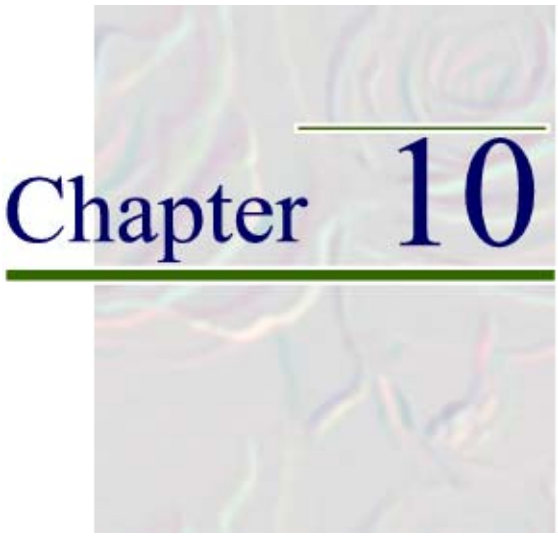



# HYPOTHESIS TESTING: CATEGORICAL DATA

## REVIEW OF KEY CONCEPTS

## SECTION 10.1 Comparison of Two Binomial Proportions

A case-control study was performed among 2982 cases, 5782 controls, from 10 geographic areas of the United States and Canada. The cases were newly diagnosed cases of bladder cancer in 1977–1978 obtained from cancer registries; the control group was a random sample of the population of the 10 study areas with a similar age, sex, and geographical distribution. The purpose of the study was to investigate the possible association between the incidence of bladder cancer and the consumption of alcoholic beverages. Let

$p_1$ = true proportion of drinkers among cases
$p_2$ = true proportion of drinkers among controls

We wish to test the hypothesis $H_0: p_1 = p_2 = p$ versus $H_1: p_1 \neq p_2$.

### 10.1.1 Two-Sample Test for Binomial Proportions (Normal-Theory Version)

In this study, if we define a drinker as a person who consumes $\geq 1$ drink/day of whiskey, then the proportion of drinkers was

$$\frac{574}{2388} = .240 \text{ for the cases } = \hat{p}_1$$
$$\frac{980}{4660} = .210 \text{ for the controls } = \hat{p}_2$$

Not all subjects provided a drinking history, which is why the sample sizes (2388, 4660) in the two groups are less than the total sample sizes in the study (2982, 5782).

We use the test statistic

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

where

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{\text{total number of drinkers over both groups}}{\text{total number of subjects over both groups}}$$

The $p$-value $= 2 \times [1 - \Phi(z)]$. We will only use this test if $n_1\hat{p}\hat{q} \geq 5$ and $n_2\hat{p}\hat{q} \geq 5$. In this case,

$$\hat{p} = \frac{574 + 980}{2388 + 4660} = \frac{1554}{7048} = .220, \quad \hat{q} = 1 - .220 = .780$$

$$n_1\hat{p}\hat{q} = 2388(.220)(.780) = 410.4 \text{ and } n_2\hat{p}\hat{q} = 4660(.220)(.780) = 800.9.$$

Thus, it is valid to use the normal-theory test. We have:

$$z = \frac{.030 - \left[\frac{1}{2(2388)} + \frac{1}{2(4660)}\right]}{\sqrt{.220(.780)\left(\frac{1}{2388} + \frac{1}{4660}\right)}} = \frac{.0298}{.0104}$$
$$= 2.852 \sim N(0, 1) \text{ under } H_0$$
$$p\text{-value} = 2 \times [1 - \Phi(2.852)] = 2(1 - .9978) = .004$$

Thus, the cases report significantly more drinking than the controls.

## SECTION 10.2 The 2 × 2 Contingency-Table Approach

Another technique for the analysis of these data is the contingency-table approach. A $2 \times 2$ *contingency table* is a table where case/control status is displayed along the rows and consumption of hard liquor along the columns, as shown in the following table. A specific row and column combination is called a *cell*, and the number of people in a given cell is called the *cell count*.

| | | Consumption of hard liquor | | |
|---|---|---|---|---|
| | | ≥1/day | <1/day | |
| Case/control status | Case | 574 | 1814 | 2388 |
| | Control | 980 | 3680 | 4660 |
| | | 1554 | 5494 | 7048 |

2388 and 4660 are called the row margins; 1554 and 5494 are called the column margins; 7048 is called the grand total. In general, we use the following notation:

| | | |
|---|---|---|
| (1, 1) cell<br>$a$ | (1, 2) cell<br>$b$ | $a + b = R_1$ |
| (2, 1) cell<br>$c$ | (2,2) cell<br>$d$ | $c + d = R_2$ |
| $a + c = C_1$ | $b + d = C_2$ | $n$ |

$(i, j)$ cell = the cell in the $i$th row, $j$th column

The entire table is referred to as the *observed* table. To test for statistical significance, we compare the observed table with what we would expect if there were no association between being a bladder-cancer case and consuming $\geq 1$ drink/day of hard liquor. We wish to test the hypothesis.

$$H_0: p_1 = p_2 = p \text{ versus } H_1: p_1 \neq p_2$$

Under $H_0$, the expected number of units in the $i$th row and $j$th column is

$$E_{ij} = \frac{R_i C_j}{n} = i\text{th row total} \times j\text{th column total}/n$$

In our case,

$$E_{11} = 2388 \times 1554/7048 = 526.5$$
$$E_{12} = 2388 \times 5494/7048 = 1861.5$$
$$E_{21} = 4660 \times 1554/7048 = 1027.5$$
$$E_{22} = 4660 \times 5494/7048 = 3632.5$$

As a check, the sum of the corresponding observed and expected row and column totals should be the same, as they are in this case.

We now wish to compare the observed and expected tables. If they are reasonably close, then we will accept $H_0$, else we will reject $H_0$. The criterion used for agreement in the $(i,\ j)$ cell is

$$\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

For the entire table, we use the test statistic

$$X^2 \equiv \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_1^2 \text{ under } H_0$$

The test statistic $X^2$ is referred to as the uncorrected chi-square statistic for $2 \times 2$ contingency tables. We will reject $H_0$ for large values of $X^2$ and accept $H_0$ for small values of $X^2$. Under $H_0$, $X^2 \sim \chi_1^2$ (only 1 *df*, because there is one independent cell in the table; all others are determined from the row and column totals).

To better approximate the chi-square distribution, we use a continuity correction (this is controversial),

$$X^2_{\text{CORR}} = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(|O_{ij} - E_{ij}| - .5)^2}{E_{ij}} \sim \chi_1^2$$
$$= \text{Yates - corrected chi - square statistic for } 2 \times 2 \text{ contingency tables}$$

Since we only reject for large values of $X^2_{\text{CORR}}$, the *p*-value $= Pr(\chi_1^2 > X^2_{\text{CORR}})$. The test procedure is referred to as the *chi-square test for* $2 \times 2$ *contingency tables*. We only use this test if all expected values are $\geq 5$.

In this example,

Observed table

| | |
|---|---|
| 574 | 1814 |
| 980 | 3680 |

Expected table

| | |
|---|---|
| 526.5 | 1861.5 |
| 1027.5 | 3632.5 |

$$X^2_{\text{CORR}} = \frac{(|574 - 526.5| - 0.5)^2}{526.5} + \frac{(|1814 - 1861.5| - .5)^2}{1861.5} + \frac{(|980 - 1027.5| - .5)^2}{1027.5} + \frac{(|3680 - 3632.5| - .5)^2}{3632.5}$$
$$= \frac{47^2}{526.5} + \frac{47^2}{1861.5} + \frac{47^2}{1027.5} + \frac{47^2}{3632.5} = 4.19 + 1.19 + 2.15 + 0.61 = 8.13 \sim \chi_1^2$$

Since $\chi^2_{1,\,.995} = 7.88$, $\chi^2_{1,\,.999} = 10.83$, and $7.88 < 8.13 < 10.83$, it follows that $1 - .999 < p < 1 - .995$ or $.001 < p < .005$.

### 10.2.1 Relationship Between the Chi-Square Test and the Two-Sample Test for Binomial Proportions

In general,

$$X^2_{\text{CORR}} = z^2_{\text{binomial}}$$

In our case,

$$X^2_{\text{CORR}} = 8.13 = (2.852)^2 = z^2$$

## SECTION 10.3 Fisher's Exact Test

Consider a study of the relationship between early age at menarche (i.e., age at which periods begin) and breast-cancer prevalence. We select 50 premenopausal breast-cancer cases and 50 premenopausal age-matched controls. We find that 5 of the cases have an age at menarche $< 11$ yrs, and 1 control has an age at menarche $< 11$ yrs. Is this a significant finding? We have the following observed and expected contingency tables:

Observed table

| | Age at menarche $< 11$ | Age at menarche $\geq 11$ | |
|---|---|---|---|
| Case | 5 | 45 | 50 |
| Control | 1 | 49 | 50 |
| | 6 | 94 | 100 |

Expected table

| | Age at menarche $< 11$ | Age at menarche $\geq 11$ |
|---|---|---|
| Case | 3.0 | 47.0 |
| Control | 3.0 | 47.0 |

We can't use the chi-square test because two of the expected values are $< 5$. Instead, we must use a method called *Fisher's exact test*. For this test, we consider the margins of the table as fixed and ask the question, How unusual is our table among all tables with the same fixed margins? Consider the following general contingency table:

| | | | |
|---|---|---|---|
| Case | $a$ | $b$ | $a + b$ |
| Control | $c$ | $d$ | $c + d$ |
| | $a + c$ | $b + d$ | $n$ |

Let

$$p_1 = Pr(\text{age at menarche} < 11 \mid \text{case}) = Pr(\text{exposed}|\text{case})$$
$$p_2 = Pr(\text{age at menarche} < 11 \mid \text{control}) = Pr(\text{exposed}|\text{control})$$

We wish to test the hypothesis

$$H_0 : p_1 = p_2 = p \text{ versus } H_1 : p_1 \neq p_2$$

Specifically we wish to assess how unusual is it to have 5 exposed cases and 1 exposed control given that there are a total of 6 exposed and 94 unexposed women and also that there are 50 cases and 50 controls. The exact binomial probability of observing our table given the fixed margins is given by:

$$Pr(a \text{ exposed cases}, c \text{ exposed controls} \mid \text{fixed margins of } a+b,\ c+d,\ a+c, \text{ and } b+d)$$
$$= \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

This is called the *hypergeometric* distribution.

Because the margins are fixed, any table is completely determined by one cell count. We usually refer to the table with cell count $= a$ in the (1, 1) cell as the "$a$" table. In our example, we observed the "5" table. Therefore,

$$Pr(5 \text{ table}) = \frac{50!50!6!94!}{100!5!45!1!49!} = \frac{50 \times 50 \times 49 \times 48 \times 47 \times 46 \times 6}{100 \times 99 \times 98 \times 97 \times 96 \times 95} = \frac{7.628 \times 10^{10}}{8.583 \times 10^{11}} = .089$$

Note that this calculation can also be done using the HYPGEOMDIST function of Excel (see appendix for details). How do we judge the significance of this particular table? We need to enumerate all tables that could have occurred with the same margins, and compute the probability of each such table. These are given as follows:

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 50 | 1 | 49 | 2 | 48 | 3 | 47 | 4 | 46 | 5 | 45 | 6 | 44 |
| 6 | 44 | 5 | 45 | 4 | 46 | 3 | 47 | 2 | 48 | 1 | 49 | 0 | 50 |
| .013 | | .089 | | .237 | | .322 | | .237 | | .089 | | .013 | |

### 10.3.1 Computation of *p*-Values with Fisher's Exact Test

There are two commonly used methods for calculation of two-tailed $p$-values, as follows:

1. $p\text{-value} = 2 \times \min\left[\sum_{i=0}^{a} Pr(i),\ \sum_{i=a}^{a+b} Pr(i),\ .5\right]$
2. $p\text{-value} = \sum_{\{i:\ Pr(i) \le Pr(a)\}} Pr(i) =$ sum of probabilities with probabilities ≤ probability of observed table.

In this case, we will use the first approach:

$$p\text{-value(2-tail)} = 2 \times \min\left[\sum_{i=0}^{5} Pr(i),\ \sum_{i=5}^{6} Pr(i),\ .5\right] = 2 \times (.987,\ .102,\ .5) = .204$$

Thus, there is no significant relationship between early menarche and breast cancer.

In general, we only need to use Fisher's exact test if at least one cell has expected value $< 5$. However, it is always a valid test, but is more tedious than the chi-square test.

## SECTION 10.4 McNemar's Test for Correlated Proportions

A case-control study was performed to study the relationship between the source of drinking water during the prenatal period and congenital malformations. Case mothers were those with malformed infants in a registry in Australia between 1951 and 1979. Controls were individually matched by hospital, maternal age (± 2 years), and date of birth (± 1 month). The suspected causal agent was groundwater nitrates. The following $2 \times 2$ table was obtained relating case-control status to the source of drinking water:

| | Source of Drinking Water | | | Percentage of |
|---|---|---|---|---|
| | Groundwater | Rainwater | Total | groundwater |
| Cases | 162 | 56 | 218 | 74.3% |
| Controls | 123 | 95 | 218 | 56.4% |
| Total | 285 | 151 | 436 | |

The corrected chi-square statistic $= X^2 = 14.63$, $p < .001$.

However, the assumptions of the $\chi^2$ test are not valid, because the women in the two samples were individually matched and are not statistically independent. We instead must analyze the data in terms of *matched pairs*. The following table gives the exposure status of case-control pairs.

| Case | Control | Frequency |
|---|---|---|
| + | + | 101 |
| + | – | 61 |
| – | + | 22 |
| – | – | 34 |

*Note*: + = groundwater, − = rainwater.

We refer to the (+, +) and (–, –) pairs as *concordant pairs*, since the case and control members of the pair have the same exposure status. We refer to the (+, –) and (–, +) pairs as *discordant pairs*. For our test, we ignore the concordant pairs and only focus on the discordant pairs. Let

$$n_A = \text{ the number of (+, –) or type A discordant pairs}$$
$$n_B = \text{ the number of (–, +) or type B discordant pairs}$$
$$n_D = n_A + n_B = \text{ total number of discordant pairs}$$

We wish to test the hypothesis $H_0: p = 1/2$ versus $H_1: p \neq 1/2$, where $p =$ prob(discordant pair is of type A). If $n_A + n_B \geq 20$, then we can use the normal-theory test. We use the test statistic

$$X^2 = \frac{\left(\left|n_A - \frac{n_D}{2}\right| - \frac{1}{2}\right)^2}{\left(\frac{n_D}{4}\right)} \sim \chi_1^2$$
$$p\text{-value} = Pr(\chi_1^2 > X^2)$$

In this case,

$$n_A = 61,\ n_B = 22,\ n_D = 83$$
$$X^2 = \frac{\left(\left|61 - \frac{83}{2}\right| - \frac{1}{2}\right)^2}{\frac{83}{4}} = 17.40 \sim \chi_1^2$$

Since $\chi^2_{1,\,.999} = 10.83 < X^2$, we obtain $p < .001$.

Thus, there is a significant association between the source of drinking water and the occurrence of congenital malformations.

The data were also analyzed separately by season of birth. The following exposure data are presented in a $2 \times 2$ table of case exposure status by control exposure status for spring births.

| | | Control + | Control – |
|---|---|---|---|
| Case | + | 30 | 14 |
| | – | 2 | 10 |

Since the number of discordant pairs $= n_A + n_B = 14 + 2 = 16 < 20$, we cannot use the normal-theory test. Instead, we must use an exact binomial test. To compute the *p*-value, we have

$$p = 2 \times \left[\sum_{k=0}^{n_A} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D}\right] \text{ if } n_A < \frac{n_D}{2}$$
$$= 2 \times \sum_{k=n_A}^{n_D} \binom{n_D}{k} \left(\frac{1}{2}\right)^{n_D} \text{ if } n_A > \frac{n_D}{2}$$
$$= 1 \text{ if } n_A = \frac{n_D}{2}$$

In this case $n_A = 14 > \frac{n_D}{2} = 8$. Therefore,

$$p\text{-value} = 2 \times \sum_{k=14}^{16} \binom{16}{k} \left(\frac{1}{2}\right)^{16}$$

From Table 1 (Appendix, text), under $n = 16$, $p = .50$ we have

$$p\text{-value} = 2(.0018 + .0002 + .000) = 2 \times .002 = .004$$

Thus, there is a significant association for the subset of spring births as well.

## SECTION 10.5 Sample Size for Comparing Two Binomial Proportions

To test the hypothesis, $H_0$: $p_1 = p_2$ versus $H_1$: $p_1 \neq p_2$, $|p_1 - p_2| = \Delta$ with significance level $\alpha$ and power $= 1 - \beta$ with an equal sample size per group, we need

$$n_1 = \frac{\left[\sqrt{2\bar{p}\bar{q}}\left(z_{1-\alpha/2}\right) + \sqrt{p_1q_1 + p_2q_2}\left(z_{1-\beta}\right)\right]^2}{\Delta^2} = n_2$$

subjects per group where $\bar{p} = (p_1 + p_2)/2$, $\bar{q} = 1 - \bar{p}$.

---

*Example*: A study is being planned among postmenopausal women to investigate the effect on breast-cancer incidence of having a family history of breast cancer. Suppose that a 5-year study is planned and it is expected that the 5-year incidence rate of breast cancer among women without a family history is 1%, while the 5-year incidence among women with a family history is 2%. If an equal number of women per group are to be studied, then how many women in each group should be enrolled to have an 80% chance of detecting a significant difference using a two-sided test with $\alpha = .05$?

In this example, $\alpha = .05$, $z_{1-.05/2} = z_{.975} = 1.96$, $1 - \beta = .8$, $z_{.8} = 0.84$, $p_1 = .01$, $q_1 = .99$, $p_2 = .02$, $q_2 = .98$, $\bar{p} = (.01 + .02)/2 = .015$, $\bar{q} = .985$, $\Delta = .01$. Therefore, we need

$$n = \frac{\left[\sqrt{2(.015)(.985)}(1.96) + \sqrt{.01(.99) + .02(.98)}(0.84)\right]^2}{.01^2}$$

$$= \frac{[.1719(1.96) + .1718(.84)]^2}{.0001} = \frac{(.4812)^2}{.0001} = 2315.5$$

Therefore, we need to study 2316 subjects in each group to have an 80% chance of finding a significant difference with this number of subjects. Over 5 years we would expect about 23 breast-cancer cases among those women without a family history and 46 cases among those women with a family history.

---

The sample-size formula can also be modified to allow for an unequal number of subjects per group—see Equation 10.14, in Chapter 10, text.

Suppose the study is performed, but only 2000 postmenopausal women per group are enrolled. How much power would such a study have? The general formula is given as follows:

$$\text{Power} = \Phi\left[\frac{\Delta}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}} - z_{1-\alpha/2}\frac{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}}\right]$$

where

$$\bar{p} = \frac{n_1p_1 + n_2p_2}{n_1+n_2},\ \bar{q} = 1-\bar{p}$$

In this example, the power is given by

$$\text{Power} = \Phi\left[\frac{.01}{\sqrt{\frac{.01(.99)}{2000}+\frac{.02(.98)}{2000}}} - 1.96\frac{\sqrt{.015(.985)\left(\frac{1}{2000}+\frac{1}{2000}\right)}}{\sqrt{\frac{.01(.99)}{2000}+\frac{.02(.98)}{2000}}}\right]$$
$$= \Phi\left[\frac{.01}{.003841} - 1.96\left(\frac{.003844}{.003841}\right)\right] = \Phi(2.604-1.962) = \Phi(0.642) = .74$$

Thus, the study would have 74% power.

## SECTION 10.6 *r* × *c* Contingency Tables

Patients with heart failure, diabetes, cancer, and lung disease who have various infections from gram-negative organisms often receive aminoglycosides. One of the side effects of aminoglycosides is nephrotoxicity (possible damage to the kidney). A study was performed comparing the nephrotoxicity (rise in serum creatinine of at least 0.5 mg/dL) for 3 aminoglycosides. The following results were obtained:

| | +* | Total | %* |
|---|---|---|---|
| Gentamicin (G) | 44 | 121 | 36.4 |
| Tobramycin (T) | 21 | 92 | 22.8 |
| Amikadn (A) | 4 | 16 | 25.0 |

* + = number of patients with a rise in serum creatinine of ≥ 0.5 mg/dL

Are there significant differences in nephrotoxicity among the 3 antibiotics?

We can represent the data in the form of a $2 \times 3$ contingency table (2 rows, 3 columns) as follows:

| | | Antibiotic | | | |
|---|---|---|---|---|---|
| | | G | T | A | |
| Nephrotoxicity | + | 44 | 21 | 4 | 69 |
| | − | 77 | 71 | 12 | 160 |
| | | 121 | 92 | 16 | 229 |

We wish to test the hypothesis $H_0$: no association between row and column classifications versus $H_1$: some association between row and column classifications. Under $H_0$, the expected number of units in the *i*th row and *j*th column is $E_{ij}$, given by

$$E_{ij} = \frac{R_iC_j}{N}$$

where $R_i = i$ th row total, $C_j = j$ th column total, and $N =$ grand total. We use the test statistic

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim \chi^2_{(r-1)\times(c-1)} \text{ under } H_0$$

$$p\text{-value} = Pr\left(\chi^2_{(r-1)\times(c-1)} > X^2\right)$$

We only use this test if no more than 1/5 of the expected values are $< 5$ and no expected value is $< 1$. We have the following expected cell counts:

$$E_{11} = \frac{69(121)}{229} = 36.5$$

$$E_{12} = \frac{69(92)}{229} = 27.7$$

etc.

The complete observed and expected tables are given as follows:

Expected table

| | | G | T | A |
|---|---|---|---|---|
| Nephrotoxicity | + | 36.5 | 27.7 | 4.8 |
| | – | 84.5 | 64.3 | 11.2 |

Observed table

| | | G | T | A |
|---|---|---|---|---|
| Nephrotoxicity | + | 44 | 21 | 4 |
| | – | 77 | 71 | 12 |

Only 1 of the 6 cells has expected value $< 5$ and no cell has an expected value $< 1$. Thus, we can use the chi-square test. We have the test statistic

$$X^2 = \frac{(44-36.5)^2}{36.5} + \frac{(21-27.7)^2}{27.7} + \frac{(4-4.8)^2}{4.8} + \frac{(77-84.5)^2}{84.5} + \frac{(71-64.3)^2}{64.3} + \frac{(12-11.2)^2}{11.2}$$

$$= \frac{7.5^2}{36.5} + \frac{6.7^2}{27.7} + \frac{0.8^2}{4.8} + \frac{7.5^2}{84.5} + \frac{6.7^2}{64.3} + \frac{0.8^2}{11.2}$$

$$= 1.56 + 1.63 + 0.14 + 0.67 + 0.70 + 0.06 = 4.76 \sim \chi^2_2 \text{ under } H_0.$$

Since $\chi^2_{2,\,.95} = 5.99 > 4.76$ it follows that $p > .05$. Thus, there are no significant differences in the rate of nephrotoxicity among the 3 antibiotics.

In the preceding example, the different antibiotics form a **nominal scale**; i.e., there is no specific ordering among the three antibiotics. For some exposures, there is an implicit ordering. For example, suppose we wish to relate the occurrence of bronchitis in the first year of life to the number of cigarettes per day smoked by the mother. If we focus on smoking mothers and categorize the amount smoked by (1–4/5–14/15–24/25–44/45+) cigarettes per day, then we might construct a $2 \times 5$ contingency table as follows:

| | | Number of cigarettes per day by mother | | | | |
|---|---|---|---|---|---|---|
| | | 1–4 | 5–14 | 15–24 | 25–44 | 45+ |
| Bronchitis in 1st year of life | + | | | | | |
| | – | | | | | |

We could perform the chi-square test for $r \times c$ tables given above (sometimes known as the "chi-square test for heterogeneity"). However, this is equivalent to testing the hypothesis $H_0: p_1 = p_2 = \ldots = p_5$ versus $H_1$: at least two of the $p_i$'s are unequal, where $p_i$ = probability of bronchitis in the $i$th smoking group. However, we would expect if there is a "dose-response" relationship between bronchitis and cigarette smoking that $p_i$ should increase as the number of cigarettes per day increases. One way to test this hypothesis is to test $H_0: p_i$ all equal versus $H_1: p_i = \alpha + \beta S_i$, where $S_i$ is a score variable attributable to the $i$th smoking group. There are different score variables that could be used. A common

choice is to use $S_i = i$; i.e., $p_i = \alpha + \beta i$. In this case, $\beta$ is interpreted as the increase in the probability of bronchitis for an increase of 1 cigarette-smoking group (e.g., from 1–4 to 5–14 cigarettes per day). To test this hypothesis, we use the **chi-square test for trend**. See Equation 10.24, in Chapter 10 of the text, for details on the test procedure. This test is often more useful for establishing dose-response relationships in $2 \times k$ tables than the chi-square test for heterogeneity.

## SECTION 10.7 Chi-Square Goodness-of-Fit Test

Let us look at the distribution of serum-cholesterol changes presented in Table 2.1 (in Chapter 2, Study Guide). How well does a normal distribution fit these data? A stem-and-leaf plot of the change scores is given as follows:

| Stem-and-leaf plot of cholesterol change | |
|---|---|
| 4 | 981 |
| 3 | 6215 |
| 2 | 7183 |
| 1 | 3969932 |
| 0 | 828 |
| –0 | 8 |
| –1 | 03 |

The arithmetic mean $= 19.5$ mg/dL, sd $= 16.8$ mg/dL, $n = 24$. Under $H_0$,

$$X_i \sim N\left(\mu,\ \sigma^2\right)$$
$$\hat{\mu} = 19.8$$
$$\hat{\sigma}^2 = 16.8^2$$

The general approach is to divide the distribution of change scores into *k* groups and compute the observed and expected number of units in each group if a normal distribution holds as shown in the table.

| Observed count | | Expected count |
|---|---|---|
| $O_1$ | | $E_1$ |
| | ⋮ | |
| $O_k$ | | $E_k$ |

We then compute the test statistic

$$X^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i^2\right)}{E_i} \sim \chi^2_{g-1-k}$$

where

$g$ = number of groups
$k$ = number of parameters estimated from the data

The *p*-value = $Pr\left(\chi^2_{g-1-k} > X^2\right)$. We will only use this test if no more than 1/5 of the expected cell counts are $< 5$. This test is referred to as the *chi-square goodness-of-fit test*.

To group the change scores, we will use the groups ($\leq$ 9/10–19/20–29/30+). The observed and expected counts for each group are given as follows:

| | Obs | Exp |
|---|---|---|
| $\leq 9$ | 6 | 6.6 |
| 10–19 | 7 | 5.4 |
| 20–29 | 4 | 5.4 |
| 30+ | 7 | 6.6 |

To compute the expected values, we employ a continuity correction. Thus, $X \leq 9$ is actually $Y \leq 9.5$, where $Y$ is the normal approximation. The expected probabilities within each group are given as follows:

$$Pr(X \leq 9) = \Phi\left(\frac{9.5 - 19.5}{16.8}\right) = \Phi\left(\frac{-10}{16.8}\right) = \Phi(-0.60) = .275$$

$$Pr(10 \leq X \leq 19) = \Phi\left(\frac{19.5 - 19.5}{16.8}\right) - \Phi\left(\frac{9.5 - 19.5}{16.8}\right) = .499 - .275 = .224$$

$$Pr(20 \leq X \leq 29) = \Phi\left(\frac{29.5 - 19.5}{16.8}\right) - .499 = .723 - .499 = .224$$

$$Pr(X \geq 30) = 1 - \Phi\left(\frac{29.5 - 19.5}{16.8}\right) = 1 - \Phi\left(\frac{10}{16.8}\right) = .277$$

The expected count within each group is

$$E_1 = 24 \times .275 = 6.6$$
$$E_2 = 24 \times .224 = 5.4$$
$$E_3 = 24 \times .224 = 5.4$$
$$E_4 = 24 \times .277 = 6.6$$

Thus, the test statistic is

$$X^2 = \frac{(6 - 6.6)^2}{6.6} + \frac{(7 - 5.4)^2}{5.4} + \frac{(4 - 5.4)^2}{5.4} + \frac{(7 - 6.6)^2}{6.6}$$
$$= 0.05 + 0.49 + 0.35 + 0.02 = 0.92 \sim \chi^2_{4-1-2} = \chi^2_1$$

In this case, there are 4 groups ($g = 4$) and 2 parameters estimated from the data ($\mu$, $\sigma^2$) ($k = 2$). Thus, $X^2$ follows a chi-square distribution with $4 - 1 - 2 = 1$ *df*. Because $X^2 < \chi^2_{1,\,.95} = 3.84$, it follows that $p > .05$. Therefore, the normal distribution provides an adequate fit.

The chi-square goodness-of-fit test can be used to test the goodness-of-fit of any probability model, not just the normal model. The general procedure is:

1. Divide the range of values into $g$ mutually exclusive and exhaustive categories
2. Compute the probabilities of obtaining values within specific categories under the probability model
3. Multiply the probabilities in step (2) by the total sample size to obtain the expected counts within each category
4. Compute $X^2 =$ chi-square goodness-of-fit test statistic and its associated $p$-value

## SECTION 10.8 The Kappa Statistic

The redness of 50 eyes were graded by 2 observers using the rating scale (0/l/2/3) by comparison with reference photographs, where a higher grade corresponds to more redness. To assess the reproducibility of the grading system, the following $2 \times 2$ table was constructed:

| | | Redness rating observer B | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | Total |
| Redness rating observer A | 0 | 15 | 2 | 1 | 0 | 18 |
| | 1 | 4 | 7 | 3 | 2 | 16 |
| | 2 | 1 | 3 | 5 | 1 | 10 |
| | 3 | 0 | 1 | 2 | 3 | 6 |
| | Total | 20 | 13 | 11 | 6 | 50 |

One measure of reproducibility for categorical data of this type is the *Kappa statistic*, which is defined by

$$\text{Kappa} = \kappa = (p_o - p_e)/(1 - p_e)$$

where

$p_o$ = observed proportion of concordant responses for observers A and B

$p_e$ = expected proportion of concordant responses for observers A and B under the assumption that the redness ratings provided by the 2 observers are independent

$$= \sum_{i=1}^{c} a_i b_i$$

and

$a_i$ = proportion of responses in category $i$ for observer A

$b_i$ = proportion of responses in category $i$ for observer B

$c$ = number of categories

Kappa varies between 0 and 1, with 1 indicating perfect reproducibility (i.e., $p_o = 1$) and 0 indicating no reproducibility at all (i.e., $p_o = p_e$). Kappa statistics of $> .75$ are considered excellent, between .4 and .75 good, and $< .4$ poor.

For the preceding data,

$$p_o = \frac{15+7+5+3}{50} = \frac{30}{50} = .60$$

$$a_1 = \frac{18}{50} = .36,\ a_2 = \frac{16}{50} = .32,\ a_3 = \frac{10}{50} = .20,\ a_4 = \frac{6}{50} = .12$$

$$b_1 = \frac{20}{50} = .40,\ b_2 = \frac{13}{50} = .26,\ b_3 = \frac{11}{50} = .22,\ b_4 = \frac{6}{50} = .12$$

$$p_e = .36(.40) + \ldots + .12(.12) = .286$$

$$\text{Kappa} = \frac{.60 - .286}{1 - .286} = \frac{.314}{.714} = .44$$

This indicates good reproducibility of the rating system.

## PROBLEMS

**Cardiovascular Disease**

In a 1985 study of the effectiveness of streptokinase in the treatment of patients who have been hospitalized after myocardial infarction, 9 of 199 males receiving streptokinase and 13 of 97 males in the control group died within 12 months [1].

**10.1** Use the normal-theory method to test for significant differences in 12-month mortality between the two groups.

**10.2** Construct the observed and expected contingency tables for these data.

**10.3** Perform the test in Problem 10.1 using the contingency-table method.

**10.4** Compare your results in Problems 10.1 and 10.3.

**Cardiovascular Disease**
In the streptokinase study in Problem 10.1, 2 of 15 females receiving streptokinase and 4 of 19 females in the control group died within 12 months.

**10.5** Why is Fisher's exact test the appropriate procedure to test for differences in 12-month mortality rates between these two groups?

**10.6** Write down all possible tables with the same row and column margins as given in the observed data.

**10.7** Calculate the probability of each of the tables enumerated in Problem 10.6.

**10.8** Evaluate whether or not there is a significant difference between the mortality rates for streptokinase and control-group females using a two-sided test based on your results in Problem 10.7.

**10.9** Test for the goodness of fit of the normal model for the distribution of survival times of mice given in Table 6.4 (Chapter 6, Study Guide).

**Pulmonary Disease**
Suppose we wish to investigate the familial aggregation of respiratory disease according to the specific type of respiratory disease. One hundred families in which the head of household or the spouse has asthma, referred to as type A families, and 200 families in which either the head of household or the spouse has non-asthmatic pulmonary disease, but neither has asthma, referred to as type B families, are identified. Suppose that in 15 of the type A families the first-born child has asthma, whereas in 3 other type A families the first-born child has some nonasthmatic respiratory disease. Furthermore, in 4 of the type B households the first-born child has asthma, whereas in 2 other type B households the first-born child has some nonasthmatic respiratory disease.

**10.10** Compare the prevalence rates of asthma in the two types of families. State all hypotheses being tested.

**10.11** Compare the prevalence rates of nonasthmatic respiratory disease in the two types of families. State all hypotheses being tested.

**Cardiovascular Disease**
A 1979 study investigated the relationship between cigarette smoking and subsequent mortality in men with a prior history of coronary disease [2]. It was found that 264 out of 1731 nonsmokers and 208 out of 1058 smokers had died in the 5-year period after the study began.

**10.12** Assuming that the age distributions of the two groups are comparable, compare the mortality rates in the two groups.

**Obstetrics**
Suppose there are 500 pairs of pregnant women who participate in a prematurity study and are paired in such a way that the body weights of the 2 women in a pair are within 5 lb of each other. One of the 2 women is given a placebo and the other drug A to see if drug A has an effect in preventing prematurity. Suppose that in 30 pairs of women, *both* women in a pair have a premature child; in 420 pairs of women, *both* women have a normal child; in 35 pairs of women, the woman taking drug A has a normal child and the woman taking the placebo has a premature child; in 15 pairs of women, the woman taking drug A has a premature child and the woman taking the placebo has a normal child.

**10.13** Assess the statistical significance of these results.

**Cancer**
Suppose we wish to compare the following two treatments for breast cancer: simple mastectomy (S) and radical mastectomy (R). Matched pairs of women who are within the same decade of age and with the same clinical condition are formed. They receive the two treatments, and their subsequent 5-year survival is monitored. The results are given in Table 10.1. We wish to test for significant differences between the treatments.

**Table 10.1** Comparison of simple and radical mastectomy in treating breast cancer

| Pair | Treatment S woman | Treatment R woman | Pair | Treatment S woman | Treatment R woman |
|---|---|---|---|---|---|
| 1 | L[a] | L | 11 | D | D |
| 2 | L | D | 12 | L | D |
| 3 | L | L | 13 | L | L |
| 4 | L | L | 14 | L | L |
| 5 | L | L | 15 | L | D |
| 6 | D[b] | L | 16 | L | L |
| 7 | L | L | 17 | L | D |
| 8 | L | D | 18 | L | D |
| 9 | L | D | 19 | L | L |
| 10 | L | L | 20 | L | D |

[a] L lived at least 5 years.
[b] D died within 5 years.

**10.14** What test should be used to analyze these data? State the hypotheses being tested.

**10.15** Conduct the test mentioned in Problem 10.14.

**Obstetrics**

**10.16** Test for the adequacy of the goodness of fit of the normal distribution when applied to the distribution of birthweights in Figure 2.6 (in Chapter 2, text). The sample mean and standard deviation for these data are 111.26 oz and 20.95 oz, respectively.

**Cardiovascular Disease**
A hypothesis has been suggested that a principal benefit of physical activity is to prevent sudden death from heart attack. The following study was designed to test this hypothesis: 100 men who died from a first heart attack and 100 men who survived a first heart attack in the age group 50–59 were identified and their wives were each given a detailed questionnaire concerning their husbands' physical activity in the year preceding their heart attacks. The men were then classified as active or inactive. Suppose that 30 of the 100 who survived and 10 of the 100 who died were physically active. If we wish to test the hypothesis, then

**10.17** Is a one-sample or two-sample test needed here?

**10.18** Which one of the following test procedures should be used to test the hypothesis?
a. Paired *t* test
b. Two-sample *t* test for independent samples
c. $\chi^2$ test for $2 \times 2$ contingency tables
d. Fisher's exact test
e. McNemar's test

**10.19** Carry out the test procedure(s) in Problem 10.18 and report a *p*-value.

**Mental Health**
A clinical trial is set up to assess the effects of lithium in treating manic-depressive patients. New patients in an outpatient service are matched according to age, sex, and clinical condition, with one patient receiving lithium and the other a placebo. Suppose the outcome variable is whether or not the patient has any manic-depressive episodes in the next 3 months. The results are as follows: In 20 cases both the lithium and placebo members of the pair have manic-depressive episodes; in 10 cases only the placebo member has an episode (the lithium member does not); in 2 cases only the lithium member has an episode (the placebo member does not); in 36 cases neither member has an episode.

**10.20** State an appropriate hypothesis to test whether lithium has any effect in treating manic-depressive patients.

**10.21** Test the hypothesis mentioned in Problem 10.20.

**Cardiovascular Disease**
In some studies heart disease has been associated with being overweight. Suppose this association is examined in a large-scale epidemiological study and it is found that of 2000 men in the age group 55–59, 200 have myocardial infarctions in the next 5 years. Suppose the men are grouped by body weight as given in Table 10.2.

**Table 10.2** Association between body weight and myocardial infarction

| Body weight (lb) | Number of myocardial infarctions | Total number of men |
|---|---|---|
| 120–139 | 10 | 300 |
| 140–159 | 20 | 700 |
| 160–179 | 50 | 600 |
| 180–199 | 95 | 300 |
| 200+ | 25 | 100 |
| Total | 200 | 2000 |

**10.22** Comment in detail on these data.

**Cerebrovascular Disease**
Atrial fibrillation (AF) is widely recognized to predispose patients to embolic stroke. Oral anticoagulant therapy has been shown to decrease the number of embolic events. However, it also increases the number of major bleeding events (i.e., bleeding events requiring hospitalization). A study is proposed in which patients with AF are randomly divided into two groups: one receives the anticoagulant warfarin, the other a placebo. The groups are then followed for the incidence of major events (i.e., embolic stroke or major bleeding events).

**10.23** Suppose that 5% of treated patients and 22% of control patients are anticipated to experience a major event over 3 years. If 100 patients are to be randomized to each group, then how much power would such a study have for detecting a significant difference if a two-sided test with $\alpha = .05$ is used?

**10.24** How large should such a study be to have an 80% chance of finding a significant difference given the same assumptions as in Problem 10.23?

**10.25** One problem with warfarin is that about 10% of patients stop taking the medication due to persistent minor bleeding (e.g., nosebleed). If we regard the probabilities in Problem 10.23 as perfect-compliance risk estimates, then recalculate the power for the study proposed in Problem 10.23 if compliance is not perfect.

**Pulmonary Disease**

Each year approximately 4% of current smokers attempt to quit smoking, and about 50% of those who try to quit are successful; that is, they are able to abstain from smoking for at least 1 year from the date they quit. Investigators have attempted to identify risk factors that might influence these two probabilities. One such variable is the number of cigarettes currently smoked per day. In particular, the investigators found that among 75 current smokers who smoked $\leq 1$ pack/day, 5 attempted to quit, whereas among 50 current smokers who smoked more than 1 pack/day, 1 attempted to quit.

**10.26** Assess the statistical significance of these results and report a *p*-value.

Similarly, a different study reported that out of 311 people who had attempted to quit smoking, 16 out of 33 with less than a high school education were successful quitters; 47 out of 76 who had finished high school but had not gone to college were successful quitters; 69 out of 125 who attended college but did not finish 4 years of college were successful quitters; and 52 out of 77 who had completed college were successful quitters.

**10.27** Do these data show an association between the number of years of education and the rate of successful quitting?

**Infectious Disease, Hepatic Disease**

Read "Foodborne Hepatitis A Infection: A Report of Two Urban Restaurant-Associated Out-Breaks" by Denes et al., in the *American Journal of Epidemiology*, **105**(2) (1977), pages 156–162, and answer the following questions based on it.

**10.28** The authors analyzed the results of Table 1 using a chi-square statistic. Is this method of analysis reasonable for this table? If not, suggest an alternative method.

**10.29** Analyze the results in Table 1 using the method suggested in Problem 10.28. Do your results agree with the authors'?

**10.30** Student's *t* test with 40 *df* was used to analyze the results in Table 2. Is this method of analysis reasonable for this table? If not, suggest an alternative method.

**10.31** The authors claim that there is a significant difference ($p = .01$) between the proportion with hepatitis A among those who did and did not eat salad. Check this result using the method of analysis suggested in Problem 10.30.

**Infectious Disease, Cardiology**

Kawasaki's syndrome is an acute illness of unknown cause that occurs predominantly in children under the age of 5. It is characterized by persistent high fever and other clinical signs and can result in death and/or coronary-artery aneurysms. In the early 1980s, standard therapy for this condition was aspirin to prevent blood clotting. A Japanese group began experimentally treating children with intravenous gamma globulin in addition to aspirin to prevent cardiac symptoms in these patients [3].

A clinical trial is planned to compare the combined therapy of gamma globulin and aspirin vs. aspirin therapy alone. Suppose the incidence of coronary-artery aneurysms over 1 year is 15% in the aspirin-treated group, based on previous experience, and the investigators intend to use a two-sided significance test with $\alpha = .05$.

**10.32** If the 1-year incidence rate of coronary aneurysms in the combined therapy group is projected to be 5%, then how much statistical power will such a study have if 125 patients are to be recruited in each treatment group?

**10.33** Answer Problem 10.32 if 150 patients are recruited in each group.

**10.34** How many patients would have to be recruited in each group to have a 95% chance of finding a significant difference?

**Obstetrics**

An issue of current interest is the effect of delayed childbearing on pregnancy outcome. In a recent paper a population of first deliveries was assessed for low-birthweight deliveries ($< 2500$ g) according to the woman's age and prior pregnancy history [4]. The data in Table 10.3 were presented.

**Table 10.3** Relationship of age and pregnancy history to low-birthweight deliveries

| Age | History[a] | *n* | Percentage low birthweight |
|---|---|---|---|
| $\geq 30$ | No | 225 | 3.56 |
| $\geq 30$ | Yes | 88 | 6.82 |
| $< 30$ | No | 906 | 3.31 |
| $< 30$ | Yes | 153 | 1.31 |

[a] History = yes if a woman had a prior history of spontaneous abortion or infertility
= no otherwise

Source: Reprinted with permission of the *American Journal of Epidemiology*, **125**(l),101–109,1987.

**10.35** What test can be used to assess the effect of age on low-birthweight deliveries among women with a negative history?

**10.36** Perform the test in Problem 10.35 and report a *p*-value.

**10.37** What test can be used to assess the effect of age on low-birthweight deliveries among women with a positive history?

**10.38** Perform the test in Problem 10.37 and report a *p*-value.

**Cancer**

A recent study looked at the association between breast-cancer incidence and alcohol consumption [5]. The data in Table 10.4 were presented for 50–54-year-old women.

**Table 10.4** Association between alcohol consumption and breast cancer in 50–54-year-old women

| | Alcohol consumption (g/day) | | | | |
|---|---|---|---|---|---|
| Group | None | < 1.5 | 1.5–4.9 | 5.0–14.9 | ≥ 15.0 |
| Breast-cancer cases | 43 | 15 | 22 | 42 | 24 |
| Total number of women | 5944 | 2069 | 3449 | 3570 | 2917 |

*Source*: Reprinted with permission of the *New England Journal of Medicine*, **316**(19), 1174–1180,1987.

**10.39** What test procedure can be used to test if there is an association between breast-cancer incidence and alcohol consumption, where alcohol consumption is coded as (drinker/nondrinker)?

**10.40** Perform the test mentioned in Problem 10.39 and report a *p*-value.

**10.41** Perform a test for linear trend based on the data in Table 10.4.

**Cancer**

A case-control study was performed looking at the association between the risk of lung cancer and the occurrence of lung cancer among first-degree relatives [6]. Lung-cancer cases were compared with controls as to the number of relatives with lung cancer. Controls were frequency matched to cases by 5-year age category, sex, vital status, and ethnicity. The following data were presented:

| | | Number of controls | Number of cases |
|---|---|---|---|
| Number of | 0 | 466 | 393 |
| relatives with | 1 | 78 | 119 |
| lung cancer | 2+ | 8 | 20 |

**10.42** What test procedure can be used to look at the association between the number of relatives with lung cancer (0/1/2+) and case-control status?

**10.43** Implement the test in Problem 10.42 and report a *p*-value. Interpret the results in one or two sentences.

**Cardiology**

A group of patients who underwent coronary angiography between Jan. 1, 1972 and Dec. 31, 1986 in a particular hospital were identified [7]. 1493 cases with confirmed coronary-artery disease were compared with 707 controls with no plaque evidence at the time of angiography. Suppose it is found that 37% of cases and 30% of controls reported a diagnosis or treatment for hypertension at the time of angiography.

**10.44** Are the proportions (37%, 30%) an example of prevalence, incidence, or neither?

**10.45** What test can be used to compare the risk of hypertension between cases and controls?

**10.46** Implement the test in Problem 10.45 and report a *p*-value.

**Ophthalmology**

A study was performed comparing the validity of different methods of reporting the ocular condition age-related macular degeneration. Information was obtained by self-report at an eye examination, surrogate (spouse) report by telephone, and by clinical determination at an eye examination [8]. The following data were reported in Tables 10.5 and 10.6.

**Table 10.5** Comparison of surrogate report by telephone to self-report at eye exam for age-related macular degeneration

| | | Self-report at eye exam | | |
|---|---|---|---|---|
| | | No | Yes | Total |
| Surrogate | No | 1314 | 12 | 1326 |
| report by | Yes | 22 | 17 | 39 |
| telephone | Total | 1336 | 29 | 1365[a] |

**Table 10.6** Comparison of surrogate report by telephone to clinical determination at an eye exam for age-related macular degeneration

| | | Clinical determination at an eye exam | | |
|---|---|---|---|---|
| | | No | Yes | Total |
| Surrogate | No | 1247 | 83 | 1330 |
| report by | Yes | 26 | 14 | 40 |
| telephone | Total | 1273 | 97 | 1370[a] |

[a] The total sample sizes in Tables 10.5 and 10.6 do not match, due to a few missing values.

**10.47** What test can be performed to compare the frequency of reporting of age-related macular degeneration by self-report vs. surrogate report if neither is regarded as a gold standard?

**10.48** Implement the test mentioned in Problem 10.47 and report a *p*-value.

**10.49** Suppose the clinical determination is considered as the gold standard. What measure(s) can be used to assess the validity of the surrogate report?

**10.50** Provide estimates and 95% CI's for these measure(s).

## SOLUTIONS

**10.1** Test the hypothesis $H_0: p_1 = p_2$ versus $H_1: p_1 \neq p_2$. The test statistic is given by

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \left(\frac{1}{2n_1} + \frac{1}{2n_2}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $\hat{p}_1 = 9/199 = .0452$, $\hat{p}_2 = 13/97 = .1340$,

$$\hat{p} = \frac{9+13}{199+97} = \frac{22}{296} = .0743,$$

$$\hat{q} = 1 - \hat{p} = .9257,$$

$$z = \frac{|.0452 - .1340| - \left[\frac{1}{2(199)} + \frac{1}{2(97)}\right]}{\sqrt{.0743(.9257)\left(\frac{1}{199} + \frac{1}{97}\right)}}$$

$$= \frac{.0811}{.0325} = 2.498 \sim N(0,\ 1) \text{ under } H_0$$

Since $z > 1.96$, reject $H_0$ at the 5% level. The *p*-value $= 2 \times [1 - \Phi(2.498)] = .013$.

**10.2** The observed table is given by

12-month mortality status-observed table

| | Dead | Alive | |
|---|---|---|---|
| Streptokinase | 9 | 190 | 199 |
| Control | 13 | 84 | 97 |
| | 22 | 274 | 296 |

The expected cell counts are obtained from the row and column margins as follows:

$$E_{11} = \frac{199 \times 22}{296} = 14.79$$

$$E_{12} = \frac{199 \times 274}{296} = 184.21$$

$$E_{21} = \frac{97 \times 22}{296} = 7.21$$

$$E_{22} = \frac{97 \times 274}{296} = 89.79$$

These values are displayed as follows:

12-month mortality status-expected table

| | Dead | Alive | |
|---|---|---|---|
| Streptokinase | 14.79 | 184.21 | 199 |
| Control | 7.21 | 89.79 | 97 |
| | 22 | 274 | 296 |

**10.3** Compute the Yates-corrected chi-square statistic as follows:

$$X^2_{\text{CORR}} = \frac{(|9 - 14.79| - .5)^2}{14.79} + \ldots + \frac{(|84 - 89.79| - .5)^2}{89.79}$$

$$= \frac{5.29^2}{14.79} + \ldots + \frac{5.29^2}{89.79} = 1.892 + 0.152 + 3.882 + 0.312$$

$$= 6.24 \sim \chi^2_1 \text{ under } H_0$$

Since $\chi^2_{1,\ .95} = 3.84 < X^2$, reject $H_0$ at the 5% level. The *p*-value $= Pr(\chi^2_1 > 6.24) = .013$ is obtained by using the CHIDIST function of Excel.

**10.4** The decisions reached in Problems 10.1 and 10.3 were the same (reject $H_0$ at the 5% level). The same *p*-value is obtained whether *z* is compared to an $N(0,\ 1)$ distribution or $X^2_{\text{CORR}} = z^2$ to a $\chi^2_1$ distribution (i.e., $p = .013$).

**10.5** Form the following observed $2 \times 2$ table:

12-month mortality status

| | Dead | Alive | |
|---|---|---|---|
| Streptokinase | 2 | 13 | 15 |
| Control | 4 | 15 | 19 |
| | 6 | 28 | 34 |

The smallest expected value

$$E_{11} = \frac{15 \times 6}{34} = 2.65 < 5.$$

Thus, Fisher's exact must be used.

**10.6**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0 | 15 | 1 | 14 | 2 | 13 | 3 | 12 |
| 6 | 13 | 5 | 14 | 4 | 15 | 3 | 16 |

| | |
|---|---|
| 4 | 11 |
| 2 | 17 |

| | |
|---|---|
| 5 | 10 |
| 1 | 18 |

| | |
|---|---|
| 6 | 9 |
| 0 | 19 |

**10.7** We use the HYPGEOMDIST function of Excel to compute the probabilities of each of the tables (see appendix) and obtain:

$$Pr(0) = .020,\ Pr(1) = .130,\ Pr(2) = .303,$$
$$Pr(3) = .328, Pr(4) = .174,\ Pr(5) = .042,\ Pr(6) = .004$$

**10.8** Our table is the "2" table. Therefore, the two-tailed $p$-value is given by

$$\begin{aligned} p &= 2 \times \min[Pr(0) + Pr(1) + Pr(2), \\ &\quad Pr(2) + Pr(3) + \ldots + Pr(7),\ .5] \\ &= 2 \times \min(.452,\ .850,\ .5) = 2 \times .452 = .905 \end{aligned}$$

Clearly, there is no significant difference in 12-month mortality status between the two treatment groups for females.

**10.9** We first compute the mean and standard deviation for the sample of survival times. We have $\bar{x} = 16.13$, $s = 2.67$, $n = 429$. We compute the probabilities under a normal model for the groups 10–12, 13–15, 16–18, 19–21, 22–24 as follows:

| Group | Probability |
|---|---|
| 10-12 | $\Phi[(12.5 - 16.13)/2.67] = \Phi(-1.36) = .087$ |
| 13-15 | $\Phi[(15.5 - 16.13)/2.67] - .087$ |
| | $= \Phi(-0.24) - .087 = .407 - .087 = .320$ |
| 16-18 | $\Phi[(18.5 - 16.13)/2.67] - .407$ |
| | $= \Phi(0.89) - .407 = .813 - .407 = .406$ |
| 19-21 | $\Phi[(21.5 - 16.13)/2.67] - .813$ |
| | $= \Phi(2.01) - .813 = .978 - .813 = .165$ |
| 22-24 | $1 - .978 = .022$ |

We now compute the observed and expected number of units in each group:

| Group | Observed number of units | Expected number of units |
|---|---|---|
| 10–12 | 45 | $429 \times .087 = 37.3$ |
| 13–15 | 121 | $429 \times .320 = 137.3$ |
| 16–18 | 184 | $429 \times .406 = 174.3$ |
| 19–21 | 69 | $429 \times .164 = 70.7$ |
| 22–24 | 10 | $429 \times .022 = 9.3$ |

We compute the chi-square goodness-of-fit statistic:

$$\begin{aligned} X^2 &= \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{(45 - 37.0)^2}{37.0} + \ldots + \frac{(10 - 9.4)^2}{9.4} \\ &= 1.611 + 1.935 + 0.544 + 0.043 + 0.034 \\ &= 4.17 \sim \chi^2_{g-1-p} = \chi^2_{5-1-2} = \chi^2_2 \text{ under } H_0 \end{aligned}$$

Since $\chi^2_{2,\ .95} = 5.99 > X^2$, it follows that $p > .05$, and we accept the null hypothesis that the normal model fits the data adequately.

**10.10** Test the hypothesis $H_0: p_A = p_B$ versus $H_1: p_A \neq p_B$, where

$p_A = Pr$ (first-born child has asthma in a type A family)
$p_B = Pr$ (first-born child has asthma in a type B family)

The observed and expected $2 \times 2$ tables are shown as follows.

Observed table

| | | Asthma + | Asthma – | |
|---|---|---|---|---|
| Type of family | A | 15 | 85 | 100 |
| | B | 4 | 196 | 200 |
| | | 19 | 281 | 300 |

Expected table

| | | Asthma + | Asthma – | |
|---|---|---|---|---|
| Type of family | A | 6.3 | 93.7 | 100 |
| | B | 12.7 | 187.3 | 200 |
| | | 19 | 281 | 300 |

*Note*: + = asthma, – = no asthma.

The $\chi^2$ test for $2 \times 2$ tables will be used, since the expected table has no expected value $< 5$. We have the following Yates-corrected chi-square statistic

$$\begin{aligned} X^2 &= \frac{(|15 - 6.3| - .5)^2}{6.3} + \ldots + \frac{(|196 - 187.3| - .5)^2}{187.3} \\ &= 16.86 \sim \chi^2_1 \text{ under } H_0 \end{aligned}$$

The $p$-value for this result is $< .001$, since

$$\chi^2_{1,\ .999} = 10.83 < 16.86$$

Thus, there is a highly significant association between the type of family and the asthma status of the child.

**10.11** The $2 \times 2$ table is shown as follows:

| | | Nonasthmatic respiratory disease status + | – | |
|---|---|---|---|---|
| Type of family | A | 3 | 97 | 100 |
| | B | 2 | 198 | 200 |
| | | 5 | 295 | 300 |

*Note*: + = nonasthmatic respiratory disease, – = no nonasthmatic respiratory disease.

There are two expected values $< 5$ ; in particular,

$$E_{11} = \frac{5(100)}{300} = 1.7 \quad E_{21} = \frac{5(200)}{300} = 3.3$$

Thus, Fisher's exact test must be used to analyze this table. We write all possible tables with the same margins as the observed table, as follows:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 100 | | 1 | 99 | | 2 | 98 | | |
| 5 | 195 | | 4 | 196 | | 3 | 197 | | |
| 3 | 97 | | 4 | 96 | | 5 | 95 | | |
| 2 | 198 | | 1 | 199 | | 0 | 200 | | |

We use the HYPGEOMDIST function of Excel to compute the probability of each table and obtain:

$$Pr(0) = .129,\ Pr(1) = .330,\ Pr(2) = .332$$
$$Pr(3) = .164,\ Pr(4) = .040,\ Pr(5) = .004$$

Thus, since the observed table is the "3" table, the two-tailed $p$-value is given by

$$2 \times \min(.164 + .040 + .004, .164 + .332 + .330 + .129, .5)$$
$$= 2 \times .208 = .416$$

The results are not statistically significant and indicate that there is no significant difference in the prevalence rate of nonasthmatic respiratory disease between households in which the parents do or do not have asthma.

**10.12** We have the following observed table:

| | | 5-year mortality incidence Dead | Alive | |
|---|---|---|---|---|
| Smoking status | Nonsmokers | 264 | 1467 | 1731 |
| | Smokers | 208 | 850 | 1058 |
| | | 472 | 2317 | 2789 |

The smallest expected value

$$\frac{472 \times 1058}{2789} = 179.1 \geq 5.$$

Thus, we can use the chi-square test for $2 \times 2$ contingency tables. We have the following test statistic:

$$X^2 = \frac{n(|ad - bc| - \frac{n}{2})^2}{(a+b)(c+d)(a+c)(b+d)}$$
$$= \frac{2789[|264(850) - 208(1467)| - \frac{2789}{2}]^2}{1731 \times 1058 \times 472 \times 2317}$$
$$= \frac{2789(79{,}341.5)^2}{1731 \times 1058 \times 472 \times 2317} = 8.77 \sim \chi_1^2 \text{ under } H_0$$

Since $\chi^2_{1,\ .995} = 7.88 < X^2 < \chi^2_{1,\ .999} = 10.83$, it follows that $.001 < p < .005$. Thus, cigarette smokers with a prior history of coronary disease have a significantly higher mortality incidence in the subsequent 5 years $(208/1058 = .197)$ than do nonsmokers with a prior history of coronary disease $(264/1731 = .153)$.

**10.13** We can use McNemar's test in this situation. We have the following table based on matched pairs:

| | | Placebo Premature | Normal | |
|---|---|---|---|---|
| Drug A | Premature | 30 | 15 | 45 |
| | Normal | 35 | 420 | 455 |
| | | 65 | 435 | 500 |

We can ignore the 30 + 420 concordant pairs and focus on the remaining 50 discordant pairs. We have the test statistic

$$X^2 = \frac{(|35 - \frac{50}{2}| - \frac{1}{2})^2}{\frac{50}{4}} = 7.22 \sim \chi_1^2 \text{ under } H_0$$

Since $\chi^2_{1,\ .99} = 6.63$, $\chi^2_{1,\ .995} = 7.88$ and $6.63 < 7.22 < 7.88$, it follows that the two-sided $p$-value is given by $.005 < p < .01$. Thus, there is a significant difference between the two treatments, with drug A women having significantly lower prematurity rates than placebo women.

**10.14** We wish to test whether or not treatment $S$ differs from treatment $R$. We will use McNemar's test for correlated proportions since we have matched pairs. The hypotheses being tested in this case are $H_0: p = 1/2$ versus $H_1: p \neq 1/2$, where $p$ = probability that a discordant pair is of type A; i.e., where the treatment $S$ woman lives for $\geq 5$ years and the treatment $R$ woman dies within 5 years, given that one woman in a matched pair survives for $\geq 5$ years and the other woman in the matched pair does not.

**10.15** We have the following $2 \times 2$ table of matched pairs:

| | | Treatment *R* woman: L | Treatment *R* woman: D |
|---|---|---|---|
| Treatment *S* woman | L | 10 | 8 |
| | D | 1 | 1 |

We must compute the *p*-value using an exact binomial test, because the number of discordant pairs (9) is too small to use the normal approximation. We refer to the exact binomial tables (Table 1, Appendix, text) and obtain

$$\begin{aligned} p\text{-value} &= 2 \times \sum_{k=8}^{9} {}_9C_k(.5)^9 \\ &= 2 \times (.0176 + .0020) = .039 \end{aligned}$$

Thus, we reject $H_0$ and conclude that treatment *S* is significantly better than treatment *R*.

**10.16** We divide the distribution of birthweights (oz) into the groups $\leq 79$, 80–89, 90–99, 100–109, 110–119, 120–129, 130–139, 140+. We will assume that each birthweight is rounded to the nearest ounce and thus 75 ounces actually represents the interval 74.5–75.5. Thus, we can compute the expected number of infants in each group under a normal model as follows:

$$\begin{aligned} E_1 &= 100Pr(X \leq 79) \\ &= 100\Phi\left(\frac{79.5 - 111.26}{20.95}\right) \\ &= 100\Phi\left(\frac{-31.76}{20.95}\right) \\ &= 100\Phi(-1.52) = 100(1 - .9352) = 6.5 \\ E_2 &= 100Pr(80 \leq X \leq 89) \\ &= 100\left[\Phi\left(\frac{89.5 - 111.26}{20.95}\right) - \Phi\left(\frac{79.5 - 111.26}{20.95}\right)\right] \\ &= 100[\Phi(-1.04) - \Phi(-1.52)] \\ &= 100(1 - .8505) - 100(1 - .9352) = 8.5 \\ &\vdots \\ E_8 &= 100Pr(X \geq 140) \\ &= 100\left[1 - \Phi\left(\frac{139.5 - 111.26}{20.95}\right)\right] \\ &= 100[1 - \Phi(1.35)] = 100(1 - .9112) = 8.9 \end{aligned}$$

We have the following table of observed and expected cell counts:

| Birthweight | Observed | Expected |
|---|---|---|
| $\leq 79$ | 5 | 6.5 |
| 80–89 | 10 | 8.5 |
| 90–99 | 11 | 13.8 |
| 100–109 | 19 | 17.9 |
| 110–119 | 17 | 18.6 |
| 120–129 | 20 | 15.5 |
| 130–139 | 12 | 10.3 |
| $\geq 140$ | 6 | 8.9 |

We can use the chi-square goodness-of-fit test because all expected values are $\geq 5$. We have the following test statistic

$$\begin{aligned} X^2 &= \frac{(5 - 6.5)^2}{6.5} + \ldots + \frac{(6 - 8.9)^2}{8.9} \\ &= 3.90 \sim \chi^2_{g-k-1} = \chi^2_5 \text{ under } H_0 \end{aligned}$$

There are 5 *df* because there are 8 groups and 2 parameters estimated from the data. Because

$$\chi^2_{5,\,.95} = 11.07 > X^2,$$

it follows that $p > .05$. Thus, the results are not statistically significant and the goodness of fit of the normal distribution is adequate.

**10.17** A two-sample test is needed here, because samples of men who survived and died, respectively, from a first heart attack are being compared.

**10.18** The observed table is shown as follows:

Observed table relating sudden death from a first heart attack and previous physical activity

| | | Physical activity: Active | Physical activity: Inactive | |
|---|---|---|---|---|
| Mortality status | Survived | 30 | 70 | 100 |
| | Died | 10 | 90 | 100 |
| | | 40 | 160 | 200 |

The smallest expected value is

$$\frac{40 \times 100}{200} = 20 \geq 5$$

Thus, the $\chi^2$ test for $2 \times 2$ contingency tables can be used here.

**10.19** The test statistic is given by

$$X^2 = \frac{n\left[|ad-bc| - \frac{n}{2}\right]^2}{(a+b)(c+d)(a+c)(b+d)}$$
$$= \frac{200(|30\times 90 - 10\times 70| - 100)^2}{100(100)(40)(160)}$$
$$= \frac{200(1900)^2}{100(100)(40)(160)} = 11.28 \sim \chi_1^2 \text{ under } H_0$$

Since $\chi^2_{1,\ .999} = 10.83 < X^2$ it follows that $p < .001$, and we can conclude that there is a significant association between physical activity and survival after a heart attack.

**10.20** This is a classic example illustrating the use of McNemar's test for correlated proportions. There are two groups of patients, one receiving lithium and one receiving placebo, but the two groups are matched on age, sex, and clinical condition and thus represent dependent samples. Let a type A discordant pair be a pair of people such that the lithium member of the pair has a manic-depressive episode and the placebo member does not. Let a type B discordant pair be a pair of people such that the placebo member of the pair has a manic-depressive episode and the lithium member does not. Let $p$ = probability that a discordant pair is of type A. Then test the hypothesis

$$H_0:\ p = \frac{1}{2} \text{ versus } H_1:\ p \neq \frac{1}{2}$$

**10.21** There are 2 type A discordant pairs and 10 type B discordant pairs, giving a total of 12 discordant pairs. Since the number of discordant pairs is < 20, an exact binomial test must be used. Under $H_0$,

$$Pr(k \text{ type A discordant pairs}) = \binom{12}{k}\left(\frac{1}{2}\right)^{12}$$

In particular, from Table 1 (Appendix, text)

$$Pr(k \le 2) = \left(\frac{1}{2}\right)^{12}\left[\binom{12}{0} + \binom{12}{1} + \binom{12}{2}\right]$$
$$= .0002 + .0029 + .0161 = .0192$$

Since a two-sided test is being performed,

$$p = 2 \times .0192 = .039$$

Thus, $H_0$ is rejected and we conclude that the placebo patients are more likely to have manic-depressive episodes when the outcome differs in the two members of a pair.

**10.22** We form the following $2 \times 5$ contingency table:

| | | Body weight | | | | | |
|---|---|---|---|---|---|---|---|
| | | 120-139 | 140-159 | 160-179 | 180-199 | 200+ | |
| MI | Yes | 10 | 20 | 50 | 95 | 25 | 200 |
| | No | 290 | 680 | 550 | 205 | 75 | 1800 |
| | | 300 | 700 | 600 | 300 | 100 | 2000 |
| | % yes | 3.3 | 2.9 | 8.3 | 31.7 | 25.0 | |

We perform the chi-square test for trend using the score statistic 1, 2, 3, 4, 5 for the five columns in the table. We have the test statistic $X_1^2 = A^2/B$, where

$$A = \sum_{i=1}^{k} x_i S_i - x\overline{S}$$
$$= 10(1) + 20(2) + \ldots + 25(5) - 200\left[\frac{300(1) + \ldots + 100(5)}{2000}\right]$$
$$= 705 - \frac{200(5200)}{2000} = 705 - 520 = 185$$

$$B = \overline{p}\,\overline{q}\left[\sum_{i=1}^{k} n_i S_i^2 - \frac{\left(\sum_{i=1}^{k} n_i S_i\right)^2}{N}\right] = \frac{200}{2000} \times \frac{1800}{2000}$$
$$\times \left[300(1^2) + \ldots + 100(5^2) - \frac{5200^2}{2000}\right]$$
$$= .09(15{,}800 - 13{,}520) = .09(2280) = 205.2$$

Thus, we have

$$X_1^2 = \frac{185^2}{205.2} = 166.8 \sim \chi_1^2 \text{ under } H_0$$

Since $X_1^2 > \chi^2_{1,\ .999} = 10.83$, we have $p < .001$ and there is a significant linear trend relating body weight and the incidence of MI.

**10.23** We use the power formula in Equation 10.15 (text, Chapter 10) as follows:

$$\text{Power} = \Phi\left[\frac{\Delta}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} - z_{1-\alpha/2}\frac{\sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}}\right]$$

where $p_1 = .05$, $p_2 = .22$, $n_1 = n_2 = 100$, $\alpha = .05$, $\overline{p} = (.05 + .22)/2 = .135$, $\overline{q} = .865$. We have

$$\text{Power} = \Phi\left[\frac{.22-.05}{\sqrt{\frac{.05(.95)+.22(.78)}{100}}} - z_{.975}\frac{\sqrt{.135(.865)\left(\frac{1}{100}+\frac{1}{100}\right)}}{\sqrt{\frac{.05(.95)+.22(.78)}{100}}}\right]$$
$$= \Phi\left(\frac{.17}{.0468} - 1.96\times\frac{.0483}{.0468}\right)$$
$$= \Phi(3.632-2.024) = \Phi(1.608) = .946$$

Thus, such a study would have a 95% chance of detecting a significant difference.

**10.24** We use the sample-size formula in Equation 10.14 (text, Chapter 10) as follows:

$$n = \frac{\left(\sqrt{2\bar{p}\bar{q}}z_{1-\alpha/2} + \sqrt{p_1q_1+p_2q_2}z_{1-\beta}\right)^2}{\Delta^2}$$
$$= \frac{\left[\sqrt{2(.135)(.865)}z_{.975} + \sqrt{.05(.95)+.22(.78)}z_{.80}\right]^2}{(.22-.05)^2}$$
$$= \frac{[.4833(1.96)+.4681(0.84)]^2}{(.17)^2}$$
$$= \frac{(1.3404)^2}{(.17)^2} = 62.2$$

Thus, we need 63 patients in each group to have an 80% probability of finding a significant difference.

**10.25** We obtain an estimate of power adjusted for noncompliance as presented in Section 10.5.3 (text, Chapter 10). We have that $\lambda_1 = .10$, $\lambda_2 = 0$. Therefore,

$$p_1^* = .05(.9) + .22(.1) = .067,\ q_1^* = .933$$
$$p_2^* = p_2 = .22,\ q_2^* = .78$$
$$\bar{p}^* = (.067+.22)/2 = .1435,\ \bar{q}^* = .8565$$
$$\Delta^* = |p_1^* - p_2^*| = .153$$

We use Equation 10.15 (text, Chapter 10) with $p_1$, $p_2$, $q_1$, $q_2$, $\bar{p}$, $\bar{q}$, and $\Delta$ replaced by $p_1^*$, $p_2^*$, $q_1^*$, $q_2^*$, $\bar{p}^*$, $\bar{q}^*$, and $\Delta^*$ as follows:

$$\text{Power} = \Phi\left[\frac{.153}{\sqrt{\frac{.067(.933)+.22(.78)}{100}}} - z_{.975}\frac{\sqrt{.1435(.8565)\left(\frac{2}{100}\right)}}{\sqrt{\frac{.067(.933)+.22(.78)}{100}}}\right]$$
$$= \Phi\left[\frac{.153}{.0484} - 1.96\left(\frac{.0496}{.0484}\right)\right]$$
$$= \Phi(3.162-2.008) = \Phi(1.154) = .876$$

Therefore, the power is reduced from 95% to 88% if lack of compliance is taken into account.

**10.26** We have the following $2\times 2$ table:

| | | Attempt to quit: Yes | Attempt to quit: No | |
|---|---|---|---|---|
| Packs/day | >1 | 1 | 49 | 50 |
| | ≤1 | 5 | 70 | 75 |
| | | 6 | 119 | 125 |

The expected number of units in the (1, 1) cell

$$6\times\frac{50}{125} = 2.4 < 5.$$

Thus, we must use Fisher's exact test to assess the significance of this table. We construct all possible tables with the same row and column margins as the observed table as follows:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 50 | 1 | 49 | 2 | 48 | 3 | 47 |
| 6 | 69 | 5 | 70 | 4 | 71 | 3 | 72 |

| | | | | |
|---|---|---|---|---|
| 4 | 46 | 5 | 45 | 6 | 44 |
| 2 | 73 | 1 | 74 | 0 | 75 |

We use the HYPGEOMDIST function of Excel to calculate the exact probability of each table as follows:

$$Pr(0) = .043,\ Pr(1) = .184,\ Pr(2) = .317,\ Pr(3) = .282,$$
$$Pr(4) = .136,\ Pr(5) = .034,\ Pr(6) = .003$$

Since we observed the "1" table, the two-tailed $p$-value is given by $p = 2\times(.043+.184) = .454$. Thus, there is no significant relationship between amount smoked and propensity to quit.

**10.27** We have the following $2\times 4$ table relating success in quitting smoking to level of education:

| | | Years of education: < 12 | 12 | > 12, < 16 | 16+ | |
|---|---|---|---|---|---|---|
| Successful quitter | Yes | 16 | 47 | 69 | 52 | 184 |
| | No | 17 | 29 | 56 | 25 | 127 |
| Percentage of successful quitters | | 33 (48) | 76 (62) | 125 (55) | 77 (68) | 311 |

We will perform the chi-square test for linear trend to detect if there is a significant association between the proportion of successful quitters and the number of years of education. We assign scores of 1, 2, 3, and 4 to the four education groups. We have the test statistic $X_1^2 = A^2/B$ where

$$A = \sum_{i=1}^{k} x_i S_i - x\bar{S}$$
$$= 16(1) + \ldots + 52(4) - 184 \times \left[\frac{33(1) + \ldots + 77(4)}{311}\right]$$
$$= 525 - 184 \times \frac{868}{311} = 525 - 513.54 = 11.46$$
$$B = \overline{pq}\left[\sum_{i=1}^{k} n_i S_i^2 - \frac{\left(\sum_{i=1}^{k} n_i S_i\right)^2}{N}\right]$$
$$= \frac{184}{311} \times \frac{127}{311} \times \left[33(1^2) + \ldots + 77(4^2) - \frac{868^2}{311}\right]$$
$$= .2416(2694 - 2422.59) = .2416(271.41) = 65.57$$

Therefore, $X_1^2 = 11.46^2/65.57 = 2.00 \sim \chi_1^2$ under $H_0$. Since $\chi^2_{1,\,.75} = 1.32$, $\chi^2_{1,\,.90} = 2.71$ and

$$1.32 < 2.00 < 2.71,$$

it follows that

$$1 - .90 < p < 1 - .75$$

or $.10 < p < .25$. Thus, there is no significant association between success in quitting smoking and number of years of education.

**10.28** The data are in the form of a $2 \times 2$ table, so the chi-square test may be an appropriate method of analysis if the expected cell counts are large enough. The smallest expected value is given by $(12 \times 22)/50 = 5.28 > 5$. Thus, this is a reasonable method of analysis.

**10.29** The observed table is given as follows:

*Association between working status and health status*

| | Ill | Well | |
|---|---|---|---|
| Worked | 10 | 12 | 22 |
| Did not work | 2 | 26 | 28 |
| | 12 | 38 | 50 |

Compute the following chi-square statistic:

$$X^2 = \frac{n\left[|ad - bc| - \frac{n}{2}\right]^2}{(a+b)(c+d)(a+c)(b+d)}$$
$$= \frac{50(|10 \times 26 - 2 \times 12| - 25)^2}{22(28)(12)(38)}$$
$$= \frac{50(211)^2}{22(28)(12)(38)} = 7.92 \sim \chi_1^2$$

Referring to the $\chi^2$ tables, we find that

$$\chi^2_{1,\,.995} = 7.88,\ \chi^2_{1,\,.999} = 10.83$$

Thus, because $7.88 < 7.92 < 10.83$, it follows that

$$.001 < p < .005.$$

The authors found a chi-square of 7.8, $p = .01$, and thus our results are somewhat more significant than those claimed in the article.

**10.30** The *t* test is not a reasonable test to use in comparing binomial proportions from two independent samples. Instead, either the chi-square test for $2 \times 2$ tables with large expected values or Fisher's exact test for tables with small expected values should be used.

**10.31** The $2 \times 2$ table is given as follows:

*Association between salad consumption and health status*

| | | Ill | Well | | Percentage Ill |
|---|---|---|---|---|---|
| Ate salad | Yes | 25 | 8 | 33 | (76) |
| | No | 3 | 6 | 9 | (33) |
| | | 28 | 14 | 42 | |

The smallest expected value $= (14 \times 9)/42 = 3 < 5$, which implies that Fisher's exact test must be used. First rearrange the table so that the smaller row total is in row 1 and the smaller column total is in column 1, as follows:

| | | Well | Ill | |
|---|---|---|---|---|
| Ate salad | No | 6 | 3 | 9 |
| | Yes | 8 | 25 | 33 |
| | | 14 | 28 | 42 |

Now enumerate all tables with the same row and column margins as follows:

| | |
|---|---|
| 0 | 9 |
| 14 | 19 |

| | |
|---|---|
| 1 | 8 |
| 13 | 20 |

| | |
|---|---|
| 2 | 7 |
| 12 | 21 |

| | |
|---|---|
| 3 | 6 |
| 11 | 22 |

| | |
|---|---|
| 4 | 5 |
| 10 | 23 |

| | |
|---|---|
| 5 | 4 |
| 9 | 24 |

| | |
|---|---|
| 6 | 3 |
| 8 | 25 |

| | |
|---|---|
| 7 | 2 |
| 7 | 26 |

| | |
|---|---|
| 8 | 1 |
| 6 | 27 |

| | |
|---|---|
| 9 | 0 |
| 5 | 28 |

Now use the HYPGEOMDIST function of Excel to compute the exact probability of each table. We have

$$Pr(0) = .015$$
$$Pr(1) = .098$$
$$Pr(2) = .242$$
$$Pr(3) = .308$$
$$Pr(4) = .221$$
$$Pr(5) = .092$$
$$Pr(6) = .022$$
$$Pr(7) = .003$$
$$Pr(8) = .0002$$
$$Pr(9) = 4.49 \times 10^{-6}$$

Since our observed table is the "6" table, the two-sided $p$-value is given by

$$p = 2 \times \min\left[\sum_{i=0}^{6} Pr(i),\ \sum_{i=6}^{9} Pr(i),\ .5\right]$$
$$= 2 \times \min(.997,\ .0252,\ .5) = .050$$

Thus, the results are on the margin of being statistically significant ($p = .05$) as opposed to the $p$-value of .01 given in the paper.

**10.32** We use the power formula in Equation 10.15 (text, Chapter 10) as follows:

$$\text{Power} = \Phi\left[\frac{\Delta}{\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}} - z_{1-\alpha/2}\frac{\sqrt{\frac{2\bar{p}\bar{q}}{n}}}{\sqrt{\frac{p_1q_1 + p_2q_2}{n}}}\right]$$

where $p_1 = .15$, $p_2 = .05$,

$$\Delta = |.15 - .05| = .10,$$

$\alpha = .05$, $n = 125$, $\bar{p} = (.15 + .05)/2 = .10$. We have

$$\text{Power} = \Phi\left\{\frac{.10}{\sqrt{\frac{.15(.85) + .05(.95)}{125}}} - z_{.975}\frac{\sqrt{\frac{2(.10)(.90)}{125}}}{\sqrt{\frac{.15(.85) + .05(.95)}{125}}}\right\}$$
$$= \Phi\left(\frac{.10}{.0374} - 1.96 \times \frac{.0379}{.0374}\right)$$
$$= \Phi(2.673 - 1.988) = \Phi(0.685) = .75$$

Thus, such a study would have 75% power.

**10.33** We let $n = 150$ and keep all other parameters the same. We have

$$\text{Power} = \Phi\left[\frac{.10}{\sqrt{\frac{.15(.85) + .05(.95)}{150}}} - 1.988\right]$$
$$= \Phi\left(\frac{.10}{.0342} - 1.988\right)$$
$$= \Phi(2.928 - 1.988) = \Phi(0.940) = .83$$

Thus, the power increases to 83% if the sample size is increased to 150 patients per group.

**10.34** We use the sample-size formula in Equation 10.14 (text, Chapter 10) as follows:

$$n = \frac{\left(\sqrt{2\bar{p}\bar{q}}\ z_{1-\alpha/2} + \sqrt{p_1q_1 + p_2q_2}\ z_{1-\beta}\right)^2}{\Delta^2}$$
$$= \frac{\left[\sqrt{2(.10)(.90)}z_{.975} + \sqrt{.15(.85) + .05(.95)}z_{.95}\right]^2}{(.10)^2}$$
$$= \frac{[.4243(1.96) + .4183(1.645)]^2}{.01} = \frac{(1.5197)^2}{.01} = 231.0$$

Thus, we would need to recruit 231 patients in each group in order to achieve 95% power.

**10.35** Form the following $2 \times 2$ table to assess age effects among women with a negative history:

Women with a negative history

| | | Low birthweight: Yes | Low birthweight: No | |
|---|---|---|---|---|
| Age | ≥ 30 | 8 | 217 | 225 |
| | < 30 | 30 | 876 | 906 |
| | | 38 | 1093 | 1131 |

The smallest expected cell count

$$E_{11} = \frac{38 \times 225}{1131} = 7.56 \geq 5.$$

Therefore, the Yates-corrected chi-square test can be used.

**10.36** The test statistic is given by

$$X^2 = \frac{1131(|8\times 876 - 30\times 217| - \frac{1131}{2})^2}{38\times 1093\times 906\times 225}$$
$$= \frac{1131(498 - 565.5)^2}{8.467\times 10^9}$$
$$= \frac{5.153\times 10^6}{8.467\times 10^9} = 0.00061 \sim \chi_1^2 \text{ under } H_0$$

Clearly, since $\chi^2_{1,\,.50} = 0.45$ and $X^2 < 0.45$, it follows that $p > .50$, and there is no significant effect of age on low-birthweight deliveries in this strata.

**10.37** Form the following $2\times 2$ contingency table among women with a positive history:

Women with a positive history

| | | Low birthweight: Yes | Low birthweight: No | |
|---|---|---|---|---|
| Age | $\geq 30$ | 6 | 82 | 88 |
| | $< 30$ | 2 | 151 | 153 |
| | | 8 | 233 | 241 |

The smallest expected value $= (8\times 88)/241 = 2.92 < 5$. Therefore, Fisher's exact test must be used to perform the test.

**10.38** First form all possible tables with the same row and column margins, as follows:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 88 | 1 | 87 | 2 | 86 | 3 | 85 | 4 | 84 |
| 8 | 145 | 7 | 146 | 6 | 147 | 5 | 148 | 4 | 149 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5 | 83 | 6 | 82 | 7 | 81 | 8 | 80 |
| 3 | 150 | 2 | 151 | 1 | 152 | 0 | 153 |

Now use the HYPGEOMDIST function of Excel to compute the exact probabilities of each table as follows:

$Pr(0) = .025,\ Pr(1) = .119,\ Pr(2) = .246,\ Pr(3) = .286,$
$Pr(4) = .204,\ Pr(5) = .091,\ Pr(6) = .025,\ Pr(7) = .004,$
$Pr(8) = .0003$

Since our table is the "6" table, a two-sided $p$-value is computed as follows:

$$p = 2\times \min\left[\sum_{k=0}^{6} Pr(k),\ \sum_{k=6}^{8} Pr(k),\ .5\right]$$
$$= 2\times \min(.025 + \ldots + .025,\ .025 + .004 + .0003,\ .5)$$
$$= 2\times \min(.996,\ .0292,\ .5) = .058$$

Thus, for women with a positive history, there is a trend toward significance, with older women having a higher incidence of low-birthweight deliveries.

**10.39** We first combine together the data from all drinking women and form the following $2\times 2$ contingency table:

| | Drinking status: Nondrinker | Drinker | |
|---|---|---|---|
| Case | 43 | 103 | 146 |
| Control | 5901 | 11,902 | 17,803 |
| | 5944 | 12,005 | 17,949 |

The smallest expected value in this table is

$$E_{11} = 146\times \frac{5944}{17{,}949} = 48.3 \geq 5.$$

Thus, we can use the Yates-corrected chi-square test for $2\times 2$ contingency tables to test this hypothesis.

**10.40** We have the test statistic

$$X^2 = \frac{n\left[|ad - bc| - \frac{n}{2}\right]^2}{(a+b)(c+d)(a+c)(b+d)}$$
$$= \frac{17{,}949\left[|43(11{,}902) - 103(5901)| - \frac{17{,}949}{2}\right]^2}{146(17{,}803)(5944)(12{,}005)}$$
$$= \frac{1.360\times 10^{14}}{1.855\times 10^{14}} = 0.73 \sim \chi_1^2 \text{ under } H_0$$

Since $\chi^2_{1,\,.50} = 0.45$, $\chi^2_{1,\,.75} = 1.32$, and

$$0.45 < 0.73 < 1.32,$$

it follows that $1 - .75 < p < 1 - .50$ or

$$.25 < p < .50.$$

Thus, there is no significant difference in breast cancer incidence between drinkers and nondrinkers.

**10.41** We use the chi-square test for linear trend using scores of 1, 2, 3, 4, 5 for the 5 alcohol-consumption groups. Compute the test statistic $X_1^2 = A^2/B$, where

$$A = 1(43) + 2(15) + 3(22) + 4(42) + 5(24)$$
$$-146\times\left[\frac{1(5944) + 2(2069) + \ldots + 5(2917)}{17{,}949}\right]$$
$$= 427 - 146\times\frac{49{,}294}{17{,}949} = 427 - 400.97 = 26.03$$
$$B = \frac{146(17{,}803)}{17{,}949^2}$$
$$\times\left[1(5944) + 4(2069) + \ldots + 25(2917) - \frac{49{,}294^2}{17{,}949}\right]$$
$$= .00807(175{,}306 - 135{,}377.93)$$
$$= .00807(39{,}928.07) = 322.14$$

Thus, $X_1^2 = 26.03^2/322.14 = 2.10 \sim \chi_1^2$ under $H_0$. Since $X_1^2 < \chi_{1,\,.95}^2 = 3.84$, it follows that $p > .05$, and there is no significant association between amount of alcohol consumption and incidence of breast cancer in this age group.

**10.42** The chi-square test for trend.

**10.43** We have the test statistic $X_1^2 = A^2/B \sim \chi_1^2$, under $H_0$. We will use scores of 0, 1, and 2 corresponding to the number of relatives with lung cancer = 0, 1, and 2+, respectively. We have the following $2 \times 3$ table:

| | Number of relatives with lung cancer | | | |
|---|---|---|---|---|
| | 0 | 1 | 2+ | Total |
| Cases | 393 | 119 | 20 | 532 |
| Controls | 466 | 78 | 8 | 552 |
| | 859 | 197 | 28 | 1084 |

For this data set,

$$A = 0(393) + 1(119) + 2(20) - \left(\frac{532}{1084}\right) \times [0(859) + 1(197) + 2(28)]$$
$$= 159 - 124.166 = 34.834$$
$$B = \left[\frac{532(552)}{1084^2}\right] \times \left\{0^2(859) + 1^2(197) + 2^2(28) - \frac{[0(859) + 1(197) + 2(28)]^2}{1084}\right\}$$
$$= .2499(309 - 59.049) = 62.467$$

Thus,

$$X_1^2 = \frac{34.834^2}{62.467} = 19.42 \sim \chi_1^2 \text{ under } H_0$$

Since $19.42 > 10.83 = \chi_{1,\,.999}^2$, it follows that $p < .001$. Since $A > 0$, we conclude that the cases have a significantly greater number of relatives with lung cancer than the controls.

**10.44** The proportions are an example of prevalence, because the subjects were asked whether they have hypertension at one point in time, viz. at the time of coronary angiography.

**10.45** We wish to test the hypothesis $H_0: p_1 = p_2$ versus $H_1: p_1 \neq p_2$, where

$p_1$ = true prevalence of hypertension among cases

$p_2$ = true prevalence of hypertension among controls

Under $H_0$, the best estimate of the common proportion $p$ is

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{1493(.37) + 707(.30)}{1493 + 707}$$
$$= \frac{552 + 212}{2200} = \frac{764}{2200} = .347$$

Since $n_1\hat{p}\hat{q} = 1493(.347)(.653) = 338.43 \geq 5$ and

$$n_2\hat{p}\hat{q} = 707(.347)(.653) = 160.26 \geq 5,$$

it follows that we can use the two-sample test for binomial proportions.

**10.46** The test statistic is

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \left[\frac{1}{2(1493)} + \frac{1}{2(707)}\right]}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$= \frac{|.37 - .30| - .001}{\sqrt{.347(.653)\left(\frac{1}{1493} + \frac{1}{707}\right)}}$$
$$= \frac{.0690}{.0217} = 3.173 \sim N(0,\ 1) \text{ under } H_0$$

The $p$-value

$$2 \times [1 - \Phi(3.173)] = 2 \times (1 - .9992) = .0015.$$

Thus, we can reject $H_0$ and conclude that the two underlying prevalence rates are not the same, with CAD cases having significantly greater rates of hypertension than controls.

**10.47** We have that each person is used as his or her own control. Thus, these are paired samples and we must use McNemar's test for correlated proportions to analyze the data.

**10.48** We test the hypothesis $H_0: p = 1/2$ versus $H_1: p \neq 1/2$, where $p$ = proportion of discordant pairs that are of type A. We have the test statistic

$$X^2 = \frac{\left(\left|12 - \frac{12+22}{2}\right| - .5\right)^2}{\frac{12+22}{4}} = \frac{(4.5)^2}{8.5}$$
$$= 2.38 \sim \chi_1^2 \text{ under } H_0$$

The $p$-value $= Pr(\chi_1^2 > 2.38) = .123$ by computer. Thus, there is no significant difference between the surrogate report by telephone and the self-report at the eye exam.

**10.49** The sensitivity and specificity are the appropriate measures.

**10.50** The sensitivity of the surrogate report

$$\begin{aligned} &Pr(\text{test } + | \text{true } +) \\ &= Pr(\text{surrogate report } + | \text{clinical determination } +) \\ &= \frac{14}{97} = .144 (\text{very poor!}) \end{aligned}$$

A 95% CI for the sensitivity is

$$.144 \pm 1.96\sqrt{\frac{.144(.856)}{97}} = .144 \pm .070 = (.074,\ .214).$$

The specificity is

$$Pr(\text{test } - | \text{true } -) = \frac{1247}{1273} = .980 \text{ (good)}.$$

A 95% CI for the specificity is

$$.980 \pm 1.96\sqrt{\frac{.980(.020)}{1273}} = .980 \pm .008 = (.972,\ .987).$$

The predictive value positive $(14/40 = .35)$ is also very poor. Thus, the surrogate report is also not an adequate substitute for a clinical examination for this particular condition.

## REFERENCES


[1] Kennedy, J. W., Ritchie, J. L., Davis, K. B., Stadius, M. L., Maynard, C., & Fritz, J. K. (1985). The western Washington randomized trial of intracoronary streptokinase in acute myocardial infarction: A 12-month follow-up report. *New England Journal of Medicine*, **312**(17), 1073–1078.

[2] The Coronary Drug Project Research Group. (1979). Cigarette smoking as a risk factor in men with a history of myocardial infarction. *Journal of Chronic Diseases*, **32**(6), 415–425.

[3] Furusko, K., Sato, K., Socda, T., et al. (1983, December 10). High dose intravenous gamma globulin for Kawasaki's syndrome [Letter]. *Lancet*, 1359.

[4] Barkan, S. E., & Bracken, M. (1987). Delayed childbearing: No evidence for increased risk of low birth weight and preterm delivery. *American Journal of Epidemiology*, **125**(l), 101–109.

[5] Willett, W., Stampfer, M. J., Colditz, G. A., Rosner, B. A., Hennekens, C. H., & Speizer, F. E. (1987). Moderate alcohol consumption and the risk of breast cancer. *New England Journal of Medicine*, **316**(19), 1174–1180.

[6] Shaw, G. L., Falk, R. T., Pickle, L. W., Mason, T J., & Buffler, P. A. (1991). Lung cancer risk associated with cancer in relatives. *Journal of Clinical Epidemiology*, **44**(4/5), 429–437.

[7] Applegate, W. B., Hughes, J. P., & Vanderzwaag, R. (1991). Case-control study of coronary heart disease risk factors in the elderly. *Journal of Clinical Epidemiology*, **44**(4/5), 409–415.

[8] Linton, K. L. P., Klein, B. E. K., & Klein, R. (1991). The validity of self-reported and surrogate-reported cataract and age-related macular degeneration in the Beaver Dam Eye Study. *American Journal of Epidemiology*, **134**(12), 1438–1446.