Review 4

A. ANOVA for testing the equality of k population means.

$$f=\frac{SSB/_{k-1}}{SSW/_{n_T-k}}=\frac{\sum_{j=1}^{k}n_j(\bar{x}_j-\bar{\bar{x}})^2/_{k-1}}{\sum_{j=1}^{k}(n_j-1)s_j^2/_{n_T-k}}=\frac{MSB}{MSW}$$

ANOVA Table:

Source	Sum of Squares	Degree of Freedom	Mean Square	F
Between	SSB	$k-1$	$MSB = SSB/_{k-1}$	$\frac{MSB}{MSW}$
Within	SSW	n_T-k	$MSW = SSW/_{n_T-k}$	
Total	SST	n_T-1		

B. χ^2 test for:

1. proportions of a multinomial population:

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-e_i)^2}{e_i}$$

2. the independence of two variables (contingency table):

$$\chi^2=\sum_{i=1}^{p}\sum_{j=1}^{m}\frac{(f_{ij}-e_{ij})^2}{e_{ij}}$$

Example 1:

The following data are from 4 different populations.

Sample 1	8.2	8.7	9.4	9.2		
Sample 2	7.7	8.4	8.6	8.1	8.0	
Sample 3	6.9	5.8	7.2	6.8	7.4	6.1
Sample 4	6.8	7.3	6.3	6.9	7.1	

Let μ_1, μ_2, μ_3 and μ_4 be the mean number of products of the 4 production lines.

(a) Provide the ANOVA table.

(b) Please test the hypothesis $H_0: \mu_1=\mu_2=\mu_3=\mu_4$ with $\alpha=0.05$.

[Solution:]

$k = 4, \alpha = 0.05, n_1 = 4, n_2 = 5, n_3 = 6, n_4 = 5, n_T = n_1 + n_2 + n_3 + n_4 = 20$

and

$$\bar{x}_1 = 8.875, \bar{x}_2 = 8.16, \bar{x}_3 = 6.7, \bar{x}_4 = 6.88, \bar{x} = 7.545$$

(a)

$$SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{j,i} - \bar{x}\right)^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{j,i}^2 - n_T\bar{x}^2 = 19.35$$

$$SSB = \sum_{j=1}^{k} n_j\left(\bar{x}_j - \bar{x}\right)^2$$

$$= 4 \cdot (8.875 - 7.545)^2 + 5 \cdot (8.816 - 7.545)^2$$
$$+6 \cdot (6.7 - 7.545)^2 + 5 \cdot (6.88 - 7.545)^2$$
$$= 15.462$$

$$SSW = SST - SSB = 19.35 - 15.462 = 3.888.$$

The ANOVA table is

Source	SS	df	MS	F
Between	15.462	$k - 1 = 3$	$15.462/3 = 5.154$	$f = 5.154/0.243$ $= 21.21$
Within	3.888	$n_T - k$ $= 16$	$3.888/16 = 0.243$	
Total	19.35	$n_T - 1$ $= 19$		

(b) We reject H_0 since

$$f = 21.21 > 3.24 = f_{3,16,0.05} = f_{k-1,n_T-k,\alpha}.$$

Example 2:

The data were selected from each of three normal populations with equal variances. The data obtained follow.

Observation	Sample 1	Sample 2	Sample 3
1	8	6	8
2	5	6	7
3	3	7	9
4	6	5	6
5	6	3	5
Sample mean	5.6	5.4	7
Sample variance	3.3	2.3	2.5

Set up the ANOVA table for this problem and test the null hypothesis that the three population means are equal at the $\alpha = 0.05$ level of significance.

[Solution:]

$$SSB = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{x})^2 = 5 \cdot (5.6 - 6)^2 + 5 \cdot (5.4 - 6)^2 + 5 \cdot (7 - 6)^2 = 7.6$$

$$SSW = \sum_{j=1}^{k} (n_j - 1)s_j^2 = 4 \cdot 3.3 + 4 \cdot 2.3 + 4 \cdot 2.5 = 32.4$$

The ANOVA table is

Source	SS	DF	MS	F
Between	7.6	2	3.8	1.41
Within	32.4	12	2.7	
Total	40	14		

Therefore,

$$f = 1.41 < f_{2,12,0.05.} = 3.89 \Rightarrow not\ \ reject\ \ H_0$$

Example 3:

Negative appeals have been recognized as an effective method of persuasion in advertising. The number of ads with guilt appeals that appeared in selected types follow.

Magazine Type	Number of Ads with Guilt Appeals
News and opinion	20
General editorial	15
Family – oriented	30
Business/financial	20
Female – oriented	20
African – American	10

Use $\alpha = 0.05$ test to determine any differences in the proportion of ads with guilt appeals among the six types of magazines.

[Solution:]

$$H_0: p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = {}^{1}/_{6}, e_i = 115 \cdot \frac{1}{6} = 19.17$$

for $i = 1, 2, \cdots, 6$. Then,

$$\chi^2=\sum_{i=1}^{6}\frac{(f_i-e_i)^2}{e_i}$$

$$=\frac{(20-19.17)^2}{19.17}+\frac{(15-19.17)^2}{19.17}+\frac{(30-19.17)^2}{19.17}$$

$$+\frac{(20-19.17)^2}{19.17}+\frac{(20-19.17)^2}{19.17}+\frac{(10-19.17)^2}{19.17}$$

$$=11.52>11.07=\chi^2_{5,0.05}=\chi^2_{k-1,\alpha}.$$

Therefore, we reject H_0.

Example 4:

The following are the number of wrong answers for the number of the students.

Number of wrong answers	0	1	2	3
Number of the students	21	31	12	0

Suppose X is the random variable representing the number of wrong answers. Please test X is distributed as $Binomial(3, 0.25)$ with $\alpha = 0.05$..

(Note: the distribution function for $Binomial(3, 0.25)$ is

$$f_X(x)=\binom{3}{x}(0.25)^x(0.75)^{3-x}, x=0,1,2,3.$$

[Solution:]

As H_0 is true, the distribution for the number of wrong answers is

$$p_1=P(X=0)=\binom{3}{0}0.25^0 0.75^3=\frac{27}{64}$$

$$p_2=P(X=1)=\binom{3}{1}0.25^1 0.75^2=\frac{27}{64}$$

$$p_3=P(X=2)=\binom{3}{2}0.25^2 0.75^1=\frac{9}{64}$$

$$p_3=P(X=3)=\binom{3}{3}0.25^3 0.75^0=\frac{1}{64}.$$

Since the sample size $n=21+31+12+0=64$, the expected numbers under H_0 are

$$e_1=np_1=64\cdot\frac{27}{64}=27, e_2=np_2=64\cdot\frac{27}{64}=27,$$

$$e_3=np_3=64\cdot\frac{9}{64}=9, e_4=np_4=64\cdot\frac{1}{64}=1.$$

Therefore,

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

$$= \frac{(21-27)^2}{27} + \frac{(31-27)^2}{27} + \frac{(12-9)^2}{9} + \frac{(0-1)^2}{1}$$

$$= 3.92$$

Since $\chi^2 = 3.92 < 7.81 = {\chi^2}_{3,0.05}$, we do *not* reject H_0.

Example 5:

Starting positions for business and engineering graduates are classified by industry as shown in the following table.

	Oil	Chemical	Electrical	Computer
Business	30	15	15	40
Engineering	30	30	20	20

Use $\alpha = 0.01$ and test for independence of degree major and industry type.

[Solution:]

The table for the expected numbers e_{ij} is

	Oil	Chemical	Electrical	Computer	Row Total
Business	$\frac{60 \cdot 100}{200} = 30$	$\frac{45 \cdot 100}{200} = 22.5$	$\frac{35 \cdot 100}{200} = 17.5$	$\frac{60 \cdot 100}{200} = 30$	100
Engineering	$\frac{60 \cdot 100}{200} = 30$	$\frac{45 \cdot 100}{200} = 22.5$	$\frac{35 \cdot 100}{200} = 17.5$	$\frac{60 \cdot 100}{200} = 30$	100
Column Total	60	45	35	60	200

Thus,

$$\chi^2 = \sum_{i=1}^{p}\sum_{j=1}^{m} \frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}} = \sum_{i=1}^{2}\sum_{j=1}^{4} \frac{\left(f_{ij} - e_{ij}\right)^2}{e_{ij}}$$

$$= \frac{(30-30)^2}{30} + \frac{(15-22,5)^2}{22.5} + \frac{(15-17.5)^2}{17.5} + \frac{(40-30)^2}{30}$$

$$+ \frac{(30-30)^2}{30} + \frac{(30-22,5)^2}{22.5} + \frac{(20-17.5)^2}{17.5} + \frac{(30-30)^2}{30}$$

$$= 12.39.$$

Since $\boldsymbol{p = 2, m = 4, \alpha = 0.01}$,

$$\boldsymbol{\chi^2 = 12.39 > 11.3449 = \chi^2_{3,0.01} = \chi^2_{(2-1)\cdot(4-1),0.01} = \chi^2_{(p-1)\cdot(m-1),\alpha}},$$

we reject $\boldsymbol{H_0}$.