16. Let the test statistic T have a t distribution when H_0 is true. Give the significance level for each of the following situations:

a. H_a: $\mu > \mu_0$, df $= 15$, rejection region $t \geq 3.733$

b. H_a: $\mu < \mu_0$, $n = 24$, rejection region $t \leq -2.500$

c. H_a: $\mu \neq \mu_0$, $n = 31$, rejection region $t \geq 1.697$ or $t \leq -1.697$

17. Answer the following questions for the tire problem in Example 8.7.

a. If $\bar{x} = 30{,}960$ and a level $\alpha = .01$ test is used, what is the decision?

b. If a level .01 test is used, what is $\beta(30{,}500)$?

c. If a level .01 test is used and it is also required that $\beta(30{,}500) = .05$, what sample size n is necessary?

d. If $\bar{x} = 30{,}960$, what is the smallest α at which H_0 can be rejected (based on $n = 16$)?

18. Reconsider the paint-drying situation of Example 8.2, in which drying time for a test specimen is normally distributed with $\sigma = 9$. The hypotheses H_0: $\mu = 75$ versus H_a: $\mu < 75$ are to be tested using a random sample of $n = 25$ observations.

a. How many standard deviations (of $\bar{X}$) below the null value is $\bar{x} = 72.3$?

b. If $\bar{x} = 72.3$, what is the conclusion using $\alpha = .01$?

c. What is α for the test procedure that rejects H_0 when $z \leq -2.88$?

d. For the test procedure of part (c), what is $\beta(70)$?

e. If the test procedure of part (c) is used, what n is necessary to ensure that $\beta(70) = .01$?

f. If a level .01 test is used with $n = 100$, what is the probability of a type I error when $\mu = 76$?

19. The melting point of each of 16 samples of a certain brand of hydrogenated vegetable oil was determined, resulting in $\bar{x} = 94.32$. Assume that the distribution of the melting point is normal with $\sigma = 1.20$.

a. Test H_0: $\mu = 95$ versus H_a: $\mu \neq 95$ using a two-tailed level .01 test.

b. If a level .01 test is used, what is $\beta(94)$, the probability of a type II error when $\mu = 94$?

c. What value of n is necessary to ensure that $\beta(94) = .1$ when $\alpha = .01$?

20. Lightbulbs of a certain type are advertised as having an average lifetime of 750 hours. The price of these bulbs is very favorable, so a potential customer has decided to go ahead with a purchase arrangement unless it can be conclusively demonstrated that the true average lifetime is smaller than what is advertised. A random sample of 50 bulbs was selected, the lifetime of each bulb determined, and the appropriate hypotheses were tested using Minitab, resulting in the accompanying output.

Variable	N	Mean	StDev	SEMean	Z	P-Value
lifetime	50	738.44	38.20	5.40	−2.14	0.016

What conclusion would be appropriate for a significance level of .05? A significance level of .01? What significance level and conclusion would you recommend?

21. The true average diameter of ball bearings of a certain type is supposed to be .5 in. A one-sample t test will be carried out to see whether this is the case. What conclusion is appropriate in each of the following situations?

a. $n = 13$, $t = 1.6$, $\alpha = .05$

b. $n = 13$, $t = -1.6$, $\alpha = .05$

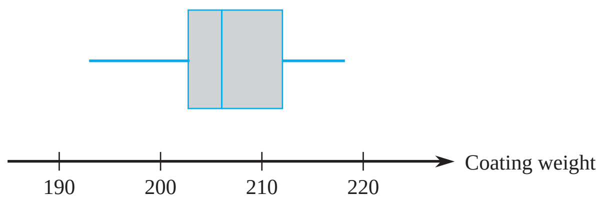
c. $n = 25$, $t = -2.6$, $\alpha = .01$

d. $n = 25$, $t = -3.9$

22. The article "The Foreman's View of Quality Control" (*Quality Engr.*, 1990: 257–280) described an investigation into the coating weights for large pipes resulting from a galvanized coating process. Production standards call for a true average weight of 200 lb per pipe. The accompanying descriptive summary and boxplot are from Minitab.

Variable	N	Mean	Median	TrMean	StDev	SEMean
ctg wt	30	206.73	206.00	206.81	6.35	1.16

Variable	Min	Max	Q1	Q3
ctg wt	193.00	218.00	202.75	212.00



a. What does the boxplot suggest about the status of the specification for true average coating weight?

b. A normal probability plot of the data was quite straight. Use the descriptive output to test the appropriate hypotheses.

23. Exercise 36 in Chapter 1 gave $n = 26$ observations on escape time (sec) for oil workers in a simulated exercise, from which the sample mean and sample standard deviation are 370.69 and 24.36, respectively. Suppose the investigators had believed *a priori* that true average escape time would be at most 6 min. Does the data contradict this prior belief? Assuming normality, test the appropriate hypotheses using a significance level of .05.

24. Reconsider the sample observations on stabilized viscosity of asphalt specimens introduced in Exercise 46 in Chapter 1 (2781, 2900, 3013, 2856, and 2888). Suppose that for a particular application it is required that true average viscosity be 3000. Does this requirement appear to have been satisfied? State and test the appropriate hypotheses.

25. The desired percentage of SiO_2 in a certain type of aluminous cement is 5.5. To test whether the true average percentage is 5.5 for a particular production facility, 16 independently obtained samples are analyzed. Suppose that the percentage of SiO_2 in a sample is normally distributed with $\sigma = .3$ and that $\bar{x} = 5.25$.

a. Does this indicate conclusively that the true average percentage differs from 5.5? Carry out the analysis using the sequence of steps suggested in the text.

b. If the true average percentage is $\mu = 5.6$ and a level $\alpha = .01$ test based on $n = 16$ is used, what is the probability of detecting this departure from H_0?

c. What value of n is required to satisfy $\alpha = .01$ and $\beta(5.6) = .01$?

26. To obtain information on the corrosion-resistance properties of a certain type of steel conduit, 45 specimens are buried in soil for a 2-year period. The maximum penetration (in mils) for each specimen is then measured, yielding a sample average penetration of $\bar{x} = 52.7$ and a sample standard deviation of $s = 4.8$. The conduits were manufactured with the specification that true average penetration be at most 50 mils. They will be used unless it can be demonstrated conclusively that the specification has not been met. What would you conclude?

27. Automatic identification of the boundaries of significant structures within a medical image is an area of ongoing research. The paper "Automatic Segmentation of Medical Images Using Image Registration: Diagnostic and Simulation Applications" (*J. of Medical Engr. and Tech.*, 2005: 53–63) discussed a new technique for such identification. A measure of the accuracy of the automatic region is the average linear displacement (ALD). The paper gave the following ALD observations for a sample of 49 kidneys (units of pixel dimensions).

1.38	0.44	1.09	0.75	0.66	1.28	0.51
0.39	0.70	0.46	0.54	0.83	0.58	0.64
1.30	0.57	0.43	0.62	1.00	1.05	0.82
1.10	0.65	0.99	0.56	0.56	0.64	0.45
0.82	1.06	0.41	0.58	0.66	0.54	0.83
0.59	0.51	1.04	0.85	0.45	0.52	0.58
1.11	0.34	1.25	0.38	1.44	1.28	0.51

a. Summarize/describe the data.

b. Is it plausible that ALD is at least approximately normally distributed? Must normality be assumed prior to calculating a CI for true average ALD or testing hypotheses about true average ALD? Explain.

c. The authors commented that in most cases the ALD is better than or of the order of 1.0. Does the data in fact provide strong evidence for concluding that true average ALD under these circumstances is less than 1.0? Carry out an appropriate test of hypotheses.

d. Calculate an upper confidence bound for true average ALD using a confidence level of 95%, and interpret this bound.

28. Minor surgery on horses under field conditions requires a reliable short-term anesthetic producing good muscle relaxation, minimal cardiovascular and respiratory changes, and a quick, smooth recovery with minimal aftereffects so that horses can be left unattended. The article "A Field Trial of Ketamine Anesthesia in the Horse" (*Equine Vet. J.*, 1984: 176–179) reports that for a sample of $n = 73$ horses to which ketamine was administered under certain conditions, the sample average lateral recumbency (lying-down) time was 18.86 min and the standard deviation was 8.6 min. Does this data suggest that true average lateral recumbency time under these conditions is less than 20 min? Test the appropriate hypotheses at level of significance .10.

29. The article "Uncertainty Estimation in Railway Track Life-Cycle Cost" (*J. of Rail and Rapid Transit*, 2009) presented the following data on time to repair (min) a rail break in the high rail on a curved track of a certain railway line.

159 120 480 149 270 547 340 43 228 202 240 218

A normal probability plot of the data shows a reasonably linear pattern, so it is plausible that the population distribution of repair time is at least approximately normal. The sample mean and standard deviation are 249.7 and 145.1, respectively.

a. Is there compelling evidence for concluding that true average repair time exceeds 200 min? Carry out a test of hypotheses using a significance level of .05.

b. Using $\sigma = 150$, what is the type II error probability of the test used in (a) when true average repair time is actually 300 min? That is, what is $\beta(300)$?

30. Have you ever been frustrated because you could not get a container of some sort to release the last bit of its contents? The article "Shake, Rattle, and Squeeze: How Much Is Left in That Container?" (*Consumer Reports*, May 2009: 8) reported on an investigation of this issue for various consumer products. Suppose five 6.0 oz tubes of toothpaste of a particular brand are randomly selected and squeezed until no more toothpaste will come out. Then each tube is cut open and the amount remaining is weighed, resulting in the following data (consistent with what the cited article reported): .53, .65, .46, .50, .37. Does it appear that the true average amount left is less than 10% of the advertised net contents?

a. Check the validity of any assumptions necessary for testing the appropriate hypotheses.

b. Carry out a test of the appropriate hypotheses using a significance level of .05. Would your conclusion change if a significance level of .01 had been used?

c. Describe in context type I and II errors, and say which error might have been made in reaching a conclusion.

31. A well-designed and safe workplace can contribute greatly to increased productivity. It is especially important that workers not be asked to perform tasks, such as lifting, that exceed their capabilities. The accompanying data on maximum weight of lift (MAWL, in kg) for a frequency of four lifts/min was reported in the article "The Effects of Speed, Frequency, and Load on Measured Hand Forces for a Floor-to-Knuckle Lifting Task" (*Ergonomics*, 1992: 833–843); subjects were randomly selected from the population of healthy males ages 18–30. Assuming that MAWL is normally distributed, does the data suggest that the population mean MAWL exceeds 25? Carry out a test using a significance level of .05.

25.8 36.6 26.3 21.8 27.2

32. The recommended daily dietary allowance for zinc among males older than age 50 years is 15 mg/day. The article "Nutrient Intakes and Dietary Patterns of Older Americans: A National Study" (*J. of Gerontology*, 1992: M145–150) reports the following summary data on intake for a sample of males age 65–74 years: $n = 115$, $\bar{x} = 11.3$, and $s = 6.43$. Does this data indicate that average daily zinc intake in the population of all males ages 65–74 falls below the recommended allowance?

33. Reconsider the accompanying sample data on expense ratio (%) for large-cap growth mutual funds first introduced in Exercise 1.53.

```
0.52 1.06 1.26 2.17 1.55 0.99 1.10 1.07 1.81 2.05
0.91 0.79 1.39 0.62 1.52 1.02 1.10 1.78 1.01 1.15
```

A normal probability plot shows a reasonably linear pattern.

a. Is there compelling evidence for concluding that the population mean expense ratio exceeds 1%? Carry out a test of the relevant hypotheses using a significance level of .01.

b. Referring back to (a), describe in context type I and II errors and say which error you might have made in reaching your conclusion. The source from which the data was obtained reported that $\mu = 1.33$ for the population of all 762 such funds. So did you actually commit an error in reaching your conclusion?

c. Supposing that $\sigma = .5$, determine and interpret the power of the test in (a) for the actual value of μ stated in (b).

34. A sample of 12 radon detectors of a certain type was selected, and each was exposed to 100 pCi/L of radon. The resulting readings were as follows:

105.6	90.9	91.2	96.9	96.5	91.3
100.1	105.0	99.6	107.7	103.3	92.4

a. Does this data suggest that the population mean reading under these conditions differs from 100? State and test the appropriate hypotheses using $\alpha = .05$.

b. Suppose that prior to the experiment a value of $\sigma = 7.5$ had been assumed. How many determinations would then have been appropriate to obtain $\beta = .10$ for the alternative $\mu = 95$?

35. Show that for any $\Delta > 0$, when the population distribution is normal and σ is known, the two-tailed test satisfies $\beta(\mu_0 - \Delta) = \beta(\mu_0 + \Delta)$, so that $\beta(\mu')$ is symmetric about μ_0.

36. For a fixed alternative value μ', show that $\beta(\mu') \to 0$ as $n \to \infty$ for either a one-tailed or a two-tailed z test in the case of a normal population distribution with known σ.

8.3 Tests Concerning a Population Proportion

Let p denote the proportion of individuals or objects in a population who possess a specified property (e.g., cars with manual transmissions or smokers who smoke a filter cigarette). If an individual or object with the property is labeled a success (S), then p is the population proportion of successes. Tests concerning p will be based on a random sample of size n from the population. Provided that n is small relative to the population size, X (the number of S's in the sample) has (approximately) a binomial distribution. Furthermore, if n itself is large [$np \geq 10$ and $n(1-p) \geq 10$], both X and the estimator $\hat{p} = X/n$ are approximately normally distributed. We first consider large-sample tests based on this latter fact and then turn to the small-sample case that directly uses the binomial distribution.

Large-Sample Tests

Large-sample tests concerning p are a special case of the more general large-sample procedures for a parameter θ. Let $\hat{\theta}$ be an estimator of θ that is (at least approximately) unbiased and has approximately a normal distribution. The null hypothesis has the form H_0: $\theta = \theta_0$ where θ_0 denotes a number (the null value) appropriate to the problem context. Suppose that when H_0 is true, the standard deviation of $\hat{\theta}$, $\sigma_{\hat{\theta}}$, involves no unknown parameters. For example, if $\theta = \mu$ and $\hat{\theta} = \overline{X}$, $\sigma_{\hat{\theta}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$, which involves no unknown parameters only if the value of σ is known. A large-sample test statistic results from standardizing $\hat{\theta}$ under the assumption that H_0 is true (so that $E(\hat{\theta}) = \theta_0$):

$$\text{Test statistic: } Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

If the alternative hypothesis is H_a: $\theta > \theta_0$, an upper-tailed test whose significance level is approximately α is specified by the rejection region $z \geq z_\alpha$. The other two alternatives, H_a: $\theta < \theta_0$ and H_a: $\theta \neq \theta_0$, are tested using a lower-tailed z test and a two-tailed z test, respectively.