1. Experimenters A, B C independently perform experiments. Their respective data result in chi square statistics and df below.

	H0	df	chi square statistic	P-value
experiment A:	H0a	12	16.3	
experiment B:	H0b	6	5.2	
experiment C:	H0c	31	39.2	
		df 49	60.7	

χ^2 COMBINED

$df = E\,\chi^2$ WHEN H_0 IS TRUE

Give the P-value for each experiment.

A: χ^2 TABLE

df		P-VALUE	.015
12		16.3	24.96

P-VALUE $P(\chi^2 \geq 16.3)$ IF H_0 IS TRUE

REJECTING H_0 IS REAL GOAL

< .015 IS NOT SO RARE COULD EASILY BE CHANCE IF H_0 IS TRUE

Give the P-value for experiments A, B, C combined.

df 49   60.7 χ^2 STAT   OFF TABLE

df 30   .0008 RARE   60.7

DOES NOT APPLY SINCE df = 49

Express H0 for the combined experiment in terms of H0a, H0b, H0c.

H_0: ALL OF H_{0a} H_{0b} H_{0c} ARE CORRECT.

2. From a large population of farms we've selected at random 50 small organic farms, 30 large organic farms, 60 small not-organic farms and 100 large not-organic farms. These farms are scored L, M, H for water usage relative to the crops they produce. We wish to use a chi square test of the hypothesis that water score is (in the population) the same for each of the four farm types.

	50	30	60	100
	small organic	**large organic**	**small not organic**	**large not organic**
L				
M				
H				

2. **Continued.**

FIXED

	50	30	60	100 **(marginal counts)**	
	small organic	**large organic**	**small not organic**	**large not organic**	TOTAL 240 FARMS
78 **L**			37 obs		
M					
H					

NO — DON'T HAVE exp COUNTS

WOULD HAVE APPLIED IF WE'D SAMPLED 240 FARMS AT RANDOM

a. Is this a test of goodness of fit, homogeneity, or independence?

NOT OVERALL RANDOM SAMPLE REQUIRES HOMOGENEITY METHOD

b. What will be the df? $(R-1)(C-1)$

$= (3-1)(4-1) = 6$ df

There are 78 farms scoring L, of which 37 are not-organic.

SMALL

c. What is the "expected count" for non-organic small farms scoring L?

78 JNO

60 $\overline{)240}$

$exp = \frac{78 \cdot 60}{240}$

d. Give the contribution of cell "not organic small farms scoring L" to the chi square statistic.

$\chi^2 = \sum \frac{(obs - exp)^2}{exp} = + \cdot + \frac{\left(37 - \frac{78 \cdot 60}{240}\right)^2}{78 \cdot 60/240} + \ldots$

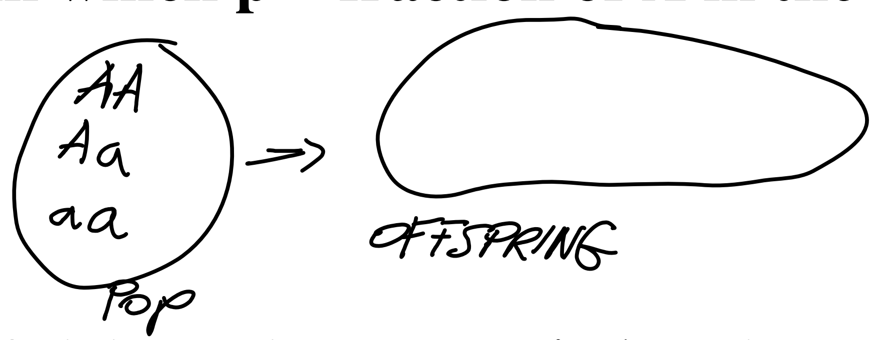
3. Mendel's model essentially says that each parent contributes a random selection of one "letter" to their offspring. So a type AA parent mated with a type aa parent will produce only type Aa offspring. If both parents are type Aa their offspring are 1/4 AA, 1/2 Aa or aA, 1/4 aa.

Pop AA, mom aa ⇒ You Aa

Pop Aa, mom Aa: aa 1/4, Aa 1/2, AA 1/4

Hardy and Weinberg proved that under random mating Mendel's model has the consequence that the generation of offspring is governed by the following model, in which p = fraction of A in the parent population's gene pool.

	AA	**Aa or aA**	**aa**
	p^2	$2pq$	q^2
eg $p = .6$	.36	.48	.16



a. A sample of 50 offspring finds 22 AA, 17 Aa or aa. Find pHAT and

	AA	Aa	aa
obs	22	17	11
exp	$50 \cdot .61^2$	$50(2) .61 .39$	$.39^2 \, 50$

50 TOTAL

$$\hat{p} = \frac{22(2) + 17(1)}{100} = .61$$

EST'd ONE THING FROM DATA

3. Continued.

Hardy and Weinberg proved that under random mating Mendel's model has the consequence that the generation of offspring is governed by the following model, in which p = fraction of A in the parent population's gene pool.

AA	Aa or aA	aa
p^2	$2\,p\,q$	q^2

$1 = (p+q)^2 = p^2 + 2pq + q^2$

a. A sample of 50 offspring finds 22 AA, 17 Aa or aa. Find pHAT and

obs	22	17	11	TOTAL 50	$\hat{p} = .61$
exp	$50 \cdot .61^2$	$50(2).61(.39)$	$50 \cdot .39^2$	TOTAL 50	$\hat{q} = .39$

b. Chi square statistic? $\left(\frac{22 - 50 \cdot .61^2}{50 \cdot .61^2}\right)^2 + \cdots +$

c. Applicable df? !! $(\#\text{CELLS} - 1) - 1$ ← PENALTY FOR EST'G p FROM DATA

$= (3-1) - 1 = 1$

d. P-value and conclusion?

4. A sample of 398 patients is selected from a large patient population. The following table classifies these patients according to insurance statusand costs of service billed. Does it appear that insurance status has something to do with cost?

	L	M	H	
insured	**68**	**136**	**80**	**284**
not insured	**42**	**72**	**0**	**114**
	110	**208**	**80**	**398**

BURED RANDOM SAMPLE OF 398 INTO TABLE.

APPLICABLE TEST IS INDEPENDENCE

a. Which chi square test applies?

b. Calculate the expected table.

	L	M	H	
insured	**78.49**	**148.42**	**57.09**	**284**
not insured	**31.51**	**59.58**	**22.91**	**114**
	110	**208**	**80**	**398**

exp $\frac{284 \cdot 80}{398}$

c. Applicable df = $(R-1)(C-1) = (2-1)(3-1) = 2$ df

χ^2 STAT ~ (NOT REAL CLOSE) df IF H_0 IS TRUE

SMALLER THAN .0001 OFF TABLE

d. If chi square = 40.6391 what is P-value? BIG

df		
2	40.63	.0001 18.42

5. Past experience suggests that restaurant menu orders occur with relative frequencies reported below. Also shown are counts from a sample of 200 orders from a new franchise in a new location.

	item 1	item 2	item 3	item 4	item 5	
	.14	**.28**	**.02**	**.53**	.03	TOT=1
obs	**33**	**74**	**6**	99	**12**	TOT 200
exp	28	56	4 LOW	106	6	TOT 200

MODEL SPECIFIED COMPLETELY

,03 (200)

a. Which chi square method applies? GOODNESS OF FIT   df = (5-1) = 4

b. Fill in the missing counts, exp counts and marginal totals.

c. Do any cells have to be merged? ITEM 3 + ITEM 5

d. Calculate the applicable chi square statistic.

$$\chi^2 = \frac{(33-28)^2}{28} + \cdots +$$

	1	2	3	4
obs	33	74	18	99
exp	28	56	10	106

e. Determine P-value.