Class 27

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

MARQUETTE UNIVERSITY

Be The Difference.

Recap Chapter 9

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

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

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.

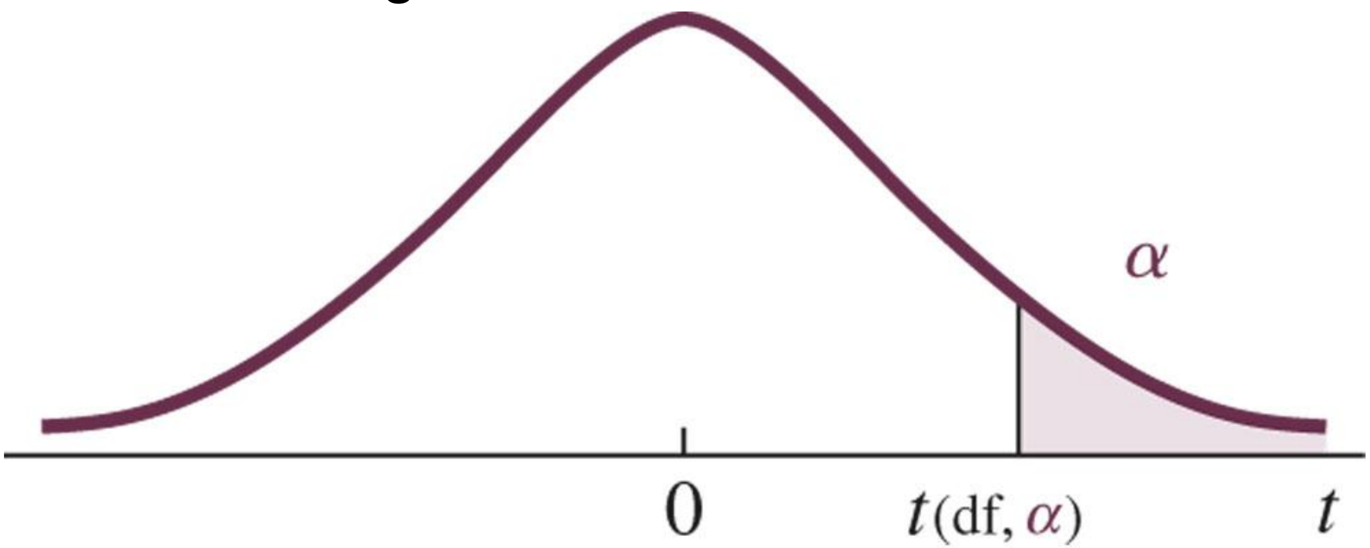
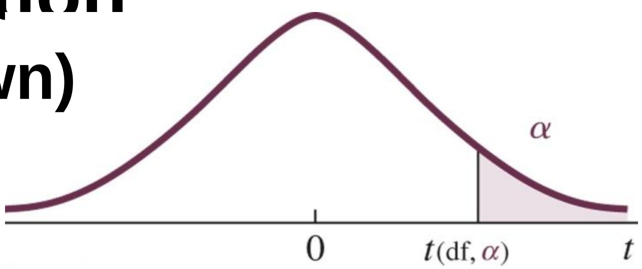


Figure from Johnson & Kuby, 2012.

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)

Recap 9.1:

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

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

σ assumed known

$$\bar{x} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

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

$\longrightarrow$

σ assumed unknown

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

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

9: Inferences Involving One Population

9.2 Inference about the Binomial Probability of Success

We talked about a Binomial experiment with two outcomes.

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \qquad \begin{array}{l} n = 1, 2, 3, \ldots \\ x = 0, 1 \ldots, n \end{array} \qquad 0 \le p \le 1$$

n = # of trials, x = # of successes, p = prob. of success

Sample Binomial Probability

i.e. number of H out of n flips

$$p' = \frac{x}{n} \tag{9.3}$$

where x is the number of successes in n trials.

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 $\sqrt{\dfrac{p(1-p)}{n}}$
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

Determining the Sample Size

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

Maximum Error of Estimate for a Proportion

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

Sample Size for 1- α Confidence Interval of p

$q^* = 1 - p^*$

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

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

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*

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

Assume n large for CLT and z.

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 χ^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.

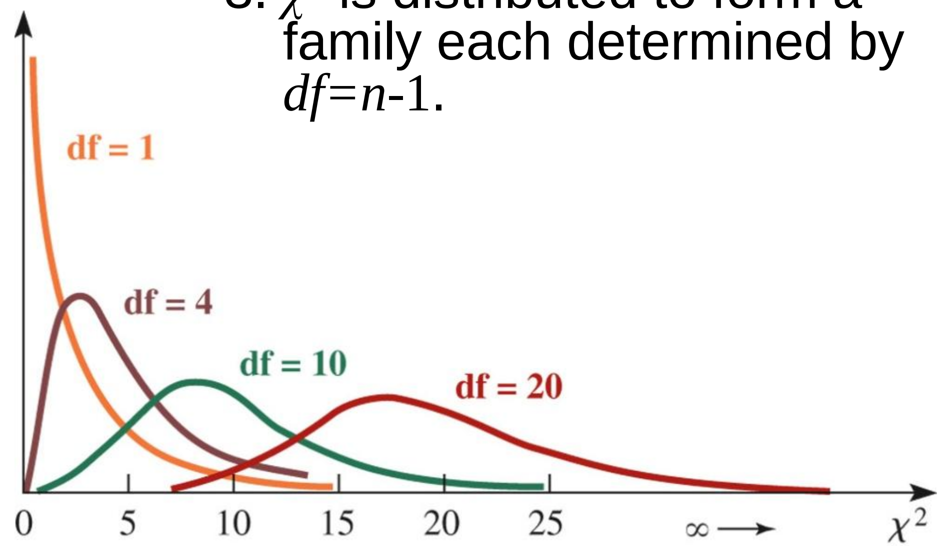


Figure from Johnson & Kuby, 2012.

9: Inferences Involving One Population

9.3 Inference about the Variance and Standard Deviation

Test Statistic for Variance (and Standard Deviation)

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

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

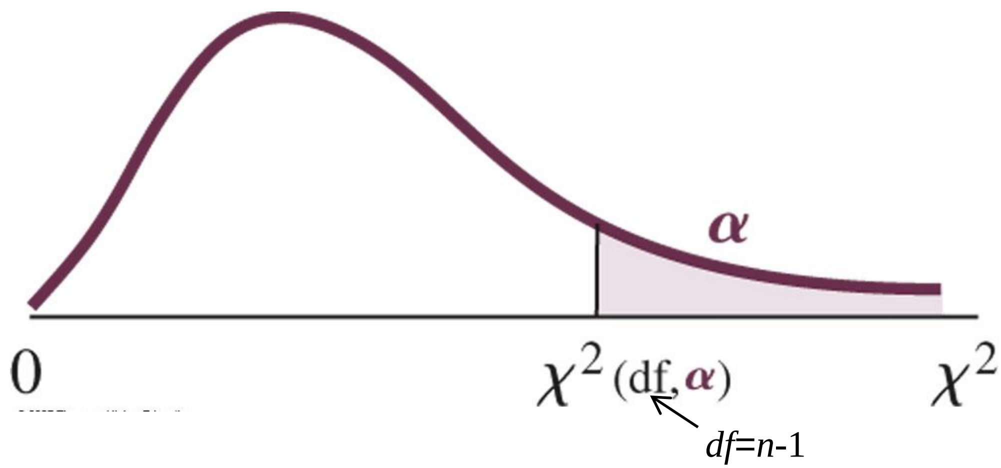
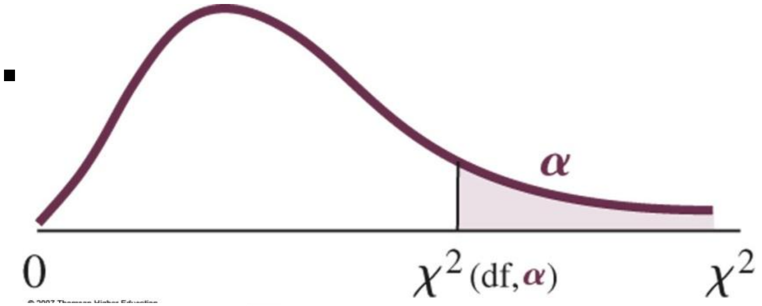


Figure from Johnson & Kuby, 2012.

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)

						Median							
a) Area to the Right	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 \ b) Area to the Left	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.

Recap Chapter 10

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

Confidence Interval Procedure

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

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \qquad s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2 \qquad \mu_{\bar{d}} = \mu_d \quad \sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}}$$

With σ_d unknown, a 1-α confidence interval for μ_d=(μ_1-μ_2) is:

Confidence Interval for Mean Difference (Dependent Samples)

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

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: $8,\ 1,\ 9,\ -1,\ 12,\ 9$

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

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

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

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

$s_d = 5.1$

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

Figure from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.2 Inference for Mean Difference Two Dependent Samples

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

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

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

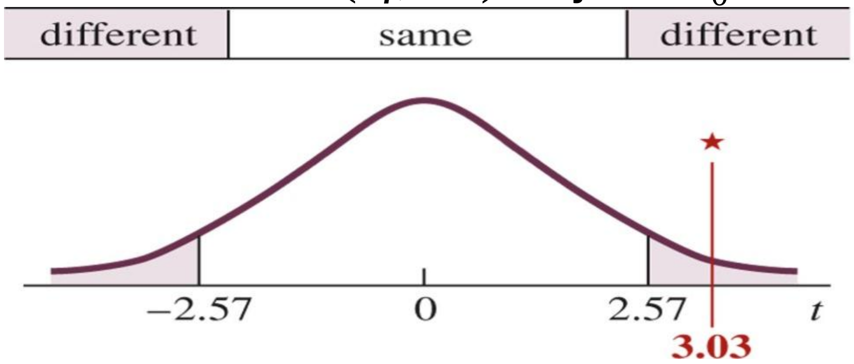
$$\alpha=.05$$

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

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

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.

10: Inferences Involving Two Populations

10.3 Inference for Mean Difference Two Independent Samples

Confidence Interval Procedure

With σ_1 and σ_2 unknown, a $1\text{-}\alpha$ confidence interval for $\mu_1-\mu_2$ is:

Confidence Interval for Mean Difference (Independent Samples)

$$(\bar{x}_1-\bar{x}_2)-t(df,\alpha/2)\sqrt{\left(\frac{s_1^2}{n_1}\right)+\left(\frac{s_2^2}{n_2}\right)} \text{ to } (\bar{x}_1-\bar{x}_2)+t(df,\alpha/2)\sqrt{\left(\frac{s_1^2}{n_1}\right)+\left(\frac{s_2^2}{n_2}\right)}$$

where df is either calculated or smaller of df_1, or df_2 (10.8)

Actually, this is for $\sigma_1 \neq \sigma_2$.

If using a computer program.

If not using a computer program.

10: Inferences Involving Two Populations

10.3 Inference Mean Difference Confidence Interval

Sample	Number	Mean	Standard Deviation
Female (f)	$n_f = 20$	$\bar{x}_f = 63.8$	$s_f = 2.18$
Male (m)	$n_m = 30$	$\bar{x}_m = 69.8$	$s_m = 1.92$

Example:

Interested in difference in mean heights between men and women. The heights of 20 females and 30 males is measured. Construct a 95% confidence interval for $\mu_m - \mu_f$, σ_m & σ_f unknown

$$(\bar{x}_m - \bar{x}_f) \pm t(df, \alpha/2)\sqrt{\left(\frac{s_m^2}{n_m}\right) + \left(\frac{s_f^2}{n_f}\right)}$$

$$\alpha = 0.05$$

$$t(19, .025) = 2.09$$

$$(69.8 - 63.8) \pm 2.09\sqrt{\left(\frac{(1.92)^2}{30}\right) + \left(\frac{(2.18)^2}{20}\right)}$$

therefore 4.75 to 7.25

Figure from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.3 Inference for Mean Difference Two Independent Samples

Hypothesis Testing Procedure

27 values

Step 1

H_0: $\mu_f = \mu_m$ vs. H_a: $\mu_f \neq \mu_m$

Step 2

$$t^* = \frac{(\bar{x}_m - \bar{x}_f) - (\mu_m - \mu_f)}{\sqrt{\left(\frac{s_m^2}{n_m}\right) + \left(\frac{s_f^2}{n_f}\right)}}$$

$df = 5$

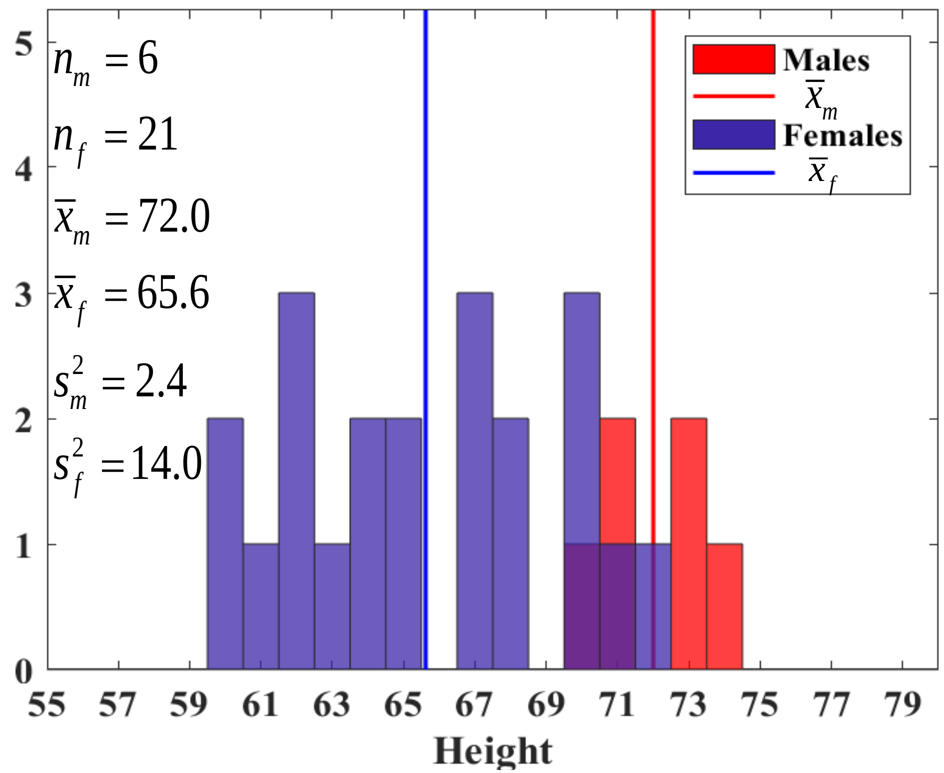
$\alpha = .05$

Step 3

$$t^* = \frac{(72.0 - 65.6) - (0)}{\sqrt{\left(\frac{2.4}{6}\right) + \left(\frac{14.0}{21}\right)}} = 6.17$$

Step 4

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



Step 5 Reject H_0 $6.17 > 2.57$, height males≠height females

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

If independent samples of size n_1 and n_2 are drawn … with p_1=P_1(success) and p_2=P_2(success) , then the sampling distribution of $p_1' - p_2'$ has these properties:

1. mean $\mu_{p_1'-p_2'} = p_1 - p_2$

2. standard error $\sigma_{p_1'-p_2'} = \sqrt{\dfrac{p_1 q_1}{n_1} + \dfrac{p_2 q_2}{n_2}}$ (10.10)

3. approximately normal dist if n_1 and n_2 are sufficiently large.
ie **I** n_1,n_2>20 **II** $n_1 p_1$, $n_1 q_1$, $n_2 p_2$, $n_2 q_2$>5 **III** sample<10% of pop

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Confidence Interval Procedure

Assumptions for … difference between two proportions p_1-p_2: The n_1 … and n_2 random observations … are selected independently from two populations that are not changing

Confidence Interval for the Difference between Two Proportions $p_1 - p_2$

$$(p_1' - p_2') - z(\alpha/2)\sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}} \quad \text{to} \quad (p_1' - p_2') + z(\alpha/2)\sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}}$$

where $p_1' = \frac{x_1}{n_1}$ and $p_2' = \frac{x_2}{n_2}$. (10.11)

$q_1' = 1 - p_1' \quad q_2' = 1 - p_2'$

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Confidence Interval Procedure

Example: Another Semester

Construct a 99% CI for proportion of female A's minus male A's difference $p_f - p_m$.

120 values

$$n_m = 52$$

$$n_f = 68$$

$$x_m = 21$$

$$x_f = 43$$

$$z(\alpha / 2) = 2.58$$

$$p'_f = \frac{x_f}{n_f} = \frac{43}{68} = .62$$

$$p'_m = \frac{x_m}{n_m} = \frac{21}{52} = .40$$

$$(p'_f - p'_m) \pm z(\alpha / 2)\sqrt{\frac{p'_f q'_f}{n_f} + \frac{p'_m q'_m}{n_m}}$$

$$(.62 - .40) \pm 2.58\sqrt{\frac{(.62)(.38)}{68} + \frac{(.40)(.60)}{52}}$$

$$-.003 \text{ to } .460$$

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

We can perform hypothesis tests on the proportion

H_0: $p_1 \geq p_2$ vs. H_a: $p_1 < p_2$

H_0: $p_1 \leq p_2$ vs. H_a: $p_1 > p_2$

H_0: $p_1 = p_2$ vs. H_a: $p_1 \neq p_2$

$$\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

when $p_1 = p_2 = p$.

Test Statistic for the Difference between two Proportions- Population Proportions **Known**

$$z^* = \frac{(p_1' - p_2') - \overbrace{(p_{10} - p_{20})}^{0}}{\sqrt{\underbrace{pq}_{p \text{ known}}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \qquad p_1' = \frac{x_1}{n_1} \quad p_2' = \frac{x_2}{n_2} \qquad (10.12)$$

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

Test Statistic for the Difference between two Proportions-Population Proportions **UnKnown**

0

$$z^* = \frac{(p_1' - p_2') - (p_{10} - p_{20})}{\sqrt{p_p' q_p' \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \qquad (10.15)$$

p_p estimated

where we assume p_1=p_2 and use pooled estimate of proportion

$$p_1' = \frac{x_1}{n_1} \quad p_2' = \frac{x_2}{n_2} \qquad \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} = pq\left[\frac{1}{n_1} + \frac{1}{n_2}\right] \qquad p_p' = \frac{x_1 + x_2}{n_1 + n_2} \qquad q_p' = 1 - p_p'$$

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Hypothesis Testing Procedure

Product	Number Defective	Number Checked
Salesperson's	$x_s = 15$	$n_s = 150$
Competitor's	$x_c = 6$	$n_c = 150$

Step 1

H_0: p_s-p_c≤0 vs. H_a: p_s-p_c>0

Step 2

$$z^* = \frac{(p'_s - p'_c) - (p_{0s} - p_{0c})}{\sqrt{p'_p q'_p \left[\frac{1}{n_s} + \frac{1}{n_c}\right]}}$$

$$\alpha = .05$$

$$p'_s = \frac{x_s}{n_s} = \frac{15}{150} \qquad p'_c = \frac{x_c}{n_c} = \frac{6}{150}$$

$$p'_p = \frac{x_s + x_c}{n_s + n_c} = \frac{15+6}{150+150}$$

Step 3

$$z^* = \frac{(.10 - .04) - (0)}{\sqrt{(.07)(.93)\left[\frac{1}{150} + \frac{1}{150}\right]}} = 2.04$$

Step 4

$$z(\alpha) = 1.65$$

Step 5 Reject H_0 $< .05$

$.02 < p - value < .023$ or $2.04 > 1.65$

Figure from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.5 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

We can perform hypothesis tests on two variances

$H_0: \sigma_1^2 \geq \sigma_2^2$ vs. $H_a: \sigma_1^2 < \sigma_2^2$

$H_0: \sigma_1^2 \leq \sigma_2^2$ vs. $H_a: \sigma_1^2 > \sigma_2^2$

$H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_a: \sigma_1^2 \neq \sigma_2^2$

Assumptions: Independent samples from normal distribution

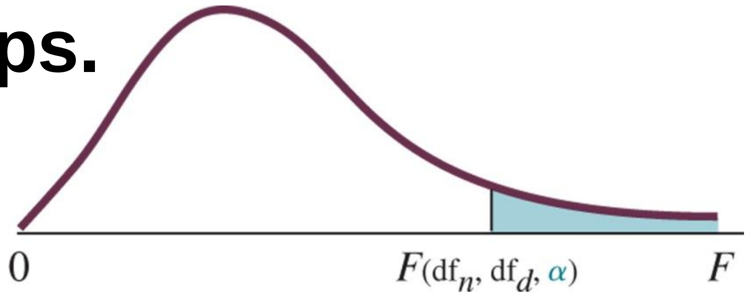
Test Statistic for Equality of Variances

$$F* = \frac{s_n^2}{s_d^2} \quad \text{with } df_n = n_n - 1 \text{ and } df_d = n_d - 1. \tag{10.16}$$

Use new table to find areas for new statistic.

10: Inferences Involving Two Pops.

10.5 Inference Ratio of Two Variances



Example: Find $F(5,8,0.05)$.

$df_n = n_n - 1$ $df_d = n_d - 1$

Table 9, Appendix B, Page 722.

$\alpha = 0.05$

Degrees of Freedom for Numerator df_n

Degrees of Freedom for Denominator df_d	1	2	3	4	5	6	7	8	9	10
1	161.	200.	216.	225.	230.	234.	237.	239.	241.	242.
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98

Figures from Johnson & Kuby, 2012.

10: Inferences Involving Two Populations

10.5 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

★

***One tailed tests*: Arrange H_0 & H_a so H_a is always "greater than"**

H_0: $\sigma_1^2 \geq \sigma_2^2$ vs. H_a: $\sigma_1^2 < \sigma_2^2$ $\longrightarrow$ H_0: $\sigma_2^2 / \sigma_1^2 \leq 1$ vs. H_a: $\sigma_2^2 / \sigma_1^2 > 1$ $\quad F^* = s_2^2 \big/ s_1^2$

H_0: $\sigma_1^2 \leq \sigma_2^2$ vs. H_a: $\sigma_1^2 > \sigma_2^2$ $\quad$ H_0: $\sigma_1^2 / \sigma_2^2 \leq 1$ vs. H_a: $\sigma_1^2 / \sigma_2^2 > 1$ $\quad F^* = s_1^2 \big/ s_2^2$

Reject H_0 if $F^* = s_n^2 / s_d^2 > F(df_n, df_d, \alpha)$.

***Two tailed tests*: put larger sample variance s^2 in numerator**

H_0: $\sigma_1^2 = \sigma_2^2$ vs. H_a: $\sigma_1^2 \neq \sigma_2^2$ $\longrightarrow$ H_0: $\sigma_n^2 / \sigma_d^2 = 1$ vs. H_a: $\sigma_n^2 / \sigma_d^2 \neq 1$

$\sigma_n^2 = \sigma_1^2$ if $s_1^2 > s_2^2$, $\sigma_n^2 = \sigma_2^2$ if $s_2^2 > s_1^2$

Reject H_0 if $F^* = s_n^2 / s_d^2 > F(df_n, df_d, \alpha/2)$.

10: Inferences Involving Two Populations

10.5 Inference for Ratio of Two Variances Two Ind. Samples

Is variance of female heights greater than that of males? $\alpha = .01$ 27 values

Step 1

$H_0: \sigma_f^2 \le \sigma_m^2$ vs. $H_a: \sigma_f^2 > \sigma_m^2$

$H_0: \sigma_f^2 / \sigma_m^2 \le 1$ vs. $H_a: \sigma_f^2 / \sigma_m^2 > 1$

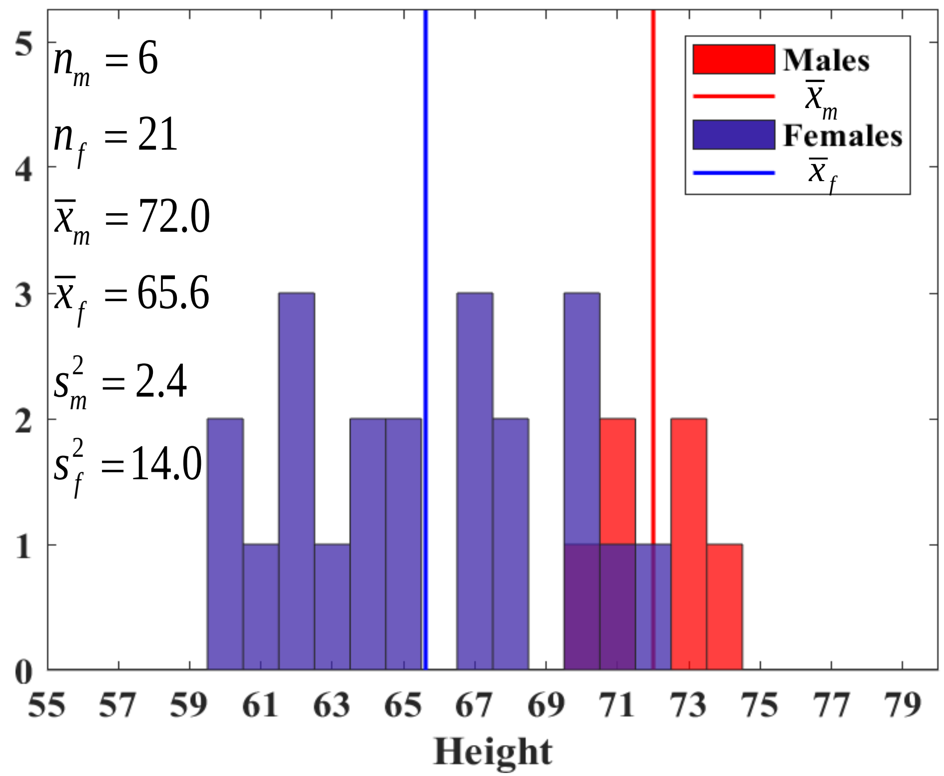
Step 2 $F^* = s_f^2 \big/ s_m^2$ $\quad df_m = 5$, $df_f = 20$, $\alpha = .01$

Step 3

$F^* = 14.0 / 2.4 = 5.83$

Step 4

$F(20,5,.01) = 9.55$

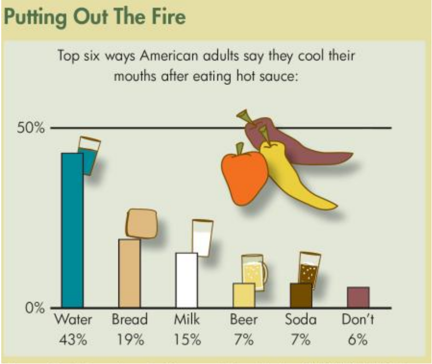


Step 5 Do not reject H_0 since $5.83 < 9.55$ and conclude σ_f^2 not $> \sigma_m^2$.

Recap Chapter 11

11: Applications of Chi-Square

11.1 Chi-Square Statistic Cooling a Great Hot Taste



Quite often we have qualitative data in categories.

Example: Cooling mouth after hot spicy food.

Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Number	73	29	35	19	20	13	11

11: Applications of Chi-Square

11.1 Chi-Square Statistic Data Setup

Example: Cooling mouth after hot spicy food.

Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Number	73	29	35	19	20	13	11

Data set up: k cells $C_1,\ldots,C_k$ that n observations sorted into

Observed frequencies in each cell $O_1,\ldots,O_k$. $O_1+\ldots+O_k=n$

Expected frequencies in each cell $E_1,\ldots,E_k$. $E_1+\ldots+E_k=n$

Cell	C_1	C_2	.	.	.	C_k
Observed	O_1	O_2	.	.	.	O_k
Expected	E_1	E_2	.	.	.	E_k

11: Applications of Chi-Square

11.1 Chi-Square Statistic Data Setup

Example: Cooling mouth after hot spicy food.

Method	Water	Bread	Milk	Beer	Soda	Nothing	Other
Number	73	29	35	19	20	13	11

Data set up: k cells $C_1,\ldots,C_k$ that n observations sorted into

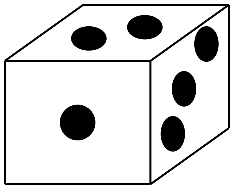
Observed frequencies in each cell $O_1,\ldots,O_k$. $O_1+\ldots+O_k=n$

Expected frequencies in each cell $E_1,\ldots,E_k$. $E_1+\ldots+E_k=n$

Cell	C_1	C_2	.	.	.	C_k
Observed	O_1	O_2	.	.	.	O_k
Expected	E_1	E_2	.	.	.	E_k

11: Applications of Chi-Square

11.2 Inferences Concerning Multinomial Experiments



Example: We work for Las Vegas Gaming Commission. We have received many complaints that dice at particular casino are loaded (weighted, not fair). We confiscate all dice and test one. We roll it n=60 times. We get following data.

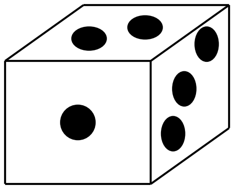
Cell, i	1	2	3	4	5	6
Observed, O_i	7	12	10	12	8	11
Expected, E_i	10	10	10	10	10	10

Expected Value for Multinomial Experiment:

$$E_i = np_i \tag{11.3}$$

11: Applications of Chi-Square

11.2 Inferences Concerning Multinomial Experiments



Example: We work for Las Vegas Gaming Commission. We have received many complaints that dice at particular casino are loaded (weighted, not fair). We confiscate all dice and test one. We roll it n=60 times. We get following data.

Cell, i	1	2	3	4	5	6
Observed, O_i	7	12	10	12	8	11
Expected, E_i	10	10	10	10	10	10

Expected Value for Multinomial Experiment:

$$E_i = np_i \qquad (11.3)$$

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

Is "Preference for math-science, social science, or humanities" … "independent of the gender of a college student?"

Sample Results for Gender and Subject Preference

	Favorite Subject Area			
Gender	Math–Science (MS)	Social Science (SS)	Humanities (H)	Total
Male (M)	37	41	44	122
Female (F)	35	72	71	178
Total	72	113	115	300

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

Is "Preference for math-science, social science, or humanities" … "independent of the gender of a college student?"

There is a Hypothesis test (of independence) to determine this.

Is Favorite Subject independent of Gender.

Similar to Example with the Die, now have rows i and columns j.

$$\chi^2 * = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Observed values, O_{ij}'s.

Gender	Favorite Subject Area: Math–Science (MS)	Social Science (SS)	Humanities (H)	Total
Male (M)	37	41	44	122
Female (F)	35	72	71	178
Total	72	113	115	300

Figure from Johnson & Kuby, 2012.

What are E_{ij}'s?

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

$$\chi^{2*} = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

D of F for Contingency Tables:

$$df = (r-1)(c-1) \qquad (11.4)$$

$r>1, c>1$

Expected Frequencies for Contingency Tables

$$E_{ij} = \frac{row\ total \times column\ total}{grand\ total} = \frac{R_i C_j}{n} \qquad (11.5)$$

Where does this formula for E_{ij}'s come from?

rows i and columns j

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

$$E_{ij} = \frac{R_i C_j}{n}$$

Where does this formula for E_{ij}'s come from?

Gender	Favorite Subject Area MS	SS	H	Total
Male	37 (29.28)	41 (45.95)	44 (46.77)	122
Female	35 (42.72)	72 (67.05)	71 (68.23)	178
Total	72	113	115	300

r=2
c=3

If Favorite Subject is independent of Gender, then

$$\chi^2 * = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} < \chi^2(2, 0.05)$$

α=0.05

$$df = (r-1)(c-1) = (2-1)(3-1)$$

$$\chi^2 * = 4.604 < \chi^2(2, 0.05) = 5.99$$

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

$$E_{ij} = \frac{R_i C_j}{n}$$

Expected Frequencies for an $r \times c$ Contingency Table

Row	Column 1	2	. . .	jth column	. . .	c	Total
1	$\frac{R_1 \times C_1}{n}$	$\frac{R_1 \times C_2}{n}$	. . .	$\frac{R_1 \times C_j}{n}$	. . .	$\frac{R_1 \times C_c}{n}$	R_1
2	$\frac{R_2 \times C_1}{n}$						R_2
⋮	⋮			⋮			⋮
i th row	$\frac{R_i \times C_1}{n}$			$\frac{R_i \times C_j}{n}$	. . .		R_i
⋮	⋮			⋮			⋮
r	$\frac{R_r \times C_1}{n}$						
Total	C_1	C_2	. . .	C_j	. . .	. . .	n

$$\chi^2 * = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \overset{?}{<} \chi^2((r-1)(c-1), \alpha)$$

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Homogeneity

$$E_{ij} = \frac{R_i C_j}{n}$$

Is the distribution within all rows the same for all rows?

Residence	Governor's Proposal: Favor	Oppose	Total
Urban	143	57	200
Suburban	98	102	200
Rural	13	87	100
Total	254	246	500

r=3
c=2

α=0.05

$$\chi^2{*} = \sum_{all\ cells} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \overset{?}{<} \chi^2((r-1)(c-1), \alpha)$$

$$df = (r-1)(c-1) = (3-1)(2-1)$$

Recap Chapter 12

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

Previously we learned about hypothesis testing for:
One Population: μ, p, and σ^2.
Two Populations: $\mu_d=\mu_1-\mu_2$, $\mu_1-\mu_2$, p_1-p_2, and σ_1^2/σ_2^2.

(We also learned about hypothesis testing for contingency tables.)

Now we are going to study hypothesis testing for three or more populations.

Three Populations: at least two of $\mu_1,\mu_2,\mu_3,\ldots$ different.

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

If we are testing for differences in means,
…why are we analyzing variance?

As it turns out, we calculate two variances and take the ratio.

If all the means are truly the same, the two variances will be the same and the ratio will be 1.

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

Hypothesis Testing Procedure

Temperature Levels		
Sample from 68°F ($i = 1$)	Sample from 72°F ($i = 2$)	Sample from 76°F ($i = 3$)
10	7	3
12	6	3
10	7	5
9	8	4
	7	
$C_1 = 41$ $\bar{x}_1 = 10.25$	$C_2 = 35$ $\bar{x}_2 = 7.0$	$C_3 = 15$ $\bar{x}_3 = 3.75$

Step 1 $H_0: \mu_1 = \mu_2 = \mu_3$ vs.

H_a: *at least two* μ*'s different*

Step 2

Sum of Squares Due to Factor

$$SS(\text{factor}) = \left(\frac{C_1^2}{k_1} + \frac{C_2^2}{k_2} + \frac{C_3^2}{k_3} + \cdots\right) - \frac{(\sum x)^2}{n}$$

$$\text{df(factor)} = c - 1$$

$$MS(\text{factor}) = \frac{SS(\text{factor})}{\text{df(factor)}}$$

Sum of Squares Due to Error

$$SS(\text{error}) = \sum(x^2) - \left(\frac{C_1^2}{k_1} + \frac{C_2^2}{k_2} + \frac{C_3^2}{k_3} + \cdots\right)$$

$$\text{df(error)} = n - c$$

$$MS(\text{error}) = \frac{SS(\text{error})}{\text{df(error)}}$$

$$F\star = \frac{MS(\text{factor})}{MS(\text{error})}$$

$$\alpha = .05$$

Shortcut for Total Sum of Squares

$$SS(\text{total}) = \sum(x^2) - \frac{(\sum x)^2}{n}$$

$$\text{df(total)} = n - 1$$

12: Analysis of Variance

12.1 Introduction to the Analysis of Variance

Hypothesis Testing Procedure

$$SS(temperature)=\left(\frac{41^2}{4}+\frac{35^2}{5}+\frac{15^2}{4}\right)-\frac{(91)^2}{13}=84.5$$

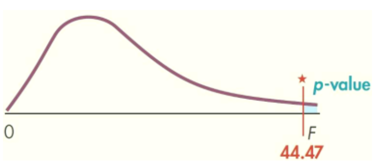
$$SS(error)=731.0-\left(\frac{41^2}{4}+\frac{35^2}{5}+\frac{15^2}{4}\right)=9.5$$

Source	df	SS	MS
Temperature	2	84.5	42.25
Error	10	9.5	0.95
Total	12	94.0	$F^*=44.47$

Step 3

$$F^*=\frac{42.25}{0.95}=44.47$$

Step 4



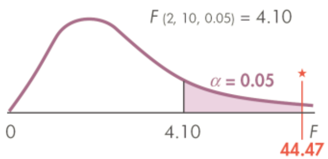
$\alpha=0.01$

Degrees of Freedom for Numerator

Degrees of Freedom for Denominator	1	2	3	4	5	6	7	8	9	10
1	4052.	5000.	5403.	5625.	5764.	5859.	5928.	5981.	6022.	6056.
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30

$$.00<p-value<.01$$

Classical approach



$\alpha=0.05$

Degrees of Freedom for Numerator

Degrees of Freedom for Denominator	1	2	3	4	5	6	7	8	9	10
1	161.	200.	216.	225.	230.	234.	237.	239.	241.	242.
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75

$$F(2,10,0.05)=4.10$$

Step 5

Decision: Reject H_0

$$p-value<\alpha$$

$$.00<p-value<.01$$

$$\alpha=0.05$$

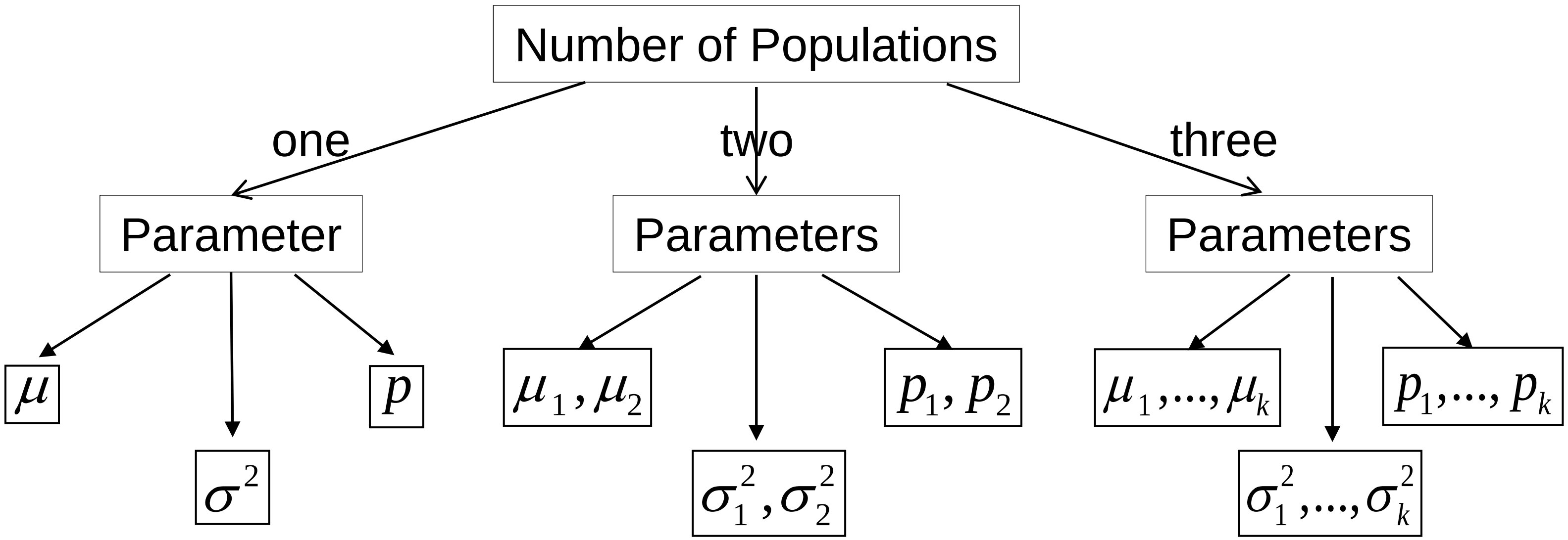
$$F^*>F_{crit}$$

$$F^*=44.47$$

$$F_{crit}=4.10$$

Statistical Inference

Statistical Inference:



Case 1: One Variable (Population)

Assumptions	σ is known	σ is unknown	$n(p_0) > 5$ and $n(1-p_0) > 5$	Data comes from Normal Population
Parameter of interest:	Mean, μ	Mean, μ	Proportion, p	Variance, σ^2
Confidence Interval Formula:	$\bar{x} \pm z(\alpha/2)\frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t(df, \alpha/2)\frac{s}{\sqrt{n}}$ with $df = n-1$	$p' \pm z(\alpha/2)\sqrt{\frac{p'q'}{n}}$	$\frac{(n-1)s^2}{\chi^2(df, \alpha/2)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2(df, 1-\alpha/2)}$ with $df = n-1$
Name of Hypothesis Test, H_0	One sample z-test, $H_0: \mu = \mu_0$	One sample t-test, $H_0: \mu = \mu_0$	One sample test of proportion, $H_0: p = p_0$	One sample test for Variance, $H_0: \sigma^2 = \sigma_0^2$
Test Statistic Formula:	$z^* = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$t^* = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ with $df = n-1$	$z^* = \frac{p' - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$	$\chi^{2*} = \frac{(n-1)s^2}{\sigma_0^2}$ with $df = n-1$
p-value:	$H_a: \mu > \mu_0$, p-value=$P(z \geq z^*)$ $H_a: \mu < \mu_0$, p-value=$P(z \leq z^*)$ $H_a: \mu \neq \mu_0$, p-value=$2 \times P(z \geq \lvert z^* \rvert)$	$H_a: \mu > \mu_0$, p-value=$P(t(df) \geq t^*)$ $H_a: \mu < \mu_0$, p-value=$P(t(df) \leq t^*)$ $H_a: \mu \neq \mu_0$, p-value=$2 \times P(t(df) \geq \lvert t^* \rvert)$	$H_a: p > p_0$, p-value=$P(z \geq z^*)$ $H_a: p < p_0$, p-value=$P(z \leq z^*)$ $H_a: p \neq p_0$, p-value=$2 \times P(z \geq \lvert z^* \rvert)$	$H_a: \sigma^2 > \sigma_0^2$, p-value=$P(\chi^2(df) \geq \chi^{2*})$ $H_a: \sigma^2 < \sigma_0^2$, p-value=$P(\chi^2(df) \leq \chi^{2*})$ $H_a: \sigma^2 \neq \sigma_0^2$, p-value=$2 \times P(\chi^2(df) \geq \lvert \chi^{2*} \rvert)$

Case 2: Two Numerical Variables (Populations)

Assumption	Dependent Samples (Paired Samples)	Independent Samples		Two Normal Populations Independent Samples
Parameter of interest:	Mean Difference, $\mu_d = \mu_1 - \mu_2$	Mean Difference, $\mu_1 - \mu_2$	Proportion Difference, $p_1 - p_2$	Ratio of variances, $\frac{\sigma_n^2}{\sigma_d^2}$
Confidence Interval Formula:	$\bar{d} \pm t(df, \alpha/2)\frac{s_d}{\sqrt{n}}$ with $df = n-1$ where $d = x_1 - x_2$	$(\bar{x}_1 - \bar{x}_2)$ $\pm t(df, \alpha/2)\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ with $df = \min(n_1 - 1, n_2 - 1)$	$(p_1' - p_2')$ $\pm z(\alpha/2)\sqrt{\frac{p_1'q_1'}{n_1} + \frac{p_2'q_2'}{n_2}}$	$\frac{1}{F(df_n, df_d, \alpha/2)}\frac{s_n^2}{s_d^2} < \frac{\sigma_n^2}{\sigma_d^2}$ $< F(df_d, df_n, \alpha/2)\frac{s_n^2}{s_d^2}$
Name of Hypothesis Test, H_0	Paired samples t-test, $H_0: \mu_1 = \mu_2$	Two independent samples t-test, $H_0: \mu_1 = \mu_2$	Two sample test of proportion $H_0: p_1 = p_2$	Two sample test for variance $H_0: \sigma_n^2 = \sigma_d^2$
Test Statistic Formula:	$t^* = \frac{\bar{d}}{s_d/\sqrt{n}}$ with $df = n-1$	$t^* = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ with $df = \min(n_1 - 1, n_2 - 1)$	$z^* = \frac{p_1' - p_2'}{\sqrt{p_p' q_p'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $p_p' = \frac{x_1 + x_2}{n_1 + n_2}$ and $q_p' = 1 - p_p'$	$F^* = \frac{s_n^2}{s_d^2}$ with $df_n = n_{num} - 1$ and $df_d = n_{den} - 1$
p-value:	$H_a: \mu_1 > \mu_2$, p-value=$P(t(df) \geq t^*)$ $H_a: \mu_1 < \mu_2$, p-value=$P(t(df) \leq t^*)$ $H_a: \mu_1 \neq \mu_2$, p-value=$2 \times P(t(df) \geq \lvert t^* \rvert)$		$H_a: p_1 > p_2$, p-value=$P(z \geq z^*)$ $H_a: p_1 < p_2$, p-value=$P(z \leq z^*)$ $H_a: p_1 \neq p_2$, p-value=$2 \times P(z \geq \lvert z^* \rvert)$	$H_a: \sigma_n^2 > \sigma_d^2$, p-value=$P(F(df_n, df_d) \geq F^*)$ $H_a: \sigma_n^2 \neq \sigma_d^2$, p-value=$2 \times P(F(df_n, df_d) \geq F^*)$

Case 3: More than Two Populations
(Multinomial Experiment, Contingency Table, Test of Homogeneity - ANOVA)

	Multinomial Experiment, Contingency Table, Homogeneity		Analysis of Variance (ANOVA)
Parameter of interest:	Probability: $p_1, p_2, \cdots, p_k$	Probability	Mean: $\mu_1, \mu_2, \cdots, \mu_c$
H_0	$H_0: p_1 = p_{10}, \cdots, p_k = p_{k0}$	H_0: Independency (Homogeneity)	$H_0: \mu_1 = \mu_2 = \cdots = \mu_c$
Test Statistic Formula:	$\chi^{2*} = \sum_{\text{all cells}} \frac{(O-E)^2}{E}$		$F^* = \frac{\text{MS(factor)}}{\text{MS(error)}} = \frac{\text{SS(factor)/df(factor)}}{\text{SS(error)/df(error)}}$
df and other items	$df = k - 1$ $E_i = n \times p_{i0}$	$df = (r-1)(c-1)$ $E_{ij} = \frac{R_i \times C_j}{n}$	$\text{df(factor)} = c - 1, \quad \text{df(error)} = n - c$ $\text{SS(factor)} = \left(\frac{C_1^2}{k_1} + \frac{C_2^2}{k_2} + \cdots + \frac{C_c^2}{k_c}\right) - \frac{(\sum x)^2}{n}$ $\text{SS(error)} = \sum (x^2) - \left(\frac{C_1^2}{k_1} + \frac{C_2^2}{k_2} + \cdots + \frac{C_c^2}{k_c}\right)$
p-value:	p-value=$P(\chi^2(df) \geq \chi^{2*})$		p-value=$P(F(df_n, df_d) \geq F^*)$

Case 1: One Variable (Population)

Assumptions	σ is known	σ is unknown	$n(p_0) > 5$ and $n(1-p_0) > 5$	Data comes from Normal Population
Parameter of interest:	Mean, μ	Mean, μ	Proportion, p	Variance, σ^2
Confidence Interval Formula:	$\bar{x} \pm z(\alpha/2)\dfrac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t(df,\alpha/2)\dfrac{s}{\sqrt{n}}$ with $df = n-1$	$p' \pm z(\alpha/2)\sqrt{\dfrac{p'q'}{n}}$	$\dfrac{(n-1)s^2}{\chi^2(df,\alpha/2)} < \sigma^2 < \dfrac{(n-1)s^2}{\chi^2(df,1-\alpha/2)}$ with $df = n-1$
Name of Hypothesis Test, H_0	One sample z-test, $H_0: \mu = \mu_0$	One sample t-test, $H_0: \mu = \mu_0$	One sample test of proportion $H_0: p = p_0$	One sample test for Variance $H_0: \sigma^2 = \sigma_0^2$
Test Statistic Formula:	$z^* = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$t^* = \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}}$ with $df = n-1$	$z^* = \dfrac{p'-p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$	$\chi^{2*} = \dfrac{(n-1)s^2}{\sigma_0^2}$ with $df = n-1$
p-value:	$H_a: \mu > \mu_0$, p-value=$P(z \geq z^*)$ $H_a: \mu < \mu_0$, p-value=$P(z \leq z^*)$ $H_a: \mu \neq \mu_0$, p-value=$2 \times P(z \geq \lvert z^* \rvert)$	$H_a: \mu > \mu_0$, p-value=$P(t(df) \geq t^*)$ $H_a: \mu < \mu_0$, p-value=$P(t(df) \leq t^*)$ $H_a: \mu \neq \mu_0$, p-value=$2 \times P(t(df) \geq \lvert t^* \rvert)$	$H_a: p > p_0$, p-value=$P(z \geq z^*)$ $H_a: p < p_0$, p-value=$P(z \leq z^*)$ $H_a: p \neq p_0$, p-value=$2 \times P(z \geq \lvert z^* \rvert)$	$H_a: \sigma^2 > \sigma_0^2$, p-value=$P(\chi^2(df) \geq \chi^{2*})$ $H_a: \sigma^2 < \sigma_0^2$, p-value=$P(\chi^2(df) \leq \chi^{2*})$ $H_a: \sigma^2 \neq \sigma_0^2$, p-value=$2 \times P(\chi^2(df) \geq \chi^{2*})$

Case 2: Two Numerical Variables (Populations)

	Dependent Samples (Paired Samples)	Independent Samples		Two Normal Populations Independent Samples
Parameter of interest:	Mean Difference, $\mu_d = \mu_1 - \mu_2$	Mean Difference, $\mu_1 - \mu_2$	Proportion Difference, $p_1 - p_2$	Ratio of variances, σ_n^2 / σ_d^2
Confidence Interval Formula:	$\bar{d} \pm t(df,\alpha/2)\dfrac{s_d}{\sqrt{n}}$ with $df = n-1$ where $d = x_1 - x_2$	$(\bar{x}_1 - \bar{x}_2)$ $\pm\, t(df,\alpha/2)\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}$ with $df = \min\,(n_1-1, n_2-1)$	$(p_1' - p_2')$ $\pm\, z(\alpha/2)\sqrt{\dfrac{p_1'q_1'}{n_1}+\dfrac{p_2'q_2'}{n_2}}$	$\dfrac{1}{F(df_n, df_d, \alpha/2)}\dfrac{s_n^2}{s_m^2} < \dfrac{\sigma_n^2}{\sigma_d^2}$ $< F(df_d, df_n, \alpha/2)\dfrac{s_n^2}{s_m^2}$ with $df_n = n_{num} - 1$ and $df_d = n_{den} - 1$
Name of Hypothesis Test, H_0	Paired samples t-test, $H_0: \mu_1 = \mu_2$	Two independent samples t-test, $H_0: \mu_1 = \mu_2$	Two sample test of proportion $H_0: p_1 = p_2$	Two sample test for variance $H_0: \sigma_n^2 = \sigma_d^2$
Test Statistic Formula:	$t^* = \dfrac{\bar{d}}{s_d/\sqrt{n}}$ with $df = n-1$	$t^* = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}}$ with $df = \min\,(n_1-1, n_2-1)$	$z^* = \dfrac{p_1' - p_2'}{\sqrt{p_p' q_p'\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ where $p_p' = \dfrac{x_1+x_2}{n_1+n_2}$ and $q_p' = 1 - p_p'$	$F^* = \dfrac{s_n^2}{s_m^2}$ with $df_n = n_{num} - 1$ and $df_d = n_{den} - 1$
p-value:	$H_a: \mu_1 > \mu_2$, p-value=$P(t(df) \geq t^*)$ $H_a: \mu_1 < \mu_2$, p-value=$P(t(df) \leq t^*)$ $H_a: \mu_1 \neq \mu_2$, p-value=$2 \times P(t(df) \geq \lvert t^* \rvert)$		$H_a: p_1 > p_2$, p-value=$P(z \geq z^*)$ $H_a: p_1 < p_2$, p-value=$P(z \leq z^*)$ $H_a: p_1 \neq p_2$, p-value=$2 \times P(z \geq \lvert z^* \rvert)$	$H_a: \sigma_n^2 > \sigma_d^2$, p-value=$P(F(df_n, df_d) \geq F^*)$ $H_a: \sigma_n^2 \neq \sigma_d^2$, p-value=$2 \times P(F(df_n, df_d) \geq F^*)$

Case 3: More than Two Populations (Multinomial Experiment, Contingency Table, Test of Homogeneity - ANOVA)

	Multinomial Experiment, Contingency Table, Homogeneity		Analysis of Variance (ANOVA)
Parameter of interest:	Probability: $p_1, p_2, \cdots, p_k$	Probability	Mean: $\mu_1, \mu_2, \cdots, \mu_c$
H_0	$H_0: p_1 = p_{10}, \cdots, p_k = p_{k0}$	H_0: Independency (Homogeneity)	$H_0: \mu_1 = \mu_2 = \cdots = \mu_c$
Test Statistic Formula:	$\chi^{2*} = \displaystyle\sum_{\text{all cells}} \dfrac{(O-E)^2}{E}$		$F^* = \dfrac{\text{MS(factor)}}{\text{MS(error)}} = \dfrac{\text{SS(factor)/df(factor)}}{\text{SS(error)/df(error)}}$
df and other items	$df = k-1$ $E_i = n \times p_{i0}$	$df = (r-1)(c-1)$ $E_{ij} = \dfrac{R_i \times C_j}{n}$	$\text{df(factor)} = c-1, \quad \text{df(error)} = n-c$ $\text{SS(factor)} = \left(\dfrac{C_1^2}{k_1}+\dfrac{C_2^2}{k_2}+\cdots+\dfrac{C_c^2}{k_c}\right) - \dfrac{(\sum x)^2}{n}$ $\text{SS(error)} = \sum(x^2) - \left(\dfrac{C_1^2}{k_1}+\dfrac{C_2^2}{k_2}+\cdots+\dfrac{C_c^2}{k_c}\right)$
p-value:	p-value=$P(\chi^2(df) \geq \chi^{2*})$		p-value=$P(F(df_n, df_d) \geq F^*)$