§11.2 Contingency Tables

A Contingency Table is categorical counts summarized in a two way table with n-rows and m-columns, $n \times m$

Two Tests

Independence row independent of column

Homogeneity Proportions are the same

Methods are the same

Categories - Level of confidence in police

Groups / Gender	Great deal	Some	little
Men			
Women			

$H_0: p_{1m} = p_{1w}, \ p_{2m} = p_{2w}, \ p_{3m} = p_{3w}$ Homogeneity

H_0: (Level of confidence in police) is independent of (Gender)

H_1: $p_{im} \neq p_{iw}$ for at least one category

H_1: Row is Dependent on Column Variable

Hypothesis Test

H_0: $O = E$ $\quad P_{1W} = P_{1M}$, $P_{2W} = P_{2M}$, $P_{3W} = P_{3M}$

Row variable is $\boxed{\text{independent}}$ of Column Variable

H_1: $O \neq E$ At least one of the proportions is Different for the ~~two~~ groups

Row Variable is $\boxed{\text{Dependent}}$ on Column Variable

Test Statistic is χ^2 Like a GOF test

Use a χ^2-Test on 84 or 2-way χ^2 test on 89

Critical Value from Table or

CV: $\chi^2_R = \text{inv}\chi^2$(① $df = (\#\text{rows} - 1)\cdot(\#\text{of col} - 1)$

② $\alpha = \text{sig.} = .05$ or $.01$

③ Rt Tailed)

For a 2×3 contingency Table

$df = (2-1)(3-1) = 1 \cdot 2 = 2$

$\chi^2_R = 5.991$

Finding Test Statistic

TS: $\chi^2 = \sum \frac{(O-E)^2}{E}$

Given: these are the observed values

	GD	S	L	
M	115	56	29	200
W	175	94	31 (36)	300
	290	150	60	500

To find the expected value = Prop in Category · #Men = # of men in this Category

What overall proportion have a great deal of confidence in the police?

$$P(GD) = \frac{290}{500} = .58$$

If M & W are the same prop with GD, then

$ex\ \#_{\text{of men with GD}} = .58 \cdot 200 = 116$

Matrix of Expected Values

	GD	S	L
M	116	60	24
W	174	90	36

$$Ex = \frac{\text{row Total} \cdot \text{Col total}}{\text{grand Total}}$$

To Find Test Statistic

$$TS: \chi^2 = \sum \frac{(O-E)^2}{E}$$

	GD	S	L	
M	115 (116)	56 (60)	29 (24)	200
W	175 (174)	94 (90)	31 (36)	300
	290	150	60	500
	.58	.30	.12	

- observed values are given
- Expected Values are the counts we would get if proportion for both groups are the Same = Proportion overall · # of people in group

$$\frac{290}{500} \cdot 200 = \text{GD Men}$$

$$.58 \cdot 300 = \text{Ex GD Women}$$

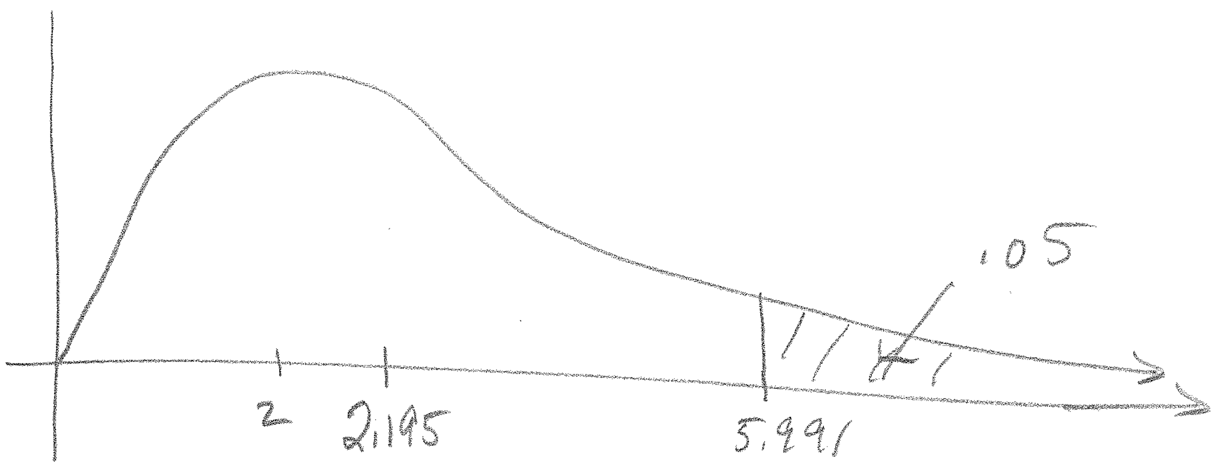
- Find Row & col totals and grand Total

$$\text{Exp Value} = \frac{\text{row Total} \cdot \text{Col Total}}{\text{Grand Total}}$$

TS: $\chi^2 = \sum \frac{(O-E)^2}{E}$

$$= \frac{(115-116)^2}{116} + \frac{4^2}{60} + \frac{5^2}{24} + \frac{1^2}{174} + \frac{4^2}{90} + \frac{5^2}{36}$$

$\chi^2 = 2.195$



fail to Reject H_0

TINSE to reject that confidence in the police is independent of gendar.

or

TINSE to show that ~~different~~ genders have different proportions of confidence in the police

Review from chapter 4

Probability of Selecting a person at random who

a) is a women and a some confidence?

$$P(W \cap S) = \frac{94}{500} = .188$$

b) has Some confidence given that they are a Women?

$$P(S|W) = \frac{94}{300} = .313$$

c) has some confidence given that they are a man?

$$P(S|M) = \frac{56}{200} = .28$$

d) is a Women or has Some confidence in police?

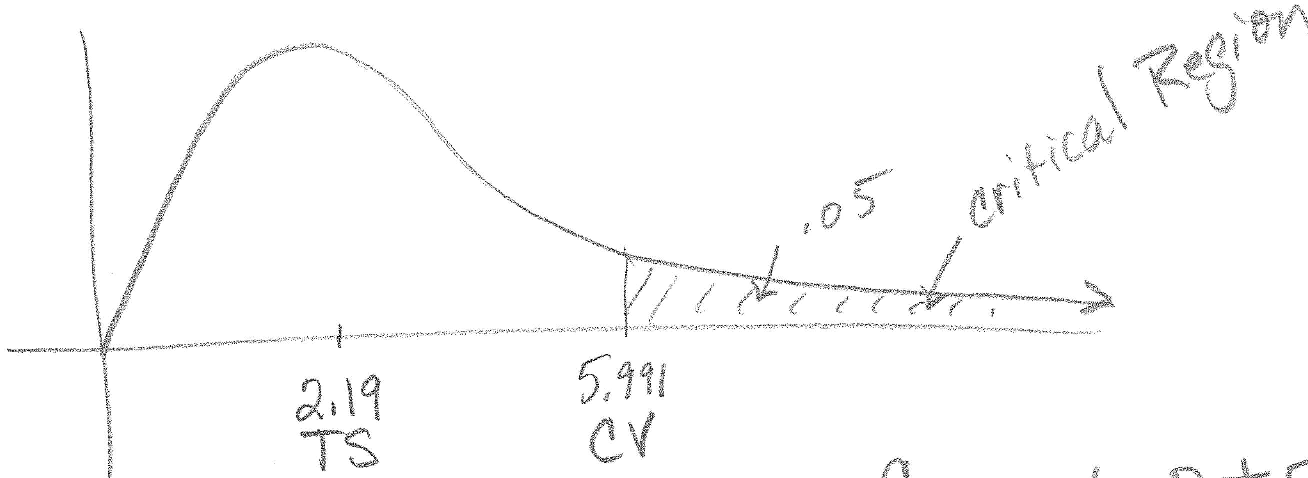
$$P(W \cup S) = P(W) + P(S) - P(W \cap S)$$

$$= \frac{300}{500} + \frac{150}{500} - .188$$

$$= .6 + .3 - .188 = \boxed{.712}$$

$$= \frac{356}{500} = .712$$

Determine Conclusion



fail to Reject $H_o: O=E$

$TS < CV$

or pvalue $> \alpha$

~~Confidence in Police is Independent of Gendar~~

TINSE to Say confidence in police is dependent on Gendar.

TINSE to Say that the Confidence in Police is different for Men & women.

fail to Support $O \neq E$

H_1: Confidence is dependent on gender

← Independence

H_1: Proportion with the Confidence in Police is -

Homogeneity

On Calculator

Put Table into a Matrix

Ex A researche claims that Obesity rates are dependent on the Number of Sugar drinks and juice consumed. (H_1)

	≤1	2	≥3
Not Obese	98	49	101
Obese	22	87	179

H_0: O = E Obesity rate is Independent of # of Sugar Drinks

H_1: O ≠ E Obesity rates are dependent on # of Sugar Drinks

To find TS: χ^2 put table in Matrix on TI

$\boxed{\text{Matrix}}$ ▷ ▷ $\boxed{\text{Edit}}$ 1: [A] = 2 X 3 [Enter Values]

Quit

$\boxed{\text{Stat}}$ $\boxed{\text{Tests}}$ χ^2 Test Calculate

TS: $\chi^2 = 77.928$ p-value $= 1.19 \times 10^{-17} \approx 0$

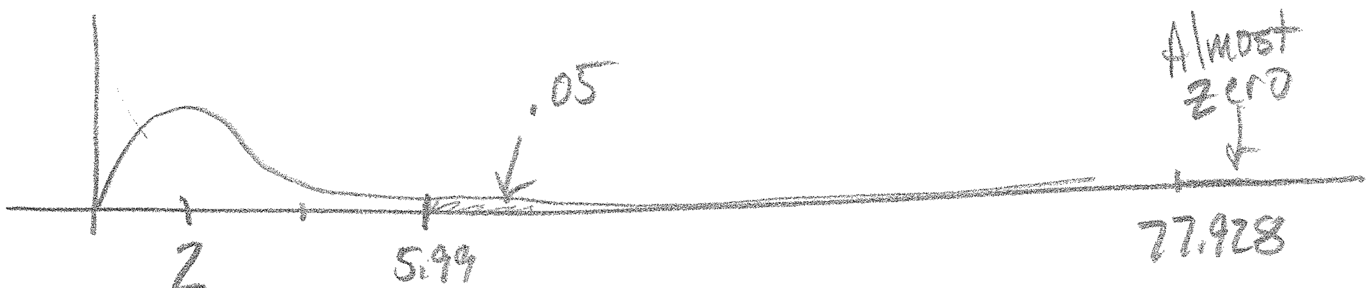
Matrix of expected $\boxed{\text{Matrix}}$ = [B] = $\begin{bmatrix} 55.5 & 62.9 & 129.5 \\ 64.5 & 73.1 & 150.4 \end{bmatrix}$

Obesity rate is Dependent on # of Sugar drinks consumed.

b) Graph and shade critical Region

$df = (\#row - 1)(\#col - 1) = (2-1)(3-1) = 2$

$\chi^2_R = 5.991$



Meaning of P-value

There is very little chance we would see Sample this large a difference in the proportion of obese kids if they came from populations with the same proportion of obese kids

7) Recent evidence suggest that obesity rates are dependent on the number of sugar or juice sweetend drinks that are consumed each day. A researcher wishes to test this claim and finds the results are shown below.

	< 1	2	> 3
Not obese	98	49	101
Obese	22	87	179

Use a 0.05 significance level to test the claim that the proportion of children who are obese is the same for all three categories.

a) State the null and alternate hypothosis.

b) Graph and shade the critical region.

c) Find the Matrix expected values, the Test Statistic, the Critical Value.

d) State the conclution of this hypothosis test.

Find proportion of obese kids in each category and overall.

$P(Ob | \le 1) = .183 = \frac{22}{120}$ Prop of obese given that they drink at most one glass a day

$\frac{87}{285} = P(2 | Ob)$

$P(Ob | 2) = .640 = \frac{87}{136}$

$P(Ob | \ge 3) = .640 = \frac{179}{280}$

$P(Ob) = \frac{288}{536} = .537$

Let Do Test This time with the Calculator

$\boxed{\text{Matrix}}$ row X Col = 2 X 3

$\boxed{\text{Edit}}$ Enter observed Values

$\boxed{\text{Stat}}$ $\boxed{\text{Tests}}$ C: χ^2-Test

TS: $\chi^2 = 77.928$

Calculator will find it $\boxed{\text{Edit}}$ Matrix: B

Note: Exp Value Matrix $= [B] = \begin{bmatrix} 55.5 & 129.5 & 62.9 \\ 64.5 & 150.4 & 73.1 \end{bmatrix}$

CV: $\chi^2 = 5.991$ Reject H_0

P-value $= 1.2 \times 10^{-17} \approx 0$

Use a χ^2 test to test the claim that in the given contingency table, the row variable and the column variable are independent.

8) Tests for adverse reactions to a new drug yielded the results given in the table. At the 0.05 significance level, test the claim that the treatment (drug or placebo) is independent of the reaction (whether or not headaches were experienced). Do this two ways first with a χ^2 test then with a two proportion z test. You should get the same answer.

	Drug	Placebo
Headaches	11	7
No headaches	73	91

a) State the null and alternate hypothosis.

b) Graph and shade the critical region.

c) Find the Matrix expected values, the Test Statistic, the Critical Value.

d) State the conclution of this hypothosis test.

Now do the problem again using a 2-PropZTest.

a) State the null and alternate hypothosis.

b) Find the critical value and then graph and shade the critical region.

c) Find the Test Statistic and p-value.

d) State the conclution of this hypothosis test.

§11.1 # 6, 8, 16

#6 $H_0: O = E$ The freq. that the Students select the 4 tires is the same

$H_1: O \neq E$ At least one tire is picked more often than the others

TS: $\chi^2 = 4.6$ CV: $\chi^2_R = 7.815$ p-value = .2035

fail to Reject H_0

~~the~~ TINSE to Reject that the 4 tires are selected at the Same rate.

IS it likely they will all randomly guess the Same tire? No, Very unlikely

$P(\text{All the Same}) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{64}$

$$= \frac{n(E)}{n(S)} = \frac{4}{4^4} = \frac{1}{64}$$

S = { LF RF RF LF,

E = { LF LF LF LF, RF ... , RR RR RR RR, RL }

§10.4

#~~19~~ c) (

#20 c) (\$2050.75, \$5419.53)

We are 95% confident that a 0.8 carat Dimond is worth between \$2050 and \$5420.

§11.2 Baseball Players Birthdays

#16

Obs L_1	Exp L_2
387	376.25
329	376.25
⋮	⋮
504	504K

way off

lots of August Birthdays

$H_0: O = E$

$H_1: O \neq E$

TS: χ^2 = use χ^2GOF-Test $\boxed{\text{Test}}$

or XZGOF Prog

$$\boxed{TS: \chi^2 = 93.072}$$

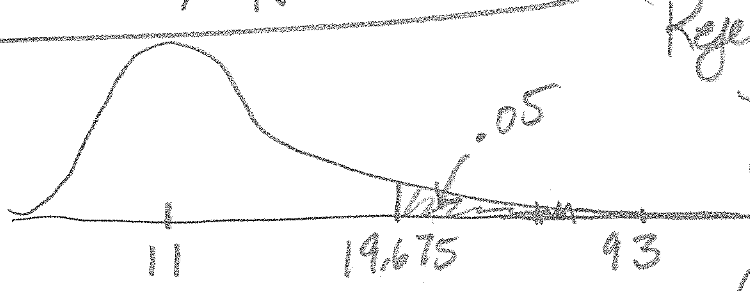
CV: χ^2_R = InvX2

df = K-1 = 11

Right Tail

Sig. level = .05 = α

$$\boxed{CV: \chi^2_R = 19.675}$$



Reject H_0 → TS > CV

We can Reject that the birth frequencies are the same in every month

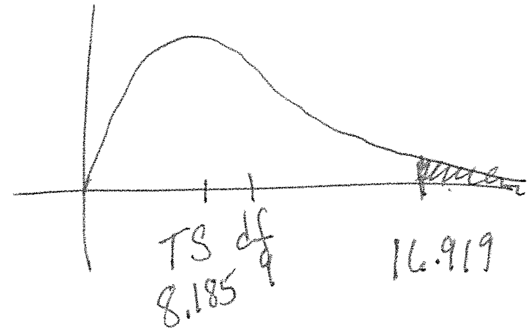
We support Gladwell's Argument that more are born after July.

11.2 #7 TS: $\chi^2 = 8.185$

Sign. Level

CV: $\chi^2 = \text{inv}\chi^2(df=9, \alpha=.05, \text{Rt tail})$

CV: $\chi^2 = 16.919$



fail to Reject $H_0: O=E$

We can not Reject that the Slot Machine is working as expected.

#16 $L_1 = 387, 329, \ldots$

$L_2 = 376.25, 376.25$

$$E = \frac{\Sigma \text{obs}}{K} = \frac{\Sigma \text{obs}}{12}$$

$df = K-1 = 12-1 = 11$

TS: $\chi^2 = 93.07$

p-value ≈ 0 $\Rightarrow$ Reject $H_0: O=E$ and Support $H_1: O \neq E$

Some of the observed # of players per Math are Not the Same as the others.

Boys with August Birthdays are More likely to Make Big league.

#10 obs # injuries 23 23 21 21 19 = L1

Exp # 21.4 21.4 21.4 21.4 21.4 = L2

$$Exp = \frac{23+23+21+21+19}{5} = \frac{\Sigma obs}{\# categories}$$

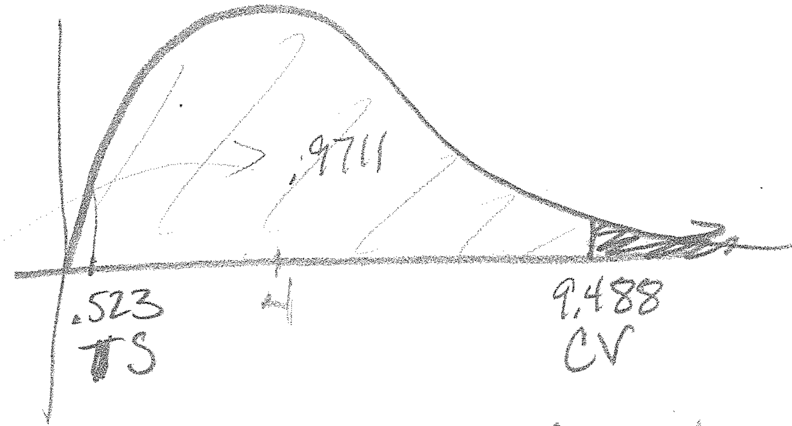
$= 21.4$

df = 4

GOF Test

TS $\chi^2 = .523$

p-value = .9711



$H_0: O = E$

$H_1: O \neq E$

there is Not a Significant difference in the prop of injuries on any particular day of the week.

#8	W	H	B	A	N	
prop in pop	.756	.091	.108	.038	.007	
L_1 = Obs	3855	60	316	54	12	$\Sigma O = \underline{4297}$
L_2 = Exp	3248.5	391.03	464.08	163.29	30.1	

$E = \text{prop} \cdot \Sigma O$
$= .756 \cdot 4297$

$\alpha = .01$
$df = 4$

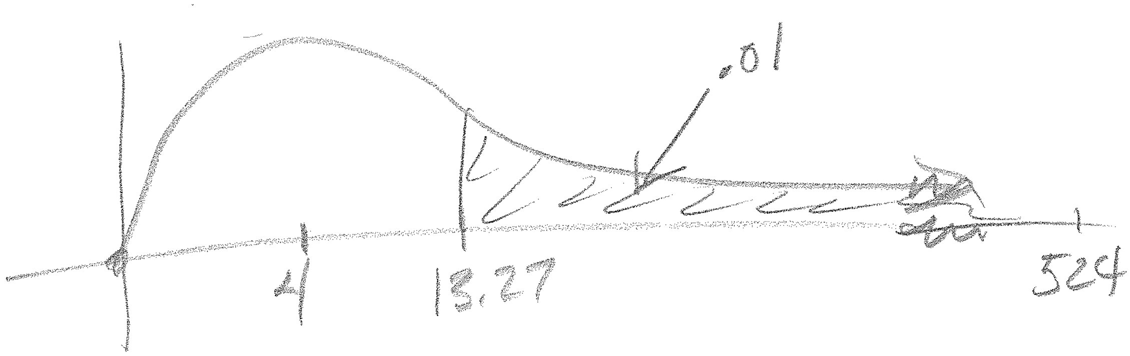
TS: $\chi^2 = 524.75$

P-value $= 0$

CV: $\chi^2 = 13.27$

Initial $O \neq E$

There is almost no chance that we would see a sample with this much variation in participation rates, if the true prop of participation were equal.



#18 $\chi^2 = .976$

p-value $= .91339$

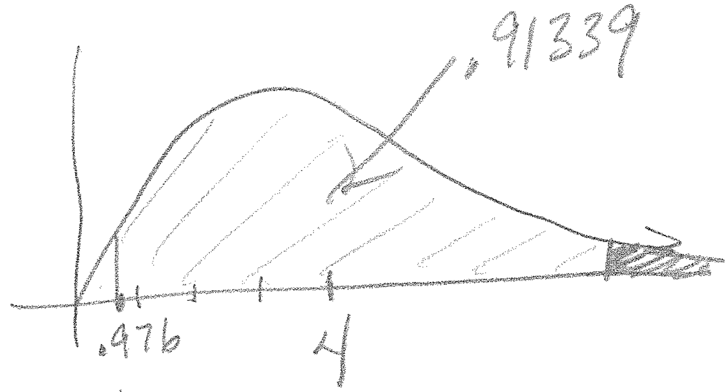
$H_0: O = E$

$H_1: O \neq E$

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

df = # categories − 1

= 5 − 1 = 4



We can't Reject O=E | O≠E

Final conclusion

TINSE to show that hits don't fit poisson.

In fact the hits are amazing close to # of hits predicted by poisson

#4 Find Meaning of P-value = .477

There is about an 48% chance of seeing a sample with this much variation in # of catered weddings or more given that the population of catered wedding has NO variation in frequency.

MML #5

	R	O	Y	B	B	G
M&M %	.14	.21	.14	.12	.23	.16
Obs	11	23	9	10	27	20
Exp	.14	21	14	12	23	16

TS: $\chi^2 = 4.648$ Fail to Reject H_0

P-Value = .46 > .05

We can't Reject that the given proportions are correct.

Count # of each color in the Lists given

Plan		Cover	Due
~~Mon~~	~~12/4~~	~~11.1~~	~~Project Rough Draft~~
Wed	12/6	11.2	Project Due, 11.1
Mon	12/11	12.1	11.2
Wed	12/13	Review	PF, 12.1
Wed.	12/20	Final	7-10 AM