52. Binomial data gathered from more than one population are often presented in a contingency table. For the case of two populations, the table might look like the following.

Frequency	1	2	Total
Success	S_1	S_2	$S = S_1 + S_2$
Failure	F_1	F_2	$F = F_1 + F_2$
Total	n_1	n_2	$n = n_1 + n_2$

where population 1 is binomial(n_1, p_1) with S_1 successes and F_1 failures, and population two is binomial(n_2, p_2) with S_2 successes and F_2 failures. Hypotheses that are typically of interest are

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

(a) Show that a test can be based on the statistic

$$Y_{n_1,n_2} = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where

$$\hat{p}_1 = \frac{S_1}{n_1}, \quad \hat{p}_2 = \frac{S_2}{n_2}, \quad \text{and} \quad \hat{p} = \frac{S_1 + S_2}{n_1 + n_2}$$

Furthermore show that as $n_1 \to \infty$ and $n_2 \to \infty$ that

$$Y_{n_1,n_2} \xrightarrow{\mathcal{D}} \chi_1^2.$$

(b) Another way of measuring departure from H_0 is by calculating an expected frequency table by conditioning on the marginal totals and filling in the table according to $p_1 = p_2$; that is,

Population	1	2	Total
Success	$\frac{n_1 S}{n}$	$\frac{n_2 S}{n}$	$S = S_1 + S_2$
Failure	$\frac{n_1 F}{n}$	$\frac{n_2 F}{n}$	$F = F_1 + F_2$
Total	n_1	n_2	$n = n_1 + n_2$

Using this expected frequency table a statistic W_{n_1,n_2} is computed by using each table of the cell and computing

$$W_{n_1,n_2} = \sum \frac{(\text{observed - expected})^2}{\text{expected}} = \frac{(S_1 - \frac{n_1 S}{n})^2}{\frac{n_1 S}{n}} + ... + \frac{(F_2 - \frac{n_2 F}{n})^2}{\frac{n_2 F}{n}}$$

Prove $W_{n_1,n_2} = Y_{n_1,n_2}$ and so $W_{n_1,n_2} \xrightarrow{\mathcal{D}} \chi_1^2$.

(c) Another statistic used for testing the hypothesis is

$$Z_{n_1,n_2} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}.$$

Show that under the null hypothesis ($p_1 = p_2$) is asymptotically standard normal and hence its square is asymptotically χ_1^2. Also show $Z_{n_1,n_2}^2 \neq Y_{n_1,n_2}$

(d) Which of the statistics might be preferred and why?

(e) A famous medical experiment was conducted by Joseph Lister in the late 1800s. Mortality associated with surgery was very high, and Lister conjectured that the use of a disinfectant, carbolic acid, would help. Over a period of several years, Lister performed 75 amputations with and without using carbolic acid. The data are given below.

	Yes	No
Yes	34	19
No	6	16

Use the statistic in (a) and (c) to test whether the presence of carbolic acid is associated with patient mortality.