TUTORIAL EXERCISE PACKAGE - 2013

MATH1015 - BIOSTATISTICS

Semester 1	Solution to Tutorial Set 12	2013

1. 1. Hypotheses:

$$H_0: \text{ performance and training are independent}$$

$$H_1: \text{ performance and training are dependent}$$

The expected counts under H_0 are:

$$E_{11} = \frac{150(120)}{200} = 90,\ E_{12} = \frac{150(80)}{200} = 60,\ E_{21} = \frac{50(120)}{200} = 30,\ E_{22} = \frac{50(80)}{200} = 20$$

2. The test statistic is:

$$\chi^2_{\text{obs}} = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(100-90)^2}{90} + \frac{(50-60)^2}{60} + \frac{(20-30)^2}{30} + \frac{(30-20)^2}{20} = 11.1$$

3. P-value: $\text{P}(\chi^2_1 > 11.1) < 0.01$, where df = (2 - 1)(2 - 1) = 1.

4. Conclusion: Since P-value $< 0.01 < 0.05$, we have very strong evidence against H_0. That is, work performance and training are related to each other.

2. 1. Hypotheses: $H_0: \ p_1 = p_2 = p_3 = p_4 = 0.25$ vs. $H_1:$ at least one $p_i \neq 0.25$

The expected counts under H_0 are: $E_1 = E_2 = E_3 = E_4 = 0.25(100) = 25$

2. The test statistic is:

$$\chi^2_{\text{obs}} = \sum_{i=1}^{4} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(35-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(20-25)^2}{25} + \frac{(15-25)^2}{25} = 10.$$

3. P-value: $\text{P}(\chi^2_3 > 10) \in (0.01, 0.025)$, where df = 4 - 1 = 3.

4. Conclusion: Since P-value $< 0.025 < 0.05$, we have strong evidence against H_0. That is, the proportions of movie type preferences are different.

3. Answer. (d)$L_{xx} = \Sigma_i {x_i}^2 - (\Sigma_i x_i)^2/n = 59 - (25)^2/12 = 6.917$,
$L_{yy} = \Sigma_i {y_i}^2 - (\Sigma_i y_i)^2/n = 15648 - (432)^2/12 = 96$ and
$L_{xy} = \Sigma_i x_i y_i - (\Sigma_i\ x_i)(\Sigma_i\ y_i)/n = 880 - (25)(432)/12 = -20$ and therefore,
$r = \dfrac{-20}{\sqrt{6.917(96)}} = -0.776$

4. Answer: (c). Explanation: See Q3 solution.

5. Answer: (d). Explanation: $b = \dfrac{L_{xy}}{L_{xx}} = \dfrac{-20}{6.917} = -2.891$

6. Answer: (a). Explanation: $a = \bar{y} - b\bar{x} = \dfrac{432}{12} - (-2.891)\dfrac{25}{12} = 36 + 2.891(2.083) = 42.02$.
Therefore, $\hat{y} = 42.02 + (-2.891)(5) = 27.568$.

7. $\Sigma_i\, x_i = 520$; $\Sigma_i\, x_i{}^2 = 38000$; $\Sigma_i\, y_i = 97$; $\Sigma_i\, y_i{}^2 = 1217$; $\Sigma_i\, x_i y_i = 5910$
$L_{xx} = \Sigma_i x_i{}^2 - (\Sigma_i x_i)^2/n = 38000 - 520^2/8 = 4200$;
$L_{yy} = \Sigma_i y_i{}^2 - (\Sigma_i y_i)^2/n = 1217 - 97^2/8 = 40.88$;
$L_{xy} = \Sigma_i x_i y_i - (\Sigma_i x_i)(\Sigma_i y_i)/n = 5910 - 520(97)/8 = -395$.

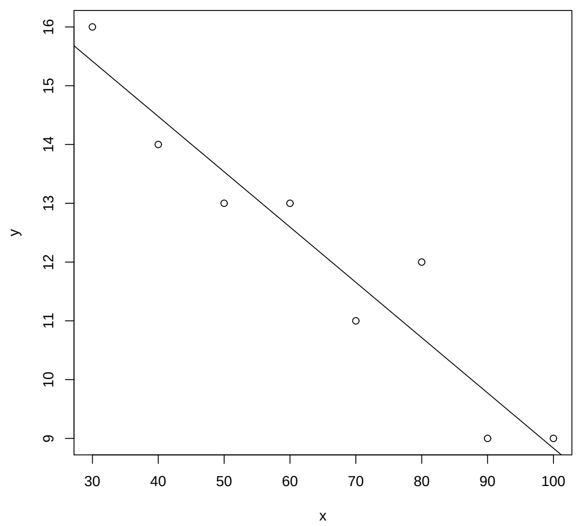
   (a) $r = \dfrac{L_{xy}}{\sqrt{L_{xx} L_{yy}}} = \dfrac{-395}{\sqrt{4200 \times 40.88}} = -0.953$.

   (b) $b = \dfrac{L_{xy}}{L_{xx}} = \dfrac{-395}{4200} = -0.0940476$;
   $a = \overline{y} - b\overline{x} = 12.125 - 65 \times (-0.0940476) = 18.261306$.

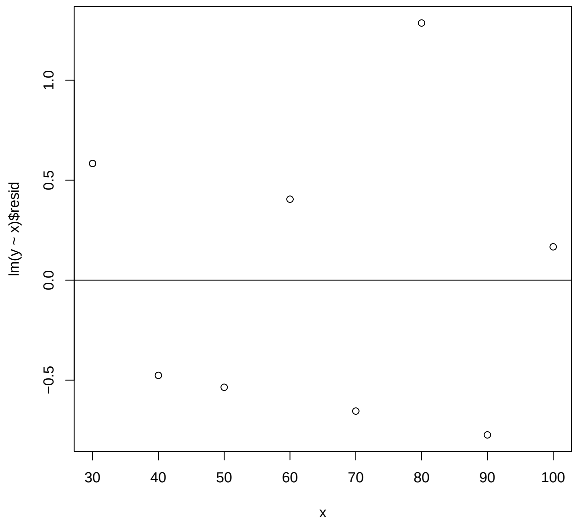
   Thus the regression line is $\hat{y} = 18.261306 - 0.0940476x$.

   (c) When $x = 85$, the predicted breathing rate is $\hat{y} = 18.261306 - 0.0940476 \times 85 = 10.267$.

   (d) Scatter plot shows that the points are negatively correlated:



   (e) The plot shows that one of the positive residuals is unusually high.



8. (a) `attach(cars)`

   (b) `cor(speed, dist)`

   `[1] 0.8068949`

   (c) `lm(dist ∼ speed)`

   `Call:`

   `lm(formula = dist ∼ speed)`

   `Coefficients:`

   `(Intercept) speed`

   `-17.579 3.932`

   (d) $\hat{\text{dist}} = -17.58 + 3.93 \times \text{speed}$

9.
```
> x = c(30,40,50,60,70,80,90,100)
> y = c(16,14,13,13,11,12,9,9)
> cor(x,y)
[1] -0.9533307

> lm(y~x)
Call:
lm(formula = y ~ x)
Coefficients:
(Intercept)            x
   18.23810     -0.09405
```

```
> 18.23810-0.09405*85
[1] 10.24385

> plot(x, y)
> abline(lm(y~x)$coeff)

> plot(x, lm(y~x)$resid)
> abline(h = 0)
```

Additional Problems for Week 12 - Solutions

1. From Chi-square table,
   (a) $a = 15.987$ (b) $a = 24.996$

   (c) $0.025 < \text{Prob} < 0.05$; (d) $0.05 < \text{Prob} < 0.1$

2. (a) The χ^2 GOF test is

   (a) **Hypothesis:**
   H_0: $p_1 = 0.25,\ p_2 = 0.25,\ p_3 = 0.25,\ p_3 = 0.25$ vs
   H_1: At least one equality does not hold.

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

   $$198 \times 0.25 = 49.5$$

   $$\chi^2_{\text{obs}} = \sum_{i=1}^{3} \frac{(O_i-E_i)^2}{E_i} = \frac{(61-49.5)^2}{49.5} + \frac{(55-49.5)^2}{49.5} + \frac{(41-49.5)^2}{49.5} + \frac{(41-49.5)^2}{49.5} = 6.202$$

   (c) **P-value:** p-value $= P(\chi^2_3 > 6.202) > 0.05$ (0.1022 from R; df=4-1=3)

   (d) **Conclusion:** Since P-value > 0.05, the data are consistent with H_0. The four colours are equally likely.

Check:

```
> chisq.test(c(61,55,41,41),p=c(1/4,1/4,1/4,1/4))

        Chi-squared test for given probabilities

data:  c(61, 55, 41, 41)
X-squared = 6.202, df = 3, p-value = 0.1022
```

(b) The χ^2 GOF test is

(a) **Hypothesis:**
H_0: $p_1 = 1/3,\ p_2 = 5/18,\ p_3 = 2/9,\ p_3 = 1/6$ vs
H_1: At least one equality does not hold.

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

$$198 \times 1/3 = 66; 198 \times 5/18 = 55; 198 \times 2/9 = 44; 198 \times 1/6 = 33$$

$\chi^2_{\text{obs}} = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(61-66)^2}{66} + \frac{(55-55)^2}{55} + \frac{(41-44)^2}{44} + \frac{(41-33)^2}{33} = 2.5227$

(c) **P-value:** p-value $= P(\chi^2_3 > 2.5227) > 0.05$ (0.4712 from R; df=4-1=3)

(d) **Conclusion:** Since P-value > 0.05, the data are consistent with H_0. The four colours are in the ratio 6:6:4:3.

Check:

```
> chisq.test(c(61,55,41,41),p=c(1/3,5/18,2/9,1/6))

        Chi-squared test for given probabilities

data:  c(61, 55, 41, 41)
X-squared = 2.5227, df = 3, p-value = 0.4712
```

3. As above.

4. As above.

5. Q10.37, 10.38: The χ^2 test for independence between "treatment" and "response" is

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

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

Treatment	+Smear	-Smear +culture	-Smear -culture
Pen	$E_{11}=\frac{200\times65}{400}=32.5$ $d_{11}^2=\frac{(40-32.5)^2}{32.5}=1.731$	$E_{12}=\frac{200\times90}{400}=45$ $d_{12}^2=\frac{(30-45)^2}{45}=5.000$	$E_{13}=\frac{200\times245}{400}=122.5$ $d_{13}^2=\frac{(130-122.5)^2}{122.5}=0.459$
Spect(low)	$E_{21}=\frac{100\times65}{400}=16.25$ $d_{21}^2=\frac{(10-16.25)^2}{16.25}=2.404$	$E_{22}=\frac{100\times90}{400}=16.25$ $d_{22}^2=\frac{(20-16.25)^2}{16.25}=0.278$	$E_{23}=\frac{100\times245}{400}=61.25$ $d_{23}^2=\frac{(70-61.25)^2}{61.25}=1.250$
Spect(high)	$E_{21}=\frac{100\times65}{400}=16.25$ $d_{21}^2=\frac{(15-16.25)^2}{16.25}=0.096$	$E_{22}=\frac{100\times90}{400}=16.25$ $d_{22}^2=\frac{(40-16.25)^2}{16.25}=13.611$	$E_{23}=\frac{100\times245}{400}=61.25$ $d_{23}^2=\frac{(45-61.25)^2}{61.25}=4.311$

$$\chi^2_{\text{obs}} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 1.731 + 5.000 + \ldots + 4.311 = 29.14$$

(c) P-value: $P(\chi^2_4 > 29.14) < 0.05$ (0.0000 from R; df=(3-1)(3-1)=4)

(d) Conclusion: Since P-value < 0.05, there is strong evidence in the data against H_0. We conclude that "treatment" and "response" are dependent at $\alpha = 0.05$.

Check:

```
> y=c(40,10,15,30,20,40,130,70,45)
> n=sum(y)
> n
[1] 400
> c=3
> r=3
> y.mat=matrix(y,r,c)
> y.mat
     [,1] [,2] [,3]
[1,]   40   30  130
[2,]   10   20   70
[3,]   15   40   45
> chisq.test(y.mat)

        Pearson's Chi-squared test
```

```
data:  y.mat
X-squared = 29.1401, df = 4, p-value = 7.322e-06
```

6. Q11.1 $\Sigma_i\ x_i = 12.6$; $\Sigma_i\ {x_i}^2 = 32.02$;
   $\Sigma_i\ y_i = 18466$; $\Sigma_i\ {y_i}^2 = 41504606$; $\Sigma_i\ x_i y_i = 27464.6$
   $L_{xx} = \Sigma_i {x_i}^2 - (\Sigma_i x_i)^2/n = 32.02 - 12.6^2/9 = 14.38$;
   $L_{yy} = \Sigma_i {y_i}^2 - (\Sigma_i y_i)^2/n = 41504606 - 18466^2/9 = 3616477.556$;
   $L_{xy} = \Sigma_i x_i y_i - (\Sigma_i x_i)(\Sigma_i y_i)/n = 27464.6 - 12.6(18466)/9 = 1612.2$.

   $$b = \frac{L_{xy}}{L_{xx}} = \frac{1612.2}{14.38} = 112.114; \qquad a = \overline{y} - b\overline{x} = 1.4 - 112.114 \times 2051.778 = 1894.818\,.$$

   Thus the regression line is $\hat{y} = 1894.818 + 112.114x$.

   Q11.3 $r = \dfrac{L_{xy}}{\sqrt{L_{xx}L_{yy}}} = \dfrac{1612.2}{\sqrt{14.38 \times 3616477.556}} = 0.2235613.$

   Hence $r^2 = 0.2235613^2 = 0.04997965$. This r^2 is very small and so the regression model provides poor fit.