HW #8 - SOLUTION

Ex 12.3.2

The null hypothesis is that data are normally distributed.

Sample mean = 127.02 and sample standard deviation = 5.08.

Standard normal rv is $z = (x-127.02)/5.08$

Class interval	O_i	z_i (lower value)	Expected rel freq	E_i	(O_i-E_i)^2/E_i
<114	0		0.0052	1.56	1.56
114-115.99	5	-2.56	0.0098	2.94	1.44
116-117.99	10	-2.17	0.0225	6.75	1.56
118-119.99	14	-1.78	0.0463	13.89	0.00
120-121.99	21	-1.38	0.0773	23.19	0.21
122-123.99	30	-0.99	0.1165	34.95	0.70
124-125.99	40	-0.59	0.1431	42.93	0.20
126-127.99	45	-0.2	0.1546	46.38	0.04
128-129.99	43	0.19	0.1471	44.13	0.03
130-131.99	42	0.59	0.1141	34.23	1.76
132-133.99	30	0.98	0.0782	23.46	1.82
134-135.99	11	1.37	0.0469	14.07	0.67
136-137.99	5	1.77	0.023	6.9	0.52
>138	4	2.16	0.0154	4.62	0.08
sum	300				10.61

Note: expected relative frequencies are computed by computing the probabilities in the corresponding class intervals. For instance:

$P(114<X<116) = P(-2.56<z<-2.17) = \text{pnorm}(-2.17)- \text{pnorm}(-2.56)= 0.0098$

This shows that the test statistic is $X^2 = 10.61$

The probability distribution is χ^2 with df= 14-3=11. Hence for α=0.05 the critical value is 19.675

Since $X^2 < 19.675$, we do not reject the null hypothesis and conclude that data are consistent with the normal distribution.

Ex 12.3.4

With the null hypothesis that data are Poisson distributes with λ=2.8, the expected frequencies are computed as:

$$E\_i = 181\,\frac{e^{-2.8}(2.8)^i}{i!}$$

Using Excel we obtain:

x	O_i	E_i	(O_i-E_i)^2/E_i
0	74	11.01	360.376
1	27	30.82	0.473
2	14	43.15	19.692
3	14	40.27	17.137
4	11	28.19	10.482
5	7	15.79	4.893
6	5	7.37	0.762
7	4	2.95	0.374
8	3	1.03	3.767
9	2	0.32	8.82
10	3	0.09	94.09
11	4	0.02	792.02
12+	13	0.01	16874.01
Sum	181	181.02	17826.52

However, we must aggregate the values of E_i to avoid cells with entries less than 1. Hence the modified table is

x	O_i	E_i	(O_i-E_i)^2/E_i
0	74	11.01	360.376
1	27	30.82	0.473
2	14	43.15	19.692
3	14	40.27	17.137
4	11	28.19	10.482
5	7	15.79	4.893
6	5	7.37	0.762
7	4	2.95	0.374
8+	25	1.47	376.64
Sum	181	181.02	790.829

This shows that the test statistic is X^2 = 790.829

The probability distribution is χ^2 with df= 9-1=8. Hence for α=0.01 the critical value is 20.090

Since X^2 > 20.090, we reject the null hypothesis and conclude that data are not consistent with a Poisson distribution with λ=2.8.

Ex 12.4.2

Contingency table:

	infected	not infected	
aqueous	14	94	108
insoluble	4	97	101
	18	191	209

We compute the test statistic: $X^2 = \frac{209*(14*97-94*4)^2}{18*191*108*101} = 5.374$

We test the null hypothesis that type of skin preparation and infection are independent with significance level α=0.05. Since X^2 is distributed like χ^2 with df=1, the critical value is 3.841.

We <u>reject the null hypothesis</u> since $X^2 > 3.841$.

Ex 12.4.4

Contingency table:

	smoking	non-smoking	
underweight	17	97	114
overweight	25	142	167
appropriate	96	816	912
	138	1055	1193

We compute the test statistic: $X^2 = \sum_{i=1}^{6} \frac{(\mathrm{O_i}-\mathrm{E_i})^2}{\mathrm{E_i}} = 4.103$

Note E_1 = (138)(114)/1193=13.19 , E_2 = (1055)(114)/1193=100.8, E_3 = (138)(167)/1193=19.32,
E_4 = (1055)(167)/1193 = 147.68, E_5 = (138)(912)/1193 = 105.50, E_6 = (1055)(912)/1193=806.50.

We test the null hypothesis that weight perception and smoking are independent with significance level α=0.05. Since X^2 is distributed like χ^2 with df=(3-1)(2-1)=2, the critical value is 5.991.

We <u>fail to reject the null hypothesis</u> since $X^2 < 5.991$.

Ex 12.5.2

Contingency table:

	hispanic	non-hispanic	
married	319	738	1057
divorced/separated	130	329	459
widowed	88	402	490
unmarried	41	95	136
	578	1564	2142

We compute the test statistic: $X^2 = \sum_{i=1}^{8} \frac{(\mathrm{O_i}-\mathrm{E_i})^2}{\mathrm{E_i}} = 26.843$

We test the null hypothesis that marital status in border counties of the Southern US are homogeneous with significance level α=0.05. Since X^2 is distributed like χ^2 with df=(4-1)(2-1)=3, the critical value is 7.815.

We <u>reject the null hypothesis</u> since $X^2 > 7.815$.

Ex 12.7.2

Classification table:

	Survival	No survival	
Conservative treatment	1751	17607	19358
Early revascularization	84	2470	2554
	1835	20077	21912

Relative Risk: $\widehat{RR} = \frac{1751/19358}{84/2554} = 2.75$

$X^2 = \frac{21912*(1751*2470-84*17607)^2}{1835*20077*2554*19358} = 97.44$

95% CI: $(2.25)^{1\pm 1.96/\sqrt{97.44}} = (2.25,3.36)$

Hence, there is evidence that the risk of dying within a year of AMI is higher among subjects receiving conservative treatment when compared to those receiving early revascularization.

Ex 12.7.3

Classification table:

	Premature birth	Regular birth	
Smoking	36	370	406
Non-smoking	168	3396	3564
	204	3766	3970

Odds Ratio: $\widehat{OR} = \frac{36*3396}{168*370} = 1.967$

$X^2 = \frac{3970*(36*3396-168*370)^2}{204*3766*406*3564} = 12.898$

95% CI: $(1.967)^{1\pm 1.96/\sqrt{12.898}} = (1.360, 2.846)$

Hence there is evidence that the odds of premature birth to occur when the mother is smoking during pregnancy is higher than the odds of premature birth to occur when the mother is not smoking during pregnancy.

QUIZ #8

Here is the distribution of the number of girls per family in a sample of 100 families of 5 children.

index	girls	frequency
1	0	2
2	1	10
3	2	31
4	3	36
5	4	17
6	5	4

(a) Test the goodness-of-fit of this data to a uniform distribution. Use \alpha =0.05

```
> observed <-c(2,10,31,36,17,4)
> expected <-c(1/6,1/6,1/6,1/6,1/6,1/6)
> chisq.test(x=observed, p=expected)
```

Chi-squared test for given probabilities
data: observed
X-squared = 59.96, df = 5, p-value = 1.239e-11

Conclusion: Since p-value is less than 0.05, **we reject the null hypothesis** that data are uniformly distributed.

(b) Test the goodness-of-fit of this data to a binomial distribution with p=0.5. Use \alpha =0.05

```
> x <- 0:5
> expected = dbinom(x, size = 5, prob = 0.5)
> observed <-c(2,10,31,36,17,4)
> chisq.test(x=observed, p=expected)
```

Chi-squared test for given probabilities

data: observed
X-squared = 3.52, df = 5, p-value = 0.6204

Conclusion: Since p-value = 0.6204, **we accept the null hypothesis** that data satisfy a binomial distribution at significance level 0.05.

2) Here is the contingency table:

	smoking	non-smoking
underweight	10	97
overweight	26	142
appropriate	121	812
	157	1051

Solution

E1= 157x107/1208 =13.91

E2= 1051x107/1208 =93.09

E3= 157x168/1208 =21.83

E4= 1051x168/1208 =146.17

E5= 157x933/1208 =121.26

E6= 1051x933/1208 =811.74

Hence: $X^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i} = 2.179466$

Since X^2 is distributed like χ^2 with df=(3-1)(2-1)=2, the critical value is

> qchisq(0.90,2) = 4.605

Since $X^2 < 4.605$, we fail to reject the null hypothesis, hence there is no sufficient evidence to reject the hypothesis that weight perception and smoking habit are independent.

Alternative solution

In R, it is sufficient to create a table and apply the test

```
> table <- cbind(c(10,26,121),c(97,142,812))
[or:  table <-matrix(c(10,26,121,97,142,812),ncol=2)]
> chisq.test(table)

        Pearson's Chi-squared test

data:  table
X-squared = 2.1753, df = 2, p-value = 0.337
```

Since p-value > 0.10, we **fail to reject the null hypothesis**, hence there is no sufficient evidence to reject the hypothesis that weight perception and smoking habit are independent.