8.5 Hypothesis Tests Concerning the Variance of a Normal Population

8.5 HYPOTHESIS TESTS CONCERNING THE VARIANCE OF A NORMAL POPULATION

Let $X_1, \ldots, X_n$ denote a sample from a normal population having unknown mean $\mu$ and unknown variance $\sigma^2$, and suppose we desire to test the hypothesis

$$H_0 : \sigma^2 = \sigma_0^2$$

versus the alternative

$$H_1 : \sigma^2 \neq \sigma_0^2$$

for some specified value $\sigma_0^2$.

To obtain a test, recall (as was shown in Section 6.5) that $(n - 1)S^2/\sigma^2$ has a chi-square distribution with $n - 1$ degrees of freedom. Hence, when $H_0$ is true

$$\frac{(n - 1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

and so

$$P_{H_0} \left\{ \chi^2_{1-\alpha/2,n-1} \leq \frac{(n - 1)S^2}{\sigma_0^2} \leq \chi^2_{\alpha/2,n-1} \right\} = 1 - \alpha$$

Therefore, a significance level $\alpha$ test is to

accept $H_0$ if $\chi^2_{1-\alpha/2,n-1} \leq \frac{(n - 1)S^2}{\sigma_0^2} \leq \chi^2_{\alpha/2,n-1}$

reject $H_0$ otherwise

The preceding test can be implemented by first computing the value of the test statistic $(n - 1)S^2/\sigma_0^2$ — call it $c$. Then compute the probability that a chi-square random variable with $n - 1$ degrees of freedom would be (a) less than and (b) greater than $c$. If either of these probabilities is less than $\alpha/2$, then the hypothesis is rejected. In other words, the $p$-value of the test data is

$$p\text{-value} = 2 \min(P(\chi^2_{n-1} < c), 1 - P(\chi^2_{n-1} < c))$$

The quantity $P(\chi^2_{n-1} < c)$ can be obtained from Program 5.8.1.A. The $p$-value for a one-sided test is similarly obtained.

EXAMPLE 8.5a A machine that automatically controls the amount of ribbon on a tape has recently been installed. This machine will be judged to be effective if the standard deviation $\sigma$ of the amount of ribbon on a tape is less than .15 cm. If a sample of 20 tapes yields a sample variance of $S^2 = .025 \text{ cm}^2$, are we justified in concluding that the machine is ineffective?
SOLUTION We will test the hypothesis that the machine is effective, since a rejection of this hypothesis will then enable us to conclude that it is ineffective. Since we are thus interested in testing

\[ H_0 : \sigma^2 \leq 0.0225 \quad \text{versus} \quad H_1 : \sigma^2 > 0.0225 \]

it follows that we would want to reject \( H_0 \) when \( S^2 \) is large. Hence, the \( p \)-value of the preceding test data is the probability that a chi-square random variable with 19 degrees of freedom would exceed the observed value of \( 19S^2/0.0225 = 19 \times 0.0258 = 21.11 \). That is,

\[
p\text{-value} = P\{X_{19}^2 > 21.111\} = 1 - 0.6693 = 0.3307 \quad \text{from Program 5.8.1.A}
\]

Therefore, we must conclude that the observed value of \( S^2 = 0.0258 \) is not large enough to reasonably preclude the possibility that \( \sigma^2 \leq 0.0225 \), and so the null hypothesis is accepted. \( \blacksquare \)

8.5.1 Testing for the Equality of Variances of Two Normal Populations

Let \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_m \) denote independent samples from two normal populations having respective (unknown) parameters \( \mu_x, \sigma_x^2 \) and \( \mu_y, \sigma_y^2 \) and consider a test of

\[ H_0 : \sigma_x^2 = \sigma_y^2 \quad \text{versus} \quad H_1 : \sigma_x^2 \neq \sigma_y^2 \]

If we let

\[
S_x^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}
\]

\[
S_y^2 = \frac{\sum_{i=1}^{m} (Y_i - \bar{Y})^2}{m - 1}
\]

denote the sample variances, then as shown in Section 6.5, \((n - 1)S_x^2/\sigma_x^2\) and \((m - 1)S_y^2/\sigma_y^2\) are independent chi-square random variables with \( n - 1 \) and \( m - 1 \) degrees of freedom, respectively. Therefore, \((S_x^2/\sigma_x^2)/(S_y^2/\sigma_y^2)\) has an \( F \)-distribution with parameters \( n - 1 \) and \( m - 1 \). Hence, when \( H_0 \) is true

\[
S_x^2/S_y^2 \sim F_{n-1,m-1}
\]

and so

\[
P_{H_0}\{F_{1,2,n-1,m-1} \leq S_x^2/S_y^2 \leq F_{\alpha/2,2,n-1,m-1}\} = 1 - \alpha
\]
Thus, a significance level $\alpha$ test of $H_0$ against $H_1$ is to

\[
\begin{align*}
\text{accept } H_0 & \text{ if } F_{1-\alpha/2,n-1,m-1} < \frac{S_1^2}{S_2^2} < F_{\alpha/2,n-1,m-1} \\
\text{reject } H_0 & \text{ otherwise}
\end{align*}
\]

The preceding test can be effected by first determining the value of the test statistic $\frac{S_1^2}{S_2^2}$, say its value is $v$, and then computing $P(F_{n-1,m-1} \leq v)$ where $F_{n-1,m-1}$ is an $F$-random variable with parameters $n-1$, $m-1$. If this probability is either less than $\alpha/2$ (which occurs when $S_1^2$ is significantly less than $S_2^2$) or greater than $1 - \alpha/2$ (which occurs when $S_1^2$ is significantly greater than $S_2^2$), then the hypothesis is rejected. In other words, the $p$-value of the test data is

\[
p\text{-value} = 2 \min(P(F_{n-1,m-1} < v), 1 - P(F_{n-1,m-1} < v))
\]

The test now calls for rejection whenever the significance level $\alpha$ is at least as large as the $p$-value.

**EXAMPLE 8.5b** There are two different choices of a catalyst to stimulate a certain chemical process. To test whether the variance of the yield is the same no matter which catalyst is used, a sample of 10 batches is produced using the first catalyst, and 12 using the second. If the resulting data is $S_1^2 = .14$ and $S_2^2 = .28$, can we reject, at the 5 percent level, the hypothesis of equal variance?

**SOLUTION** Program 5.8.3, which computes the $F$ cumulative distribution function, yields that

\[
P(F_{9,11} \leq .5) = .1539
\]

Hence,

\[
p\text{-value} = 2 \min(\{.1539, .8461\} = .3074
\]

and so the hypothesis of equal variance cannot be rejected.

---

### 8.6 HYPOTHESIS TESTS IN BERNOULLI POPULATIONS

The binomial distribution is frequently encountered in engineering problems. For a typical example, consider a production process that manufactures items that can be classified in one of two ways — either as acceptable or as defective. An assumption often made is that each item produced will, independently, be defective with probability $p$; and so the number of defects in a sample of $n$ items will thus have a binomial distribution with parameters $(n, p)$. We will now consider a test of

\[
H_0 : p \leq p_0 \quad \text{versus} \quad H_1 : p > p_0
\]

where $p_0$ is some specified value.
If we let $X$ denote the number of defects in the sample of size $n$, then it is clear that we wish to reject $H_0$ when $X$ is large. To see how large it need be to justify rejection at the $\alpha$ level of significance, note that

$$P(X \geq k) = \sum_{i=k}^{n} P(X = i) = \sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i}$$

Now it is certainly intuitive (and can be proven) that $P(X \geq k)$ is an increasing function of $p$ — that is, the probability that the sample will contain at least $k$ errors increases in the defect probability $p$. Using this, we see that when $H_0$ is true (and so $p \leq p_0$),

$$P(X \geq k) \leq \sum_{i=k}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i}$$

Hence, a significance level $\alpha$ test of $H_0 : p \leq p_0$ versus $H_1 : p > p_0$ is to reject $H_0$ when

$$X \geq k^*$$

where $k^*$ is the smallest value of $k$ for which $\sum_{i=k}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha$. That is,

$$k^* = \min \left\{ k : \sum_{i=k}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i} \leq \alpha \right\}$$

This test can best be performed by first determining the value of the test statistic — say, $X = x$ — and then computing the $p$-value given by

$$p\text{-value} = P(B(n, p_0) \geq x)$$

$$= \sum_{i=x}^{n} \binom{n}{i} p_0^i (1-p_0)^{n-i}$$

**EXAMPLE 8.6a** A computer chip manufacturer claims that no more than 2 percent of the chips it sends out are defective. An electronics company, impressed with this claim, has purchased a large quantity of such chips. To determine if the manufacturer's claim can be taken literally, the company has decided to test a sample of 300 of these chips. If 10 of these 300 chips are found to be defective, should the manufacturer's claim be rejected?

**SOLUTION** Let us test the claim at the 5 percent level of significance. To see if rejection is called for, we need to compute the probability that the sample of size 300 would have resulted in 10 or more defectives when $p$ is equal to .02. (That is, we compute the $p$-value.) If this probability is less than or equal to .05, then the manufacturer's claim
should be rejected. Now

\[ P_{.02}\{X \geq 10\} = 1 - P_{.02}\{X < 10\} \]
\[ = 1 - \sum_{i=0}^{9} \binom{300}{i}(.02)^i(.98)^{300-i} \]
\[ = .0818 \quad \text{from Program 3.1} \]

and so the manufacturer’s claim cannot be rejected at the 5 percent level of significance.

When the sample size \( n \) is large, we can derive an approximate significance level \( \alpha \) test of \( H_0: p \leq p_0 \) versus \( H_1: p > p_0 \) by using the normal approximation to the binomial. It works as follows: Because when \( n \) is large \( X \) will have approximately a normal distribution with mean and variance

\[ E[X] = np, \quad \text{Var}(X) = np(1 - p) \]

it follows that

\[ \frac{X - np}{\sqrt{np(1 - p)}} \]

will have approximately a standard normal distribution. Therefore, an approximate significance level \( \alpha \) test would be to reject \( H_0 \) if

\[ \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} \geq z_\alpha \]

Equivalently, one can use the normal approximation to approximate the \( p \)-value.

**EXAMPLE 8.6b** In Example 8.6a, \( np_0 = 300(.02) = 6 \), and \( \sqrt{np_0(1 - p_0)} = \sqrt{5.88} \). Consequently, the \( p \)-value that results from the data \( X = 10 \) is

\[ p\text{-value} = P_{.02}\{X \geq 10\} \]
\[ = P_{.02}\{X \geq 9.5\} \]
\[ = P_{.02}\left\{ \frac{X - 6}{\sqrt{5.88}} \geq \frac{9.5 - 6}{\sqrt{5.88}} \right\} \]
\[ \approx P(Z \geq 1.443) \]
\[ = .0745 \quad \text{■} \]
Suppose now that we want to test the null hypothesis that $p$ is equal to some specified value; that is, we want to test

$$H_0 : p = p_0 \quad \text{versus} \quad H_1 : p \neq p_0$$

If $X$, a binomial random variable with parameters $n$ and $p$, is observed to equal $x$, then a significance level $\alpha$ test would reject $H_0$ if the value $x$ was either significantly larger or significantly smaller than what would be expected when $p$ is equal to $p_0$. More precisely, the test would reject $H_0$ if either

$$P(\text{Bin}(n, p_0) \geq x) \leq \alpha/2 \quad \text{or} \quad P(\text{Bin}(n, p_0) \leq x) \leq \alpha/2$$

In other words, the $p$-value when $X = x$ is

$$p\text{-value} = 2 \min(P(\text{Bin}(n, p_0) \geq x), P(\text{Bin}(n, p_0) \leq x))$$

**EXAMPLE 8.6c**  Historical data indicate that 4 percent of the components produced at a certain manufacturing facility are defective. A particularly acrimonious labor dispute has recently been concluded, and management is curious about whether it will result in any change in this figure of 4 percent. If a random sample of 500 items indicated 16 defectives (3.2 percent), is this significant evidence, at the 5 percent level of significance, to conclude that a change has occurred?

**SOLUTION**  To be able to conclude that a change has occurred, the data need to be strong enough to reject the null hypothesis when we are testing

$$H_0 : p = .04 \quad \text{versus} \quad H_1 : p \neq .04$$

where $p$ is the probability that an item is defective. The $p$-value of the observed data of 16 defectives in 500 items is

$$p\text{-value} = 2 \min(P(X \leq 16), P(X \geq 16))$$

where $X$ is a binomial $(500, .04)$ random variable. Since $500 \times .04 = 20$, we see that

$$p\text{-value} = 2P(X \leq 16)$$

Since $X$ has mean 20 and standard deviation $\sqrt{20(.04)} = 4.38$, it is clear that twice the probability that $X$ will be less than or equal to 16 — a value less than one standard deviation lower than the mean — is not going to be small enough to justify rejection. Indeed, it can be shown that

$$p\text{-value} = 2P(X \leq 16) = .432$$

and so there is not sufficient evidence to reject the hypothesis that the probability of a defective item has remained unchanged. ■
8.6.1 Testing the Equality of Parameters in Two Bernoulli Populations

Suppose there are two distinct methods for producing a certain type of transistor; and suppose that transistors produced by the first method will, independently, be defective with probability \( p_1 \), with the corresponding probability being \( p_2 \) for those produced by the second method. To test the hypothesis that \( p_1 = p_2 \), a sample of \( n_1 \) transistors is produced using method 1 and \( n_2 \) using method 2.

Let \( X_1 \) denote the number of defective transistors obtained from the first sample and \( X_2 \) for the second. Thus, \( X_1 \) and \( X_2 \) are independent binomial random variables with respective parameters \( (n_1, p_1) \) and \( (n_2, p_2) \). Suppose that \( X_1 + X_2 = k \) and so there have been a total of \( k \) defectives. Now, if \( H_0 \) is true, then each of the \( n_1 + n_2 \) transistors produced will have the same probability of being defective, and so the determination of the \( k \) defectives will have the same distribution as a random selection of a sample of size \( k \) from a population of \( n_1 + n_2 \) items of which \( n_1 \) are white and \( n_2 \) are black. In other words, given a total of \( k \) defectives, the conditional distribution of the number of defective transistors obtained from method 1 will, when \( H_0 \) is true, have the following hypergeometric distribution:

\[
P_{H_0}[X_1 = i \mid X_1 + X_2 = k] = \binom{n_1}{i} \binom{n_2}{k-i} / \binom{n_1 + n_2}{k}, \quad i = 0, 1, \ldots, k \tag{8.6.1}
\]

Now, in testing

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

it seems reasonable to reject the null hypothesis when the proportion of defective transistors produced by method 1 is much different than the proportion of defectives obtained under method 2. Therefore, if there is a total of \( k \) defectives, then we would expect, when \( H_0 \) is true, that \( X_1/n_1 \) (the proportion of defective transistors produced by method 1) would be close to \( (k - X_1)/n_2 \) (the proportion of defective transistors produced by method 2). Because \( X_1/n_1 \) and \( (k - X_1)/n_2 \) will be farthest apart when \( X_1 \) is either very small or very large, it thus seems that a reasonable significance level \( \alpha \) test of Equation 8.6.1 is as follows. If \( X_1 + X_2 = k \), then one should

- reject \( H_0 \) if either \( P(X \leq x_1) \leq \alpha/2 \) or \( P(X \geq x_1) \leq \alpha/2 \)
- accept \( H_0 \) otherwise

* See Example 5.3b for a formal verification of Equation 8.6.1.
where $X$ is a hypergeometric random variable with probability mass function

$$P(X = i) = \binom{n_1}{i} \binom{n_2}{k - i} \binom{n_1 + n_2}{k} i = 0, \ldots, k$$

(8.6.2)

In other words, this test will call for rejection if the significance level is at least as large as the $p$-value given by

$$p\text{-value} = 2 \min(P[X \leq x_1], P[X \geq x_1])$$

(8.6.3)

This is called the Fisher-Irwin test.

COMPUTATIONS FOR THE FISHER-IRWIN TEST

To utilize the Fisher-Irwin test, we need to be able to compute the hypergeometric distribution function. To do so, note that with $X$ having mass function Equation 8.6.2,

$$\frac{P[X = i + 1]}{P[X = i]} = \frac{\binom{n_1}{i + 1} \binom{n_2}{k - i - 1}}{\binom{n_1}{i} \binom{n_2}{k - i}} = \frac{(n_1 - i)(k - i)}{(i + 1)(n_2 - k + i + 1)}$$

(8.6.4)

(8.6.5)

where the verification of the final equality is left as an exercise.

Program 8.6.1 uses the preceding identity to compute the $p$-value of the data for the Fisher-Irwin test of the equality of two Bernoulli probabilities. The program will work best if the Bernoulli outcome that is called unsuccessful (or defective) is the one whose probability is less than .5. For instance, if over half the items produced are defective, then rather than testing that the defect probability is the same in both samples, one should test that the probability of producing an acceptable item is the same in both samples.

EXAMPLE 8.6d Suppose that method 1 resulted in 20 unacceptable transistors out of 100 produced; whereas method 2 resulted in 12 unacceptable transistors out of 100 produced. Can we conclude from this, at the 10 percent level of significance, that the two methods are equivalent?

SOLUTION Upon running Program 8.6.1, we obtain that

$$p\text{-value} = .1763$$

Hence, the hypothesis that the two methods are equivalent cannot be rejected.
The ideal way to test the hypothesis that the results of two different treatments are identical is to randomly divide a group of people into a set that will receive the first treatment and one that will receive the second. However, such randomization is not always possible. For instance, if we want to study whether drinking alcohol increases the risk of prostate cancer, we cannot instruct a randomly chosen sample to drink alcohol. An alternative way to study the hypothesis is to use an observational study that begins by randomly choosing a set of drinkers and one of nondrinkers. These sets are followed for a period of time and the resulting data is then used to test the hypothesis that members of the two groups have the same risk for prostate cancer.

Our next sample illustrates another way of performing an observational study.

**EXAMPLE 8.6e** In 1970, the researchers Herbst, Ulfelder, and Poskanzer (H-U-P) suspected that vaginal cancer in young women, a rather rare disease, might be caused by one’s mother having taken the drug diethylstilbestrol (usually referred to as DES) while pregnant. To study this possibility, the researchers could have performed an observational study by searching for a (treatment) group of women whose mothers took DES when pregnant and a (control) group of women whose mothers did not. They could then observe these groups for a period of time and use the resulting data to test the hypothesis that the probabilities of contracting vaginal cancer are the same for both groups. However, because vaginal cancer is so rare (in both groups) such a study would require a large number of individuals in both groups and would probably have to continue for many years to obtain significant results. Consequently, H-U-P decided on a different type of observational study. They uncovered 8 women between the ages of 15 and 22 who had vaginal cancer. Each of these women (called cases) was then matched with 4 others, called referents or controls. Each of the referents of a case was free of the cancer and was born within 5 days in the same hospital and in the same type of room (either private or public) as the case. Arguing that if DES had no effect on vaginal cancer then the probability, call it \( p_c \), that the mother of a case took DES would be the same as the probability, call it \( p_r \), that the mother of a referent took DES, the researchers H-U-P decided to test

\[
H_0 : p_c = p_r \quad \text{against} \quad H_1 : p_c \neq p_r
\]

Discovering that 7 of the 8 cases had mothers who took DES while pregnant, while none of the 32 referents had mothers who took the drug, the researchers (see Herbst, A., Ulfelder, H., and Poskanzer, D., “Adenocarcinoma of the Vagina: Association of Maternal Stilbestrol Therapy with Tumor Appearance in Young Women,” *New England Journal of Medicine*, 284, 878–881, 1971) concluded that there was a strong association between DES and vaginal cancer. (The \( p \)-value for these data is approximately 0.)

When \( n_1 \) and \( n_2 \) are large, an approximate level \( \alpha \) test of \( H_0 : p_1 = p_2 \), based on the normal approximation to the binomial, is outlined in Problem 63.
8.7 TESTS CONCERNING THE MEAN OF A POISSON DISTRIBUTION

Let \( X \) denote a Poisson random variable having mean \( \lambda \) and consider a test of

\[
H_0 : \lambda = \lambda_0 \quad \text{versus} \quad H_1 : \lambda \neq \lambda_0
\]

If the observed value of \( X \) is \( x \), then a level \( \alpha \) test would reject \( H_0 \) if either

\[
P_{\lambda_0}(X \geq x) \leq \alpha/2 \quad \text{or} \quad P_{\lambda_0}(X \leq x) \leq \alpha/2 \tag{8.7.1}
\]

where \( P_{\lambda_0} \) means that the probability is computed under the assumption that the Poisson mean is \( \lambda_0 \). It follows from Equation 8.7.1 that the \( p \)-value is given by

\[
p\text{-value} = 2 \min(P_{\lambda_0}(X \geq x), P_{\lambda_0}(X \leq x))
\]

The calculation of the preceding probabilities that a Poisson random variable with mean \( \lambda_0 \) is greater (less) than or equal to \( x \) can be obtained by using Program 5.2.

EXAMPLE 8.7a Management’s claim that the mean number of defective computer chips produced daily is not greater than 25 is in dispute. Test this hypothesis, at the 5 percent level of significance, if a sample of 5 days revealed 28, 34, 32, 38, and 22 defective chips.

SOLUTION Because each individual computer chip has a very small chance of being defective, it is probably reasonable to suppose that the daily number of defective chips is approximately a Poisson random variable, with mean, say, \( \lambda \). To see whether or not the manufacturer’s claim is credible, we shall test the hypothesis

\[
H_0 : \lambda \leq 25 \quad \text{versus} \quad H_1 : \lambda > 25
\]

Now, under \( H_0 \), the total number of defective chips produced over a 5-day period is Poisson distributed (since the sum of independent Poisson random variables is Poisson) with a mean no greater than 125. Since this number is equal to 154, it follows that the \( p \)-value of the data is given by

\[
p\text{-value} = P_{125}(X \geq 154) \\
= 1 - P_{125}(X \leq 153) \\
= .0066 \quad \text{from Program 5.2}
\]

Therefore, the manufacturer’s claim is rejected at the 5 percent (as it would be even at the 1 percent) level of significance.
REMARK
If Program 5.2 is not available, one can use the fact that a Poisson random variable with mean \( \lambda \) is, for large \( \lambda \) approximately normally distributed with a mean and variance equal to \( \lambda \).

8.7.1 TESTING THE RELATIONSHIP BETWEEN TWO Poisson PARAMETERS
Let \( X_1 \) and \( X_2 \) be independent Poisson random variables with respective means \( \lambda_1 \) and \( \lambda_2 \), and consider a test of

\[ H_0 : \lambda_2 = c \lambda_1 \quad \text{versus} \quad H_1 : \lambda_2 \neq c \lambda_1 \]

for a given constant \( c \). Our test of this is a conditional test (similar in spirit to the Fisher-Irwin test of Section 8.6.1), which is based on the fact that the conditional distribution of \( X_1 \) given the sum of \( X_1 \) and \( X_2 \) is binomial. More specifically, we have the following proposition.

PROPOSITION 8.7.1

\[ P[X_1 = k | X_1 + X_2 = n] = \binom{n}{k} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k} \]

Proof

\[
P[X_1 = k | X_1 + X_2 = n] \\
= \frac{P[X_1 = k, X_1 + X_2 = n]}{P[X_1 + X_2 = n]} \\
= \frac{P[X_1 = k, X_2 = n-k]}{P[X_1 + X_2 = n]} \\
= \frac{P[X_1 = k]P[X_2 = n-k]}{P[X_1 + X_2 = n]} \quad \text{by independence} \\
= \frac{\exp\{-\lambda_1\} \lambda_1^k/k! \exp\{-\lambda_2\} \lambda_2^{n-k}/(n-k)!}{\exp\{-\lambda_1 - \lambda_2\}(\lambda_1 + \lambda_2)^n/n!} \\
= \frac{n!}{(n-k)!k!} \left[ \frac{\lambda_1}{\lambda_1 + \lambda_2} \right]^k \left[ \frac{\lambda_2}{\lambda_1 + \lambda_2} \right]^{n-k} \quad \square
\]

It follows from Proposition 8.7.1 that, if \( H_0 \) is true, then the conditional distribution of \( X_1 \) given that \( X_1 + X_2 = n \) is the binomial distribution with parameters \( n \) and \( p = 1/(1+c) \). From this we can conclude that if \( X_1 + X_2 = n \), then \( H_0 \) should be rejected if the observed value of \( X_1 \), call it \( x_1 \), is such that either

\[ P[\text{Bin}(n, 1/(1 + c)) \geq x_1] \leq \alpha/2 \]
or
\[ P(\text{Bin}(n, 1/(1 + c)) \leq x_1) \leq \alpha/2 \]

**EXAMPLE 8.7b**  An industrial concern runs two large plants. If the number of accidents during the last 8 weeks at plant 1 were 16, 18, 9, 22, 17, 19, 24, 8 while the number of accidents during the last 6 weeks at plant 2 were 22, 18, 26, 30, 25, 28, can we conclude, at the 5 percent level of significance, that the safety conditions differ from plant to plant?

**SOLUTION**  Since there is a small probability of an industrial accident in any given minute, it would seem that the weekly number of such accidents should have approximately a Poisson distribution. If we let \( X_1 \) denote the total number of accidents during an 8-week period at plant 1, and let \( X_2 \) be the number during a 6-week period at plant 2, then if the safety conditions did not differ at the two plants we would have that

\[ \lambda_2 = \frac{3}{4} \lambda_1 \]

where \( \lambda_i \equiv E[X_i], i = 1, 2 \). Hence, as \( X_1 = 133, X_2 = 149 \) it follows that the \( p \)-value of the test of

\[ H_0 : \lambda_2 = \frac{3}{4} \lambda_1 \quad \text{versus} \quad H_1 : \lambda_2 \neq \frac{3}{4} \lambda_1 \]

is given by

\[ p\text{-value} = 2 \min \{ P(\text{Bin}(282, \frac{3}{4}) \geq 133), P(\text{Bin}(282, \frac{3}{4}) \leq 133) \} \]

\[ = 9.408 \times 10^{-4} \]

Thus, the hypothesis that the safety conditions at the two plants are equivalent is rejected.

**EXAMPLE 8.7c**  In an attempt to show that proofreader A is superior to proofreader B, both proofreaders were given the same manuscript to read. If proofreader A found 28 errors, and proofreader B found 18, with 10 of these errors being found by both, can we conclude that A is the superior proofreader?

**SOLUTION**  To begin, we need a model. So let us assume that each manuscript error is independently found by proofreader A with probability \( P_A \) and by proofreader B with probability \( P_B \). To see if the data prove that A is the superior proofreader, we need to check if it would lead to rejecting the hypothesis that B is at least as good. That is, we need to test the null hypothesis

\[ H_0 : P_A \leq P_B \]

against the alternative hypothesis

\[ H_1 : P_A > P_B \]
To determine a test, note that each error can be classified as being of one of 4 types: it is type 1 if it is found by both proofreaders; it is type 2 if found by A but not by B; it is type 3 if found by B but not by A; and it is type 4 if found by neither. Thus, under our independence assumptions, it follows that each error will independently be type \( i \) with probability \( p_i \), where

\[
p_1 = P_A P_B, \quad p_2 = P_A (1 - P_B), \quad p_3 = (1 - P_A) P_B, \quad p_4 = (1 - P_A) (1 - P_B)
\]

Now, if we do our analysis under the assumption that \( N \), the total number of errors in the manuscript, is a random variable that is Poisson distributed with some unknown mean \( \lambda \), then it follows from the results of Section 5.2 that the numbers of errors of types 1, 2, 3, 4 are independent Poisson random variables with respective means \( \lambda p_1, \lambda p_2, \lambda p_3, \lambda p_4 \). Now, because

\[
\frac{x}{1-x} = \frac{1}{1/x - 1}
\]

is an increasing function of \( x \) in the region \( 0 \leq x \leq 1 \),

\[
P_A > P_B \iff \frac{P_A}{1 - P_A} > \frac{P_B}{1 - P_B} \iff P_A (1 - P_B) > (1 - P_A) P_B
\]

In other words, \( P_A > P_B \) if and only if \( p_2 > p_3 \). As a result, it suffices to use the data to test

\[
H_0 : p_2 \leq p_3 \quad \text{versus} \quad H_1 : p_2 > p_3
\]

Therefore, with \( N_2 \) denoting the number of errors of type 2 (that is, the number of errors found by A but not by B), and \( N_3 \) the number of errors of type 3 (that is, the number found by B but not by A), it follows that we need to test

\[
H_0 : E[N_2] \leq E[N_3] \quad \text{versus} \quad H_1 : E[N_2] > E[N_3]
\]

where \( N_2 \) and \( N_3 \) are independent Poisson random variables. Now, by Proposition 8.7.1, the conditional distribution of \( N_2 \) given \( N_2 + N_3 \) is binomial \( (n, p) \) where \( n = N_2 + N_3 \) and \( p = (E[N_2])/(E[N_2] + E[N_3]) \). Because Equation 8.7.2 is equivalent to

\[
H_0 : p \leq 1/2 \quad \text{versus} \quad H_1 : p > 1/2
\]

it follows that the \( p \)-value that results when \( N_2 = n_2, N_3 = n_3 \) is

\[
p\text{-value} = P(\text{Bin}(n_2 + n_3, .5) \geq n_2)
\]

For the data given, \( n_2 = 18, n_3 = 8 \), yielding that

\[
p\text{-value} = P(\text{Bin}(26, .5) \geq 18) = .0378
\]

Consequently, at the 5 percent level of significance, the null hypothesis is rejected leading to the conclusion that \( A \) is the superior proofreader.
Problems

1. Consider a trial in which a jury must decide between the hypothesis that the defendant is guilty and the hypothesis that he or she is innocent.
   
   (a) In the framework of hypothesis testing and the U.S. legal system, which of the hypotheses should be the null hypothesis?
   
   (b) What do you think would be an appropriate significance level in this situation?

2. A colony of laboratory mice consists of several thousand mice. The average weight of all the mice is 32 grams with a standard deviation of 4 grams. A laboratory assistant was asked by a scientist to select 25 mice for an experiment. However, before performing the experiment the scientist decided to weigh the mice as an indicator of whether the assistant’s selection constituted a random sample or whether it was made with some unconscious bias (perhaps the mice selected were the ones that were slowest in avoiding the assistant, which might indicate some inferiority about this group). If the sample mean of the 25 mice was 30.4, would this be significant evidence, at the 5 percent level of significance, against the hypothesis that the selection constituted a random sample?

3. A population distribution is known to have standard deviation 20. Determine the p-value of a test of the hypothesis that the population mean is equal to 50, if the average of a sample of 64 observations is (a) 52.5; (b) 55.0; (c) 57.5.

4. In a certain chemical process, it is very important that a particular solution that is to be used as a reactant have a pH of exactly 8.20. A method for determining pH that is available for solutions of this type is known to give measurements that are normally distributed with a mean equal to the actual pH and with a standard deviation of .02. Suppose 10 independent measurements yielded the following pH values:

   8.18  8.17  
   8.16  8.15  
   8.17  8.21  
   8.22  8.16  
   8.19  8.18

   (a) What conclusion can be drawn at the $\alpha = .10$ level of significance?
   
   (b) What about at the $\alpha = .05$ level of significance?

5. The mean breaking strength of a certain type of fiber is required to be at least 200 psi. Past experience indicates that the standard deviation of breaking strength
is 5 psi. If a sample of 8 pieces of fiber yielded breakage at the following pressures,

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>210</td>
<td>198</td>
</tr>
<tr>
<td>195</td>
<td>202</td>
</tr>
<tr>
<td>197.4</td>
<td>196</td>
</tr>
<tr>
<td>199</td>
<td>195.5</td>
</tr>
</tbody>
</table>

would you conclude, at the 5 percent level of significance, that the fiber is unacceptable? What about at the 10 percent level of significance?

6. It is known that the average height of a man residing in the United States is 5 feet 10 inches and the standard deviation is 3 inches. To test the hypothesis that men in your city are “average,” a sample of 20 men have been chosen. The heights of the men in the sample follow:

<table>
<thead>
<tr>
<th>Man</th>
<th>Height in</th>
<th>Inches</th>
<th>Man</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
<td>70.4</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>68.1</td>
<td>76</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>69.2</td>
<td>72.5</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>72.8</td>
<td>74</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>71.2</td>
<td>71.8</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>72.2</td>
<td>69.6</td>
<td>16</td>
</tr>
<tr>
<td>7</td>
<td>70.8</td>
<td>75.6</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>74</td>
<td>70.6</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>66</td>
<td>76.2</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>70.3</td>
<td>77</td>
<td>20</td>
</tr>
</tbody>
</table>

What do you conclude? Explain what assumptions you are making.

7. Suppose in Problem 4 that we wished to design a test so that if the pH were really equal to 8.20, then this conclusion will be reached with probability equal to .95. On the other hand, if the pH differs from 8.20 by .03 (in either direction), we want the probability of picking up such a difference to exceed .95.

(a) What test procedure should be used?
(b) What is the required sample size?
(c) If $\bar{x} = 8.31$, what is your conclusion?
(d) If the actual pH is 8.32, what is the probability of concluding that the pH is not 8.20, using the foregoing procedure?

8. Verify that the approximation in Equation 8.3.7 remains valid even when $\mu_1 < \mu_0$.

9. A British pharmaceutical company, Glaxo Holdings, has recently developed a new drug for migraine headaches. Among the claims Glaxo made for its drug, called
somatrapan, was that the mean time it takes for it to enter the bloodstream is less than 10 minutes. To convince the Food and Drug Administration of the validity of this claim, Glaxo conducted an experiment on a randomly chosen set of migraine sufferers. To prove its claim, what should they have taken as the null and what as the alternative hypothesis?

10. The weights of salmon grown at a commercial hatchery are normally distributed with a standard deviation of 1.2 pounds. The hatchery claims that the mean weight of this year's crop is at least 7.6 pounds. Suppose a random sample of 16 fish yielded an average weight of 7.2 pounds. Is this strong enough evidence to reject the hatchery's claims at the

(a) 5 percent level of significance;
(b) 1 percent level of significance?
(c) What is the $p$-value?

11. Consider a test of $H_0 : \mu \leq 100$ versus $H_1 : \mu > 100$. Suppose that a sample of size 20 has a sample mean of $\bar{X} = 105$. Determine the $p$-value of this outcome if the population standard deviation is known to equal

(a) 5; (b) 10; (c) 15.

12. An advertisement for a new toothpaste claims that it reduces cavities of children in their cavity-prone years. Cavities per year for this age group are normal with mean 3 and standard deviation 1. A study of 2,500 children who used this toothpaste found an average of 2.95 cavities per child. Assume that the standard deviation of the number of cavities of a child using this new toothpaste remains equal to 1.

(a) Are these data strong enough, at the 5 percent level of significance, to establish the claim of the toothpaste advertisement?
(b) Do the data convince you to switch to this new toothpaste?

13. There is some variability in the amount of phenobarbital in each capsule sold by a manufacturer. However, the manufacturer claims that the mean value is 20.0 mg. To test this, a sample of 25 pills yielded a sample mean of 19.7 with a sample standard deviation of 1.3. What inference would you draw from these data? In particular, are the data strong enough evidence to discredit the claim of the manufacturer? Use the 5 percent level of significance.

14. Twenty years ago, entering male high school students of Central High could do an average of 24 pushups in 60 seconds. To see whether this remains true today, a random sample of 36 freshmen was chosen. If their average was 22.5 with a sample standard deviation of 3.1, can we conclude that the mean is no longer equal to 24? Use the 5 percent level of significance.

15. The mean response time of a species of pigs to a stimulus is .8 seconds. Twenty-eight pigs were given 2 oz of alcohol and then tested. If their average response time
was 1.0 seconds with a standard deviation of .3 seconds, can we conclude that alcohol affects the mean response time? Use the 5 percent level of significance.

16. Suppose that team A and team B are to play a National Football League game and that team A is favored by \( f \) points. Let \( S(A) \) and \( S(B) \) denote the scores of teams A and B, and let \( X = S(A) - S(B) - f \). That is, \( X \) is the amount by which team A beats the point spread. It has been claimed that the distribution of \( X \) is normal with mean 0 and standard deviation 14. Use data from randomly chosen football games to test this hypothesis.

17. A medical scientist believes that the average basal temperature of (outwardly) healthy individuals has increased over time and is now greater than 98.6 degrees Fahrenheit (37 degrees Celsius). To prove this, she has randomly selected 100 healthy individuals. If their mean temperature is 98.74 with a sample standard deviation of 1.1 degrees, does this prove her claim at the 5 percent level? What about at the 1 percent level?

18. Use the results of a Sunday's worth of NFL professional football games to test the hypothesis that the average number of points scored by winning teams is less than or equal to 28. Use the 5 percent level of significance.

19. Use the results of a Sunday's worth of major league baseball scores to test the hypothesis that the average number of runs scored by winning teams is at least 5.6. Use the 5 percent level of significance.

20. A car is advertised as having a gas mileage rating of at least 30 miles/gallon in highway driving. If the miles per gallon obtained in 10 independent experiments are 26, 24, 20, 25, 27, 25, 28, 30, 26, 33, should you believe the advertisement? What assumptions are you making?

21. A producer specifies that the mean lifetime of a certain type of battery is at least 240 hours. A sample of 18 such batteries yielded the following data.

\[
\begin{align*}
237 & \quad 242 & \quad 232 \\
242 & \quad 248 & \quad 230 \\
244 & \quad 243 & \quad 254 \\
262 & \quad 234 & \quad 220 \\
225 & \quad 236 & \quad 232 \\
218 & \quad 228 & \quad 240
\end{align*}
\]

Assuming that the life of the batteries is approximately normally distributed, do the data indicate that the specifications are not being met?

22. Use the data of Example 2.3i of Chapter 2 to test the null hypothesis that the average noise level directly outside of Grand Central Station is less than or equal to 80 decibels.
23. An oil company claims that the sulfur content of its diesel fuel is at most .15 percent. To check this claim, the sulfur contents of 40 randomly chosen samples were determined; the resulting sample mean and sample standard deviation were .162 and .040. Using the 5 percent level of significance, can we conclude that the company’s claims are invalid?

24. A company supplies plastic sheets for industrial use. A new type of plastic has been produced and the company would like to claim that the average stress resistance of this new product is at least 30.0, where stress resistance is measured in pounds per square inch (psi) necessary to crack the sheet. The following random sample was drawn off the production line. Based on this sample, would the claim clearly be unjustified?

30.1 32.7 22.5 27.5
27.7 29.8 28.9 31.4
31.2 24.3 26.4 22.8
29.1 33.4 32.5 21.7

Assume normality and use the 5 percent level of significance.

25. It is claimed that a certain type of bipolar transistor has a mean value of current gain that is at least 210. A sample of these transistors is tested. If the sample mean value of current gain is 200 with a sample standard deviation of 35, would the claim be rejected at the 5 percent level of significance if

(a) the sample size is 25;
(b) the sample size is 64?

26. A manufacturer of capacitors claims that the breakdown voltage of these capacitors has a mean value of at least 100 V. A test of 12 of these capacitors yielded the following breakdown voltages:

96, 98, 105, 92, 111, 114, 99, 103, 95, 101, 106, 97

Do these results prove the manufacturer’s claim? Do they disprove them?

27. A sample of 10 fish were caught at lake A and their PCB concentrations were measured using a certain technique. The resulting data in parts per million were

Lake A: 11.5, 10.8, 11.6, 9.4, 12.4, 11.4, 12.2, 11, 10.6, 10.8

In addition, a sample of 8 fish were caught at lake B and their levels of PCB were measured by a different technique than that used at lake A. The resultant data were

Lake B: 11.8, 12.6, 12.2, 12.5, 11.7, 12.1, 10.4, 12.6
If it is known that the measuring technique used at lake A has a variance of .09 whereas the one used at lake B has a variance of .16, could you reject (at the 5 percent level of significance) a claim that the two lakes are equally contaminated?

28. A method for measuring the pH level of a solution yields a measurement value that is normally distributed with a mean equal to the actual pH of the solution and with a standard deviation equal to .05. An environmental pollution scientist claims that two different solutions come from the same source. If this were so, then the pH level of the solutions would be equal. To test the plausibility of this claim, 10 independent measurements were made of the pH level for both solutions, with the following data resulting.

<table>
<thead>
<tr>
<th>Measurements of Solution A</th>
<th>Measurements of Solution B</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.24</td>
<td>6.27</td>
</tr>
<tr>
<td>6.31</td>
<td>6.25</td>
</tr>
<tr>
<td>6.28</td>
<td>6.33</td>
</tr>
<tr>
<td>6.30</td>
<td>6.27</td>
</tr>
<tr>
<td>6.25</td>
<td>6.24</td>
</tr>
<tr>
<td>6.26</td>
<td>6.31</td>
</tr>
<tr>
<td>6.24</td>
<td>6.28</td>
</tr>
<tr>
<td>6.29</td>
<td>6.29</td>
</tr>
<tr>
<td>6.22</td>
<td>6.34</td>
</tr>
<tr>
<td>6.28</td>
<td>6.27</td>
</tr>
</tbody>
</table>

(a) Do the data disprove the scientist’s claim? Use the 5 percent level of significance.
(b) What is the p-value?

29. The following are the values of independent samples from two different populations.

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>122, 114, 130, 165, 144, 133, 139, 142, 150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 2</td>
<td>108, 125, 122, 140, 132, 120, 137, 128, 138</td>
</tr>
</tbody>
</table>

Let $\mu_1$ and $\mu_2$ be the respective means of the two populations. Find the $p$-value of the test of the null hypothesis

$$H_0 : \mu_1 \leq \mu_2$$

versus the alternative

$$H_1 : \mu_1 > \mu_2$$
when the population standard deviations are \( \sigma_1 = 10 \) and
(a) \( \sigma_2 = 5 \); (b) \( \sigma_2 = 10 \); (c) \( \sigma_2 = 20 \).

30. The data below give the lifetimes in hundreds of hours of samples of two types of
electronic tubes. Past lifetime data of such tubes have shown that they can often be
modeled as arising from a lognormal distribution. That is, the logarithms of the
data are normally distributed. Assuming that variance of the logarithms is equal
for the two populations, test, at the 5 percent level of significance, the hypothesis
that the two population distributions are identical.

<table>
<thead>
<tr>
<th></th>
<th>Type 1</th>
<th>Type 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>32, 84, 37, 42, 78, 62, 59, 74</td>
<td>39, 111, 55, 106, 90, 87, 85</td>
</tr>
</tbody>
</table>

31. The viscosity of two different brands of car oil is measured and the following data resulted:

<table>
<thead>
<tr>
<th></th>
<th>Brand 1</th>
<th>Brand 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10.62, 10.58, 10.33, 10.72, 10.44, 10.74</td>
<td>10.50, 10.52, 10.58, 10.62, 10.55, 10.51, 10.53</td>
</tr>
</tbody>
</table>

Test the hypothesis that the mean viscosity of the two brands is equal, assuming
that the populations have normal distributions with equal variances.

32. It is argued that the resistance of wire A is greater than the resistance of wire B.
You make tests on each wire with the following results.

<table>
<thead>
<tr>
<th>Wire A</th>
<th>Wire B</th>
</tr>
</thead>
<tbody>
<tr>
<td>.140 ohm</td>
<td>.135 ohm</td>
</tr>
<tr>
<td>.138</td>
<td>.140</td>
</tr>
<tr>
<td>.143</td>
<td>.136</td>
</tr>
<tr>
<td>.142</td>
<td>.142</td>
</tr>
<tr>
<td>.144</td>
<td>.138</td>
</tr>
<tr>
<td>.137</td>
<td>.140</td>
</tr>
</tbody>
</table>

What conclusion can you draw at the 10 percent significance level? Explain what
assumptions you are making.

In Problems 33 through 40, assume that the population distributions are normal
and have equal variances.

33. Twenty-five men between the ages of 25 and 30, who were participating in a well-
known heart study carried out in Framingham, Massachusetts, were randomly
selected. Of these, 11 were smokers and 14 were not. The following data refer to
readings of their systolic blood pressure.
Use these data to test the hypothesis that the mean blood pressures of smokers and non-smokers are the same.

34. In a 1943 experiment (Whitlock and Bliss, “A Bioassay Technique for Anti-helminthics,” *Journal of Parasitology*, 29, pp. 48–58) 10 albino rats were used to study the effectiveness of carbon tetrachloride as a treatment for worms. Each rat received an injection of worm larvae. After 8 days, the rats were randomly divided into two groups of 5 each; each rat in the first group received a dose of 0.032 cc of carbon tetrachloride, whereas the dosage for each rat in the second group was 0.063 cc. Two days later the rats were killed, and the number of adult worms in each rat was determined. The numbers detected in the group receiving the 0.032 dosage were

\[ 421, 462, 400, 378, 413 \]

whereas they were

\[ 207, 17, 412, 74, 116 \]

for those receiving the 0.063 dosage. Do the data prove that the larger dosage is more effective than the smaller?

35. A professor claims that the average starting salary of industrial engineering graduating seniors is greater than that of civil engineering graduates. To study this claim, samples of 16 industrial engineers and 16 civil engineers, all of whom graduated in 1993, were chosen and sample members were queried about their starting salaries. If the industrial engineers had a sample mean salary of $47,700 and a sample standard deviation of $2,400, and the civil engineers had a sample mean
salary of $46,400 and a sample standard deviation of $2,200, has the professor’s claim been verified? Find the appropriate p-value.

36. In a certain experimental laboratory, a method A for producing gasoline from crude oil is being investigated. Before completing experimentation, a new method B is proposed. All other things being equal, it was decided to abandon A in favor of B only if the average yield of the latter was clearly greater. The yield of both processes is assumed to be normally distributed. However, there has been insufficient time to ascertain their true standard deviations, although there appears to be no reason why they cannot be assumed equal. Cost considerations impose size limits on the size of samples that can be obtained. If a 1 percent significance level is all that is allowed, what would be your recommendation based on the following random samples? The numbers represent percent yield of crude oil.

<table>
<thead>
<tr>
<th>A</th>
<th>23.2, 26.6, 24.4, 23.5, 22.6, 25.7, 25.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>25.7, 27.7, 26.2, 27.9, 25.0, 21.4, 26.1</td>
</tr>
</tbody>
</table>

37. A study was instituted to learn how the diets of women changed during the winter and the summer. A random group of 12 women were observed during the month of July and the percentage of each woman’s calories that came from fat was determined. Similar observations were made on a different randomly selected group of size 12 during the month of January. The results were as follows:

<table>
<thead>
<tr>
<th>July</th>
<th>32.2, 27.4, 28.6, 32.4, 40.5, 26.2, 29.4, 25.8, 36.6, 30.3, 28.5, 32.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>30.5, 28.4, 40.2, 37.6, 36.5, 38.8, 34.7, 29.5, 29.7, 37.2, 41.5, 37.0</td>
</tr>
</tbody>
</table>

Test the hypothesis that the mean fat percentage intake is the same for both months. Use the (a) 5 percent level of significance and (b) 1 percent level of significance.

38. To learn about the feeding habits of bats, 22 bats were tagged and tracked by radio. Of these 22 bats, 12 were female and 10 were male. The distances flown (in meters) between feedings were noted for each of the 22 bats, and the following summary statistics were obtained.

<table>
<thead>
<tr>
<th></th>
<th>Female Bats</th>
<th>Male Bats</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>$\bar{X}$</td>
<td>180</td>
<td>$\bar{Y}$ = 136</td>
</tr>
<tr>
<td>$S_x$</td>
<td>92</td>
<td>$S_y$ = 86</td>
</tr>
</tbody>
</table>

Test the hypothesis that the mean distance flown between feedings is the same for the populations of both male and of female bats. Use the 5 percent level of significance.
39. The following data summary was obtained from a comparison of the lead content of human hair removed from adult individuals that had died between 1880 and 1920 with the lead content of present-day adults. The data are in units of micrograms, equal to one-millionth of a gram.

<table>
<thead>
<tr>
<th></th>
<th>1880–1920</th>
<th>Today</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>30</td>
<td>100</td>
</tr>
<tr>
<td>Sample mean</td>
<td>48.5</td>
<td>26.6</td>
</tr>
<tr>
<td>Sample standard deviation</td>
<td>14.5</td>
<td>12.3</td>
</tr>
</tbody>
</table>

(a) Do the above data establish, at the 1 percent level of significance, that the mean lead content of human hair is less today than it was in the years between 1880 and 1920? Clearly state what the null and alternative hypotheses are.

(b) What is the \( p \)-value for the hypothesis test in part (a)?

40. Sample weights (in pounds) of newborn babies born in two adjacent counties in Western Pennsylvania yielded the following data.

\[
\begin{aligned}
&n = 53, \quad m = 44 \\
&\bar{X} = 6.8, \quad \bar{Y} = 7.2 \\
&S^2 = 5.2, \quad S^2 = 4.9
\end{aligned}
\]

Consider a test of the hypothesis that the mean weight of newborns is the same in both counties. What is the resulting \( p \)-value?

41. To verify the hypothesis that blood lead levels tend to be higher for children whose parents work in a factory that uses lead in the manufacturing process, researchers examined lead levels in the blood of 33 children whose parents worked in a battery manufacturing factory. (Morton, D., Saah, A., Silberg, S., Owens, W., Roberts, M., and Saah, M., “Lead Absorption in Children of Employees in a Lead-Related Industry,” *American Journal of Epidemiology*, **115**, 549–555, 1982.) Each of these children were then matched by another child who was of similar age, lived in a similar neighborhood, had a similar exposure to traffic, but whose parent did not work with lead. The blood levels of the 33 cases (sample 1) as well as those of the 33 controls (sample 2) were then used to test the hypothesis that the average blood levels of these groups are the same. If the resulting sample means and sample standard deviations were

\[
\begin{aligned}
&\bar{x}_1 = .015, \quad s_1 = .004, \quad \bar{x}_2 = .006, \quad s_2 = .006
\end{aligned}
\]

find the resulting \( p \)-value. Assume a common variance.
42. Ten pregnant women were given an injection of pitocin to induce labor. Their systolic blood pressures immediately before and after the injection were:

<table>
<thead>
<tr>
<th>Patient</th>
<th>Before</th>
<th>After</th>
<th>Patient</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>134</td>
<td>140</td>
<td>6</td>
<td>140</td>
<td>138</td>
</tr>
<tr>
<td>2</td>
<td>122</td>
<td>130</td>
<td>7</td>
<td>118</td>
<td>124</td>
</tr>
<tr>
<td>3</td>
<td>132</td>
<td>135</td>
<td>8</td>
<td>127</td>
<td>126</td>
</tr>
<tr>
<td>4</td>
<td>130</td>
<td>126</td>
<td>9</td>
<td>125</td>
<td>132</td>
</tr>
<tr>
<td>5</td>
<td>128</td>
<td>134</td>
<td>10</td>
<td>142</td>
<td>144</td>
</tr>
</tbody>
</table>

Do the data indicate that injection of this drug changes blood pressure?

43. A question of medical importance is whether jogging leads to a reduction in one's pulse rate. To test this hypothesis, 8 nonjogging volunteers agreed to begin a 1-month jogging program. After the month their pulse rates were determined and compared with their earlier values. If the data are as follows, can we conclude that jogging has had an effect on the pulse rates?

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulse Rate Before</td>
<td>74</td>
<td>86</td>
<td>98</td>
<td>102</td>
<td>78</td>
<td>84</td>
<td>79</td>
<td>70</td>
</tr>
<tr>
<td>Pulse Rate After</td>
<td>70</td>
<td>85</td>
<td>90</td>
<td>110</td>
<td>71</td>
<td>80</td>
<td>69</td>
<td>74</td>
</tr>
</tbody>
</table>

44. If $X_1, \ldots, X_n$ is a sample from a normal population having unknown parameters $\mu$ and $\sigma^2$, devise a significance level $\alpha$ test of

$$H_0 = \sigma^2 \leq \sigma_0^2$$

versus the alternative

$$H_1 = \sigma^2 > \sigma_0^2$$

for a given positive value $\sigma_0^2$.

45. In Problem 44, explain how the test would be modified if the population mean $\mu$ were known in advance.

46. A gun-like apparatus has recently been designed to replace needles in administering vaccines. The apparatus can be set to inject different amounts of the serum, but because of random fluctuations the actual amount injected is normally distributed with a mean equal to the setting and with an unknown variance $\sigma^2$. It has been decided that the apparatus would be too dangerous to use if $\sigma$ exceeds .10. If a random sample of 50 injections resulted in a sample standard deviation of .08, should use of the new apparatus be discontinued? Suppose the level of significance is $\alpha = .10$. Comment on the appropriate choice of a significance level for this problem, as well as the appropriate choice of the null hypothesis.
47. A pharmaceutical house produces a certain drug item whose weight has a standard deviation of .5 milligrams. The company's research team has proposed a new method of producing the drug. However, this entails some costs and will be adopted only if there is strong evidence that the standard deviation of the weight of the items will drop to below .4 milligrams. If a sample of 10 items is produced and has the following weights, should the new method be adopted?

| Weight (mg) | 5.728 | 5.731 |
|           | 5.722 | 5.719 |
|           | 5.727 | 5.724 |
|           | 5.718 | 5.726 |
|           | 5.723 | 5.722 |

48. The production of large electrical transformers and capacitors requires the use of polychlorinated biphenyls (PCBs), which are extremely hazardous when released into the environment. Two methods have been suggested to monitor the levels of PCB in fish near a large plant. It is believed that each method will result in a normal random variable that depends on the method. Test the hypothesis at the $\alpha = .10$ level of significance that both methods have the same variance, if a given fish is checked 8 times by each method with the following data (in parts per million) recorded.

| Method 1 | 6.2, 5.8, 5.7, 6.3, 5.9, 6.1, 6.2, 5.7 |
|         | Method 2 | 6.3, 5.7, 5.9, 6.4, 5.8, 6.2, 6.3, 5.5 |

49. In Problem 31, test the hypothesis that the populations have the same variances.

50. If $X_1, \ldots, X_n$ is a sample from a normal population with variance $\sigma_x^2$, and $Y_1, \ldots, Y_n$ is an independent sample from normal population with variance $\sigma_y^2$, develop a significance level $\alpha$ test of

$$H_0 : \sigma_x^2 < \sigma_y^2 \quad \text{versus} \quad H_1 : \sigma_x^2 > \sigma_y^2$$

51. The amount of surface wax on each side of waxed paper bags is believed to be normally distributed. However, there is reason to believe that there is greater variation in the amount on the inner side of the paper than on the outside. A sample of 75 observations of the amount of wax on each side of these bags is obtained and the following data recorded.

<table>
<thead>
<tr>
<th>Wax in Pounds per Unit Area of Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outside Surface</td>
</tr>
<tr>
<td>$\bar{x} = .948$</td>
</tr>
<tr>
<td>$\sum x_i^2 = 91$</td>
</tr>
</tbody>
</table>
Conduct a test to determine whether or not the variability of the amount of wax on the inner surface is greater than the variability of the amount on the outer surface ($\alpha = .05$).

52. In a famous experiment to determine the efficacy of aspirin in preventing heart attacks, 22,000 healthy middle-aged men were randomly divided into two equal groups, one of which was given a daily dose of aspirin and the other a placebo that looked and tasted identical to the aspirin. The experiment was halted at a time when 104 men in the aspirin group and 189 in the control group had had heart attacks. Use these data to test the hypothesis that the taking of aspirin does not change the probability of having a heart attack.

53. In the study of Problem 52, it also resulted that 119 from the aspirin group and 98 from the control group suffered strokes. Are these numbers significant to show that taking aspirin changes the probability of having a stroke?

54. A standard drug is known to be effective in 72 percent of the cases in which it is used to treat a certain infection. A new drug has been developed and testing has found it to be effective in 42 cases out of 50. Is this strong enough evidence to prove that the new drug is more effective than the old one? Find the relevant $p$-value.

55. Three independent news services are running a poll to determine if over half the population supports an initiative concerning limitations on driving automobiles in the downtown area. Each wants to see if the evidence indicates that over half the population is in favor. As a result, all three services will be testing

$$H_0 : p \leq .5 \quad \text{versus} \quad H_1 : p > .5$$

where $p$ is the proportion of the population in favor of the initiative.

(a) Suppose the first news organization samples 100 people, of which 56 are in favor of the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis and so establish that over half the population favors the initiative?

(b) Suppose the second news organization samples 120 people, of which 68 are in favor of the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis?

(c) Suppose the third news organization samples 110 people, of which 62 are in favor of the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis?

(d) Suppose the news organizations combine their samples, to come up with a sample of 330 people, of which 186 support the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis?
56. According to the U.S. Bureau of the Census, 25.5 percent of the population of those age 18 or over smoked in 1990. A scientist has recently claimed that this percentage has since increased, and to prove her claim she randomly sampled 500 individuals from this population. If 138 of them were smokers, is her claim proved? Use the 5 percent level of significance.

57. An ambulance service claims that at least 45 percent of its calls involve life-threatening emergencies. To check this claim, a random sample of 200 calls was selected from the service’s files. If 70 of these calls involved life-threatening emergencies, is the service’s claim believable at the

(a) 5 percent level of significance;
(b) 1 percent level of significance?

58. A standard drug is known to be effective in 75 percent of the cases in which it is used to treat a certain infection. A new drug has been developed and has been found to be effective in 42 cases out of 50. Based on this, would you accept, at the 5 percent level of significance, the hypothesis that the two drugs are of equal effectiveness? What is the \( p \)-value?

59. Do Problem 58 by using a test based on the normal approximation to the binomial.

60. In a recently conducted poll, 54 out of 200 people surveyed claimed to have a firearm in their homes. In a similar survey done earlier, 30 out of 150 people made that claim. Is it possible that the proportion of the population having firearms has not changed and the foregoing is due to the inherent randomness in sampling?

61. Let \( X_1 \) denote a binomial random variable with parameters \( (n_1, p_1) \) and \( X_2 \) an independent binomial random variable with parameters \( (n_2, p_2) \). Develop a test, using the same approach as in the Fisher-Irwin test, of

\[
H_0 : p_1 \leq p_2
\]

versus the alternative

\[
H_1 : p_1 > p_2
\]

62. Verify that Equation 8.6.5 follows from Equation 8.6.4.

63. Let \( X_1 \) and \( X_2 \) be binomial random variables with respective parameters \( n_1, p_1 \) and \( n_2, p_2 \). Show that when \( n_1 \) and \( n_2 \) are large, an approximate level \( \alpha \) test of \( H_0 : p_1 = p_2 \) versus \( H_1 : p_1 \neq p_2 \) is as follows:

\[
\text{reject } H_0 \text{ if } \frac{|X_1/n_1 - X_2/n_2|}{\sqrt{\frac{X_1 + X_2}{n_1 + n_2} \left( 1 - \frac{X_1 + X_2}{n_1 + n_2} \right) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} > z_{\alpha/2}
\]
Hint: (a) Argue first that when \( n_1 \) and \( n_2 \) are large

\[
\frac{X_1 - X_2}{\sqrt{\frac{n_1}{p_1(1-p_1)} + \frac{n_2}{p_2(1-p_2)}}} \sim N(0, 1)
\]

where \( \sim \) means “approximately has the distribution.”

(b) Now argue that when \( H_0 \) is true and so \( p_1 = p_2 \), their common value can be best estimated by \( (X_1 + X_2) / (n_1 + n_2) \).

64. Use the approximate test given in Problem 63 on the data of Problem 60.

65. Patients suffering from cancer must often decide whether to have their tumors treated with surgery or with radiation. A factor in their decision is the 5-year survival rates for these treatments. Surprisingly, it has been found that patient’s decisions often seem to be affected by whether they are told the 5-year survival rates or the 5-year death rates (even though the information content is identical). For instance, in an experiment a group of 200 male prostate cancer patients were randomly divided into two groups of size 100 each. Each member of the first group was told that the 5-year survival rate for those electing surgery was 77 percent, whereas each member of the second group was told that the 5-year death rate for those electing surgery was 23 percent. Both groups were given the same information about radiation therapy. If it resulted that 24 members of the first group and 12 of the second group elected to have surgery, what conclusions would you draw?

66. The following data refer to Larry Bird’s results when shooting a pair of free throws in basketball. During two consecutive seasons in the National Basketball Association, Bird shot a pair of free throws on 338 occasions. On 251 occasions he made both shots; on 34 occasions he made the first shot but missed the second one; on 48 occasions he missed the first shot but made the second one; on 5 occasions he missed both shots.

(a) Use these data to test the hypothesis that Bird’s probability of making the first shot is equal to his probability of making the second shot.

(b) Use these data to test the hypothesis that Bird’s probability of making the second shot is the same regardless of whether he made or missed the first one.

67. In the nineteen seventies, the U.S. Veterans Administration (Murphy, 1977) conducted an experiment comparing coronary artery bypass surgery with medical drug therapy as treatments for coronary artery disease. The experiment involved 596 patients, of whom 286 were randomly assigned to receive surgery, with the remaining 310 assigned to drug therapy. A total of 252 of those receiving surgery, and a total of 270 of those receiving drug therapy were still alive three years after
treatment. Use these data to test the hypothesis that the survival probabilities are equal.

68. Test the hypothesis, at the .05 level of significance, that the yearly number of earthquakes felt on a certain island has mean 52 if the readings for the last 8 years are 46, 62, 60, 58, 47, 50, 59, 49. Assume an underlying Poisson distribution and give an explanation to justify this assumption.

69. The following table gives the number of fatal accidents of U.S. commercial airline carriers in the 16 years from 1980 to 1995. Do these data disprove, at the 5 percent level of significance, the hypothesis that the mean number of accidents in a year is greater than or equal to 4.5? What is the p-value? (Hint: First formulate a model for the number of accidents.)

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Departures (millions)</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>1980 ....</td>
</tr>
<tr>
<td>1981 ....</td>
</tr>
<tr>
<td>1982 ....</td>
</tr>
<tr>
<td>1983 ....</td>
</tr>
<tr>
<td>1984 ....</td>
</tr>
<tr>
<td>1985 ....</td>
</tr>
<tr>
<td>1986 ....</td>
</tr>
<tr>
<td>1987 ....</td>
</tr>
</tbody>
</table>

Source: National Transportation Safety Board

70. For the following data, sample 1 is from a Poisson distribution with mean $\lambda_1$ and sample 2 is from a Poisson distribution with mean $\lambda_2$. Test the hypothesis that $\lambda_1 = \lambda_2$.

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>24, 32, 29, 33, 40, 28, 34, 36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 2</td>
<td>42, 36, 41</td>
</tr>
</tbody>
</table>

71. A scientist looking into the effect of smoking on heart disease has chosen a large random sample of smokers and of nonsmokers. She plans to study these two groups for 5 years to see if the number of heart attacks among the members of the smokers' group is significantly greater than the number among the nonsmokers. Such a result, the scientist feels, should be strong evidence of an association between
smoking and heart attacks. Given that

1. Older people are at greater risk of heart disease than are younger people; and
2. As a group, smokers tend to be somewhat older than nonsmokers;

would the scientist be justified in her conclusion? Explain how the experimental design can be improved so that meaningful conclusions can be drawn.

72. A researcher wants to analyze the average yearly increase in a stock over a 20 year period. To do so, she plans to randomly choose 100 stocks from the listing of current stocks, discarding any that were not in existence 20 years ago. She will then compare the current price of each stock with its price 20 years ago to determine its percentage increase. Do you think this is a valid method to study the average increase in the price of a stock?