10.1.8 In trials of a medical screening test for a particular illness, 23 cases out of 324 positive results turned out to be "false-positive" results. The screening test is acceptable as long as p, the probability of a positive result being incorrect, is no larger than 10%. Calculate a p-value for the hypotheses

$$H_0 : p \geq 0.1 \quad \text{versus} \quad H_A : p < 0.1$$

Construct a 99% upper confidence bound on p. Do you think that the screening test is acceptable? (**With** $\alpha = 0.01$)

10.2.2 Suppose that $x = 261$ is an observation from a $B(302, p_A)$ random variable, and that $y = 401$ is an observation from a $B(454, p_B)$ random variable.
(a) Compute a two-sided 99% confidence interval for $p_A - p_B$.
(b) Compute a two-sided 90% confidence interval for $p_A - p_B$.
(c) Compute a one-sided 95% confidence interval that provides an upper bound for $p_A - p_B$.
(d) Calculate the p-value for the test of the hypotheses

$$H_0 : p_A = p_B \quad \text{versus} \quad H_A : p_A \neq p_B$$

For every test in problems below, do it with the significance level 0.05:

10.2.12 Recall from Problem 10.1.16 that in a particular day, 22 out of 542 visitors to a website followed a link provided by an advertiser. After the advertisements were modified, it was found that 64 out of 601 visitors to the website on a day followed the link. Is there any evidence that the modifications to the advertisements attracted more customers?

10.3.6 **Taste Tests for Soft Drink Formulations**
A beverage company has three formulations of a soft drink product. DS 10.3.6 gives the results of some taste tests where participants are asked to declare which formulation they like best. Is it plausible that the three formulations are equally popular?

10.3.10 Each of 205 consumers was asked to choose which of three products they preferred. Product A was chosen by 83 of the consumers, product B was chosen by 75 of the consumers, and product C was chosen by 47 of the consumers. Is there sufficient evidence to conclude that the three products do not have equal probabilities of being chosen?

10.3.12 An experiment was performed to investigate how long batteries remain charged under certain storage conditions. A total of 125 batteries were charged to the same level and stored in the designated conditions. After 24 hours all 125 batteries were tested and it was found that 12 of them had charges that had dropped below the threshold level. After an additional 24 hours the remaining 113 batteries were tested and it was found that 53 of them had charges that had dropped below the threshold level. Finally, after an additional 24 hours the remaining 60 batteries were tested and it was found that 39 of them had charges that had dropped below the threshold level. It is claimed that for these batteries under these storage conditions the time in hours until the charge drops below the threshold level has a Weibull distribution with parameters $\lambda = 0.065$ and $a = 0.45$. Are the results of this experiment consistent with that claim?

10.4.2 **Fertilizer Comparisons**

Seedlings are grown without fertilizer or with one of two kinds of fertilizer. After a certain period of time a seedling's growth is classified into one of four categories, as given in DS 10.4.2. Test whether the seedlings' growth can be taken to be the same for all three sets of growing conditions.

10.4.6 Show that for a 2×2 contingency table the Pearson chi-square statistic can be written

$$X^2 = \frac{n(x_{11}x_{22} - x_{12}x_{21})^2}{x_{1.}x_{.1}x_{2.}x_{.2}}$$

10.4.10 **Asphalt Load Testing**

An experiment was conducted to compare three types of asphalt. Samples of each type of asphalt were subjected to repeated loads at high temperatures, and the resulting cracking was analyzed. For type A, 57 samples were tested, of which 9 had severe cracking, 17 had medium cracking, and 31 had minor cracking. For type B, 49 samples were tested, of which 4 had severe cracking, 9 had medium cracking, and 36 had minor cracking. For type C, 90 samples were tested, of which 15 had severe cracking, 19 had medium cracking, and 56 had minor cracking. Does this experiment provide any evidence that the three types of asphalt are different with respect to cracking?

10.7.20 Clinical Trial

Patients were diagnosed as being either condition A or condition B before undergoing a treatment. The treatment was successful for 56 out of 94 patients classified as condition A, and the treatment was successful for 64 out of 153 patients classified as condition B.

(a) Perform a hypothesis test to assess whether there is sufficient evidence to conclude that the chance of success for patients with condition A is better than 50%.

(b) Construct a two-sided 99% confidence interval for the difference between the success probabilities for patients with condition A and with condition B.

(c) Perform a chi-square goodness of fit test to investigate whether there is sufficient evidence to conclude that the success probabilities are different for patients with condition A and with condition B. What is your conclusion?