Semester 1	Solution to Tutorial Set 11	2013

1. (a) $a = 9.236$; (b) $a = 18.307$

   (c) $0.025 < \text{Prob} < 0.05$; (d) $0.025 < \text{Prob} < 0.05$

2. (d)

3. (a) $\chi^2_{\text{obs}} = \sum_{i=1}^{3} \frac{(18-25)^2}{25} + \frac{(55-50)^2}{50} + \frac{(27-25)^2}{25} = 1.96 + 0.50 + 0.16 = 2.62$

4. P-value $= P(\chi^2_2 \geq 2.62) > 0.1$
   Thus the data are consistent with the null hypothesis H_0.

5. The χ^2 GOF test is

   (a) **Hypothesis:**
   H_0: $p_1 = 0.45,\ p_2 = 0.40,\ p_3 = 0.15$ vs
   H_1: At least one equality does not hold.

   (b) **Test statistic:** The expected frequencies are respectively:

   $$200 \times 0.45 = 90; \quad 200 \times 0.40 = 80; \quad 200 \times 0.15 = 30$$

   $\chi^2_{\text{obs}} = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(102-90)^2}{90} + \frac{(82-80)^2}{80} + \frac{(16-30)^2}{30} = 8.183$

   (c) **P-value:** $0.01 <$ p-value $= P(\chi^2_2 > 8.183) < 0.025$ (0.016711 from R; df=3-1=2)

   (d) **Conclusion:** Since P-value < 0.05, there is strong evidence in the data against H_0. There is significant change in the mortality rates over the past ten.

6. The χ^2 test for independence between "level of job satisfaction" and "gender" is

   (a) Hypotheses:
   $H_0 : p_{ij} = p_i \times p_j$, i.e. "level of job satisfaction" and "gender" are independent.
   H_1: Not all equalities hold, i.e. "level of job satisfaction" and "gender" are dependent.

   (b) Test statistic: The calculation of the expected frequencies E_{ij} and squared standardized residuals d^2_{ij} under H_0 are:

Gender	Very interesting	Fairly interesting	Not interesting
Male	$E_{11}=\frac{120\times105}{200}=63$ $d^2_{11}=\frac{(70-63)^2}{63}=0.778$	$E_{12}=\frac{120\times75}{200}=45$ $d^2_{12}=\frac{(41-45)^2}{45}=0.356$	$E_{13}=\frac{120\times20}{200}=12$ $d^2_{13}=\frac{(9-12)^2}{12}=0.750$
Female	$E_{21}=\frac{80\times105}{200}=42$ $d^2_{21}=\frac{(35-42)^2}{42}=1.167$	$E_{22}=\frac{80\times75}{200}=30$ $d^2_{22}=\frac{(34-30)^2}{30}=0.533$	$E_{23}=\frac{80\times20}{200}=8$ $d^2_{23}=\frac{(11-8)^2}{8}=1.125$

$$\chi^2_{\text{obs}} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 0.778 + 0.356 + \ldots + 1.125 = 4.708$$

(c) P-value: $P(\chi^2_2 > 4.708) < 0.10$ (0.09497 from R; df=(3-1)(2-1)=2)

(d) Conclusion: Since P-value < 0.10, there is sufficient evidence in the data against H_0. We conclude that "level of job satisfaction" and "gender" are dependent at $\alpha = 0.10$.

7. Refer to Q1.

   (c) `1 - pchisq(38.5, 25)` Answer is 0.04131387

   (d) `1 - pchisq(22.1, 12)` Answer is 0.03641422

8. Refer to Q5.

```
> chisq.test(c(102,82,16),p=c(0.45,0.4,0.15))

        Chi-squared test for given probabilities

data:  c(102, 82, 16)
X-squared = 8.1833, df = 2, p-value = 0.01671
```

9. Refer to Q6.

```
> x=c(70,35,41,34,9,11)
> x.mat=matrix(x,2,3)
> x.mat
     [,1] [,2] [,3]
[1,]   70   41    9
[2,]   35   34   11
> chisq.test(x.mat)

        Pearson's Chi-squared test

data:  x.mat
X-squared = 4.7083, df = 2, p-value = 0.09497
```

10.
```
> dat=read.table(file="lead-IQ.txt")
> x=dat[,1]
> y=dat[,2]
> tab=table(x,y)
> tab
   y
x    1  2
  1 46 11
  2 33 12
  3 15  7
> chisq.test(tab)

        Pearson's Chi-squared test

data:  tab
X-squared = 1.5922, df = 2, p-value = 0.4511
```

The test statistic is 1.5922, the df is 2 ((3-1)(2-1)=2) and the P-value is 0.4511. Since the P-value is greater than 0.05, the data are consistent with H_0. The two factors of `distance` and `IQ` are independent.

Additional Problems for Week 11 - Solutions

1. Complete the following table to calculate the test statistics χ^2_{obs}:

Restaurant	Observed freq. O_i	Expected freq. $E_i = np_i$	$(O_i - E_i)$	$\frac{(O_i-E_i)^2}{E_i}$
A	48	$150 \times \frac{1}{3} = 50$	-2	$\frac{(-2)^2}{50} = 0.08$
B	62	$150 \times \frac{1}{3} = 50$	12	$\frac{12^2}{50} = 2.88$
C	40	$150 \times \frac{1}{3} = 50$	-10	$\frac{(-10)^2}{50} = 2.00$
Total	150	150	0	$\chi^2_{\text{obs}} = 4.96$

The χ^2 GOF test is

1. **Hypothesis:** H_0: $p_1 = p_2 = p_3 = \frac{1}{3}$ vs
   H_1: At least one equality does not hold.
2. **Test statistic:** $\chi^2_{\text{obs}} = \sum_i \frac{(O_i-E_i)^2}{E_i} = 4.96$
3. **P-value:** P-value $= P(\chi^2_2 > 4.96) > 0.05$ (0.08374 from R; df=3-1=2)
4. **Conclusion:** Since P-value > 0.05, the data are consistent with H_0.
   There is no preference among the three restaurants.

```
> x=c(48,62,40)
> p=c(1/3,1/3,1/3)
```

```
> chisq.test(x,p=p)

        Chi-squared test for given probabilities

data:  x
X-squared = 4.96, df = 2, p-value = 0.08374
```

2. The data are

	Duration of IUD use				
	< 3	$\geq 3, < 8$	$\geq 18, \leq 36$	> 36	Total
Cases	10	23	20	36	89
Controls	53	200	168	219	640
Total	63	223	188	255	729

The χ^2 test for independence between "duration of IUD" and "infertility" is

(a) Hypotheses:
$H_0 : p_{ij} = p_i \times p_j$, i.e. "duration of IUD" and "infertility" are independent.
H_1: Not all equalities hold, i.e. "duration of IUD" and "infertility" are dependent.

(b) Test statistic: The calculation of the expected frequencies E_{ij} and squared residuals d_{ij}^2 under H_0 are:

	< 3	$\geq 3, < 8$	$\geq 18, \leq 36$	> 36
Cases	$E_{11}=\frac{89\times 63}{729}=7.69$	$E_{12}=\frac{89\times 223}{729}=27.22$	$E_{13}=\frac{89\times 188}{729}=22.95$	$E_{14}=\frac{89\times 255}{729}=31.13$
	$d_{11}^2=\frac{(10-7.69)^2}{7.69}=0.693$	$d_{12}^2=\frac{(23-27.22)^2}{27.22}=0.656$	$d_{13}^2=\frac{(20-22.95)^2}{22.95}=0.380$	$d_{14}^2=\frac{(36-31.13)^2}{31.13}=0.761$
Control	$E_{21}=\frac{640\times 63}{729}=55.31$	$E_{22}=\frac{640\times 223}{729}=195.78$	$E_{23}=\frac{640\times 188}{729}=165.05$	$E_{24}=\frac{640\times 255}{729}=223.87$
	$d_{21}^2=\frac{(53-55.31)^2}{55.31}=0.096$	$d_{22}^2=\frac{(200-195.78)^2}{195.78}=0.091$	$d_{23}^2=\frac{(168-165.05)^2}{165.05}=0.053$	$d_{24}^2=\frac{(219-223.87)^2}{223.87}=0.106$

$$\chi^2_{\text{obs}} = \sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = 0.693 + 0.656 + \ldots + 0.106 = 2.836$$

(c) P-value: $P(\chi^2_3 > 2.836) > 0.25$ (0.4176 from R; df=(4-1)(2-1)=3)

(d) Conclusion: Since P-value > 0.05, the data are consistent with H_0. We conclude that "duration of IUD" and "fertility" are independent.

```
> x=c(10,53,23,200,20,168,36,219)
> x.mat=matrix(y,2,4)
> x.mat
     [,1] [,2] [,3] [,4]
[1,]   10   23   20   36
```

```
[2,]   53  200  168  219
> chisq.test(x.mat)

        Pearson's Chi-squared test

data:  x.mat
X-squared = 2.8358, df = 3, p-value = 0.4176
```

3. The data are

	Disease			
	sex-linked	recessive	dominant	Total
English	46	25	54	125
Swiss	1	99	10	110
Total	47	124	64	235

The χ^2 test for independence between "ethnic origin" and "genetic type" is

(a) Hypotheses:
$H_0 : p_{ij} = p_i \times p_j$, i.e. "ethnic origin" and "genetic type" are independent.
H_1: Not all equalities hold, i.e. "ethnic origin" and "genetic type" are dependent.

(b) Test statistic: The calculation the expected frequencies E_{ij} and squared residuals d_{ij}^2 under H_0 are:

	sex-linked	recessive	dominant
Eng	$E_{11}=\frac{125\times47}{235}=25.00$	$E_{12}=\frac{125\times124}{235}=65.96$	$E_{13}=\frac{125\times64}{235}=34.04$
	$d_{11}^2=\frac{(46-25)^2}{25}=17.64$	$d_{12}^2=\frac{(25-65.96)^2}{65.96}=25.43$	$d_{13}^2=\frac{(54-34.04)^2}{34.04}=11.70$
Swiss	$E_{21}=\frac{110\times47}{235}=22.00$	$E_{22}=\frac{110\times124}{235}=58.04$	$E_{23}=\frac{110\times64}{235}=29.96$
	$d_{21}^2=\frac{(1-22)^2}{22}=20.05$	$d_{22}^2=\frac{(99-58.04)^2}{58.04}=28.90$	$d_{23}^2=\frac{(10-29.96)^2}{29.96}=13.30$

$$\chi^2_{\text{obs}} = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = 17.64 + 25.43 + \ldots + 13.30 = 117.02$$

(c) P-value: $P(\chi^2_2 > 117.02) < 0.0000$ (0.0000 from R; df=(3-1)(2-1)=2)

(d) Conclusion: Since P-value < 0.05, there is very strong evidence in the data against H_0. We conclude that "ethnic origin" and "genetic type" are dependent.

```
> x=c(46,1,25,99,54,10)
> x.mat=matrix(y,2,3)
> x.mat
     [,1] [,2] [,3]
```

```
[1,]   46   25   54
[2,]    1   99   10
> chisq.test(x.mat)

        Pearson's Chi-squared test

data:  x.mat
X-squared = 117.0157, df = 2, p-value < 2.2e-16
```