Class 24

Daniel B. Rowe, Ph.D.

Department of Mathematical and Statistical Sciences

Agenda:

Recap Chapter 10.4-10.5

Lecture Chapter 11.1-11.3

Recap Chapter 10.4-10.5

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Interested in comparisons between proportions $p_1 - p_2$.

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 $p_1 - p_2$

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

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

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Confidence Interval Procedure

$\alpha = 0.01$

Example:

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**

$$z^* = \frac{(p_1' - p_2') - \overbrace{(p_{0,1} - p_{0,2})}^{0}}{\sqrt{\underbrace{pq}_{p \text{ known}}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

$$p_1' = \frac{x_1}{n_1} \quad p_2' = \frac{x_2}{n_2} \qquad (10.12)$$

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**

$$z^* = \frac{(p_1' - p_2') - \overbrace{(p_{0,1} - p_{0,2})}^{0}}{\sqrt{p_p' q_p' \left[\dfrac{1}{n_1} + \dfrac{1}{n_2} \right]}} \qquad (10.15)$$

p_p estimated

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

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

10: Inferences Involving Two Populations

10.4 Inference for Difference between Two Proportions

Is proportion of Salesman's defectives less than Competitor's? $\alpha = .05$

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 $\swarrow < .05$

$.02 < p-value < .023$ 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

Actually $\longleftarrow$ ignore

$$F^* = \frac{\left[(n_n - 1)s_n^2 / \sigma^2\right] / (n_n - 1)}{\left[(n_d - 1)s_d^2 / \sigma^2\right] / (n_d - 1)}$$

Test Statistic for Equality of Variances

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

Use new table to find areas for new statistic.

10: Inferences Involving Two Populations

10.5 Inference for Ratio of Two Variances Two Ind. Samples

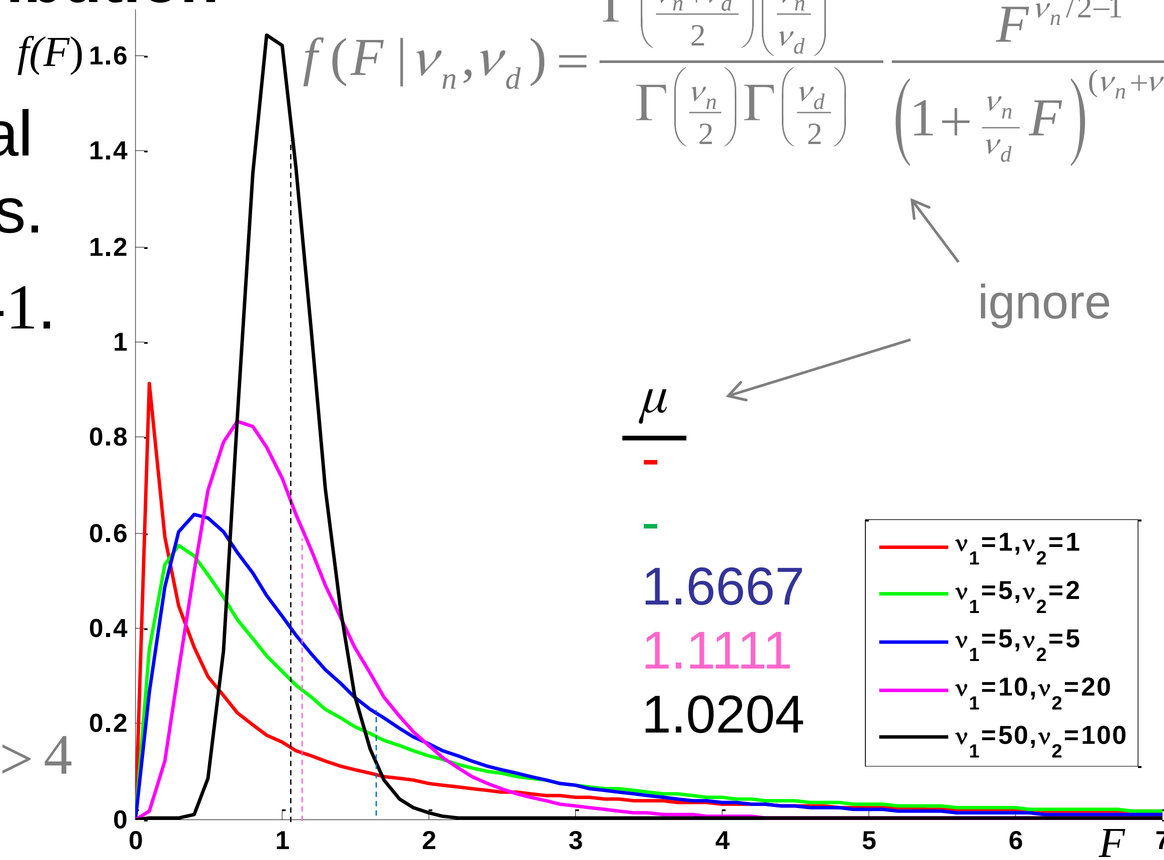
Properties of F distribution

1. F is non-negative
2. F is nonsymmetrical
3. F is a family of dists.

$df_n = \nu_n = n_n - 1, df_d = \nu_d = n_d - 1.$

$$f(F\mid\nu_n,\nu_d)=\frac{\Gamma\left(\frac{\nu_n+\nu_d}{2}\right)\left(\frac{\nu_n}{\nu_d}\right)^{\nu_n/2}}{\Gamma\left(\frac{\nu_n}{2}\right)\Gamma\left(\frac{\nu_d}{2}\right)}\frac{F^{\nu_n/2-1}}{\left(1+\frac{\nu_n}{\nu_d}F\right)^{(\nu_n+\nu_d)/2}}$$

ignore



ignore

$$\mu=\frac{\nu_d}{\nu_d-2}\quad,\ \nu_d>2$$

$$\sigma^2=\frac{2\nu_d^2(\nu_n+\nu_d-2)}{\nu_n(\nu_d-2)^2(\nu_d-4)}\ ,\ \nu_2>4$$

10: Inferences Involving Two Populations

10.5 Inference for Ratio of Two Variances Two Ind. Samples

Hypothesis Testing Procedure

Test Statistic for Equality of Variances

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

Will also need critical values.

$$P\left(F > F(df_n, df_d, \alpha)\right) = \alpha$$

Table 9
Appendix B
Page 722

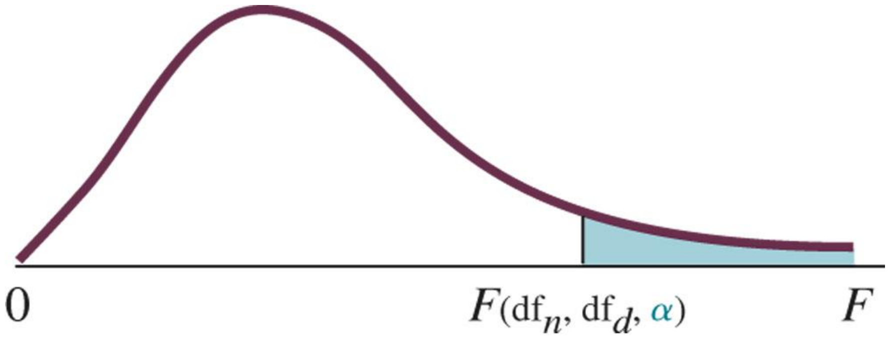


Figure from Johnson & Kuby, 2012.

10: Inferences Involving Two Pops.

10.5 Inference Ratio of Two Variances

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

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

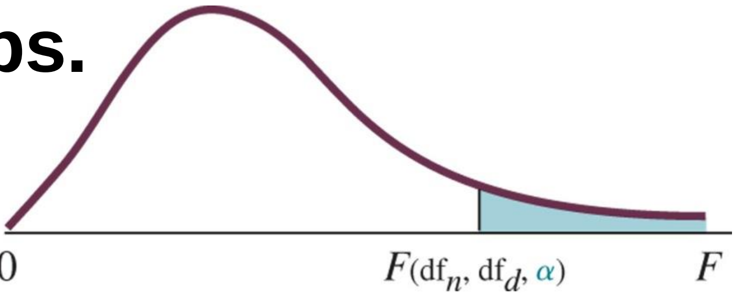


Table 9, Appendix B, Page 722.

$\alpha = 0.05$

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: $\sigma_2^2 / \sigma_1^2 > 1$ $\quad F^* = s_2^2 / s_1^2$

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

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

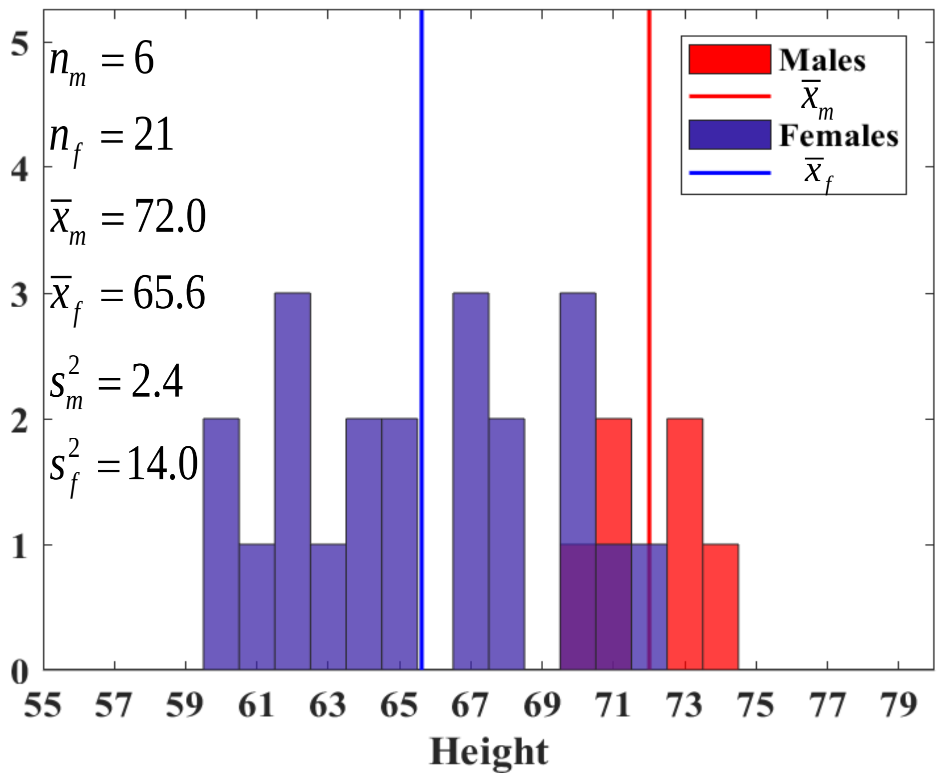
Step 2 $F^* = s_f^2 / s_m^2$ $\quad df_m = 5$, $df_f = 20$, $\alpha = .01$

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 $> \sigma_m^2$.

Chapter 10: Inferences Involving Two Populations

Questions?

Homework: Read Chapter 10.4-10.5

WebAssign

Chapter 10# 83, 85, 91, 98, 99, 101, 111, 113, 115, 117, 119, 125, 133

Lecture Chapter 11.1-11.3

Chapter 11: Applications of Chi-Square

Daniel B. Rowe, Ph.D.

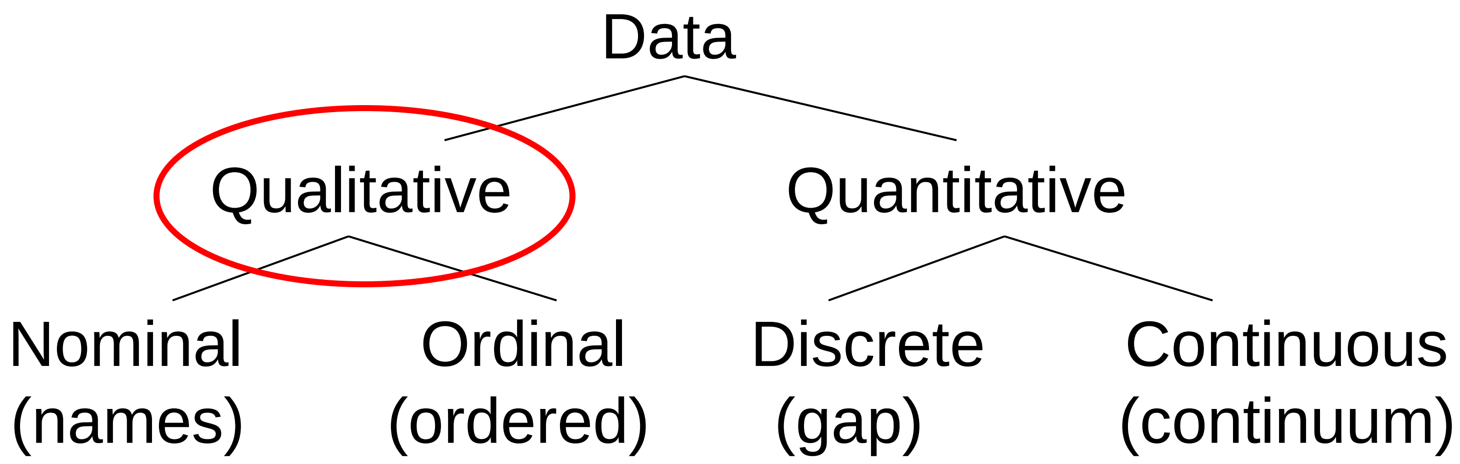
Department of Mathematical and Statistical

Recall

1: Statistics

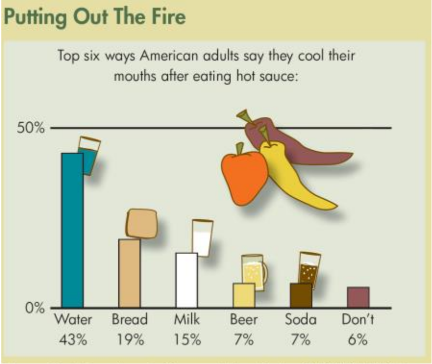
1.2 What is Statistics?

Data: The set of values collected from the variable from each of the elements that belong to the sample.



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 $O_1,\ldots,O_k$. $O_1+\ldots+O_k=n$

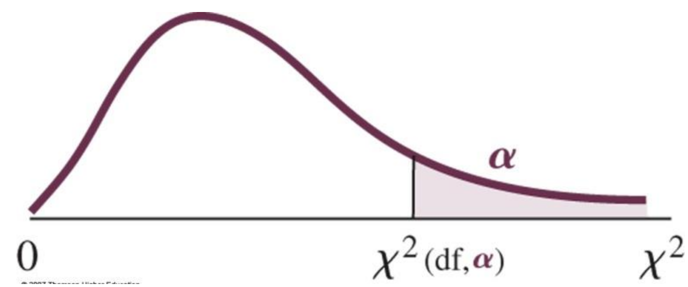
Expected frequencies in each cell $E_1,\ldots,E_k$. $E_1+\ldots+E_k=n$

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

Outline of Test Procedure



When we have observed cell frequencies $O_1,\ldots,O_k$, we can test to see if they match with some expected cell frequencies $E_1,\ldots,E_k$.

Test Statistic for Chi-Square

$$\chi^2 * = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad df = k-1 \qquad (11.1)$$

If the O_i's are different from E_i's then $\chi^2 *$ is "large."

Go through 5 hypothesis testing steps as before.

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.1 Chi-Square Statistic

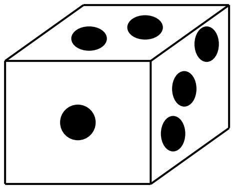
Assumption for using the chi-square statistic to make inferences based upon enumerative data: … a random sample drawn from a population where each individual is classified according to the categories

$$\chi^2* = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad df = k-1$$

observed cell frequencies $O_1,\ldots,O_k$,
expected cell frequencies $E_1,\ldots,E_k$.

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.

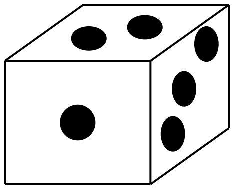
Cell, i	1	2	3	4	5	6
Observed, O_i	7	12	10	12	8	11
Expected, E_i	10	10	10	10	10	10

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 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

$E_i = 60(1/6)$

Is the die fair? Need to go through the hypothesis testing procedure to determine if it is fair.

11: Applications of Chi-Square

11.2 Inferences Concerning Multinomial Experiments

Example: Is the die fair? $\chi^2 * = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$

Calculating χ^2

D of F for Mult:
$df = k - 1$ (11.2)

Number	Observed (O)	Expected (E)	O − E	(O − E)²	$\frac{(O-E)^2}{E}$
1	7	10	−3	9	0.9
2	12	10	2	4	0.4
3	10	10	0	0	0.0
4	12	10	2	4	0.4
5	8	10	−2	4	0.4
6	11	10	1	1	0.1
Total	60	60	0 ck		$\chi^2 * = 2.2$

$O_1 + \ldots + O_k = n$ $E_1 + \ldots + E_k = n$

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.2 Inference for Mean Difference Two Dependent Samples

Cell	1	2	3	4	5	6
Observed	7	12	10	12	8	11
Expected	10	10	10	10	10	10

$k = 6$

Observed different than expected?

Step 1 Fill in.

Step 2

χ^2

Step 3 **Step 4**

Step 5

χ^2

Figures from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.2 Inference for Mean Difference Two Dependent Samples

Cell	1	2	3	4	5	6
Observed	7	12	10	12	8	11
Expected	10	10	10	10	10	10

$k=6$

Observed different than expected?

Step 1

H_0:Die fair $p_i\text{'s}=\frac{1}{6}$ H_a:Die not fair $p_i\text{'s}\neq\frac{1}{6}$

Step 2

Step 3 **Step 4**

Step 5

Figures from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.2 Inference for Mean Difference Two Dependent Samples

Cell	1	2	3	4	5	6
Observed	7	12	10	12	8	11
Expected	10	10	10	10	10	10

$k = 6$

Observed different than expected?

Step 1

H_0:Die fair $p_i\text{'s} = \frac{1}{6}$ H_a:Die not fair $p_i\text{'s} \neq \frac{1}{6}$

Step 2

$$\chi^2* = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad df = k - 1 \qquad \alpha = .05$$

Step 3 **Step 4**

Step 5

Figures from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.2 Inference for Mean Difference Two Dependent Samples

Cell	1	2	3	4	5	6
Observed	7	12	10	12	8	11
Expected	10	10	10	10	10	10

$k = 6$

Observed different than expected?

Step 1

H_0:Die fair $p_i\text{'s} = \frac{1}{6}$ H_a:Die not fair $p_i\text{'s} \neq \frac{1}{6}$

Step 2

$$\chi^2{}^* = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad df = k - 1 \quad \alpha = .05$$

Step 3 **Step 4**

$$\chi^2{}^* = 2.2$$

Step 5

Figures from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.2 Inference for Mean Difference Two Dependent Samples

Cell	1	2	3	4	5	6
Observed	7	12	10	12	8	11
Expected	10	10	10	10	10	10

$k = 6$

Observed different than expected?

Step 1

H_0:Die fair $p_i\text{'s} = \frac{1}{6}$ H_a:Die not fair $p_i\text{'s} \neq \frac{1}{6}$

Step 2

$$\chi^2 * = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad df = k-1 \quad \alpha = .05$$

Step 3

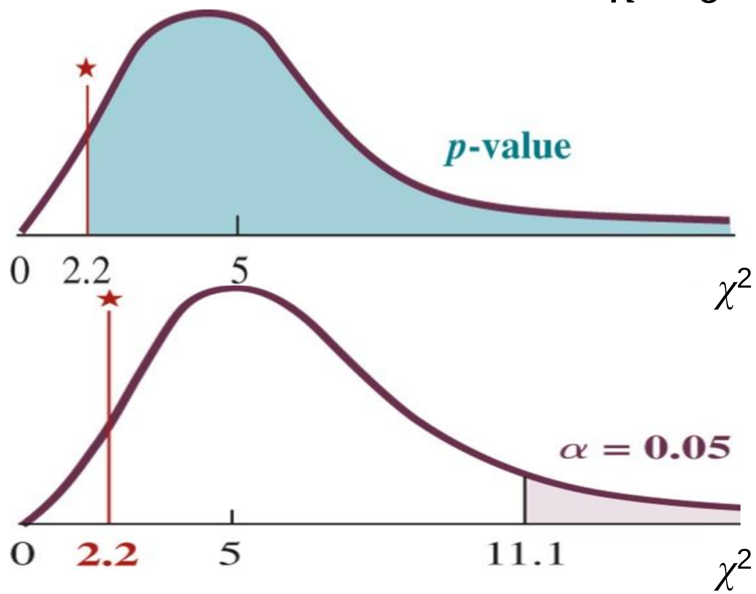
$\chi^2 * = 2.2$

Step 4

$\chi^2(df, \alpha) = 11.1$

$df = 5$

Step 5



Figures from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.2 Inference for Mean Difference Two Dependent Samples

Cell	1	2	3	4	5	6
Observed	7	12	10	12	8	11
Expected	10	10	10	10	10	10

$k = 6$

Observed different than expected?

Step 1

H_0:Die fair $p_i\text{'s} = \frac{1}{6}$ H_a:Die not fair $p_i\text{'s} \neq \frac{1}{6}$

Step 2

$$\chi^2{*} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad df = k - 1 \quad \alpha = .05$$

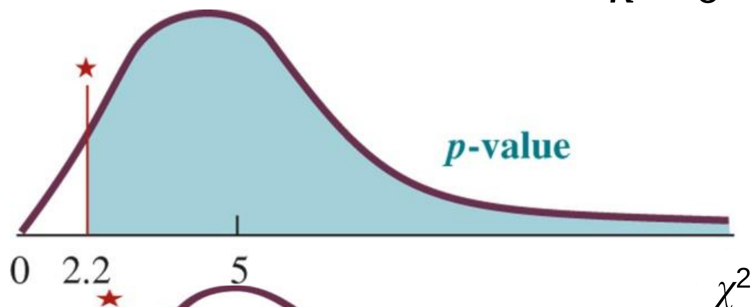
Step 3

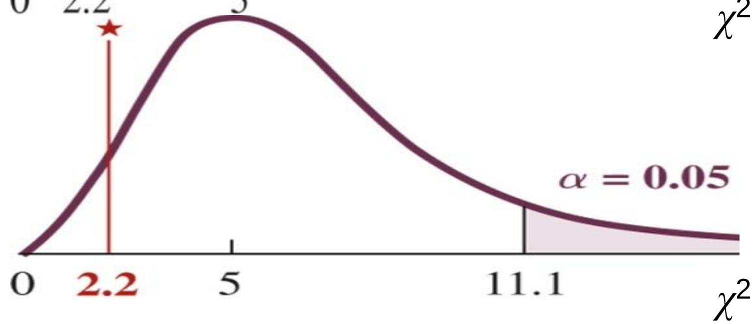
$$\chi^2{*} = 2.2$$

Step 4

$$\chi^2(df, \alpha) = 11.1$$
$$df = 5$$

Step 5





Figures from Johnson & Kuby, 2012.

Since .05<p-$value$=.82 **or** because $\chi^2{*}<\chi^2(df,\alpha)$, fail to reject H_0

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data

Bivariate data: The values of two different variables that are obtained from the same population element.

Qualitative-Qualitative
Qualitative-Quantitative
Quantitative-Quantitative

When Qualitative-Qualitative
Cross-tabulation tables or **contingency tables**
Sometimes called r by c ($r \times c$)

3: Descriptive Analysis and Bivariate Data

3.1 Bivariate Data: two qualitative

Example:
Construct a 2×3 table.

Name	Gender	Major	Name	Gender	Major	Name	Gender	Major
Adams	M	LA	Feeney	M	T	McGowan	M	BA
Argento	F	BA	Flanigan	M	LA	Mowers	F	BA
Baker	M	LA	Hodge	F	LA	Ornt	M	T
Bennett	F	LA	Holmes	M	T	Palmer	F	LA
Brand	M	T	Jopson	F	T	Pullen	M	T
Brock	M	BA	Kee	M	BA	Rattan	M	BA
Chun	F	LA	Kleeberg	M	LA	Sherman	F	LA
Crain	M	T	Light	M	BA	Small	F	T
Cross	F	BA	Linton	F	LA	Tate	M	BA
Ellis	F	BA	Lopez	M	T	Yamamoto	M	LA

	Major		
Gender	LA	BA	T
M	||||| (5)	||||| | (6)	||||| || (7)
F	||||| | (6)	|||| (4)	|| (2)

M = male
F = female
LA = liberal arts
BA = business admin
T = technology

Figures from Johnson & Kuby, 2012.

11: Applications of Chi-Square
11.3 Inferences Concerning Contingency Tables

Example:
Construct a 2×3 table.

Each in group of 300 students identified as male or female and asked if preferred classes in math-science, social science, or humanities.

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?”

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

4: Probability

4.5 Independent Events

Independent events: Two events are independent if the occurrence or nonoccurence of one gives us no information about the likeliness of occurrence for the other.

In algebra: $P(A) = P(A \mid B) = P(A \mid \text{not } B)$

In words:

1. Prob of A unaffected by knowledge that B has occurred, not occurred, or no knowledge.
2. …
3. …

4: Probability

4.5 Independent Events

Two events A and B are independent if the probability of one is not "influenced" by the occurrence or nonoccurrence of the other.

Two Events A and B are independent if:

1. $P(A) = P(A|B)$
2. $P(B) = P(B|A)$
3. $P(A \text{ and } B) = P(A) \cdot P(B)$

Examples:?

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

If Favorite Subject (column variable) is independent of Gender (row variable), then

$$P(A) = P(A \mid B)$$

$$P(A \text{ and } B) = P(A) \cdot P(B)$$

$$P(MS \mid M) = P(MS \mid F) = P(MS)$$

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

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

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

	MS	SS	H	Total
Male	29.28	45.95	46.77	122.00
Female	42.72	67.05	68.23	178.00
Total	72.00	113.00	115.00	300.00

$$P(M) = 122/300$$

$$P(F) = 178/300$$

$$P(MS) = 72/300$$

$$P(SS) = 113/300$$

$$P(H) = 115/300$$

If Favorite Subject is independent of Gender, then

$$P(M \;\; and \;\; MS) = P(M)P(MS) = (122/300)(72/300)$$

$$E(M \;\; and \;\; MS) = nP(M)P(MS) = 300(122/300)(72/300)$$

$$E(M \;\; and \;\; MS) = 122 \times 72/300$$

Figure from Johnson & Kuby, 2012.

11: Applications of Chi-Square

11.3 Inferences Concerning Contingency Tables

Test of Independence

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

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

	Favorite Subject Area			
Gender	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
⋮	⋮			⋮			⋮
ith 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

If so, then $P(F \text{ and } Urban) = P(F)P(U)$

$$E(F \text{ and } Urban) = nP(F)P(U)$$

$$E(F \text{ and } Urban) = 500(254/500)(200/500)$$

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

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

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

Chapter 11: Applications of Chi-Square

Questions?

Homework: Read Chapter 11
WebAssign
Chapter 11 # 3, 5, 11, 15, 21, 49, 53