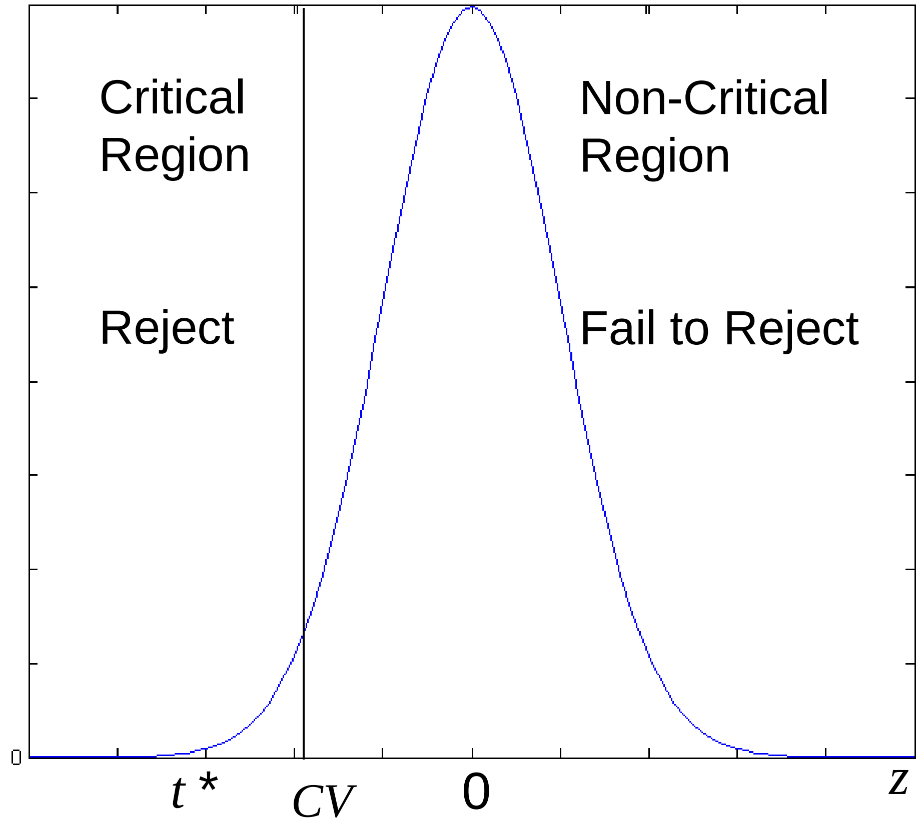
9.2 Inference about the Binomial Probability of Success

There are three possible hypothesis pairs for the proportion.

H_0: $\mu_d \geq \mu_{d0}$ vs. H_a: $\mu_d < \mu_{d0}$

Reject H_0 if $t^* = \dfrac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$ less than $-t(df,\alpha)$

data indicates $\mu_d < \mu_{d0}$ because $\bar{d}$ is "a lot" smaller than μ_{d0}



9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

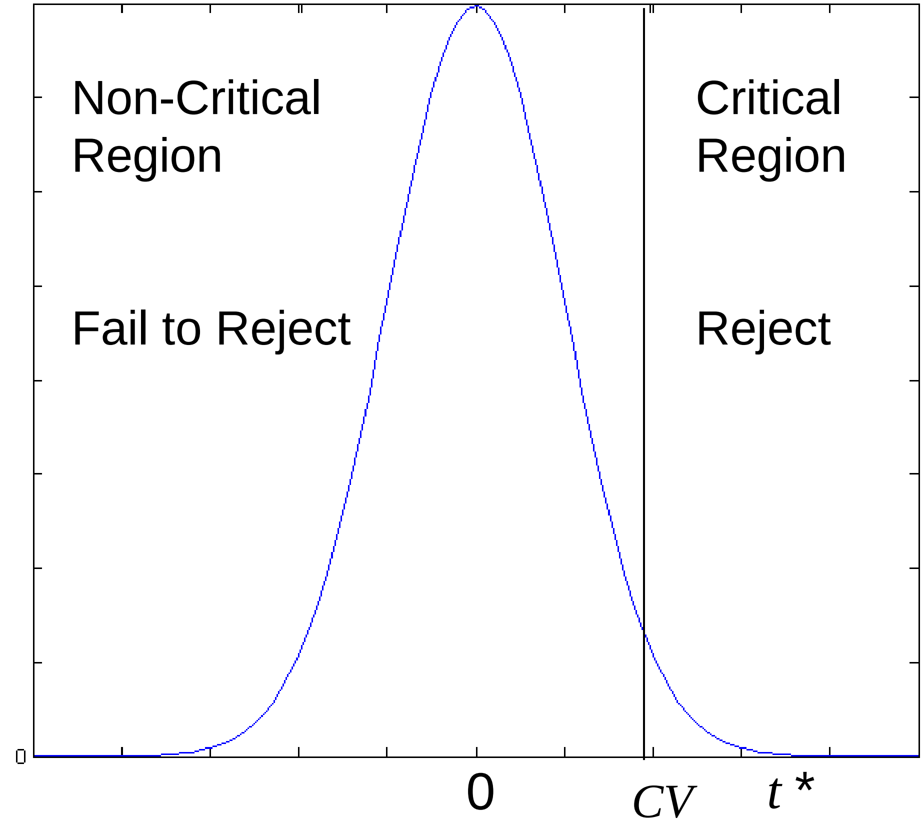
There are three possible hypothesis pairs for the proportion.

H_0: $\mu_d \leq \mu_{d0}$ vs. H_a: $\mu_d > \mu_{d0}$

Reject H_0 if greater then

$$t^* = \frac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}} \qquad t(df, \alpha)$$

data indicates $\mu_d > \mu_{d0}$ because $\bar{d}$ is "a lot" smaller than μ_{d0}



9: Inferences Involving One Population

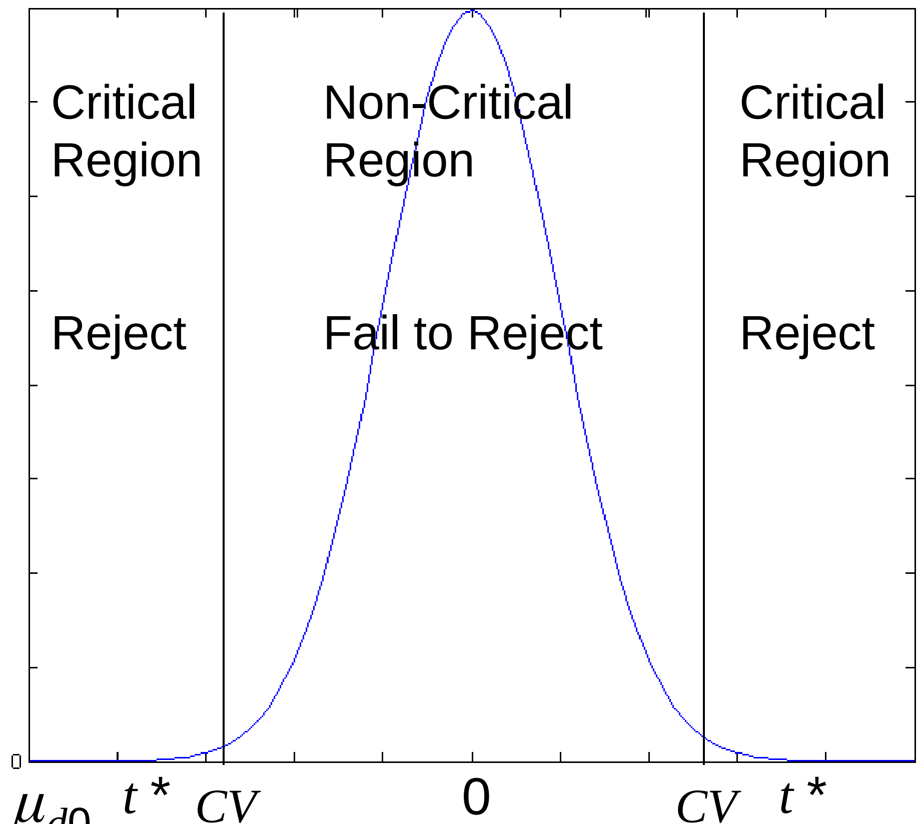
9.2 Inference about the Binomial Probability of Success

There are three possible hypothesis pairs for the proportion.

H_0: $\mu_d = \mu_{d0}$ vs. H_a: $\mu_d \neq \mu_{d0}$

Reject H_0 if $t^* = \dfrac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$ less than $-t(df, \alpha/2)$

or if $t^* = \dfrac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$ is greater than $t(df, \alpha/2)$



data indicates $\mu_d \neq \mu_{d0}, \bar{d}$ far from μ_{d0}

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

$n = 6 \quad 8,\ 1,\ 9,\ -1,\ 12,\ 9$

Example:

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Test mean difference of Brand B minus Brand A is zero.

Step 1

Step 5

Step 2

Step 3

Step 4

Figures from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

$n = 6$ $8,\ 1,\ 9,\ -1,\ 12,\ 9$

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Example:

Test mean difference of Brand B minus Brand A is zero.

Step 1 H_0: $\mu_d=0$ vs. H_a: $\mu_d \neq 0$

Step 5

Step 2

Step 3

Step 4

Figures from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

$n=6$ $8,\ 1,\ 9,\ -1,\ 12,\ 9$

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Example:

Test mean difference of Brand B minus Brand A is zero.

Step 1 H_0: $\mu_d=0$ vs. H_a: $\mu_d \neq 0$

Step 2

$df=5$

$\alpha=.05$

$$t^* = \frac{\bar{d}-\mu_{d0}}{s_d / \sqrt{n}}$$

Step 3

Step 4

Step 5

Figures from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

$n = 6 \quad 8, 1, 9, -1, 12, 9$

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Example:

Test mean difference of Brand B minus Brand A is zero.

Step 1 H_0: $\mu_d=0$ vs. H_a: $\mu_d \neq 0$

Step 2

$df = 5$

$\alpha = .05$

$$t^* = \frac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$$

Step 3 $\bar{d} = 6.3$

$s_d = 5.1$

$$t^* = \frac{6.3 - 0}{5.1 / \sqrt{6}} = 3.03$$

Step 4

Step 5

Figures from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

$n = 6 \quad 8,\ 1,\ 9,\ -1,\ 12,\ 9$

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Example:

Test mean difference of Brand B minus Brand A is zero.

Step 1 $H_0\colon \mu_d = 0$ vs. $H_a\colon \mu_d \neq 0$

Step 2

$df = 5 \qquad t^* = \dfrac{\bar{d} - \mu_{d0}}{s_d / \sqrt{n}}$

$\alpha = .05$

Step 3 $\bar{d} = 6.3 \qquad t^* = \dfrac{6.3 - 0}{5.1 / \sqrt{6}} = 3.03$

$s_d = 5.1$

Step 4 $t(df, \alpha / 2) = 2.57$

Step 5

Figures from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

$n=6 \quad 8,\ 1,\ 9,\ -1,\ 12,\ 9$

Car	1	2	3	4	5	6
Brand A	125	64	94	38	90	106
Brand B	133	65	103	37	102	115

Example:

Test mean difference of Brand B minus Brand A is zero.

Step 1 H_0: $\mu_d=0$ vs. H_a: $\mu_d \neq 0$

Step 2

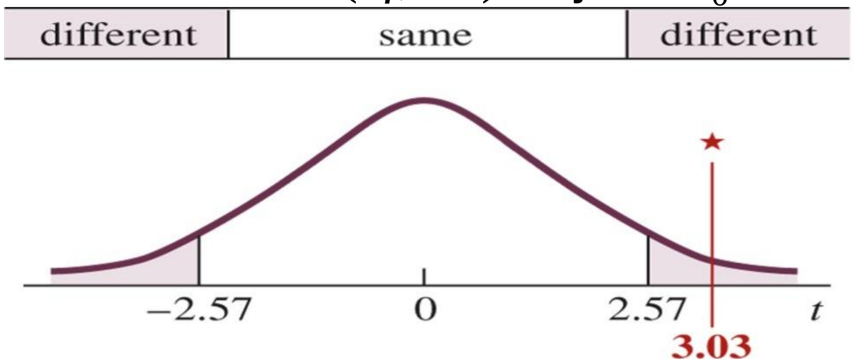
$$df=5 \qquad t^*=\frac{\bar{d}-\mu_{d0}}{s_d/\sqrt{n}}$$
$$\alpha=.05$$

Step 3 $\bar{d}=6.3$, $s_d=5.1$

$$t^*=\frac{6.3-0}{5.1/\sqrt{6}}=3.03$$

Step 4 $t(df,\alpha/2)=2.57$

Step 5 Since $t^*>t(df,\alpha/2)$, reject H_0



Conclusion: Significant difference in tread wear at .05 level.

Figures from Johnson & Kuby, 2012.

Chapter 10: Inferences Involving Two Populations

Questions?

Homework: Read Chapter 10.1-10.2
WebAssign
Chapter 10 #13, 15, 23, 25, 29, 31, 35