THE UNIVERSITY OF SYDNEY

STAT2012 STATISTICAL TESTS

Semester 2	**Solution to Tutorial Week 13**	2015

Tutorial questions

1. We have $\bar{x} = 26.8$ and $s = 6.38$. Since the sample size is less than 80, we employ the minimum number of intervals 4. Hence the 3 cut-off points in the standardized (Z) scale is -1,0,1. The cut-off points in the data scale is

$$\bar{x} - s = 26.8 - 6.38 = 20.42 \quad \bar{x} = 26.8 \quad \bar{x} + s = 26.8 + 6.38 = 33.18$$

The sorted data and marks to show the cut-off points are shown below:

```
14 16 18 20 | 21 21 21 21 22 23 23 23 24 24 24 25 25 26 26 26
26 26 26 26 | 27 27 29 29 30 30 31 32 32 32 33 | 35 36 36 40 46
```

The calculation of the test statistics is summarized in the following table:

Interval i	Stand. int. i	Obs. f. y_i	Exp. prob. p_i	Exp. f. np_i	Chi-square $\frac{(y_i-np_i)^2}{np_i}$
$X \leq$20.42	$Z \leq$-1	4	Φ(-1)=0.159	40(0.159)=6.3	$\frac{(4-6.3)^2}{6.3} = 0.87$
20.42< $X \leq$26.80	-1< $Z \leq$0	20	Φ(0)-Φ(-1)=0.341	40(0.341)=13.7	$\frac{(20-13.7)^2}{13.7} = 2.95$
26.80< $X \leq$33.18	0< $Z \leq$1	11	Φ(1)-Φ(0)=0.341	40(0.341)=13.7	$\frac{(11-13.7)^2}{13.7} = 0.52$
$X >$33.18	$Z >$1	5	1-Φ(1)=0.159	40(0.159)=6.3	$\frac{(5-6.3)^2}{6.3} = 0.29$
Total		40	1.000	40.0	4.62

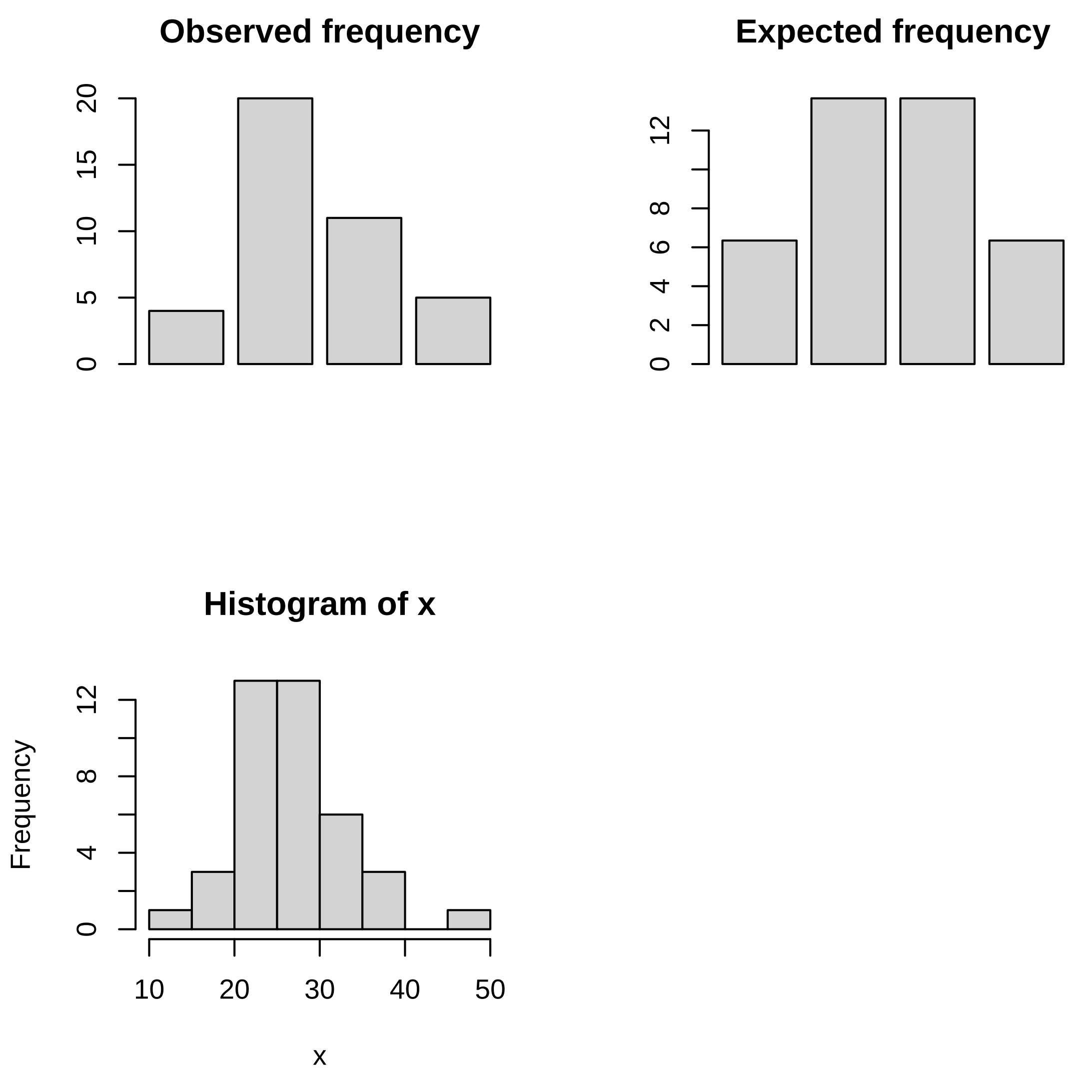
The Chi-square goodness-of-fit test for the normality of the data is

1. **Hypotheses:** H_0 : The population has a normal distribution.
   H_1 : The population does not have a normal distribution.
2. **Test statistic:** $\chi_0^2 = 4.6216$
3. **Assumption:** $E_i = np_{i0} \geq 5$. Under H_0, $\chi_0^2 \sim \chi_{k-1-h}^2$ approx.
4. P**-value:** p-value $= \Pr(\chi_1^2 \geq 4.6216) = 0.0316$
5. **Decision:** Since the p-value < 0.05, we reject H_0. There is strong evidence in the data against H_0 that the population has a normal probability distribution.

In R,

```
> x=c(26,21,25,20,21,29,26,23,22,24,24,30,23,32,26,24,32,16,36,26,21,31,
      26,23,32,35,40,30,14,26,46,27,33,25,27,21,26,18,29,36)
> x=sort(x)
> x
 [1] 14 16 18 20 21 21 21 21 22 23 23 23 24 24 24 25 25 26 26 26 26 26 26 26 27
[26] 27 29 29 30 30 31 32 32 32 33 35 36 36 40 46
> n=length(x)
> k=6
> if (n<=80) k=4 else if (n<=220) k=5
> k
[1] 4
> xm=mean(x)
> xsd=sd(x)
> c(xm,xsd)
[1] 26.800000  6.377846
> zlow=-2   #set the lowest class marks for z
> if (n<=80) zlow=-1 else if (n<=220) zlow=-1.5
> zlow
[1] -1
> int=c()  #set class marks for x
> for (i in 1:k-1) {int[i]=xm+(zlow+i-1)*xsd}
> int
[1] 20.42215 26.80000 33.17785
> y=c()   #count class freq.
> y[1]=length(x[x<=int[1]])
> y[k]=length(x[x>int[k-1]])
> for(i in 2:k-1)
+ {y[i]=length(x[x<=int[i]])-length(x[x<=int[i-1]])}
> y
[1]  4 20 11  5
> sum(y)
[1] 40
> p=c()   #calculate expected prob.
> for(i in 2:k-1)
+ {p[i]=pnorm(zlow+i-1)-pnorm(zlow+i-2)}
> p[1]=pnorm(zlow)
> p[k]=1-pnorm(zlow+k-2)
> p
[1] 0.1586553 0.3413447 0.3413447 0.1586553
> sum(p)
[1] 1
> ey=n*p
> ey
[1]  6.34621 13.65379 13.65379  6.34621
> chi=(y-ey)^2/(ey)
> chi
[1] 0.8673999 2.9496853 0.5157982 0.2855691
```

```
> chi=sum(chi)
> chi
[1] 4.618453
> p.value=1-pchisq(chi,k-1-2)
> p.value
[1] 0.03162976
> par(mfrow=c(2,2))
> barplot(y,col="lightgray",main="Observed frequency")
> barplot(ey,col="lightgray",main="Expected frequency")
> hist(x,col="lightgray")
```



2. Calculation of the test statistic is summarized in the following table:

View		16-34 $(j=1)$	35-54 $(j=2)$	55 & over $(j=3)$	Mar. prob. $p_{i\cdot}$
Low	$O_i = y_{ij}$	8	12	21	41
$(i=1)$	$E_i = np_{i\cdot}p_{\cdot j}$	$81(0.321)(0.51) = 13.16$	$81(0.333)(0.51) = 13.67$	$81(0.346)(0.51) = 14.17$	$\frac{41}{81} = 0.506$
	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{(-5.16)^2}{13.16} = 2.024$	$\frac{(-1.67)^2}{13.67} = 0.203$	$\frac{(6.83)^2}{14.2} = 3.289$	
High	$O_i = y_{ij}$	18	15	7	40
$(i=2)$	$E_i = np_{i\cdot}p_{\cdot j}$	$81(0.321)(0.49) = 12.84$	$81(0.333)(0.49) = 13.33$	$81(0.346)(0.49) = 13.83$	$\frac{40}{81} = 0.494$
	$\frac{(O_i - E_i)^2}{E_i}$	$\frac{(5.16)^2}{12.84} = 2.074$	$\frac{(1.67)^2}{13.33} = 0.208$	$\frac{(-6.83)^2}{13.8} = 3.371$	11.169
	$\sum_i O_i$	26	27	28	81
	$\sum_i E_i$	26	27	28	81
	Mar. prob. $p_{\cdot j}$	$\frac{26}{81} = 0.321$	$\frac{27}{81} = 0.333$	$\frac{28}{81} = 0.346$	

Let $p_{1j}, j = 1, 2, 3$, denote the probability of a person who is a Low violence viewer in the age group 16-34, 35-54, 55 and Over respectively, and $p_{2j}, j = 1, 2, 3$, denote the probability of a person who is a High violence viewer in the age group 16-34, 35-54, 55 and Over respectively.

The Chi-square test for independence between factors is

1. **Hypotheses:** $H_0 : p_{ij} = p_{i.}p_{.j},\ i = 1, 2;\ j = 1, 2, 3,$ vs H_1 : Not all are same.
2. **Test statistic:**

$$\begin{aligned}\chi_0^2 &= \sum_{i=1}^{2}\sum_{j=1}^{3} \frac{(y_{ij} - y_{i.}y_{.j}/n)^2}{y_{i.}y_{.j}/n} \\ &= \frac{(8-13.16)^2}{13.16} + \ldots + \frac{(7-13.83)^2}{13.83} = 11.169.\end{aligned}$$

3. **Assumption:** $E_i = np_{i0} \geq 5$. Under H_0, $\chi_0^2 \sim \chi^2_{(r-1)(c-1)}$ approx.
4. **P-value:** p-value $= \Pr(\chi_2^2 \geq 11.169) < 0.01$ ($\chi^2_{2,0.99} = 9.21,\ 0.0037$ from R).
5. **Decision:** Since the p-value < 0.05, we reject H_0. There is strong evidence in the data that the view of violence is dependent on age of viewer.

3. The calculation of the test statistic is summarized in the following table:

x	Obs. f. y_i	Prod. $x \times y_i$	Exp. prob. $p_i = \lambda^x e^{-\lambda} x!$	Exp. f. np_i	Chi-square $\frac{(y_i - np_i)^2}{np_i}$
0	7	0	$\frac{2.823^0 e^{-2.823}}{0!} = 0.0594$	130 (0.0594)= 7.72	$\frac{(-0.725)^2}{7.72} = 0.068$
1	18	18	$\frac{2.823^1 e^{-2.823}}{1!} = 0.1678$	130 (0.1678)= 21.81	$\frac{(-3.808)^2}{21.81} = 0.665$
2	38	76	$\frac{2.823^2 e^{-2.823}}{2!} = 0.2368$	130 (0.2368)= 30.78	$\frac{(7.217)^2}{30.78} = 1.692$
3	29	87	$\frac{2.823^3 e^{-2.823}}{3!} = 0.2228$	130 (0.2228)= 28.97	$\frac{(0.032)^2}{28.97} = 0.000$
4	16	64	$\frac{2.823^4 e^{-2.823}}{4!} = 0.1573$	130 (0.1573)= 20.44	$\frac{(-4.444)^2}{20.44} = 0.966$
5	12	60	$\frac{2.823^5 e^{-2.823}}{5!} = 0.0888$	130 (0.0888)= 11.54	$\frac{(0.457)^2}{11.54} = 0.018$
6	8	48	$\frac{2.823^6 e^{-2.823}}{6!} = 0.0418$	130 (0.0418)= 5.43	$\frac{(2.569)^2}{5.43} = 1.215$
≥ 7	2	14	$\frac{2.823^7 e^{-2.823}}{7!} = 0.0254$	130 (0.0254)= 3.30	$\frac{(-1.297)^2}{3.30} = 0.510$
Sum	130	367	1.0000	130.00	

Now $\hat{\lambda} = \frac{367}{130} = 2.8231$. Note that $E_{\geq 7} = 3.30 < 5$ which violates the assumption. We combine the last two classes so that

$$O_{\geq 6} = 8 + 2 = 10,\ E_{\geq 6} = 5.43 + 3.30 = 8.73,\ \chi^2_{\geq 6} = \frac{(10 - 8.73)^2}{8.73} = 0.185$$

and $\chi^2_0 = 0.068 + 0.665 + 1.692 + 0.000 + 0.966 + 0.018 + 0.185 = 3.594$.

The chi-square goodness-of-fit test for the Poisson distribution is

1. **Hypothesis:** H_0: Follow a Poisson distribution vs
   H_1: Do not follow a Poisson distribution
2. **Test statistic:** $\chi^2_0 = \sum_i \frac{(n_i - np_i)^2}{np_i} = 3.594$
3. **Assumption:** $E_i = np_{i0} \geq 5$. Under H_0, $\chi^2_0 \sim \chi^2_{k-1-h}$ approx.
4. **P-value:** $\Pr(\chi^2_5 > 3.594) > 0.1$ ($\chi^2_{5,0.9} = 9.236$, 0.2380 from R).
5. **Conclusion:** Since the p-value > 0.05, we accept H_0. The data are consistent with the null hypothesis that the data follow a Poisson distribution.

Extra problems

1. Let $p_{1j}, j = 1, 2, 3$ denote the proportion of High IQ in the income group A, B, C respectively.

   Let $p_{2j}, j = 1, 2, 3$ denote the proportion of Moderate/low IQ in the income group A, B, C respectively.

   The calculation of test statistic is summarized in the following table:

Income		High IQ ($j = 1$)	Low IQ ($j = 1$)	Mar. prob. $p_{i\cdot}$
A	$O_i = y_{ij}$	14	18	32
($i = 1$)	$E_i = np_{i\cdot}p_{\cdot j}$	$103(0.602)(0.31) = 19.26$	$103(0.398)(0.31) = 12.74$	$\frac{32}{103} = 0.311$
	$\frac{(O_i-E_i)^2}{E_i}$	$\frac{(-5.26)^2}{19.26} = 1.438$	$\frac{(5.26)^2}{12.7} = 2.174$	
B	$O_i = y_{ij}$	25	8	33
($i = 2$)	$E_i = np_{i\cdot}p_{\cdot j}$	$103(0.602)(0.32) = 19.86$	$103(0.398)(0.32) = 13.14$	$\frac{33}{103} = 0.320$
	$\frac{(O_i-E_i)^2}{E_i}$	$\frac{(5.14)^2}{19.86} = 1.328$	$\frac{(-5.14)^2}{13.1} = 2.008$	
C	$O_i = y_{ij}$	23	15	38
($i = 3$)	$E_i = np_{i\cdot}p_{\cdot j}$	$103(0.602)(0.37) = 22.87$	$103(0.398)(0.37) = 15.13$	$\frac{38}{103} = 0.369$
	$\frac{(O_i-E_i)^2}{E_i}$	$\frac{(0.13)^2}{22.87} = 0.001$	$\frac{(-0.13)^2}{15.1} = 0.001$	6.949
	$\sum_i O_i$	62	41	103
	$\sum_i E_i$	62.00	41.00	103
	Mar. prob. $p_{\cdot j}$	$\frac{62}{103} = 0.602$	$\frac{41}{103} = 0.398$	

   The Chi-square test for the homogeneity of the distribution of IQ in the three income groups is

   1. **Hypotheses:**

$$\begin{aligned} H_0 :& \quad p_{ij} = p_{i\cdot}p_{\cdot j}, \quad i = 1, 2, 3, \; j = 1, 2 \quad \text{vs} \\ H_1 :& \quad \text{At least one of the equalities does not hold.} \end{aligned}$$

   2. **Test statistic:** We have that, with $n_{i.}n_{.j}/n$ in brackets,

$$\begin{aligned} \chi_0^2 &= \sum_{i=1}^{2}\sum_{j=1}^{3} \frac{(y_{ij} - y_{i.}y_{.j}/n)^2}{y_{i.}y_{.j}/n} \\ &= \frac{(14 - 19.26)^2}{19.26} + ... + \frac{(15 - 15.13)^2}{15.13} = 6.95 \end{aligned}$$

   3. **Assumption:** $E_{ij} = np_{i.}p_{.j} \geq 5$. Under H_0, $\chi_0^2 \sim \chi^2_{(r-1)(c-1)}$ approx.

4. ***P*-value:** $0.025 < p\text{-value} = \Pr(\chi^2_2 \geq 6.949) < 0.05$
   $(\chi^2_{2,0.95} = 5.991,\ \chi^2_{2,0.975} = 7.378,\ 0.0309$ (from R).
5. **Decision:** Since the p-value < 0.05, we reject H_0. There is strong evidence in the data against H_0 that the distribution of IQ is the same in the three income groups.