ST 430
Homework #11

10.4.6 Below is the set of observed and expected frequencies, the latter based on the null hypothesis that the states' SAT scores are normally distributed with $\bar{y} = 949.4$ and $s = 68.4$. With $t = 4$ classes and two estimated parameters, H_0 should be rejected if $d_1 \geq \chi^2_{.95,4-1-2} = 3.841$. For these data,

$$d_1 = \frac{(18-12.2)^2}{12.2} + \frac{(10-13.7)^2}{13.7} + \frac{(6-13.5)^2}{13.5} + \frac{(17-11.6)^2}{11.6} = 10.44,$$ suggesting that the normality assumption is unwarranted.

Range	Frequency	Probability	Expected Frequency
≤ 900	18	0.2389	12.2
901-950	10	0.2691	13.7
950-1000	6	0.2654	13.6
≥1001	17	0.2266	11.6
	51	1.0000	51.0

10.4.8 The table below gives the observed frequencies for 100 supposedly random choices from the [0, 1] interval, as well as the expected values of 10 for each category. With 10 classes and no parameters estimated, H_0 should be rejected if $d_1 \geq \chi^2_{.95,10-1} = 16.919$. For these data,

$$d_1 = \frac{(12-10)^2}{10} + \frac{(9-10)^2}{10} + \cdots + \frac{(8-10)^2}{10} = 1.8$$

We can accept the null hypothesis that the data come from a uniform pdf over [0, 1].

Interval	Observed Frequency	Expected Frequency
.000-.099	12	10
.100-.199	9	10
.200-.299	11	10
.300-.399	8	10
.400-.499	11	10
.500-.599	10	10
.600-.699	11	10
.700-.799	9	10
.800-.899	11	10
.900-.999	8	10
	100	100

10.4.10 Take $\hat{\lambda}$ to be the mean of the data or 0.363. The model to be fit, then, is the Poisson pdf with parameter 0.363. The table gives the observed frequencies, the estimated probabilities and the estimated frequencies. Note that the last three classes should be collapsed, giving a total of three classes. With one parameter estimated, we should reject H_0 if $d_1 \geq \chi^2_{.95,3-1-1} = 3.841$. The data gives

$$d_1 = \frac{(82-78.6)^2}{78.6} + \frac{(25-28.5)^2}{28.5} + \frac{(6-5.9)^2}{5.9} = 0.58$$

and we can accept the Poisson model for these data.

No. of years	Frequency	$\hat{p}_i$	$113 \cdot \hat{p}_i$
0	82	0.6956	78.6
1	25	0.2525	28.5
2	4	0.0458	5.2
3	0	0.0055	0.6
4	2	0.0006	0.1
		1.0000	113.0

10.5.2 At the $\alpha = .05$ level, H_0: Type of company and importance of work force are independent is rejected if $d_2 \geq \chi^2_{.95,(2-1)(2-1)} = 3.841$. But $d_2 = \frac{(168-163.79)^2}{163.79} + \ldots + \frac{(26-21.79)^2}{21.79} = 1.54$, so H_0 is not rejected.

	Manufacturing	Other	
Important	168 (163.79)	73 (77.21)	241
Not Important	42 (46.21)	26 (21.79)	68
	210	99	309

10.5.6 Let $\alpha = 0.05$. To test H_0: Children's blood pressures are independent of their parent's blood pressures versus H_1: Children's blood pressures are not independent of their parent's blood pressures, reject the null hypothesis if $d_2 \geq \chi^2_{.95,(3-1)(3-1)} = 9.488$. Here,

$d_2 = \frac{(14-11.12)^2}{11.12} + \ldots + \frac{(12-8.83)^2}{8.83} = 3.81$, so H_0 would not be rejected. Based on these data, attempts to use one group to screen for high-risk individuals in the other group are not likely to be successful.

		Child's blood pressure			
		Lower	Middle	Upper	
Father's blood Pressure	Lower	14 (11.12)	11 (11.48)	8 (10.40)	33
	Middle	11 (10.45)	11 (10.78)	9 (9.77)	31
	Upper	6 (9.43)	10 (9.74)	12 (8.83)	28
		31	32	29	92

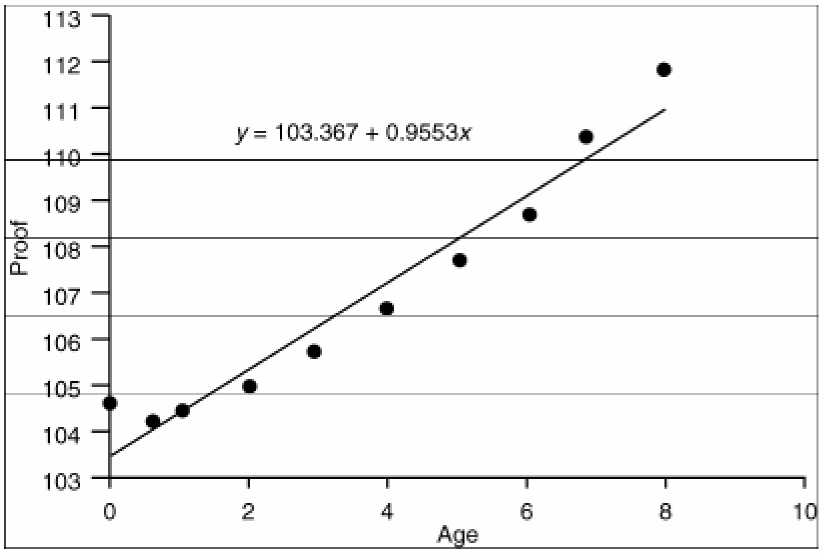
10.5.8 The null hypothesis that enrollment rates are independent of racial groups is rejected at the $\alpha = 0.05$ level if $d_2 \geq \chi^2_{.95,(4-1)(2-1)} = 7.815$. For these data,

$d_2 = \frac{(2592-2622.49)^2}{2622.49} + \ldots + \frac{(399-379.63)^2}{379.63} = 10.29$, implying that the differences in enrollment rates from race to race are statistically significant.

	Admitted	Enrolled	
White	2592 (2622.49)	1481 (1450.51)	4073
Af.-Amer.	159 (152.60)	78 (84.40)	237
Hispanic	800 (756.55)	375 (418.45)	1175
Asian	667 (686.37)	399 (379.63)	1066
	4218	2333	6551

11.2.2 $b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{10(3973.35) - (36.5)(1070)}{10(204.25) - (36.5)^2} = 0.9953$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n} = \frac{1070 - 0.9953(36.5)}{10} = 103.367$$



11.2.4 In the first graph, all of the residuals are positive. The residuals in the second graph alternate from positive to negative. Neither graph would normally occur from linear models.

11.2.6. The problem here is the gap in x values, leaving some doubt as to the x–y relationship.

11.2.12 Using Cramer's rule we obtain

$$b = \frac{\begin{vmatrix} n & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i y_i \end{vmatrix}}{\begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix}} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)}$$

which is essentially the form of b in Theorem 11.2.1. The first row of the matrix equation is $na + \left(\sum_{i=1}^{n} x_i\right)b = \sum_{i=1}^{n} y_i$. Solving this equation for a in terms of b gives the expression in Theorem 11.2.1 for a.