R.R. - We reject H_0 if $X^2 > X^2_{\alpha}(k-1)$

Assumptions:

① A random sample taken from

② multinomial distr.

③ n is lrg enough for X^2 to apply (since X^2 is very sensitive to departures from normality we need $n > 30$ so that C.L.T. applies; i.e we have approx. normality)

(Ex1) Handout - Heating Gas Payments

	Accnts Fully paid	1 mnth behind	2 mnth behind	>2 mnths behind	Total
Observed	287	49	30	34	400
Expected np_{io}	$np_{10}=400(.8)$ $=320$	$np_{20}=400(.1)$ $=40$	$np_{30}=400(.06)$ $=24$	$np_{40}=400(.04)$ $=16$	400

H_0: $p_1=.8$, $p_2=.1$, $p_3=.06$, $p_4=.04$; $\alpha=.01$

H_a: @ least one $p_i \neq p_{io}$

Assume: ① random sample ✓ (given)

② from multinomial dist.

③ n lrg ($n=400$) for X^2 to apply

Test stat. $X^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} = \frac{(287-320)^2}{320} + \frac{(49-40)^2}{40} + \frac{(30-24)^2}{24} + \frac{(34-16)^2}{16} = \underline{27.178}$

R.R. We reject H_0 if $X^2 > X^2_{\alpha}(k-1) = X^2_{.01}(3) = 11.34$

Since $X^2 = 27.178 > 11.34$, we reject H_0 and conclude that @ 1% significance level there is an evidence that this winters pattern of pay.'s differs from the historical norm.

end of email

3/29 - don't have to do follow (only w/ experimental design)

(Ex2) A researcher designs an experiment in which a rat is attracted to the end of a ramp that divides, leading to drs of 3 different colors. The rat is sent down the ramp 90 times and its choice of dr colour is observed. The results are, the green dr chosen 20 times, red door chosen 39 times, + blue door 31 times. Do this data provide evidence to suggest that the rat prefers one door color over another $\alpha = .05$

	Door Colour			Tot.
	Green	Red	Blue	
observed count O_i	20	39	31	90
expected count $E_i = np_{i0} = 90(1/3) = 30$	30	30	30	90

$H_0: p_1 = p_2 = p_3 = 1/3$ (each door colour is equally likely to be selected)

H_a: @ least one $p \neq 1/3$; $\alpha = .05$

Assume: ① ✓ sample (here even though we have only rat only, the rat is sent down the ramp and is randomly selecting a dr color, 90 times. Assuming no incentive
② multinomial distr
③ n is lrg for X^2 to apply $(n = 90)$ ✓

test stat: $X^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = \frac{(20-30)^2}{30} + \frac{(39-30)^2}{30} + \frac{(31-30)^2}{30}$
$= 6.067$

R.R. We reject H_0 if $X^2 > X^2_{\alpha(K-1)} = X^2_{.05,2} = 5.99$

Since $X^2 = 6.067 > 5.99$ we reject H_0 and conclude that @ 5% level of significance, There is an evidence that the rat prefers one door colour over another.

(III) <u>r x c contingency tables</u> (eg 3)

- these are tests for relationships btwn 2 categorical variables
- Each element in the population is classified (grouped) according to 2 criteria of interest
  e.g.(1) emp. is classified by <u>education level</u> (HS, BA, MSO, PhD) and <u>job performance</u> (high, average, low)

  e.g.(2) people are classified by <u>inc.</u> + <u>political leaning</u>

  e.g.(3) people classified by <u>inc.</u> + ~~nationality~~

  ect.

- In general we have 2 classifications i.e. 2 categorical variables of interest one w/ r possible categories + the other one w/ c possible categories

<u>Var. 1</u> <u>Var. 2</u>

	1	2	...	j	...	c	Tot
1	O_{11}	O_{12}	...	O_{1j}		O_{1c}	R_1
2	O_{21}	O_{22}	...	O_{2j}		r	R_2
⋮	⋮	⋮	...	⋮		⋮	⋮
i	O_{i1}	O_{i2}	...	O_{ij}		O_{ic}	R_i
⋮	⋮	⋮					
r	O_{r1}	O_{r2}		O_{rj}		O_{rc}	R_r
Tots	C_1	C_2		C_j		C_c	n

where O_{ij} is the observed count of the cell (ij) i.e in the cell belonging to the category of the 1st variable + cat.) of the 2nd variable

$R_i = \sum_{j=1}^{c} O_{ij}$, $C_j = \sum_{i=1}^{r} O_{ij}$, i = counts # of rows
j = counts # of columns

R_i: Tot. # of observ's in row i

C_j: Tot. # of observ's in col. j

$$\sum_{i=1}^{r} R_i = \sum_{j=1}^{c} C_j = \sum_{i=1}^{r} \sum_{j=1}^{c} O_{ij}$$

We use r x c contengency tables for 2 different tests
① Test of independence
② Test of homogeneity

Case A) Test of independence

- A single random sample of size n is obtained from a single population + then classified by 2 characteristics of interest
- We are interested in whether the 2 characteristics are related or not

i.e. H_0: 2 characteristics are indep. ; α
H_a: they are dependent

test stat. $X^2_{(r-1)(c-1)} = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$ — Estimated Expected

where $\hat{E}_{ij}$ is the estimated expected count, $\hat{E}_{ij} = \frac{R_i C_j}{n}$

RR - We reject H_0 if $X^2_{(r-1)(c-1)} > X^2_{\alpha;(r-1)(c-1)}$

Assume: ① a random sample from
② population classified by 2 characteristics
③ n is lrg enough for X^2 to apply

Ex) A certain cola co. sells 4 types of cola thruout North America. To help determine if the same marketing apprch used in the US can be used in Canada + Mexico one of the firms mrkting analyst wants to ascertain if there is an association btwn the

type of cola preferred + the nationality of the consumer. A random sample of 250 cola drinkers from the 3 countries was interviewed and then classified according to the type of cola preferred + nationality. Is there evidence of an association btwn cola preference and nationality?
Use $\alpha = .01$

Cola Pref.

Nationality	A	B	C	D	Tot
N_1	72 (48.3)	8 (12.88)	12 (19.32)	23 (34.5)	115
N_2	26 (35.7)	10 (9.52)	16 (14.28)	33 (25.5)	85
N_3	7 (21)	10 (5.6)	14 (8.4)	19 (15)	50
Tot.	105	28	42	75	n=250

where $\hat{E}_{11} = \frac{R_1 C_1}{n} = \frac{115 \times 105}{250} = 48.3$

$\hat{E}_{12} = \frac{R_1 C_2}{n} = \frac{115 \times 28}{250} = 12.88$, ect.

Assume
1. random sample ✓
2. from pop. w/ 2 characteristics ✓
3. n=250 (lrg for X^2 to apply)

H_0: cola preference and nationality are independent
H_a: " " " " are related ; $\alpha = .01$

test stat: $X^2_{(r-1)(c+1)} = \sum_{i=1}^{3}\sum_{j=1}^{4}\frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} =$

$= \frac{(72-48.3)^2}{48.3} + \frac{(8-12.88)^2}{12.88} + \dots + \frac{(19-15)^2}{15}$

$= 42.75$

RR - We reject H_0 if $X^2(r-1)(c-1) > X^2_{\alpha}(r-1)(c-1) =$ 2×3
$X^2_{.01;6} = 16.8119$

Since $X^2_{(r-1)(c-1)} = 42.75 > 16.8119$ we reject H_0 and conclude that @ 1% level of significance there is evidence that the cola preference is related to the nationality

(Case B) X^2 - test of Homogeneity

- several random samples taken from several pop.'s w/ multinomial distr.'s
- we are interested in whether category proportions are the same across all pop.'s or not

i.e. - H_0: category proportions are the same for all pop.'s
H_a: they differ ; $\alpha =$

Test stat: $X^2(r-1)(c-1) = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$

RR - We reject H_0 if $X^2(r-1)(c-1) > X^2_{\alpha;(r-1)(c-1)}$

Assume: ① m random samples that are taken from
② m multinomial pop.'s
③ $n_1, n_2 \ldots, n_m$ are lrg enough for X^2 to apply (>30)

(Ex) A survey of voter sentiment was conducted in 3 city political wards to compare the proportions of voters favouring the 3 candidates A, B, + C. Independent random samples of 200 voters were polled in each of the 3 wards w/ results shown below. Do the data provide sufficient evident to indicate that the proportions of voters favouring A, B, + C differ in 3 wards. Use $\alpha = .05$

Ward

Canidate	1	2	3	Tot
A	108 (102.33)	87 (102.33)	112 (102.33)	307
B	52 (47.33)	51 (47.33)	39 (47.33)	142
C	40 (50.33)	62 (50.33)	49 (50.33)	151
Tot.	200	200	200	600

$\hat{E}_{1j} = \frac{R_1 C_1}{n} = \frac{307 \times 200}{600} = \hat{E}_{12} = \hat{E}_{13}$

$\hat{E}_{21} = \frac{R_2 C_1}{n} = \hat{E}_{22} = \hat{E}_{23}$ ect

H_0: the proportions of voters favoring canidates A, B, + C are the same for 3 wards

H_a: proportions differ ; $\alpha = .05$

Assume ① 3 random samples from
② 3 pop.'s that are multinomial pop.'s
③ Each sample size must be lrg enough for X^2 to apply
$n_1 = n_2 = n_3 = 200$

test stat. $X^2_{(r-1)(c-1)} = \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$

$= \frac{(108 - 102.33)^2}{102.33} + \ldots + \frac{(49 - 50.33)^2}{50.33}$

$= 10.597$

R.R - We reject H_0 if $X^2_{(r-1)(c-1)} > X^2_{\alpha;(r-1)(c-1)} = X^2_{.05;4}$ (2×2)

$= 9.4877$

Since X^2 ~~calc~~ $10.597 > 9.4877$ we reject H_0 and conclude that @ 5% level of significance the proportions of voters favoring canidates A, B, C differ across the 3 wards

* Exam - covers all material w/ emphasis on last part (after test 2)
  - show all work for partial credit
  - All form. shts provided
  - Final mark based on term + final or final only, which ever higher
  - min 40% to pass
  - not on Exam.
    - Variable selection (4) techniques
      max r^2, min. r^2, stepwise regression
    - multicolinearity (relationship btwn x's)
  - 20 m/c questions
  -