7.0 A

CONTINGENCY TABLES

ONE-DIMENSIONAL

V	W	Z
n_V	n_W	n_Z

$n = n_V + n_W + n_Z$

n_i ARE OBSERVED COUNTS

IN THIS CASE, WE WANT TO TEST WHETHER V, W, Z FOLLOW A THEORETICAL DISTRIBUTION OR NOT.

$$\begin{cases} H_0: & n_i = n_{i\,GIVEN} \\ H_1: & n_i \neq n_{i\,GIVEN} \end{cases}$$

OR EQUIV.

$$\begin{cases} H_0: & p_i = p_{i\,GIVEN} \\ H_1: & p_i \neq p_{i\,GIVEN} \end{cases}$$

WHERE $p_i = \frac{n_i}{n}$

TWO DIMENSIONAL

	V	W	Z	
A	n_{AV}	n_{AW}	n_{AZ}	$r_A = n_{AV} + n_{AW} + n_{AZ}$
B	n_{BV}	n_{BW}	n_{BZ}	$r_B = n_{BV} + n_{BW} + n_{BZ}$
	$c_V = n_{AV} + n_{BV}$	$c_W = n_{AW} + n_{BW}$	$c_Z = n_{AZ} + n_{BZ}$	$n = r_A + r_B = c_V + c_W + c_Z$

n_{ij} ARE OBSERVED COUNTS $\quad p_{ij} = \frac{n_{ij}}{n}$ ARE OBSERVED PROB.

IN THIS CASE, WE WANT TO TEST WHETHER n_{ij} ARE DISTRIBUTED FOLLOWING SIMPLY THE MARGINAL VALUES,

$n_{ij} = \frac{r_i \cdot c_j}{n}$ OR NOT.

7.0 B

FOR ONE-DIMENSIONAL, STATISTICS IS

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - n_{i\,given})^2}{n_{i\,given}}$$

D.F. = NUMBER OF COLUMNS − 1 = K − 1

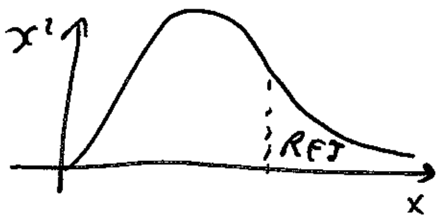
FOR TWO-DIMENSIONAL, STATISTICS IS

$$\chi^2 = \sum_{\substack{i=1\\j=1}}^{H,K} \frac{\left(n_{ij} - \frac{r_i c_j}{n}\right)^2}{\frac{r_i c_j}{n}}$$

D.F. = (NUMBER OF COL − 1)(NUM. ROWS − 1) = (K − 1)(H − 1)

$\chi^2 = 0$ IS PERFECT VALUE FOR H_0

SO REJECTION AREA IS



WARNING: THEORETICAL VALUES INSIDE EACH CELL, THAT IS $n_{i\,given}$ OR $\frac{r_i \cdot c_j}{n}$ MUST ALL BE LARGER OR EQUAL TO 5

7.1

TAKE 150 BUYERS AND ASK THEM WHICH PRODUCT DO THEY PREFER.

ANSWERS ARE A: 61 B: 53 C: 36

DO BUYERS HAVE A PREFERENCE?

A	B	C
61	53	36

THE PERFECTLY EQUAL PREFERENCE TABLE WOULD HAVE BEEN

A	B	C
50	50	50

$$\begin{cases} H_0: n_A = 50 \text{ AND } n_B = 50 \text{ AND } n_C = 50 \\ H_1: n_A \neq 50 \text{ OR } n_B \neq 50 \text{ OR } n_C \neq 50 \end{cases}$$

OUR STATISTICS $= \sum_{i=1}^{3} \frac{(n_i - n_{i\,given})^2}{n_{i\,given}} = \frac{(n_A - 50)^2}{50} + \frac{(n_B - 50)^2}{50} +$

$+ \frac{(n_C - 50)^2}{50} = \frac{(61 - 50)^2}{50} + \frac{(53-50)^2}{50} + \frac{(36-50)^2}{50} = \frac{11^2 + 9 + 14^2}{50} =$

$= \frac{326}{50} = 6{,}52$

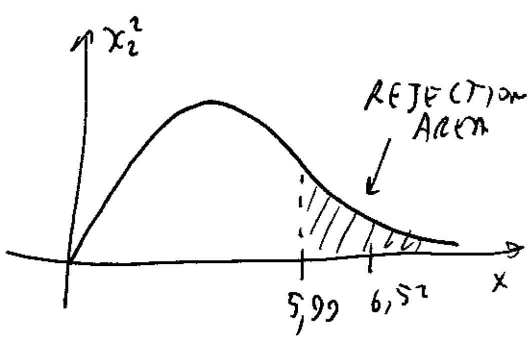
$\chi^2_2 = 6{,}52$

2 DEGREES OF FREEDOM

WITH $\alpha = 0{,}05$ χ^2_2 CRITICAL VALUE $= 5{,}99$

$\Downarrow$

WE REJECT H_0



WITH $\alpha = 0{,}01$ χ^2_2 CRITICAL VALUE $= 9{,}21$

$\Downarrow$

WE DO NOT REJECT H_0

SINCE $n_{i\,given} \geq 5 \ \forall i$

THE TEST'S ASSUMPTION IS SATISFIED □

7.2 A MULTINOMIAL EXPERIMENT WITH K = 3 AND n = 310 PRODUCED THESE DATA

$n_1 = 62 \quad n_2 = 79 \quad n_3 = 169$

HW

TEST WITH $\alpha = 0{,}05$ THE HYPOTHESIS THAT $p_1 = p_2 = 0{,}25$, $p_3 = 0{,}5$

$H_0: \; p_1 = 0{,}25 \quad p_2 = 0{,}25 \quad p_3 = 0{,}5$

2 DEGREES OF FREEDOM = (K - 1)

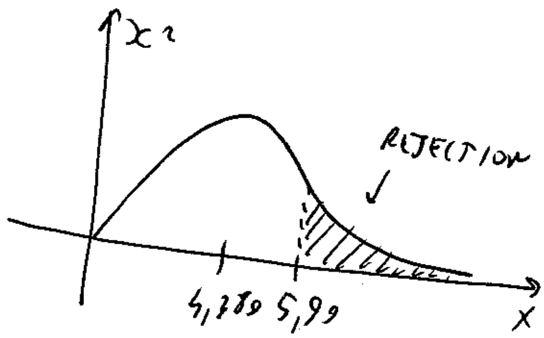
$$\chi^2_2 = \frac{(62 - 310 \cdot 0{,}25)^2}{310 \cdot 0{,}25} + \frac{(79 - 310 \cdot 0{,}25)^2}{310 \cdot 0{,}25} + \frac{(169 - 310 \cdot 0{,}5)^2}{310 \cdot 0{,}5} =$$

$$= \frac{240{,}25}{77{,}5} + \frac{2{,}25}{77{,}5} + \frac{196}{155} \approx 3{,}1 + 0{,}029 + 1{,}26 = 4{,}389$$

χ^2_2 CRITICAL VALUE FOR 0,05 IS 5,99

$\Downarrow$

WE DO NOT REJECT H_0



FROM ANOTHER POINT OF VIEW, THE AREA $\int_{4{,}389}^{+\infty} \chi^2_2(s)\,ds =$

$\approx 0{,}12$

$\Downarrow$

IF H_0 IS TRUE, WE HAVE A 12% PROBABILITY OF RANDOMLY GETTING SUCH A BAD RESULT.

$310 \cdot 0{,}25 \geq 5$ AND $310 \cdot 0{,}5 \geq 5 \Rightarrow$ THE TEST'S ASSUMPTION IS SATISF.

D

7.3 A 4 CAR MANUFACTORS SELL CARS DIVIDED IN 3 CATEGORIES (LARGE, MEDIUM, SMALL)

TEST THE HYPOTHESIS THAT CAR SIZE AND MANUFACTOR ARE INDEPENDENT, WHEN 1000 RANDOM BUYERS ANSWERS IN THIS WAY.

	FIAT	MERCEDES	VW	CITROEN
SMALL	157	65	181	10
INTERMEDIATE	126	82	141	46
LARGE	58	45	60	28

WE EVALUATE MARGINAL NUMBERS

$c_F = 341$ $c_M = 192$ $c_V = 382$ $c_C = 84$

$n_S = 413$ $n_I = 395$ $n_L = 191$ $n = 999$

WE COMPUTE THE EXPECTED NUMBERS AS

$$\hat{E}N_{ij} = \frac{n_i \cdot c_j}{n}$$

	F	M	VW	C	
SM	140,933	79,796	158,179	34,632	413
IN	135,036	76,032	151,668	33,464	395
LA	65,131	36,672	73,153	16,044	191
	341	192	382	84	999

H_0: THE OBSERVED TABLE DIFFERS FROM THE EXPECTED ONE ONLY DUE TO RANDOM ERRORS

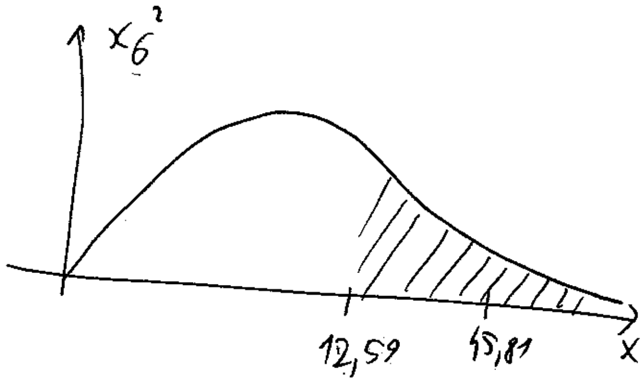
7.3 B

$$\chi^2_{11} = \sum_{i,j} \frac{(n_{ij} - \hat{E} N_{ij})^2}{\hat{E} N_{ij}} =$$

$$= \frac{(157 - 140,833)^2}{140,833} + \frac{(65 - 79,296)^2}{79,296} + \dots\dots + \frac{(28 - 16,044)^2}{16,044} = 45,81$$

χ^2_6 CRITICAL VALUE WITH $\alpha = 0,05$ IS 12,59

D.F. $= (R-1) \cdot (C-1) = 3 \cdot 2 = 6$



WITH $\alpha = 0,05$ H_0 IS REJECTED

<u>REMEMBER</u> TO CHECK THAT EVERY $\hat{E} N_{ij} \geq 5$

D

(7-4) IS IT TRUE THAT THE NUMBER OF CAR CRASHES IS INDEPENDENT FROM DRIVER'S AGE?

HW: A SURVEY OF 4194 DRIVERS PRODUCED:

	21-30	31-50	51 OR MORE
0	748	1607	1392
1-2	105	158	147
MORE 2	9	16	12

WE CALCULATE MARGINALS

$c_{YOUNG} = 862 \qquad c_{MID.} = 1781 \qquad c_{OLD} = 1551$

$r_{NO} = 3747 \qquad r_{FEW} = 410 \qquad r_{MANY} = 37$

$n = 4194$

	Y	M	O	
0	770,13	1591,18	1385,69	3747
1-2	84,27	174,11	151,62	410
>2	7,60	15,71	13,68	37
	862	1781	1551	4194

$\hat{E} N_{ij} \geq 5$ $\boxed{OK}$

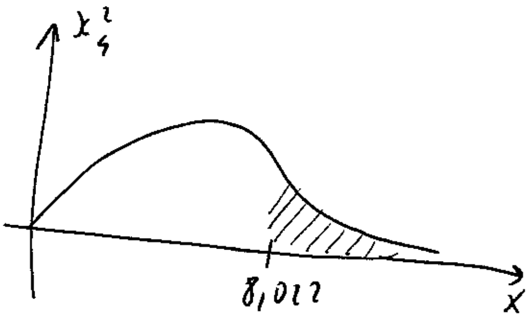
$$\chi^2_4 = \frac{(748 - 770{,}13)^2}{770{,}13} + \frac{(1607 - 1591{,}18)^2}{1591{,}18} + \dots + \frac{(12 - 13{,}68)^2}{13{,}68} =$$

$= 8{,}022$

THE AREA AFTER 8,022 IS 8%

WE ACCEPT INDEPENDENCE WITH $\alpha = 0{,}05$

WE REJECT WITH $\alpha = 0{,}1$



□

7.5

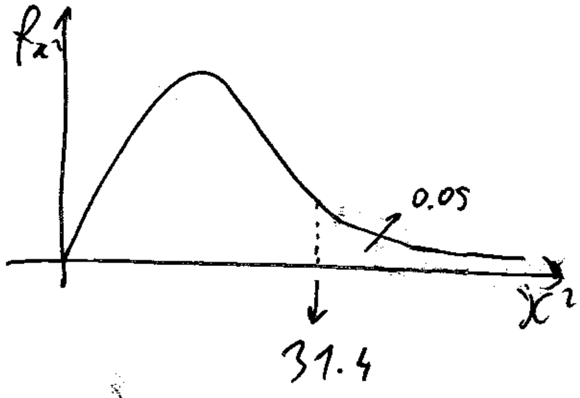
FIND A REJECTION REGION FOR A TEST OF INDEPENDENCE OF TWO CLASSIFICATIONS WITH 5 ROWS AND 6 COLUMNS WITH $\alpha = 0.05$.

χ^2 IS THE STATISTICS

$$\chi^2 = \sum_{\substack{i=1\to 5\\ j=1\to 6}} \frac{\left(n_{ij} - \frac{n_i n_j}{n}\right)^2}{\frac{n_i n_j}{n}}$$

WHERE $n_i = \sum_{j=1\to 6} n_{ij}$

$n_j = \sum_{i=1\to 5} n_{ij}$



D.F. $= 4 \cdot 5 = 20$

REJECTION FOR $\chi^2 > 31.4$

$\square$

7.6

A MULTINOMIAL EXPERIMENT WITH $n = 310$ PRODUCES

$n_A = 71$ $n_B = 57$ $n_3 = 182$

$\alpha = 0.1$ DO THESE DATA CONTRADICT HYPOTHESIS THAT $p_A = p_B = 0.2$ $p_C = 0.6$?

A	B	C
71	57	182

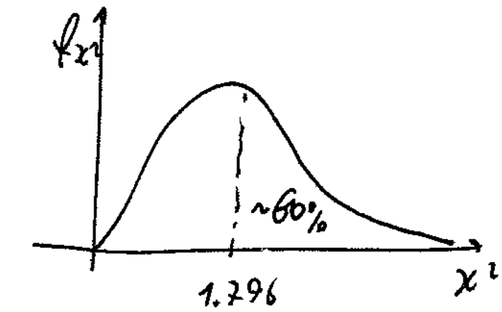
$\hat{n}_A = p_A \cdot n = 0.2 \cdot 310 = 62$

$\hat{n}_B = 62$

$\hat{n}_C = p_C \cdot n = 0.6 \cdot 310 = 186$

$$\chi^2 = \frac{(71-62)^2}{62} + \frac{(57-62)^2}{62} + \frac{(182-186)^2}{186} = \frac{9^2}{62} + \frac{5^2}{62} + \frac{4^2}{186} =$$

$$= \frac{81+25}{62} + \frac{16}{186} = 1.71 + 0.086 = 1.796$$



$k = 3$ D.F. $= 2$

FOR $\alpha = 0.1$ CRITICAL VALUE $= 4.61$

$\Downarrow$

WE DO NOT REJECT H_0 ($p_A = p_B = 0.2$ $p_C = 0.6$)

IT IS NOT ENOUGH TO CONTRADICT IT !

□