Brock Francom
A02052161

Homework Chi Square Tests

- If you use a different chi-square table than I do, then your boundaries for the p-value might be different.
  - But they should agree and we should both be right.
  - For example, suppose the true p-value is .0034, but I might be able to tell from my table that the p-value is between .005 and .001 and you might be able to narrow it down more to say that it is between .005 and .0025.
- In my other class we check "conditions" instead of "assumptions". So when you see the word "condition" you know it means "assumption".

Chi Square Goodness of Fit

1. One proposed M&M color distribution is listed below. Tommy took a sample of M&Ms and found the colors as listed in the other table.

Color	Theoretical Percentage
brown	13%
yellow	14%
red	13%
orange	20%
blue	24%
green	16%

Color	Tommy's Data
brown	61
yellow	59
red	49
orange	77
blue	141
green	88

(a) Fill in the table of observed and expected counts.

Color	Observed Count	Expected Count
brown	61	61.75
yellow	59	66.5
red	49	61.75
orange	77	95
blue	141	114
green	88	76

(b) Conduct a hypothesis test to see if the proposed color distribution is a good fit for Tommy's data.

assumptions ex ≥ 5 ✓

$\alpha = .05$

H_0: proposed M&M distribution is a good fit for data.

H_A: " " not a good "

test stat:

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp} = 15.2$$

df = 5

Pval: between .005 and .01

small, reject null.

we found convincing evidence that the proposed M&M color distribution is not a good fit for data.

2. In 2005, a national survey reported how often students use campus services. You take a sample of students at your university to determine if your university has a different distribution than the national percentages.

Campus Service	National Percentages
never	43%
sometimes	35%
often	15%
very often	7%

Campus Service	Your Counts
never	79
sometimes	83
often	36
very often	12

(a) Fill in the table of observed and expected counts.

	Observed Count	Expected Count
Nev	79	90.3
	83	73.5
	36	31.5
	12	14.7

(b) Conduct a hypothesis test to determine if the national distribution describes the data from your university.

conditional counts ≥ 5 ✓
$\alpha = .05$

H_0: national distribution is a good fit for data
H_A: " " is not a good "

test stat
$\chi^2 = 3.7807$
df = 3

pval: greater than .25
big pval. fail to reject

we didn't find evidence that the national distribution is not a good fit for the data.

3. We have a data set of 500 values and we want to know if it can be described by the standard normal distribution. We can break our values into categories. (We could have picked other categories as well.)

Category	Observed Counts
$z \leq -.6$	139
$-.6 < z \leq -.1$	102
$-.1 < z \leq .1$	41
$.1 < z \leq .6$	78
$z > .6$	140

(a) Use the standard normal table to find the probabilities for each category.

Category	Theoretical Probability
$z \leq -.6$	.2743
$-.6 < z \leq -.1$	.1859
$-.1 < z \leq .1$	.0796
$.1 < z \leq .6$	.1859
$z > .6$	.2743

(b) Fill in the table of observed and expected counts.

Category	Observed Count	Expected Count
$z \leq -.6$	139	137.2
$-.6 < z \leq -.1$	102	93
$-.1 < z \leq .1$	41	39.8
$.1 < z \leq .6$	78	93
$z > .6$	140	137.2

(c) Conduct a hypothesis test to determine if the data can be described by the normal distribution.

all counts ≥ 5 ✓
$\alpha = .05$
H_0: normal distribution is good fit for data
H_A: " " is not good "

test stat:
$\chi^2 = 3.4601$
df = 4

pval = greater than .25
big pval, fail to reject H_0.
we didn't find evidence that the normal distribution is not a good fit for our data.

Chi Square Test of Independence

4. A study was conducted on the ongoing fright symptoms after watching horror movies as a child. They noted whether the students had symptoms when they were awake and if the students had symptoms at bedtime.

		Waking Symptoms		
		Yes	No	Total
Bedtime Symptoms	Yes	36 40	33 29	69
	No	33 29	17 21	50
	Total	69	50	119

Conduct a hypothesis test to determine if there is a relationship between bedtime symptoms and waking symptoms.

expected vals > 5 ✓

$\alpha = .05$

Ho: Bedtime and waking ~~are~~ are independant

HA: " " are not independent

test stat

$$\chi^2 = \sum \frac{(obs - exp)^2}{exp} = 2.3$$

df = 1

Pval = between .15 and .1

big pval. fail to reject.

we didn't find evidence that bedtime and waking symptoms are dependant.

5. If you are still confused about degrees of freedom and what they really mean, here is a great explanation: https://blog.minitab.com/blog/statistics-and-quality-data-analysis/what-are-degrees-of-freedom-in-statistics

6. A study looked at the gender of U.S. college students that are 20-24 years old and their status as full-time or part-time students. Does gender affect if students are full or part time? You can use the fact that $\chi^2 = 27.79$.

	Male	Female	**Total**
Full Time	2719 2638.73	2991 3071.27	**5710**
Part Time	435 515.27	680 599.73	**1115**
Total	**3154**	**3671**	**6825**

expected counts > 5 ✓

$\alpha = .05$

Ho: They are independant

HA: They are dependant

test stat:

$$\chi^2 = \frac{(obs - exp)^2}{exp} = 27.79$$

df = 1

Pval = less than .0005
small reject null.

We found extremely strong evidence that they are dependant.

7. A midwestern university undertook a program to try to decrease the percentage of students getting a D, F, or W (withdraw) grade in one of their introductory classes. The first year they taught the traditional class, they introduced a few changes, the final year they completely revamped the course.

Observed Counts	DFW	Pass	**Total**
Year 1	1018	1390	**2408**
Year 2	579	1746	**2325**
Year 3	423	1703	**2126**
Total	**2020**	**4839**	**6859**

(a) I've filled in a few of the expected counts for you. Fill in the rest.

Expected Counts	DFW	Pass	Total
Year 1	709.16	1698.84	2408
Year 2	684.72	1640.28	2325
Year 3	626.11	1499.89	2126
Total	2020	4839	6859

(b) Conduct a hypothesis test to determine if the year (teaching style) affects the grade. Use $\chi^2 = 307.17$ as your test statistic to save time.

expected counts > 5 ✓

$\alpha = .05$

H_0: They are independent
H_A: They are dependant

test stat:

$$\chi^2 = \frac{(obs-exp)^2}{exp} = 307.17$$

df = 2

Pval = less than .0005
small, reject null

we have extreemly strong evidence that they are dependant.

A Mix of Problems

8. The market shares of the U.S. automobile industry in 1990 are listed in the table below. Shaylee took a sample of 1000 newly bought cars and she wants to know if the market shares are different now.

Manufacturer	Percentage
GM	36%
Japanese	26%
Ford	21%
Chrysler	9%
Other	8%

Manufacturer	Shaylee's Counts
GM	391
Japanese	202
Ford	275
Chrysler	53
Other	79

(a) Fill in the table of observed and expected counts. I've done a few for you.

	Observed Count	Expected Count
GM	391	360
Japanese	202	260
Ford	275	210
Chrysler	53	90
Other	79	80

(b) Conduct a hypothesis test to determine if the market shares are now different.

expected > 5 ✓

$\alpha = .05$

H_0: distribution of market shares is a good fit for data.

H_A: " " is not a good " "

test stat:

$$\chi^2 = \frac{(obs - exp)^2}{exp} = 50.95$$

df = 4

pval = less than .0005

small, reject null

we found extremely strong evidence that the 1960 distribution is not a good fit for the data.

9. Below is a table comparing satisfaction rating and fund type. Conduct a test to see if the fund type affects the customer satisfaction rating. (Does satisfaction rating depend on fund type?)

(a) Fill in the rest of the expected counts.

Observed counts

fund type	High	Med	Low	Total
Bond	15	12	3	**30**
Stock	24	4	2	**30**
Tax Def	1	24	15	**40**
Total	**40**	**40**	**20**	**100**

(Customer Satisfaction)

EXPECTED COUNTS

fund type	High	Med	Low	Total
Bond	12	12	6	**30**
Stock	12	12	6	**30**
Tax Def	16	16	8	**40**
Total	**40**	**40**	**20**	**100**

(Customer Satisfaction)

(b) Conduct the hypothesis test. You can use the fact that $\chi^2 = 46.44$.

expected > 5 ✓

$\alpha = .05$

H_0: They are independant.

H_A: They are dependant.

test stat:

$$\chi^2 = \frac{(obs - exp)^2}{exp} = 46.44$$

df = 4

Pval = smaller .0005

small, reject null

we have extremly strong evidence that they are dependant.

10. Jacob thinks that the probability of a randomly chosen tv viewer in the ares choosing to watch a channel are as given in the table. To test this, he takes a sample of 2000 random viewers. You can use the fact that $\chi^2 = 300.61$.

TV channel	Theoretical Probability
WDUX	.15
WWTY	.19
WACO	.22
WTJW	.16
Others	.28

TV Channel	Jacob's Data
WDUX	182
WWTY	536
WACO	354
WTJW	151
Others	777

(a) Fill in the table of observed and expected counts.

TV Channel	Observed Count	Expected Count
WDUX	182	300
WWTY	536	380
WACO	354	440
WTJW	151	320
Others	777	560

(b) Conduct a hypothesis test to determine if his proposed probabilities are correct. Use $\chi^2 = 300.61$ to save time.

expected > 5 ✓

$\alpha = .05$

H_0: theoretical probabilities are correct.

H_A: at least one is incorrect

test stat:

$\chi^2 = 300.61$

df = 4

pval = less than .0005

Small, reject H_0

we find extremely strong evidence that at least one of the proposed probabilites is wrong.

11. A study looked at the age of U.S. college students and their status as full-time or part-time students.

Age	Full Time	Part Time	**Total**
15-19 years	3553	329	**3882**
20-24 years	5710	1215	**6925**
25-34 years	1825	1864	**3689**
35 and over	901	1983	**2884**
Total	**11989**	**5391**	**17380**

- Conduct a hypothesis test to determine if there is a relationship between student's age and their status as full time or part time students.I have filled in some of the expected values for you.

Age	Full Time	Part Time	**Total**
15-19 years	2677.87	1204.13	**3882**
20-24 years	4776.97	2148.03	**6925**
25-34 years	2544.73	1144.27	**3689**
35 and over	1989.43	894.57	**2884**
Total	**11989**	**5391**	**17380**

- The χ^2 test statistic is 4085.58.

expected > 5 ✓

$\alpha = .05$

Ho: They are independant

Ha: They are dependant.

$\chi^2 = 4085.58$

df = 3

Pval = basically 0,

Small, reject Ho

we found extremly strong evidence that they are dependant

12. Are gender and voting preference dependent? Do they affect each other? I've filled in some of the expected counts for you and you can use the fact that $\chi^2 = .10$.

Observed counts

		Republican	Democrat	Independent	Total
Gender	Male	200	150	50	
	Female	247	187	66	
	Total				

Expected counts

		Republican	Democrat	Independent	Total
Gender	Male	198.67	149.78	51.56	400
	Female	248.33	187.22	64.44	500
	Total	447	337	116	900

expected > 5

$\alpha = .05$

H_0: They are independant

H_A: They are dependant

$\chi^2 = .10$

df = 2

Pval = bigger than .25

big, fail to reject null.

we didnt find evidence that they are dependant

13. We want to know how people choose a number between 1 and 10. Conduct a hypothesis test with the sample data collected in class last year to determine if the choices are equally likely. Use $\alpha = .10$.

Number	Number of Students	Theoretical Probability	Expected Counts
One	2	1/10	2.7
Two	3		
Three	4		
Four	3		
Five	2		
Six	1		
Seven	8		
Eight	2		
Nine	2		
Ten	0		

Cant do it, expected counts less than 5