HW9 Solution

10.1.3 To check whether or not a relationship exists between Y and X we calculate the conditional distributions of Y given X. These are given in the following table.

	$Y = 1$	$Y = 2$	$Y = 3$
$X = 1$	$0.15/0.73 = .20548$	$0.18/0.73 = .24658$	$0.40/0.73 = .54795$
$X = 2$	$0.12/0.27 = .44444$	$0.09/0.27 = .33333$	$0.06/0.27 = .22222$

The conditional distribution of Y given $X = x$ does change as we change x, so we conclude that X and Y are related.

10.2.2 First note that the predictor variable, X (received vitamin C or not), is deterministic. The estimated conditional distributions of Y given X are recorded in the following table.

	No cold	Cold
Placebo	.22143	.77857
Vitamin C	.12230	.87770

Under the null hypothesis of no relationship between taking vitamin C and the incidence of the common cold, the MLE's are given by

$$\hat{\theta}_1 = \frac{48}{279} = .17204, \ \hat{\theta}_2 = \frac{231}{279} = .82796.$$

Then the estimates of the expected counts $n_i\theta_j$ are given in the following table.

	No cold	Cold
Placebo	24.086	115.91
Vitamin C	23.914	115.09

The Chi-squared statistic is equal to $X_0^2 = 4.8105$ and, with $X^2 \sim \chi^2(1)$, the P-value equals $P\left(X^2 > 4.8105\right) = .02829$. Therefore, we have evidence against the null hypothesis of no relationship between taking vitamin C and the incidence of the common cold.

10.2.3 The estimated conditional distributions of Y (second digit) given X (first digit) are recorded in the following table.

	Second digit 0	Second digit 1
First digit 0	0.489796	0.510204
First digit 1	0.500000	0.500000

Under the null hypothesis of no relationship between the digits, the MLE's are given by

$$\hat{\theta}_{.1} = \frac{495}{1000} = .495, \; \hat{\theta}_{.2} = \frac{505}{1000} = .505$$

for the Y probabilities and

$$\hat{\theta}_{1.} = \frac{490}{1000} = .49, \; \hat{\theta}_{.2} = \frac{510}{1000} = .51$$

for the X probabilities. Then the estimates of the expected counts $n_i \theta_{i\cdot} \theta_{\cdot j}$ are given in the following table.

	Second digit 0	Second digit 1
First digit 0	242.55	247.45
First digit 1	252.45	257.55

The Chi-squared statistic is then equal to $X_0^2 = .10409$ and, with $X^2 \sim \chi^2(1)$, the P-value equals $P\left(X^2 > 0.104092\right) = .74698$. Therefore, we have no evidence against the null hypothesis of no relationship between the two digits.

10.2.5

(a) First, note that the predictor variable, X (gender), is not random. The estimated conditional distributions of Y given X are given in the following table.

	$Y=$ fair	$Y=$ red	$Y=$ medium	$Y=$ dark	$Y=$ jet black
$X=$ m	0.281905	0.0566667	0.404286	0.240000	0.0171429
$X=$ f	0.305104	0.0544027	0.379697	0.252944	0.0078519

Under the null hypothesis of no relationship between hair color and gender, the MLE's are given by

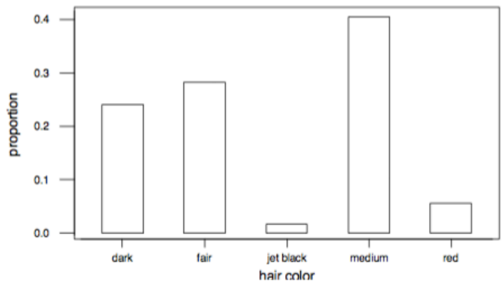
$$\hat{\theta}_1 = \frac{1136}{3883} = .292557,\ \hat{\theta}_2 = \frac{216}{3883} = .055627,\ \hat{\theta}_3 = \frac{1526}{3883} = .0.392995,$$
$$\hat{\theta}_4 = \frac{955}{3883} = .245944,\ \hat{\theta}_5 = \frac{50}{3883} = 0.012877.$$

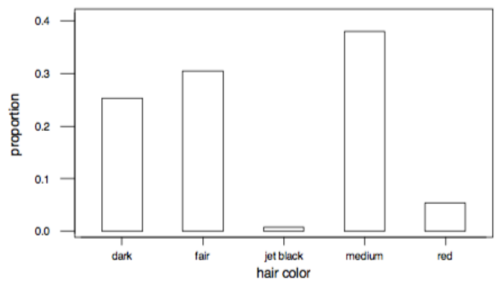
Then the estimates of the expected counts $n_i\theta_j$ are given in the following table.

	$Y=$ fair	$Y=$ red	$Y=$ medium	$Y=$ dark	$Y=$ jet black
$X=$ m	614.370	116.817	825.290	516.482	27.041
$X=$ f	521.630	99.183	700.710	438.518	22.959

The Chi-squared statistic is then equal to $X_0^2 = 10.4674$ and, with $X^2 \sim \chi^2(4)$, the P-value equals $P\left(X^2 > 10.4674\right) = .03325$. Therefore, we have some evidence against the null hypothesis of no relationship between hair color and gender.

(b) The appropriate bar plots are the two conditional distributions and these are plotted as follows for males and then females.





(c) The standardized residuals are given in the following table. They all look reasonable, so nothing stands out as an explanation of why the model of independence doesn't fit. Overall, it looks like a large sample size has detected a small difference.

	$Y = \text{fair}$	$Y = \text{red}$	$Y = \text{medium}$	$Y = \text{dark}$	$Y = \text{jet black}$
$X = \text{m}$	-1.07303	0.20785	1.05934	-0.63250	1.73407
$X = \text{f}$	1.16452	-0.22557	-1.14966	0.68642	-1.88191