Suggested Problems Ch10

Retinitis pigmentosa is a disease that manifests itself via different genetic modes of inheritance. Cases have been documented with a dominant, recessive, and sex-linked mode of inheritance. It has been conjectured that mode of inheritance is related to the ethnic origin of the individual. Cases of the disease have been surveyed in an English and a Swiss population with the following results: Of 125 English cases, 46 had sex-linked disease, 25 had recessive disease, and 54 had dominant disease. Of 110 Swiss cases, 1 had sex-linked disease, 99 had recessive disease, and 10 had dominant disease.

*10.15 Do these data show a significant association between ethnic origin and genetic type?

10.15 We display the data in the form of a 2 × 3 contingency table as follows:

		genetic type			
		Sex linked	**recessive**	**dominant**	
ethnicity	**Engl.**	46	25	54	125
	Swiss	1	99	10	110
		47	124	64	235

We then compute the table of expected values where

$$E_{ij} = \frac{(\text{Row Total})_i \times (\text{Column Total})_j}{n}$$

We have

		genetic type		
		Sex linked	recessive	dominant
ethnicity	Engl.	25.00	65.96	34.04
	Swiss	22.00	58.04	29.96

Since none of the expected values are small, we can use the chi-square test.
We have the test statistic

$$X^2 = \sum_{i,j} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

$$= \frac{\left(46-25\right)^2}{25} + \cdots + \frac{\left(10-29.96\right)^2}{29.96} = 117.02$$

Under the null hypothesis of no association between ethnicity and genetic type, we have that $X^2 \sim \chi_2^2$. Thus, we compare X^2 with $\chi^2_{2,.95} = 5.99$. Clearly, $p < .05$. Indeed, $x^2_{2,.999} = 13.81 < X^2$ and thus, $p < .001$. Therefore, we reject H_0 and conclude that there is a significant association between ethnicity and genetic type.

TABLE 10.22 Number of children with otorrhea at 2 weeks of follow-up

Group	Number of children	Number of children with otorrhea at 2 weeks
Antibiotic ear drops	76	4
Oral antibiotics	77	34
Observation	75	41

10.21 What test can be used to compare the prevalence of otorrhea for the ear drop group vs. the observation group? State the hypotheses to be tested.

10.22 Perform the test in Problem 10.21 and report a *p*-value (two-tailed). Interpret your results in words.

10.21 We wish to test the hypothesis:

$H_0: p_1 = p_2$ vs. $H_1: p_1 \neq p_2$, where

$p_1 = Prob$(otorrhea) for the ear drop group,

$p_2 = Prob$(otorrhea) for the observation group.

10.22 We form the 2 x 2 table relating outcome to group as follows:

Observed Table

Group	Otorrhea 2 weeks = yes	Otorrhea 2 weeks = no	Total
Ear drop	4	72	76
Observation	41	34	75
Total	45	106	151

The expected counts under the null hypothesis are as follows:

$E_{11} = 76(45)/151 = 22.65$,

$E_{12} = 76 - 22.65 = 53.35$,

$E_{21} = 45 - 22.65 = 22.35$,

$E_{22} = 75 - 22.35 = 52.65$.

Thus, the expected table is as follows:

Expected Table

Group	Otorrhea 2 weeks = yes	Otorrhea 2 weeks = no	Total
Ear drop	22.65	53.35	76
Observation	22.35	52.65	75
Total	45	106	151

Since all expected counts are >= 5, we can use the chi-square test for 2 x 2 tables.

We have the chi-square statistic:

$$X^2_{corr} = \frac{\left(|4-22.65|-0.5\right)^2}{22.65} + \frac{\left(|72-53.35|-0.5\right)^2}{53.35} + \frac{\left(|41-22.35|-0.5\right)^2}{22.35} + \frac{\left(|34-52.65|-0.5\right)^2}{52.65}$$

$$= \frac{18.15^2}{22.65} + \frac{18.15^2}{53.35} + \frac{18.15^2}{22.35} + \frac{18.15^2}{52.65}$$

$$= 14.54 + 6.17 + 14.74 + 6.26$$

$$= 41.71 \sim \chi_1^2 \text{ under } H_0.$$

Since $\chi^2_{1,.999} = 10.83 < 41.71$, it follows that $p < 0.001$.

Thus, there is a highly significant difference in prevalence between the 2 groups with the ear drop group having a significantly lower prevalence of otorrhea at two weeks than the observation group.

TABLE 10.24 **Five-year survival rates for breast cancer by stage at diagnosis, age at diagnosis, and race, SEER Cancer data, 1999–2005**

	Caucasian females		African American females	
Stage	<50 (n = 53,060)	50+ (n = 174,080)	<50 (n = 8063)	50+ (n = 16,300)
Localized	96.5*	99.6	91.6	94.9
Regional	84.6	85.0	71.3	72.6
Distant	33.2	22.5	15.0	16.4
Unstaged	76.7	53.5	49.7	42.2

10.23 Test whether the distribution of stage of disease is significantly different between Caucasian and African American women with breast cancer who are younger than 50 years of age. Please provide a p-value (two-tailed). Ignore the unstaged cases in your analysis.

The 5-year survival rates by stage of disease, age at diagnosis, and race are provided in Table 10.24.

10.24 Test whether the 5-year survival rate for breast cancer is significantly different between African American and Caucasian women who are younger than 50 years of age and have localized disease. Provide a p-value (two-tailed).

10.23 We will use a Chi-Sq test for heterogeneity. Because Table 10.23 does not give count data, we must calculate the counts for each cell by multiplying the given probability by the appropriate n for each group. Below we show the observed and expected counts for each cell.

	Caucasian		African-American		
Stage	Obs. Proportion	Obs. Count	Obs. Proportion	Obs. Count	Total
Localized	0.54	28652.4	0.46	3708.98	32361.38
Regional	0.41	21754.6	0.46	3708.98	25463.58
Distant	0.03	1591.8	0.07	564.41	2156.21
Total	1	51998.8	1.01	7982.37	59981.17

Expected Count

	Caucasian	African-American
Localized	28055.44	4305.94
Regional	22075.45	3388.13
Distant	1869.31	286.90

We then calculate the test statistic $X^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

$$X^2 = \frac{(28652.4 - 28{,}055.44)^2}{28{,}055.44} + \ldots + \frac{(564.41 - 286.95)^2}{286.95}$$

$$= 12.70 + \ldots + 268.28 = 440.15 \sim \chi_2^2 \text{ under } H_0.$$

Since $\chi^2_{2,0.999} = 13.82$, it follows that p<0.001. Therefore, there is a significant difference in distribution of disease stage between young Caucasian and African-American women. African-American women under 50 have more advanced disease than Caucasian women under 50.

10.24 We will use a two-sample test for binomial proportions. From 10.23, we estimate that 28,652 young Caucasian women have localized disease, and that 28,652 x 0.965 = 27,649 survived. Among African-American women under age 50, we estimate that 3,709 had localized disease, of which 3,709x0.916 = 3,397 survived. We use the test statistic

$$Z = \frac{\left|\hat{p}_1 - \hat{p}_2\right| - \left(\frac{1}{2N_1} + \frac{1}{2N_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} \sim N(0,1)$$

where

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

$$= \frac{28652(.965) + 3709(.916)}{28652 + 3709} = .959$$

The test statistic is then

$$Z = \frac{|.965 - .916| - \left(\frac{1}{2(28652)} + \frac{1}{2(3709)}\right)}{\sqrt{.959(.041)(1/28652 + 1/3709)}}\ .$$

$$= \frac{.0488}{.0035} = 13.94 \sim N(0,1) \text{ under } H_0$$

The 2-tailed p-value $= 2 \times \left[1 - \Phi(13.94)\right] < 0.001$.

Thus, there is a highly significant difference in the survival rate between Caucasian and African-American women under age 50 with localized cancer. The survival rate is significantly higher for Caucasian than African-American women with localized breast cancer under age 50.