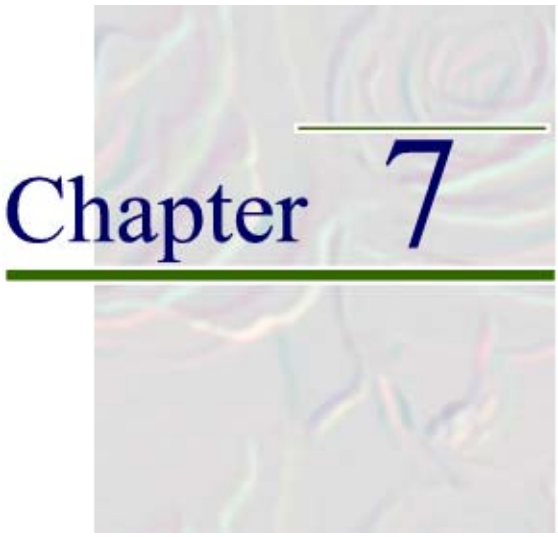


HYPOTHESIS TESTING: ONE SAMPLE INFERENCE

REVIEW OF KEY CONCEPTS

SECTION 7.1 Fundamentals of Hypothesis Testing

Let us focus on the birthweight example mentioned in section 6.10.2. Suppose instead of dichotomizing the outcome, we compute differences in birthweight $x_{i1} - x_{i2}$, i.e., (drug – placebo) $= x_i$ for all 50 pairs, not just discordant pairs. Assume $X_i \sim N(\mu, \sigma^2)$. If $\bar{x} = 8$ oz and $s = 30$ oz, then is $\bar{x}$ far enough from zero for us to be convinced that the drug is effective in increasing mean birthweight? We set up two hypotheses:

1. H_0 = null hypothesis, $\mu = \mu_0 = 0$.
2. H_1 = alternative hypothesis, $\mu = \mu_1 > \mu_0$.

Under H_0, the mean birthweight for the drug and placebo babies are the same (i.e., the drug is ineffective). Under H_1, the mean birthweight for the drug babies is greater than for placebo babies (i.e., the drug is effective in increasing mean birthweight). How can we use the data to decide which of these two hypotheses is likely to be correct?

What types of errors can we make? If we decide that H_1 is true and H_0 is actually true, then we call this a *type I error*. If we decide that H_0 is true and H_1 is actually true, then we call this a *type II error*. The probabilities of these two types of errors are called α and β, respectively. Also, the *power of a test* is defined as $1 - \beta$ = the probability of deciding that H_1 is true when H_1 is actually true.

We usually set $\alpha = .05$ by convention. Another name for the decision that H_1 is true is that we reject H_0; another name for the decision that H_0 is true is that we accept H_0. We will use hypothesis tests that for a given α error will minimize the type II error (or equivalently maximize the power).

SECTION 7.2 One-Sample *t* Test

For the birthweight example, we will base our test on $\bar{x}$ and will reject H_0 for large values of $\bar{x}$ (that is, if $\bar{x} > c$). How large should *c* be? This depends on the type I error.

We refer to this method of hypothesis testing as the *critical-value method* and refer to the set of values of $\bar{x}$ for which we reject H_0 as the *critical region*.

Another way of implementing the critical-value method is by using a *test statistic*. In general, we reject H_0 if the test statistic is in the critical region and accept H_0 otherwise. For the birthweight example, the test statistic is

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

and we reject H_0 if $t > t_{n-1,\,1-\alpha}$ and accept H_0 if $t \le t_{n-1,\,1-\alpha}$. The value $t_{n-1,\,1-\alpha}$ is referred to as the critical value. The name of this test is the *one-sample t test*. In our case, $\mu_0 = 0$, $s = 30$, $n = 50$, $t_{n-1,\,1-\alpha} = t_{n-1,\,0.95} = t_{49,\,0.95} = 1.677$ (using the TINV function of Excel). Therefore, we reject H_0 if

$$t = \frac{\bar{x} - 0}{\frac{30}{\sqrt{50}}} > 1.677$$

and accept H_0 otherwise. Since $\bar{x} = 8$ oz and $s = 30$ oz, it follows that

$$t = \frac{8 - 0}{\frac{30}{\sqrt{50}}} = \frac{8}{4.243} = 1.886 > 1.677.$$

Therefore, we reject H_0.

What if we conducted our test at the 1% significance level? Since $t_{49,\,.99} = 2.405 =$ critical value and $t = 1.886 < 2.405$, this implies that we would accept H_0 in this case. Instead of conducting significance tests at many different α levels, we can use a different method of hypothesis testing called the *p-value method*. A ***p*-value** is defined as the α level at which we would be indifferent between accepting and rejecting H_0 based on the observed data. In this case, p-value $= Pr(t_{n-1} > t)$.

If we use an $\alpha \le p$, then we accept H_0.
If we use an $\alpha > p$, then we reject H_0.

In our case,

$Pr(t_{49} > 1.886) = .033$ (based on the TDIST function of Excel using TDIST(1.886, 49, 1)) $= p$-value

Refer to the Appendix for a discussion of the TINV and TDIST functions of Excel.

SECTION 7.3 Guidelines for Assessing Statistical Significance

How can we interpret the p-value? If $p < .05$, then the results are statistically significant. Different levels of significance are sometimes noted. Specifically, if

$.01 \le p < .05$ then the results are considered significant
$.001 \le p \le .01$ then the results are considered highly significant
$p < .001$ then the results are considered very highly significant

If $p \ge .05$, then the results are not statistically significant however, if $.05 \le p \le .10$ then a trend toward statistical significance is sometimes noted.

Statistical Significance Versus Clinical Significance

Suppose $\bar{x} = 1$ oz, $n = 10{,}000$; then

$$p\text{-value} = Pr\left(t_{9999} > \frac{1}{\frac{30}{\sqrt{10{,}000}}}\right) \cong Pr[N(0,\ 1) > 3.33] = 1 - \Phi(3.33) = 1 - .99957 \approx .0004 < .001$$

The results are statistically significant, but not very important.

Conversely, suppose $\bar{x} = 4$ oz., $n = 50$; then p-value $= Pr\left[t_{49} > 4 \Big/ \left(30 \Big/ \sqrt{50}\right)\right] = Pr\left(t_{49} > .943\right) = .175$.

The results are not statistically significant but could possibly be important if confirmed by a larger study.

SECTION 7.4 Two-Sided Alternatives

A randomized study was conducted among patients with unstable angina (i.e., angina severe enough to require hospitalization). The study was based on a comparison of two drugs: Nifedipine and Propranolol. The patients were not matched.

Primary outcome measure—relief of pain within 48 hours
Secondary outcome measure(s)—change in heart rate and blood pressure from baseline to 48 hours.

For the secondary outcome measures, interest is in changes in these parameters within each drug group separately and also in a comparison of the change between the two drug groups. We now look at changes in heart rate within each drug group separately. We will compare changes between the two drug groups in Chapter 8. The results are as follows:

Nifedipine group: mean change in heart rate (heart rate 48 hours – heart-rate baseline) $= \bar{x} = +3.0$ beats per minute. $sd = 10.0$, $n = 61$
Propranolol group: mean change in heart rate $= \bar{x} = -2.2$ beats per minute, $sd = 10.0$, $n = 61$

Suppose we are uncertain whether either of these drugs raises, lowers, or has no effect on heart rate. Thus, we wish to test $H_0: \mu = 0$ is $H_1: \mu \neq 0$ and assume $X \sim N(\mu,\ \sigma^2)$. In this case, we will reject H_0 if $\bar{x}$ or equivalently, the test statistic $t = (\bar{x} - \mu_0) \Big/ \left(s \Big/ \sqrt{n}\right)$ is either too small or too large. This type of alternative is called a *two-sided alternative*, because we are interested in deviations from the null hypothesis where either $\mu < \mu_0 = 0$ or $\mu > \mu_0 = 0$. In contrast, the alternative hypothesis for the birthweight example was a one-sided alternative (i.e., $\mu > \mu_0$).

The specific test criteria for a level of significance $= \alpha$ are

Reject H_0 if $t < t_{n-1,\ \alpha/2} = -t_{n-1,\ 1-\alpha/2}$ or $t > t_{n-1,\ 1-\alpha/2}$
Accept H_0 otherwise.

To compute the p-value, if $t > 0$, p-value $= 2 \times Pr(t_{n-1} > t)$. If $t < 0$, p-value $= 2 \times Pr(t_{n-1} < t)$.

The name of this procedure is the *one-sample t test*.

Nifedipine Group

Assess if mean heart rate has changed significantly over 48 hours among patients in the nifedipine group. Our hypotheses are $H_0: \mu = 0$ versus $H_1: \mu \neq 0$. The test statistic is

$$t = \frac{3.0}{\frac{10}{\sqrt{61}}} = \frac{3}{1.28} = 2.343 \sim t_{60} \text{ under } H_0.$$

Since $t_{60,\ .99} = 2.390$ and $t_{60,\ .975} = 2.000$ and $2.000 < 2.343 < 2.390$, it follows that $.01 < p/2 < .025$, or $.02 < p < .05$. Thus, there is a significant increase in mean heart rate for nifedipine users. The exact *p*-value $= 2 \times Pr(t_{60} > 2.343)$. Using Excel we can obtain the 2-sided *p*-value by computing

TDIST(2.343, 60, 2) = .022.

Propranolol Group

Test the same hypotheses as for the nifedipine group. The test statistic is

$$t = \frac{-2.2}{\frac{10}{\sqrt{61}}} = \frac{-2.2}{1.28} = -1.718 \sim t_{60}$$

Since $t_{60,\ .95} = 1.671$ and $t_{60,\ .975} = 2.000$ and $1.671 < 1.718 < 2.000$, it follows that $.025 < p/2 < .05$, or $.05 < p < .10$. Thus, there is no significant change in mean heart rate for propranolol users. The exact *p*-value $= 2 \times Pr(t_{60} < -1.718) = .091$. Using Excel we can obtain the 2-sided *p*-value (.091) by computing TDIST(1.718, 60, 2). Thus, there is no significant change in mean heart rate over 48 hours for the propranolol group.

Notice that it is possible for results to be significant using a one-sided test, while they would not be significant using a two-sided test. For example, if we tested the hypothesis $H_0: \mu = 0$ versus $H_1: \mu < 0$, the one-tailed *p*-value for the propranolol group would be $p = .045$, which is statistically significant.

Generally, most research results are reported using two-sided *p*-values because the direction of the alternative hypothesis is not known in advance. A strong case needs to be made for using a one-sided test and should be documented in a protocol before a study begins. All tests for the remainder of this study guide will be two-sided.

Finally, in some case, σ^2 is known, and we can replace the critical value $t_{n-1,\ 1-\alpha/2}$ by $z_{1-\alpha/2}$. The test statistic is given by

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

where we reject H_0 if $z > z_{1-\alpha/2}$ or $z < -z_{1-\alpha/2}$. Alternatively, the *p*-value is given by $2\Phi(z)$ if $z < 0$, or $2 \times [1 - \Phi(z)]$ if $z > 0$. This is known as the *one-sample z test*.

SECTION 7.5 The Power of a Test

Suppose we wish to test the hypothesis $H_0: \mu = \mu_0$ versus $H_1: \mu = \mu_1 > \mu_0$. We will assume that α is known.

The power of the test

$$= \text{Prob}(\text{reject null hypothesis given } H_1 \text{ true})$$
$$= \Phi\left(\frac{\sqrt{n}\,|\mu_1 - \mu_0|}{\sigma} - z_{1-\alpha}\right)$$

The power depends on

1. n; as $n \uparrow$, power $\uparrow$
2. $|\mu_1 - \mu_0|$; as $|\mu_1 - \mu_0| \uparrow$, power $\uparrow$
3. σ, as $\sigma \uparrow$, power $\downarrow$
4. α; as $\alpha \downarrow$, power $\downarrow$

It is important for a study to have adequate power to achieve its goals. Consider the birthweight example. Suppose $\mu_1 - \mu_0 = 4$ oz, $\alpha = .05$, $\sigma = 30$ oz, $n = 50$. What power does such a study have?

$$\text{Power} = \Phi\left(\frac{\sqrt{50}(4)}{30} - 1.645\right) = \Phi(-0.702) = .24$$

Therefore we will be able to detect a significant difference 24% of the time based on a sample of size 50 matched pairs if the underlying treatment effect = 4 oz. To compute power for a two-sided alternative, substitute $z_{1-\alpha/2}$ for $z_{1-\alpha}$ in the preceding formula.

SECTION 7.6 Sample Size

Another way of looking at the same question is to ask how many subjects do we need to achieve a given level of power. Specifically, if we wish to use the one-sided alternative given in Section 7.5 with a type I error of α and a power of $1-\beta$, then the required sample size is

$$n = \frac{\sigma^2\left(z_{1-\beta} + z_{1-\alpha}\right)^2}{(\mu_1 - \mu_0)^2}$$

Sample size is related to

1. α; as $\alpha \downarrow$, $n \uparrow$
2. β; as $\beta \downarrow$, $n \uparrow$
3. σ; as $\sigma \uparrow$, $n \uparrow$
4. $|\mu_1 - \mu_0|$; as $|\mu_1 - \mu_0| \uparrow$, $n \downarrow$

Suppose $\mu_1 - \mu_0 = 8$ oz, $\alpha = .05$, $\beta = .20$, $\sigma = 30$. How many subjects are needed? In this case, $z_{1-\beta} = z_{.8} = 0.84$, $z_{1-\alpha} = z_{.95} = 1.645$ and

$$n = \frac{900(0.84 + 1.645)^2}{64} = \frac{900(6.175)}{64} = 86.8$$

Thus, we need 87 matched pairs or 174 subjects to achieve 80% power.

For a two-sided test, we substitute $z_{1-\alpha/2}$ for $z_{1-\alpha}$ in the preceding sample-size formula.

SECTION 7.7 Relationship Between Hypothesis Testing and Confidence Intervals

Notice that from Chapter 6, another method for making the decision about whether $\mu = \mu_0$ (that is, whether the null hypothesis is true), would be to construct a 95% CI for μ and note if μ_0 falls in this interval.

From a *CI approach*, we decide that μ_0 is a likely value for μ if

$$\bar{x} - \frac{t_{n-1,\,1-\alpha/2}s}{\sqrt{n}} < \mu_0 < \bar{x} + \frac{t_{n-1,\,1-\alpha/2}s}{\sqrt{n}} \tag{1}$$

and decide it is an unlikely value, otherwise.

From a *hypothesis-testing approach*, we will accept H_0 if

$$t_{n-1,\,\alpha/2} < t < t_{n-1,\,1-\alpha/2} \tag{2}$$

and reject H_0 otherwise, where $t = (\bar{x} - \mu_0) \Big/ \left(s/\sqrt{n}\right)$.

Note that Equations (1) and (2) are equivalent algebraically. Thus, the two criteria are the same. We will accept (reject) H_0 at level α using a two-sided test if and only if the two-sided $100\% \times (1-\alpha)$ CI for μ does (does not) contain μ_0.

7.7.1 Relative Advantages of Hypothesis-Testing Versus Confidence-Interval Approaches

Hypothesis-testing approach

1. We obtain an exact *p*-value.
2. We can distinguish between $\mu = \mu_0$ and $\mu \neq \mu_0$.

Confidence-interval approach

1. Only known if $p < .05$ or $p > .05$.
2. We obtain an interval within which the parameter is likely to fall.

The two methods complement each other.

SECTION 7.8 One-Sample χ^2 Test

A new eye chart, the ETDRS (Early Treatment for Diabetic Retinopathy Scale) chart, has been used to measure visual acuity in several research studies. Different charts are used for the left and right eyes of an individual. A measure of acuity is the number of letters read correctly. One measure of variability is the number of letters read with the left eye minus the number of letters read with the right eye. The variability should increase with the new chart versus the standard eye chart (the Snellen chart), where the same chart is used for each eye. Let x_i = the number of letters read correctly with the left eye minus the number of letters read correctly with the right eye for the *i*th subject. We assume $x_i \sim N(0,\, \sigma^2)$. Suppose σ is known to be 1.5 based on previous studies with the Snellen chart. We collect data on 20 subjects using the new chart and find $s = 2.0$. Is there significantly more variability with the new chart?

Let $\sigma_0^2 =$ true variance for the Snellen chart (known), $\sigma^2 =$ true variance for the ETDRS chart (unknown). We wish to test the hypothesis $H_0: \sigma^2 = \sigma_0^2 = 2.25$ versus $H_1: \sigma^2 \neq \sigma_0^2$. We base our test on the test statistic

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}, \text{ under } H_0$$

Critical-Value Method

Reject H_0 if

$$X^2 < c_1 = \chi^2_{n-1,\ \alpha/2} \text{ or } X^2 > c_2 = \chi^2_{n-1,\ 1-\alpha/2}$$

Accept H_0 otherwise.

p-Value Method

$$\begin{aligned} p\text{-value} &= 2 \times Pr\left(\chi^2_{n-1} > X^2\right) \text{ if } s^2 > \sigma_0^2, \\ &= 2 \times Pr\left(\chi^2_{n-1} < X^2\right) \text{ if } s^2 < \sigma_0^2 \end{aligned}$$

In this case, $X^2 = 19(4.0)/2.25 = 33.78$. Since $c_2 = \chi^2_{19,\ .975} = 32.85$, $c_1 = \chi^2_{19,\ .025} = 8.91$ and $X^2 > c_2$, it follows that $p < .05$. The p-value $= 2 \times Pr(\chi^2_{19} > 33.78)$. Since $\chi^2_{19,\ .99} = 36.19$, $\chi^2_{19,\ .975} = 32.85$ and $32.85 < 33.78 < 36.19$, it follows that $.02 < p < .05$. Thus, there is significantly more variability between left and right eyes using the ETDRS chart compared with the Snellen chart. This may actually indicate that the ETDRS chart is more valid, because we are obtaining independent assessments of acuity for the left and right eyes.

SECTION 7.9 One-Sample Inference for the Binomial Distribution

The incidence of myocardial infarction (MI) in 50–59-year-old postmenopausal women is on the order of 2 events per 1000 women per year. Suppose in a large group-health practice, 1000 women are identified who have been taking postmenopausal hormones (PMH) for at least 1 year as of 1985, of whom 3 have an MI from 1985 to 1990. Is this an unusual event. We wish to test the hypothesis: $H_0: p = p_0$ versus $H_1: p \neq p_0$ where $p_0 =$ overall population incidence rate, and $p =$ true incidence rate for PMH users.

We use a normal-theory method if $np_0q_0 \geq 5$ and an exact method otherwise.

Normal-Theory Method

We use the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0,\ 1) \text{ under } H_0$$

where $\hat{p} =$ sample proportion of events (or more generally, successes). If $z \leq 0$, then the p-value $= 2\Phi(z)$; if $z > 0$, then the p-value $= 2[1 - \Phi(z)]$.

In this case, $n = 1000$, $\hat{p} = 3/1000 = .003$, $p_0 = 5(2/1000) = .01$, $q_0 = .99$, $np_0q_0 = 9.9 > 5$. Thus, we can use the normal-theory test. The test statistic is

$$z = \frac{.003 - .01}{\sqrt{.01(.99)/1000}} = \frac{-0.007}{0.0031} = -2.225$$

Since $z < 0$, the p-value $= 2\Phi(-2.225) = 2 \times [1 - \Phi(2.225)] = 2 \times [1 - .98695] = .026$.

Thus, there are significantly fewer MI's than expected among postmenopausal estrogen users. This could form the basis for a better study using a control group of postmenopausal women who never used estrogens.

Exact Method

If $np_0q_0 < 5$, then we have to use the exact method. The exact method can be used at any time, but is more difficult computationally than the normal-theory test. To obtain the p-value using the exact method,

$$\text{If } \hat{p} < p_0, \text{ then } p\text{-value} = 2 \times \sum_{k=0}^{x} \binom{n}{k} p_0^k q_0^{n-k}$$

$$\text{If } \hat{p} \geq p_0, \text{ then } p\text{-value} = 2 \times \sum_{k=x}^{n} \binom{n}{k} p_0^k q_0^{n-k}$$

where $x =$ observed number of events (successes).

We now apply the exact method to the MI data for comparison with the normal-theory method. In our case, $\hat{p} = .003 < p_0 = .01$. Thus,

$$p\text{-value} = 2 \times \sum_{k=0}^{3} \binom{1000}{k} (.01)^k (.99)^{1000-k}$$

Using the BINOMDIST function of Excel, we have:

$$Pr(X \leq 3 | n = 1000,\ p = .01) = .010073$$
$$2 \times Pr(X \leq 3) = .020 = p\text{-value}.$$

Thus, the results are also significant using exact methods. Note that we only need to use exact methods if $np_0q_0 < 5$.

SECTION 7.10 One-Sample Inference for the Poisson Distribution

The usual annual death rate due to asthma in England over the period 1862–1962 for people aged 5–34 was approximately 1 death per 100,000 people. Suppose that in 1963, 20 deaths due to asthma were observed among 1 million people in this age group living in Britain. Is this number of deaths inconsistent with the preceding 100-year rate?

Since n is large (1 million) and $p =$ the probability of death due to asthma is small, we can model the number of asthma deaths in 1963 by a Poisson distribution with parameter $= \mu$. Specifically, we wish to test the hypothesis $H_0: \mu = \mu_0$ versus $H_1: \mu \neq \mu_0$, where

μ = true expected number of deaths due to asthma in 1963

μ_0 = expected number of deaths due to asthma in 1963 if the asthma death rate from 1862-1962 holds in 1963

To obtain a *p*-value,

$$p = \min\left[2 \times \sum_{k=0}^{x} \frac{e^{-\mu_0}\mu_0^k}{k!},\ 1\right] \text{ if } x < \mu_0$$

$$p = \min\left[2 \times \left(1 - \sum_{k=0}^{x-1} \frac{e^{-\mu_0}\mu_0^k}{k!}\right),\ 1\right] \text{ if } x \geq \mu_0$$

where $x =$ observed number of deaths (or in general, events).

In this example, $x = 20$, $\mu_0 = 1{,}000{,}000(1/100{,}000) = 10$. Therefore, the *p*-value is given by

$$p = 2 \times \left(1 - \sum_{k=0}^{19} \frac{e^{-10}10^k}{k!}\right) = 2 \times \left(\sum_{k=20}^{\infty} \frac{e^{-10}10^k}{k!}\right)$$

We refer to Table 2 (see the appendix in the text) under $\mu = 10$. Since $Pr(X = 25 | \mu = 10) = .0000$, we compute

$$\begin{aligned} p &= 2 \times [Pr(20) + Pr(21) + Pr(22) + Pr(23) + Pr(24)] \\ &= 2 \times (.0019 + .0009 + .0004 + .0002 + .0001) \\ &= .007 \end{aligned}$$

Thus, there are significantly more asthma deaths in 1963 than would be expected based on the asthma death rate from 1862 to 1962.

PROBLEMS

Nutrition

As part of a dietary-instruction program, ten 25–34-year-old males adopted a vegetarian diet for 1 month. While on the diet, the average daily intake of linoleic acid was 13 g with standard deviation $= 4$ g.

7.1 If the average daily intake among 25–34-year-old males in the general population is 15 g, then, using a significance level of .05, test the hypothesis that the intake of linoleic acid in this group is lower than that in the general population.

7.2 Compute a *p*-value for the hypothesis test in Problem 7.1.

As part of the same program, eight 25–34-year-old females report an average daily intake of saturated fat of 11 g with standard deviation $= 11$ g while on a vegetarian diet.

7.3 If the average daily intake of saturated fat among 25–34-year-old females in the general population is 24 g, then, using a significance level of .01, test the hypothesis that the intake of saturated fat in this group is lower than that in the general population.

7.4 Compute a *p*-value for the hypothesis test in Problem 7.3.

7.5 What is the relationship between your answers to Problems 7.3 and 7.4?

Suppose we are uncertain what effect a vegetarian diet will have on the level of linoleic-acid intake in Problem 7.1.

7.6 What are the null and alternative hypotheses in this case?

7.7 Compare the mean level of linoleic acid in the vegetarian population with that of the general population under the hypotheses in Problem 7.6. Report a *p*-value.

7.8 Suppose $s = 5$ based on a sample of 20 subjects. Test the null hypothesis $H_0: \sigma^2 = 16$ versus $H_1: \sigma^2 \neq 16$ using the critical-value method based on a significance level of .05.

7.9 Answer Problem 7.8 using the *p*-value method.

7.10 Suppose we are uncertain what effect a vegetarian diet will have on saturated fat intake. Compute a two-sided *p*-value for the hypothesis testing situation mentioned in Problem 7.3.

7.11 Compute a lower one-sided 95% CI for the true mean intake of linoleic acid in the vegetarian population depicted in Problem 7.1.

7.12 How does your answer to Problem 7.11 relate to your answer to Problem 7.1?

Nutrition
A food-frequency questionnaire was mailed to 20 subjects to assess the intake of various food groups. The sample standard deviation of vitamin-C intake over the 20 subjects was 15 (exclusive of vitamin-C supplements). Suppose we know from using an in-person diet-interview method in a large previous study that the standard deviation is 20.

7.13 What hypotheses can be used to test if there are any differences between the standard deviations of the two methods?

7.14 Perform the test described in Problem 7.13 and report a *p*-value.

The sample standard deviation for ln (vitamin-A intake) exclusive of supplements based on the 20 subjects using the food-frequency questionnaire was 0.016. Suppose the standard deviation from the diet-interview method is known to be 0.020 based on a large previous study.

7.15 Test the hypothesis that the variances using the two methods are the same. Use the critical-value method with a $\alpha = .05$.

7.16 Report a *p*-value corresponding to the test in Problem 7.15.

Occupational Health
Suppose that 28 cancer deaths are noted among workers exposed to asbestos in a building-materials plant from 1981–1985. Only 20.5 cancer deaths are expected from statewide cancer-mortality rates.

7.17 What is the estimated SMR for total cancer mortality?

7.18 Is there a significant excess or deficit of total cancer deaths among these workers?

In the same group of workers, 7 deaths due to leukemia are noted. Only 4.5 are expected from statewide rates.

7.19 What is the estimated SMR for leukemia?

7.20 Is there a significant excess or deficit of leukemia deaths among these workers?

Infectious Disease
Suppose the annual incidence of diarrhea (defined as 1+ episodes per year) in a Third-World country is 5% in children under the age of 2.

7.21 If 10 children out of 108 under the age of 2 in a poor rural community in the country have 1+ episodes of diarrhea in a year, then test if this represents a significant departure from the overall rate for the country using the critical-value method.

7.22 Report a *p*-value corresponding to your answer to Problem 7.21.

Ophthalmology
Suppose the distribution of systolic blood pressure in the general population is normal with a mean of 130 mm Hg and a standard deviation of 20 mm Hg. In a special subgroup of 85 people with glaucoma, we find that the mean systolic blood pressure is 135 mm Hg with a standard deviation of 22 mm Hg.

7.23 Assuming that the standard deviation of the glaucoma patients is the same as that of the general population, test for an association between glaucoma and high blood pressure.

7.24 Answer Problem 7.23 without making the assumption concerning the standard deviation.

Occupational Health

7.25 Suppose it is known that the average life expectancy of a 50-year-old man in 1945 was 18.5 years. Twenty men aged 50 who have been working for at least 20 years in a potentially hazardous industry were ascertained in 1945. On follow-up in 1985 all the men have died, with an average lifetime of 16.2 years and a standard deviation of 7.3 years since 1945. Assuming that life expectancy of 50-year-old men is approximately normally distributed, test if the underlying life expectancy for workers in this industry is shorter than for comparably aged men in the general population.

Cancer

7.26 An area of current interest in cancer epidemiology is the possible role of oral contraceptives (OC's) in the development of breast cancer. Suppose that in a group of 1000 premenopausal women ages 40–49 who are current users of OC's, 15 subsequently develop breast cancer over the next 5 years. If the expected 5-year incidence rate of breast cancer in this group is 1.2% based on national incidence rates, then test the hypothesis that there is an association between current OC use and the subsequent development of breast cancer.

Cancer
A group of investigators wishes to explore the relationship between the use of hair dyes and the development of breast cancer in females. A group of 1000 beauticians 30–39 years of age is identified and followed for 5 years. After 5 years, 20 new cases of breast cancer have occurred. Assume that breast-cancer incidence over this time period for an average woman in this age group is 7/1000. We wish to test the hypothesis that using hair dyes increases the risk of breast cancer.

7.27 Is a one-sided or two-sided test appropriate here?

7.28 Test the hypothesis.

Renal Disease
The level of serum creatinine in the blood is considered a good indicator of the presence or absence of kidney disease. Normal

people generally have low concentrations of serum creatinine, whereas diseased people have high concentrations. Suppose we want to look at the relation between analgesic abuse and kidney disorder. In particular, suppose we look at 15 people working in a factory who are known to be "analgesic abusers" (i.e., they take more than 10 pills per day) and we measure their creatinine levels. The creatinine levels are

0.9, 1.1, 1.6, 2.0, 0.8,
0.7, 1.4, 1.2, 1.5, 0.8,
1.0, 1.1, 1.4, 2.2, 1.4

7.29 If we assume that creatinine levels for normal people are normally distributed with mean 1.0, then can we make any comment about the levels for analgesic abusers via some statistical test?

7.30 Suppose the standard deviation of serum creatinine in the general population $= 0.40$. Can we compare the variance of serum creatinine among analgesic abusers versus the variance of serum creatinine in the general population?

Otolaryngology
Otitis media is an extremely common disease of the middle ear in children under 2 years of age. It can cause prolonged hearing loss during this period and may result in subsequent defects in speech and language. Suppose we wish to design a study to test the latter hypothesis, and we set up a study group consisting of children with 3 or more episodes of otitis media in the first 2 years of life. We have no idea what the size of the effect will be. Thus, we set up a pilot study with 20 cases, and find that 5 of the cases have speech and language defects at age 3.

7.31 If we regard this experience as representative of what would occur in a large study and if we *know* that 15% of all normal children have speech and language defects by age 3, then how large of a study group is needed to have an 80% chance of detecting a significant difference using a one-sided test at the 5% level?

7.32 Suppose only 50 cases can be recruited for the study group. How likely are we to find a significant difference if the true proportion of affected children with speech and language defects at age 3 is the same as that in the pilot study?

Epidemiology
Height and weight are often used in epidemiological studies as possible predictors of disease outcomes. If the people in the study are assessed in a clinic, then heights and weights are usually measured directly. However, if the people are interviewed at home or by mail, then a person's self-reported height and weight are often used instead. Suppose we conduct a study on 10 people to test the comparability of these two methods. The data for weight are given in Table 7.1.

7.33 Should a one-sided or two-sided test be used here?

Table 7.1 A comparison of self-reported and measured weight (lb) for 10 subjects

Subject number	Self–reported weight	Measured weight	Difference
1	120	125	–5
2	120	118	+2
3	135	139	–4
4	118	120	–2
5	120	125	–5
6	190	198	–8
7	124	128	–4
8	175	176	–1
9	133	131	+2
10	125	125	0

7.34 Which test procedure should be used to test the preceding hypothesis?

7.35 Conduct the test in Problem 7.34 using the critical-value method with $\alpha = .05$.

7.36 Compute the *p*-value for the test in Problem 7.34.

7.37 Is there evidence of digit preference among the self-reported weights? Specifically, compare the observed proportion of self-reported weights whose last digit is 0 or 5 with the expected proportion based on chance and report a *p*-value.

Hypertension
Several studies have been performed relating urinary potassium excretion to blood-pressure level. These studies have tended to show an inverse relationship between these two variables, with the higher the level of potassium excretion, the lower the BP level. Therefore, a treatment trial is planned to look at the effect of potassium intake in the form of supplement capsules on changes in DBP level. Suppose that in a pilot study, 20 people are given potassium supplements for 1 month. The data are as follows:

Mean change (1 month-baseline)	–3.2
sd change	8.5
n	20

7.38 What test should be used to assess if potassium supplements have any effect on DBP level?

7.39 Perform the test in Problem 7.38 using a two-sided test and report the *p*-value.

7.40 Derive a 95% CI for the true mean change based on the preceding data. What is the relationship of your results here and in Problem 7.39?

7.41 How many subjects need to be studied to have an 80% chance of detecting a significant treatment effect using a two-sided test with an α level of .05 if the mean and sd of the pilot study are assumed to be the population mean and sd?

Hospital Epidemiology
A study was conducted to identify characteristics that would predict 1-year survival for patients admitted to the medical service at New York Hospital [1]. One factor that was considered was the physician's estimate of the patients' severity of illness at the time of admission.

Suppose that it is expected, based on previous studies in this hospital, that 2/3 of admitted patients will survive for at least 1 year.

7.42 If 47% of 136 patients deemed severely ill survive at least 1 year, then what test can be used to test the hypothesis that the severity-of-illness rating is predictive of 1-year mortality?

7.43 Perform the test in Problem 7.42 using the critical-value method based on an α level pf .05.

7.44 Provide a 95% CI for the 1-year survival rate among severely ill patients.

Suppose that it is expected, based on previous studies, that among *all patients who survive hospitalization*, 75% will survive for 1 year. Furthermore, of the 136 severely ill patients, 33 die during hospitalization, and an additional 39 die during the 1st year, but after hospitalization.

7.45 Test the hypothesis that severity of illness is predictive of 1-year mortality among patients who are discharged from the hospital. Report a *p*-value.

Cancer
Radiotherapy is a common treatment for breast cancer, with about 25% of women receiving this form of treatment. Assume that the figure 25% is based on a very large sample and is known *without error*. One hypothesis is that radiotherapy applied to the contralateral breast may be a risk factor for development of breast cancer in the opposite breast 5 or more years after the initial cancer. Suppose that 655 women are identified who developed breast cancer in the opposite breast 5 or more years after the initial cancer [2].

7.46 If 206 of the women received radiotherapy after their initial diagnosis, then test the hypothesis that radiotherapy is associated with the development of breast cancer in the opposite breast. Please report a *p*-value.

7.47 Provide a 95% CI for the true proportion of women with contralateral breast cancer who received radiotherapy treatment.

7.48 Suppose that our *p*-value in Problem 7.46 = .03 (this is not necessarily the actual *p*-value in Problem 7.46). If we conduct a test using the critical-value method with $\alpha = .05$, then would we accept or reject H_0 and why? (Do not actually conduct the test.)

7.49 Suppose we wanted a 99% CI instead of a 95% CI in Problem 7.47. Would the length of the 99% CI be narrower, wider, or the same as the CI in Problem 7.47? (Do not actually construct the interval.)

Renal Disease
To compare two methods for the determination of uric acid, 23 blood samples from 23 individuals were divided and analyzed both by the colorimetric method and the uricase method. Suppose the sample means of the colorimetric and uricase assessments were 6.26 and 6.20 mg/dL, respectively, the sample standard deviation of the paired difference between repeated assessments is 0.50 mg/dL, and it is reasonable to assume that this paired difference has a normal distribution.

7.50 What is the best estimate for the mean difference between approaches?

7.51 Construct a 95% CI for the mean difference.

7.52 What can be concluded about the difference between assessments by the two methods?

Cardiovascular Disease
The relationship between serum cholesterol and coronary-heart-disease mortality has been a subject of much debate over the past 30 years. Some data relevant to this question were presented in the Whitehall study [3]. It was shown there that among 3615 men 40–64 years of age who were in the top quintile of serum cholesterol at baseline, 194 died from coronary heart disease over the next 10 years. Suppose the incidence rate of coronary-heart-disease mortality over 10 years among men in this age group in Great Britain is 4%.

7.53 What test can be performed to compare the incidence rate of coronary-heart-disease mortality among men in the top cholesterol quintile vs general-population incidence rates?

7.54 Implement the test in Problem 7.53 and report a *p*-value.

7.55 Construct a 95% CI for the true incidence rate in the group of 40–64-year-old men in the top quintile of serum cholesterol.

Cancer
It is well established that exposure to ionizing radiation at or after puberty increases a woman's risk of breast cancer. However, it is uncertain whether such exposure early in life is also carcinogenic. A study was performed [4] in a cohort of 1201 women who received x-ray treatment in infancy for an enlarged thymus gland and were followed prospectively for 36 years. It was found that 22 breast cancers occurred over a 36-year period among the 1201 women.

7.56 If the expected incidence rate of breast cancer in this age group over this time period is 1 event per 200 women based on New York State cancer-incidence rates, then test the hypothesis that irradiation has an effect on breast-cancer incidence. Please report a *p*-value and construct a 95% CI for the incidence rate among the exposed group. It was also found that 6 breast cancers occurred over a 36-year period among a subgroup of 138 women who were exposed to a high radioactive dose (0.50–1.99 gray, where a gray is a unit of radiation absorbed by the breast tissue).

7.57 Test the hypothesis that the high-risk subgroup is at excess risk for breast cancer, and report a *p*-value.

SOLUTIONS

7.1 Test the hypothesis $H_0: \mu = 15 = \mu_0$ versus $H_1: \mu < 15$. Reject H_0 if

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < t_{9,\ .05} = -1.833$$

and accept H_0 if $t \geq -1.833$. Since

$$t = \frac{13 - 15}{\frac{4}{\sqrt{10}}} = -1.581 > -1.833,$$

we accept H_0 at the 5% level.

7.2 The *p*-value is given by $Pr(t_9 < -1.581)$. Since $t_{9,\ .10} = -1.383$ and $t_{9,\ .05} = -1.833$ it follows that $.05 < p < .10$.

7.3 Test the hypothesis $H_0: \mu = 24 = \mu_0$ versus $H_1: \mu < 24$. Reject H_0 if

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} < t_{7,\ .01} = -2.998$$

and accept H_0 otherwise. Since

$$t = \frac{11 - 24}{\frac{11}{\sqrt{8}}} = -3.343 < -2.998,$$

reject H_0 at the 1% level.

7.4 The *p*-value is given by $Pr(t_7 < -3.343)$. Since $t_{7,\ .005} = -3.499$ and $t_{7,\ .01} = -2.998$ it follows that $.005 < p < .01$.

7.5 Since H_0 was rejected at the 1% level in Problem 7.3, *p* must be $< .01$, which is the case.

7.6 The hypotheses to be tested are $H_0: \mu = 15$ versus $H_1: \mu \neq 15$. This alternative is two-sided in contrast to the one-sided alternative in Problem 7.1.

7.7 The rejection region is given by $t < t_{9,\ .025} = -2.262$ or $t > t_{9,\ .975} = 2.262$. We have that $t = -1.581$. Since $-2.262 < -1.581 < 2.262$, we accept H_0 using a two-sided test at the 5% level. The exact *p*-value is given by $2 \times Pr(t_9 < -1.581) = \text{TDIST}(1.581,\ 9,\ 2) = .148$.

7.8 We have the test statistic

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1} \text{ under } H_0$$

We reject H_0 if $X^2 > \chi^2_{n-1,\ 1-\alpha/2}$ or

$$X^2 < \chi^2_{n-1,\ \alpha/2}.$$

In this case,

$$X^2 = \frac{19(25)}{16} = 29.69 \sim \chi^2_{19} \text{ under } H_0$$

The critical values are $\chi^2_{19,\ .025} = 8.91$ and

$$\chi^2_{19,\ .975} = 32.85.$$

Since $8.91 < 29.69 < 32.85$, it follows that we accept H_0 at the 5% level.

7.9 The two-sided *p*-value is given by

$$p = 2 \times Pr\left(\chi^2_{19} > 29.69\right)$$

Since $\chi^2_{19,\ .90} = 27.20$, $\chi^2_{19,\ .95} = 30.14$ and

$$27.20 < 29.69 < 30.14,$$

it follows that

$$1 - .95 < \frac{p}{2} < 1 - .90$$

or

$$.05 < \frac{p}{2} < .10$$

or

$$.10 < p < .20.$$

Using the CHIDIST function of Excel, the exact p-value $= 2 \times Pr(\chi^2_{19} > 29.69) = .112$.

7.10 p-value $= 2 \times Pr(t_7 < -3.343)$
$= \text{TDIST}(3.343,\ 7,\ 2) = .012.$

7.11 A lower one-sided 95% CI for the true mean intake of linoleic acid is given by

$$\mu < \bar{x} + t_{9,\ .95}\frac{s}{\sqrt{n}}$$
$$= 13 + 1.833\frac{4}{\sqrt{10}}$$
$$= 13 + 2.32 = 15.32$$

7.12 This interval contains 15.0 $g =$ mean of the general population. Thus, this answer corresponds to the answer to Problem 7.1, where we accepted H_0 using a one-sided significance test at the 5% level.

7.13 Test the hypothesis $H_0: \sigma^2 = \sigma_0^2$ versus $H_1: \sigma^2 \neq \sigma_0^2$. Use the one-sample χ^2 test for the variance of a normal distribution to test these hypotheses.

7.14 We have the test statistic

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{19(15)^2}{20^2}$$
$$= 10.69 \sim \chi^2_{19} \text{ under } H_0$$

Since $\chi^2_{19,\ .05} = 10.12$, $\chi^2_{19,\ .10} = 11.65$, and

$$10.12 < 10.69 < 11.65,$$

it follows that $2 \times .05 < p < 2 \times .10$, or $.10 < p < .20$. Thus, there is no significant difference between the variances using the two methods. The exact p-value $= .13$.

7.15 Test the hypothesis $H_0: \sigma^2 = \sigma_0^2 = (0.020)^2$ versus $H_1: \sigma^2 \neq \sigma_0^2$. Reject H_0 if

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\ 1-\alpha/2}$$

or $X^2 < \chi^2_{n-1,\ \alpha/2}$ and accept H_0 otherwise. We have the test statistic

$$X^2 = \frac{19(0.016)^2}{(0.020)^2} = 12.16 \sim \chi^2_{19} \text{ under } H_0$$

Since $\chi^2_{19,\ .025} = 8.91$, $\chi^2_{19,\ .975} = 32.85$ and

$$8.91 < 12.16 < 32.85$$

it follows that H_0 is accepted at the 5% level and we conclude that the variances of the two methods are not significantly different.

7.16 We have the test statistic $X^2 = 12.16 \sim \chi^2_{19}$ under H_0. Since $\chi^2_{19,\ .10} = 11.65$, $\chi^2_{19,\ .25} = 14.56$, and

$$11.65 < 12.16 < 14.56,$$

it follows that $2 \times .10 < p < 2 \times .25$, or

$$.20 < p < .50.$$

The exact p-value $= .24$.

7.17 SMR $= 100\% \times 28/20.5 = 137\%$.

7.18 Since the expected number of cancer deaths is ≥ 10, we can use the large-sample test procedure in (Equation 7.49), see text, Chapter 7. We have the test statistic

$$X^2 = \frac{(28-20.5)^2}{20.5} = \frac{7.5^2}{20.5} = \frac{56.25}{20.5}$$
$$= 2.74 \sim \chi^2_1 \text{ under } H_0$$

Since $\chi^2_{1,\ .90} = 2.71$, $\chi^2_{1,\ .95} = 3.84$ and

$$2.71 < 2.74 < 3.84,$$

it follows that $1 - .95 < p < 1 - .90$ or $.05 < p < .10$. The exact p-value, obtained by computer, is

$$p = Pr(\chi^2_1 > 2.74) = .098.$$

7.19 SMR $= 100\% \times 7/4.5 = 156\%$.

7.20 Since $E = 4.5 < 10$, we must use the small-sample method in (Equation 7.48), see text, Chapter 7, to test these hypotheses. Since $O > E$, the p-value is given by

$$p = 2 \times \left[1 - \sum_{k=0}^{6} \frac{e^{-4.5}(4.5)^k}{k!}\right]$$

We refer to the exact Poisson tables (Table 2, Appendix, text) under $\mu = 4.5$ and obtain

$$Pr(0) = .0111,\ Pr(1) = .0500,\ \ldots,\ Pr(6) = .1281.$$

Therefore,

$$p = 2 \times [1 - (.0111 + .0500 + \ldots + .1281)]$$
$$= 2 \times (1 - .8310) = .338$$

There is no significant excess or deficit of leukemia deaths in this group of workers.

7.21 We wish to test the hypothesis $H_0: p = p_0 = .05$ versus $H_1: p \neq p_0$. Since

$$np_0q_0 = 108(.05)(.95) = 5.13 \geq 5,$$

we can use the normal-theory test. If we use the critical-value method with a 5% significance level, then the critical values are $z_{.025} = -1.96$ and $z_{.975} = 1.96$. We reject H_0 if the test statistic z is ≤ -1.96 or ≥ 1.96. The test statistic is given by

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{\frac{10}{108} - .05}{\sqrt{\frac{.05(.95)}{108}}} = \frac{.0426}{.0210}$$
$$= 2.031$$

Since $z > 1.96$, it follows that we reject H_0 and conclude that there is a significant excess rate of diarrhea in the rural community.

7.22 Since $\hat{p} > p_0$, the p-value

$$= 2 \times [1 - \Phi(z)] = 2 \times [1 - \Phi(2.031)] = .042.$$

7.23 Test the hypothesis $H_0: \mu = 130 = \mu_0$ versus $H_1: \mu \neq 130$, where σ is assumed to be 20 mm Hg. Use the test statistic

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{135 - 130}{\frac{20}{\sqrt{85}}} = 2.305 \sim N(0, 1) \text{ under } H_0$$

Thus, the two-tailed p-value equals

$$2 \times [1 - \Phi(2.305)] = 2 \times (1 - .9894) = .021.$$

Therefore, there is a significant association between glaucoma and high blood pressure.

7.24 Perform the one-sample t test here, since the standard deviation is not assumed known. Use the test statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{135 - 130}{\frac{22}{\sqrt{85}}} = 2.095 \sim t_{84} \text{ under } H_0$$

Since

$$t_{60,\ .975} = 2.000,\ t_{60,\ .99} = 2.390$$
$$t_{120,\ .975} = 1.980,\ t_{120,\ .99} = 2.358$$

it follows that if there were 60 or 120 *df*, then

$$2 \times (1 - .99) < p < 2 \times (1 - .975)$$

or

$$.02 < p < .05$$

Since there are 84 *df* we must also have $.02 < p < .05$. Thus, there is a significant association between glaucoma and high blood pressure in this case as well.

7.25 We wish to test the hypothesis $H_0: \mu = \mu_0 = 18.5$ versus $H_1: \mu < \mu_0$. We will use a one-sample t test. We have the test statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{16.2 - 18.5}{\frac{7.3}{\sqrt{20}}} = \frac{-2.3}{1.632}$$
$$= -1.409 \sim t_{19} \text{ under } H_0$$

Since $t_{19,\ .9} = 1.328$, $t_{19,\ .95} = 1.729$ and

$$1.328 < 1.409 < 1.729,$$

it follows that $1 - .95 < p < 1 - .90$ or $.05 < p < .1$. Thus, the results are not statistically significant, and we cannot conclude that workers in this industry have significantly shorter lifetimes.

7.26 We wish to test the hypothesis $H_0: p = p_0 = .012$ versus $H_1: p \neq p_0$. We use the normal-theory test since $np_0q_0 = 1000(.012)(.988) = 11.86 \geq 5$. We have the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{.015 - .012}{\sqrt{\frac{.012(.988)}{1000}}} = \frac{.003}{.00344}$$
$$= 0.87 \sim N(0, 1) \text{ under } H_0$$

Thus,

$$p = 2[1 - \Phi(0.87)]$$
$$= 2(1 - .8082) = .38$$

and there is no significant association between current OC use and the subsequent development of breast cancer.

7.27 Test the hypothesis $H_0: p = p_0$ versus $H_1: p > p_0$. A one-sided alternative is more appropriate here, since we wish to detect only if hair dyes increase the risk of breast cancer rather than decrease the risk.

7.28 Since $np_0q_0 = 1000(.007)(.993) = 6.95 \geq 5$, use the large-sample test. We have the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{\frac{20}{1000} - .007}{\sqrt{\frac{.007 \times .993}{1000}}} = \frac{.013}{.0026}$$
$$= 4.93 \sim N(0, 1) \text{ under } H_0$$

The p-value is given by

$$p = 1 - \Phi(z) = 1 - \Phi(4.93) < .001.$$

Thus, the results are very highly statistically significant, and we conclude that extensive occupational exposure to hair dyes significantly increases the risk of breast cancer.

7.29 We have that $\bar{x} = 1.273$, $s = 0.435$, and $n = 15$. We must test the hypothesis $H_0: \mu = 1.0$ versus $H_1: \mu \neq 1.0$. We perform the one-sample t test using the test statistic

$$t = \frac{\bar{x} - 1.0}{\frac{0.435}{\sqrt{n}}} \sim t_{14} \text{ under } H_0$$

Hence, we have

$$t = \frac{1.273 - 1.0}{\frac{0.435}{\sqrt{15}}} = \frac{0.273}{0.112} = 2.434$$

Thus, the p-value

$$= 2 \times Pr\left(t_{14} > 2.434\right) = \text{TDIST}\left(2.434,\ 14,\ 2\right) = .029 .$$

Therefore, we reject the null hypothesis at a significance level of $p = .029$.

7.30 We can test for whether $\sigma = 0.40$ by performing a one-sample χ^2 test. In particular, we test the hypothesis $H_0: \sigma = 0.40$ versus $H_1: \sigma \neq 0.40$. We have the test statistic

$$X^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \text{ under } H_0$$

Thus,

$$X^2 = \frac{14(0.435)^2}{(0.40)^2} = 16.56 \sim \chi^2_{14} \text{ under } H_0$$

Since $\chi^2_{14,\ .975} = 26.12$, $\chi^2_{14,\ .025} = 5.63$ and

$$5.63 < 16.56 < 26.12$$

it follows that $p > .05$. Thus, we can accept the null hypothesis that $\sigma = 0.40$.

7.31 We use the sample-size formula in (Equation 7.46) (text, Chapter 7) using $z_{1-\alpha}$, instead of $z_{1-\alpha/2}$ for a one-sided test. In this case, $p_0 = .15$, $q_0 = .85$, $p_1 = .25$, $q_1 = .75$, $\alpha = .05$. Therefore,

$$n = \frac{p_0 q_0 \left(z_{1-\alpha} + z_{1-\beta}\sqrt{\frac{p_1 q_1}{p_0 q_0}}\right)^2}{\left(p_1 - p_0\right)^2}$$
$$= \frac{.15(.85)\left(z_{.95} + z_{.80}\sqrt{\frac{0.25(0.75)}{0.15(0.85)}}\right)^2}{\left(.25 - .15\right)^2}$$
$$= \frac{.1275\left[1.645 + .84\left(1.213\right)\right]^2}{.01}$$
$$= 12.75\left(7.095\right) = 90.5$$

Therefore, we need to study 91 children with 3+ episodes of otitis media in the first two years of life in order to achieve an 80% power.

7.32 We must compute the power of the study. We refer to the power formula in (Equation 7.45) (text, Chapter 7) using z_α instead of $z_{\alpha/2}$ for a one-sided test. We have

$$\text{Power} = \Phi\left[\sqrt{\frac{p_0 q_0}{p_1 q_1}}\left(z_\alpha + \frac{|p_1 - p_0|\sqrt{n}}{\sqrt{p_0 q_0}}\right)\right]$$

where $n = 50$, $\alpha = .05$, $p_0 = .15$, $q_0 = .85$, $p_1 = .25$, $q_1 = .75$. Thus

$$\text{Power} = \Phi\left[\sqrt{\frac{.15(.85)}{.25(.75)}}\left(z_{.05} + \frac{|.25 - .15|\sqrt{50}}{\sqrt{.15(.85)}}\right)\right]$$
$$= \Phi\left[\sqrt{\frac{.1275}{.1875}}\left(-1.645 + \frac{.10\sqrt{50}}{\sqrt{.1275}}\right)\right]$$
$$= \Phi[0.8246(-1.645 + 1.9803)]$$
$$= \Phi(0.276) = .61$$

Thus, there is a 61% chance of finding a significant difference if we enroll 50 children in the study.

7.33 Use a two-sided test, because there is no reason to expect self-reported weights to be higher or lower than actual weights.

7.34 Use a one-sample t test to test the hypothesis $H_0: \mu = 0$ versus $H_1: \mu \neq 0$, where μ is the true mean difference (self-reported weight minus measured weight).

7.35 We have $\bar{x} = -2.5$, $s = 3.27$ based on the 10 difference scores. The rejection region is given by $t > t_{9,\ .975} = 2.262$ or $t < t_{9,\ .025} = -2.262$. We have the test statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{-2.5 - 0}{\frac{3.27}{\sqrt{10}}}$$
$$= \frac{-2.5}{1.035} = -2.414$$

Since $t = -2.414 < -2.262$, we reject H_0 at the 5% level.

7.36 Since $\bar{x} < \mu_0 = 0$, the p-value is obtained from

$$2 \times Pr(t_9 < -2.414) = 2 \times Pr(t_9 > 2.414)$$

Since $t_{9,\ .975} = 2.262$, $t_{9,\ .99} = 2.821$, and

$$2.262 < 2.414 < 2.821,$$

it follows that $2 \times (1 - .99) < p < 2 \times (1 - .975)$, or $.02 < p < .05$. The exact p-value, obtained by computer, is $p = 2 \times Pr(t_9 < -2.414) = .039$.

7.37 There are 7 reported weights out of 10 that end in 0 or 5 ($\hat{p} = .7$) compared with 2 expected ($p_0 = .2$). Use a one-sided hypothesis test of the form H_0: $p = p_0$ versus H_1: $p > p_0$, since p would only be expected to be larger than p_0 due to digit preference. Since $np_0q_0 = 10 \times .2 \times .8 = 1.6 < 5$, use the exact method. The p-value is given by

$$p = \sum_{k=7}^{10} {}_{10}C_k (.2)^k (.8)^{10-k}$$

Refer to Table 1 (Appendix, text) under $n = 10$, $p = .2$ and obtain

$$p = .0008 + .0001 + .0000 + .0000 = .0009.$$

Thus, there is a significant excess of 0's and 5's among the self-reported weights.

7.38 Use the one-sample t test to test the hypothesis H_0: $\mu = 0$ versus H_1: $\mu \neq 0$ where μ is the true mean change in DBP after taking potassium supplements.

7.39 We have the test statistic

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{-3.2}{\frac{8.5}{\sqrt{20}}}$$
$$= \frac{-3.2}{1.90} = -1.684 \sim t_{19} \text{ under } H_0$$

Note from the t tables that

$$t_{19,\ .95} = 1.729,\ t_{19,\ .90} = 1.328.$$

Since $1.328 < 1.684 < 1.729$, it follows that

$$2 \times (1 - .95) < p < 2 \times (1 - .90),$$

or $.1 < p < .2$. The exact p-value, obtained by computer, is $p = 2 \times Pr(t_{19} < -1.684) = .11$. Thus, there is not a significant change in mean diastolic blood pressure based on the pilot-study results.

7.40 A two-sided 95% CI is given by

$$\bar{x} \pm t_{19,\ .975} \frac{s}{\sqrt{n}} = -3.2 \pm 2.093\left(\frac{8.5}{\sqrt{20}}\right)$$
$$= -3.2 \pm 2.093(1.90)$$
$$= -3.2 \pm 4.0 = (-7.2,\ 0.8)$$

The 95% CI includes the null value for $\mu(0)$, which is consistent with the nonsignificant results in Problem 7.39.

7.41 Use the sample-size formula

$$n = \frac{\sigma^2 \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\Delta^2}$$

In this case, $\sigma = 8.5$, $\alpha = .05$, $\beta = 1 - .80 = .20$, $\Delta = 3.2$. Therefore,

$$n = \frac{8.5^2 (z_{.975} + z_{.80})^2}{3.2^2} = \frac{8.5^2 (1.96 + 0.84)^2}{3.2^2}$$
$$= 55.3$$

Thus, 56 subjects are needed in order to have an 80% chance of detecting a significant difference using a two-sided test with an α level = .05.

7.42 We wish to test the hypothesis H_0: $p = p_0$ versus H_1: $p \neq p_0$, where $p =$ the true one-year survival rate among patients deemed to be severely ill at admission to the hospital and $p_0 = 2/3$. We use the one-sample binomial test using the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0,\ 1) \text{ under } H_0.$$

7.43 If we use an α level of .05, then the critical values for this test are given by $z_{.025} = -1.96$ and $z_{.975} = 1.96$, where we reject H_0 if $z < -1.96$ or $z > 1.96$ and accept H_0 otherwise. We have

$$z = \frac{0.47 - \frac{2}{3}}{\sqrt{\frac{(2/3)(1/3)}{136}}} = \frac{-.1967}{.0404} = -4.865$$

Since $-4.865 < -1.96$, we reject H_0 at the 5% level.

7.44 The 95% CI is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = .47 \pm 1.96 \times \sqrt{\frac{.47(.53)}{136}}$$
$$= .47 \pm .084 = (.386,\ .554).$$

7.45 We use the same test as in Problem 7.42. In this case, the one-year survival rate post-hospitalization

$$1 - \frac{39}{136 - 33} = 1 - \frac{39}{103} = .621 = \hat{p}$$

while $p_0 = .75$. We have the test statistic

$$z = \frac{.621 - .75}{\sqrt{\frac{.75(.25)}{103}}}$$
$$= \frac{-.129}{.0427} = -3.015 \sim N(0,\ 1) \text{ under } H_0$$

We have that

$$p = 2 \times \Phi(-3.015) = 2 \times [1 - \Phi(3.015)]$$
$$= 2 \times (1 - .9987) = .003.$$

Thus, severely ill patients who survive the hospitalization are significantly less likely to survive the remainder of the first year after hospitalization compared with the expected survival rate of .75 for all hospitalized patients.

7.46 We wish to test the hypothesis $H_0: p = .25 = p_0$ versus $H_1: p \neq .25$, where $p =$ proportion of women who received radiotherapy among women with breast cancer who subsequently develop contralateral breast cancer. We have that $\hat{p} = 206/655 = .315$. Since

$$np_0q_0 = 655(.25)(.75) = 122.8 \geq 5$$

we can use the normal-theory test. We have the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{.315 - .25}{\sqrt{\frac{.25(.75)}{655}}}$$
$$= \frac{.065}{.0169} = 3.81 \sim N(0,\ 1) \text{ under } H_0$$

The p-value $= 2 \times [1 - \Phi(3.81)] < .001$. Thus, women with breast cancer who subsequently develop contralateral breast cancer were significantly more likely to have received radiotherapy previously than the average breast-cancer patient.

7.47 A 95% CI for p is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = .315 \pm 1.96\sqrt{\frac{.315(.685)}{655}}$$
$$= .315 \pm .036 = (.279,\ .350).$$

This interval excludes $p_0 = .25$.

7.48 We would reject H_0 because the p-value represents the α level at which we would be indifferent between accepting and rejecting H_0. For all α levels $> p$, we reject H_0; for all α levels $\leq p$, we accept H_0. Since $\alpha = .05 > .03 = p$, we would reject H_0.

7.49 The 99% CI would be wider than the 95% CI in Problem 7.47. As the level of confidence increases, the width of the CI must increase.

7.50 The best estimate of the mean difference

$$6.26 - 6.20 = 0.06 \text{ mg/dL}.$$

7.51 A 95% CI for the mean difference is given by

$$0.06 \pm t_{22,\ .975}\left(\frac{0.50}{\sqrt{23}}\right) = 0.06 \pm 2.074\left(\frac{0.50}{\sqrt{23}}\right)$$
$$= 0.06 \pm 0.22 = (-0.16,\ 0.28)$$

7.52 Since the 95% CI includes zero, we conclude that there is no significant difference between the methods; i.e., *on average* we obtain the same value using either method. However, there may be a large difference for specific people.

7.53 We can use the one-sample binomial test. The hypotheses are $H_0: p = p_0$ versus $H_1: p \neq p_0$, where $p_0 =$ incidence rate of coronary heart disease among 40–64-year-old men in the general population and $p =$ incidence rate of coronary heart disease among 40–64-year-old men in the top quintile of serum cholesterol.

7.54 In this example, $n = 3615$,

$$\hat{p} = \frac{194}{3615} = .0537.$$

Since $np_0q_0 = 3615(.04)(.96) = 138.8 \geq 5$, we can use the normal-theory test. We have the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}} = \frac{\frac{194}{3615} - .04}{\sqrt{\frac{.04(.96)}{3615}}}$$
$$= \frac{.0137}{.0033} = 4.19 \sim N(0,\ 1) \text{ under } H_0$$

The p-value $= 2 \times [1 - \Phi(4.19)] < .001$. Thus, men in the top quintile of serum cholesterol at baseline have a significantly higher incidence rate of coronary-heart disease mortality over 10 years than comparably aged men in the general population.

7.55 A 95% CI for the true incidence rate of coronary-heart-disease mortality among 40–64-year-old men in the top quintile of serum cholesterol is

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = .0537 \pm 1.96\sqrt{\frac{.0537(.9563)}{3615}}$$
$$= .0537 \pm .0073 = (.046,\ .061)$$

7.56 We wish to test the hypothesis $H_0: p = p_0$ versus $H_1: p \neq p_0$ where $p =$ incidence rate of breast cancer among irradiated women, $p_0 =$ incidence rate of breast cancer from New York State incidence rates $= 1/200 = .005$. Since $np_0q_0 = 1201(.005)(.995) = 5.97 \geq 5$, we can use the normal-theory test. We have the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}}$$

where $\hat{p} = 22/1201 = .0183$, $n = 1201$. Thus,

$$z = \frac{.0183 - .005}{\sqrt{\frac{.005(.995)}{1201}}}$$
$$= \frac{.0133}{.0020} = 6.54 \sim N(0,\ 1)$$

under H_0. The p-value $= 2 \times [1 - \Phi(6.54)] < .001$. Thus, there is a highly significant excess risk among the irradiated group. A 95% CI for p is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = .0183 \pm 1.96\sqrt{\frac{.0183(.9817)}{1201}}$$
$$= .0183 \pm 1.96(.0039)$$
$$= .0183 \pm .0076 = (.011,\ .026)$$

7.57 In this case, $np_0q_0 = 138(.005)(.995) = 0.69 < 5$. Thus, we must use the small sample test based on exact binomial probabilities. The p-value is given by

$$p = 2 \times \sum_{k=6}^{138} \binom{138}{k} p_0^k q_0^{138-k}$$
$$= 2 \times \sum_{k=6}^{138} \binom{138}{k} (.005)^k (.995)^{138-k}$$
$$= 2 \times \left[1 - \sum_{k=0}^{5} \binom{138}{k} (.005)^k (.995)^{138-k}\right]$$

We use the BINOMDIST function of Excel to evaluate the p-value. We have

$$\Pr(X \leq 5 | n = 138,\ p = .005) = .999924$$

Thus, the p-value $= 2 \times (1 - .999924) < .001$. Thus, there is a significant excess risk in this subgroup as well.

REFERENCES

[1] Pompei, P., Charlson, M. E., & Douglas, R. G., Jr. (1988). Clinical assessments as predictors of one-year survival after hospitalization: Implications for prognostic stratification. *Journal of Clinical Epidemiology*, **41**(3), 275–284.

[2] Boice, J. D., Jr., Harvey, E. B., Blettner, M., Stovall, M., Flannery, J. (1992). Cancer in the contralateral breast after radiotherapy for breast cancer. *New England Journal of Medicine*, **326**(12),781–785.

[3] Rose, G. & Shipley, M. (1986). Plasma cholesterol concentration and death from coronary heart disease: 10-year results of the Whitehall study, *British Medical Journal* (Clinical Research Edition), **293**(6542), 306–307.

[4] Hildreth, N. G., Shore, R. E., & Dvoretsky, P. M. (1989). The risk of breast cancer after irradiation of the thymus in infancy. *New England Journal of Medicine*, **321**(19), 1281–1284.