

MA 116 Elements of Statistics
 Selected Review Exercises for Final Exam
 Spring 2010
 Organized by Dr. Paul Duty (MA 116 Course Chair)

I. Exercises for the Normal Distribution

1. **Three great hitters.** Three landmarks of baseball achievement are Ty Cobb's batting average of .420 in 1911, Ted Williams's .406 in 1941, and George Brett's .390 in 1980. These batting averages cannot be compared directly because the distribution of major league batting averages has changed over the years. The distributions are quite symmetric and (except for outliers such as Cobb, Williams, and Brett) reasonably Normal. While the mean batting average has been held roughly constant by rule changes and the balance between hitting and pitching, the standard deviation has dropped over time. Here are the facts:

Decade	Mean	Std. dev.
1910s	.266	.0371
1940s	.267	.0326
1970s	.261	.0317

Compute the standardized batting averages for Cobb, Williams, and Brett to compare how far each stood above his peers.¹⁰

2. **How hard do locomotives pull?** An important measure of the performance of a locomotive is its "adhesion," which is the locomotive's pulling force as a multiple of its weight. The adhesion of one 4400-horsepower diesel locomotive model varies in actual use according to a Normal distribution with mean $\mu = 0.37$ and standard deviation $\sigma = 0.04$.

- (a) What proportion of adhesions measured in use are higher than 0.40?
- (b) What proportion of adhesions are between 0.40 and 0.50?

3. **Are we getting smarter?** When the Stanford-Binet "IQ test" came into use in 1932, it was adjusted so that scores for each age group of children followed roughly the Normal distribution with mean $\mu = 100$ and standard deviation $\sigma = 15$. The test is readjusted from time to time to keep the mean at 100. If present-day American children took the 1932 Stanford-Binet test, their mean score would be about 120. The reasons for the increase in IQ over time are not known but probably include better childhood nutrition and more experience in taking tests.¹¹

- (a) IQ scores above 130 are often called "very superior." What percent of children had very superior scores in 1932?
- (b) If present-day children took the 1932 test, what percent would have very superior scores? (Assume that the standard deviation $\sigma = 15$ does not change.)

4. **Deciles.** The *deciles* of any distribution are the points that mark off the lowest 10% and the highest 10%. On a density curve, these are the points with areas 0.1 and 0.9 to their left under the curve.

- (a) What are the deciles of the standard Normal distribution? 65
- (b) The heights of young women are approximately Normal with mean ~~65~~ inches and standard deviation 2.7 inches. What are the deciles of this distribution?

II. Exercises for Confidence Intervals and Hypothesis Tests for a Population Mean μ (σ Known)

1. Polling women. A *New York Times* poll on women's issues interviewed 1025 women randomly selected from the United States, excluding Alaska and Hawaii. The poll found that 47% of the women said they do not get enough time for themselves.

(a) The poll announced a margin of error of ± 3 percentage points for 95% confidence in its conclusions. What is the 95% confidence interval for the percent of all adult women who think they do not get enough time for themselves?

(b) Explain to someone who knows no statistics why we can't just say that 47% of all adult women do not get enough time for themselves.

2. Surveying hotel managers. A study of the career paths of hotel general managers sent questionnaires to an SRS of 160 hotels belonging to major U.S. hotel chains. There were 114 responses. The average time these 114 general managers had spent with their current company was 11.78 years. Give a 99% confidence interval for the mean number of years general managers of major-chain hotels have spent with their current company. (Take it as known that the standard deviation of time with the company for all general managers is 3.2 years.)

3. Cellulose in hay. An agronomist examines the cellulose content of alfalfa hay. Suppose that the cellulose content in the population has standard deviation $\sigma = 8$ mg/g. A sample of 15 cuttings has mean cellulose content $\bar{x} = 145$ mg/g.

(a) Give a 90% confidence interval for the mean cellulose content in the population.

(b) A previous study claimed that the mean cellulose content was $\mu = 140$ mg/g, but the agronomist believes that the mean is higher than that figure. State H_0 and H_a and carry out a significance test to see if the new data support this belief.

(c) The statistical procedures used in (a) and (b) are valid when several assumptions are met. What are these assumptions?

4. Filling cola bottles. Bottles of a popular cola are supposed to contain 300 milliliters (ml) of cola. There is some variation from bottle to bottle because the filling machinery is not perfectly precise. The distribution of the contents is Normal with standard deviation $\sigma = 3$ ml. An inspector who suspects that the bottler is underfilling measures the contents of six bottles. The results are

299.4 297.7 301.0 298.9 300.2 297.0

Is this convincing evidence that the mean content of cola bottles is less than the advertised 300 ml?

(a) State the hypotheses that you will test.

(b) Calculate the test statistic.

(c) Find the P -value and state your conclusion.

5. **Anemia.** Hemoglobin is a protein in red blood cells that carries oxygen from the lungs to body tissues. People with less than 12 grams of hemoglobin per deciliter of blood (g/dl) are anemic. A public health official in Jordan suspects that the mean μ for all children in Jordan is less than 12. He measures a sample of 50 children. Suppose that the "simple conditions" hold: the 50 children are an SRS from all Jordanian children and the hemoglobin level in this population follows a Normal distribution with standard deviation $\sigma = 1.6$ g/dl.

(a) We seek evidence *against* the claim that $\mu = 12$. What is the sampling distribution of \bar{x} in many samples of size 50 if in fact $\mu = 12$? Make a sketch of the Normal curve for this distribution. (Sketch a Normal curve, then mark the axis using what you know about locating the mean and standard deviation on a Normal curve.)

(b) The sample mean was $\bar{x} = 11.3$. Mark this outcome on the sampling distribution. Also mark the outcome $\bar{x} = 11.8$ g/dl of a different study of 50 children in another country. Explain carefully from your sketch why one of these outcomes is good evidence that μ is lower than 12, and also why the other outcome is not good evidence for this conclusion.

6. **What's the P -value?** A test of the null hypothesis $H_0: \mu = 0$ gives test statistic $z = 1.8$.

(a) What is the P -value if the alternative is $H_a: \mu > 0$?

(b) What is the P -value if the alternative is $H_a: \mu < 0$?

(c) What is the P -value if the alternative is $H_a: \mu \neq 0$?

7. **P and significance.** The P -value for a significance test is 0.078.

(a) Do you reject the null hypothesis at level $\alpha = 0.05$? Explain your answer.

(b) Do you reject the null hypothesis at level $\alpha = 0.01$? Explain your answer.

III. Exercises for Thinking about Inference

1. **Confidence level and margin of error.** The National Assessment of Educational Progress (NAEP) gave its test of quantitative skills to a sample of 1077 women of ages 21 to 25 years. Their mean quantitative score was 275. Individual NAEP scores have a Normal distribution with standard deviation $\sigma = 60$.

(a) Give a 95% confidence interval for the mean score μ in the population of all young women.

(b) Give the 90% and 99% confidence intervals for μ .

(c) What are the margins of error for 90%, 95%, and 99% confidence? How does increasing the confidence level affect the margin of error of a confidence interval?

2. **Planning a sample.** A study of the career paths of hotel general managers sent questionnaires to an SRS of 160 hotels belonging to major U.S. hotel chains. There were 114 responses. The average time these 114 general managers had spent with their current company was 11.78 years. Give a 99% confidence interval for the mean number of years general managers of major-chain hotels have spent with their current company. (Take it as known that the standard deviation of time with the company for all general managers is 3.2 years.) How large a sample of the hotel managers would be needed to estimate the mean μ within ± 1 year with 99% confidence?

3. **Explaining significance.** A social psychologist reports: "In our sample, ethnocentrism was significantly higher ($P < 0.05$) among church attenders than among nonattenders." Explain what this means in language understandable to someone who knows no statistics. Do not use the word "significance" in your answer.

4. **Prayer in the schools?** A *New York Times*/CBS News poll asked the question "Do you favor an amendment to the Constitution that would permit organized prayer in public schools?" Sixty-six percent of the sample answered "Yes." The article describing the poll says that it "is based on telephone interviews conducted from Sept. 13 to Sept. 18 with 1,664 adults around the United States, excluding Alaska and Hawaii ... the telephone numbers were formed by random digits, thus permitting access to both listed and unlisted residential numbers." The article gives the margin of error as 3 percentage points. Opinion polls customarily announce margins of error for 95% confidence, so we are 95% confident that the percent of all adults who favor prayer in the schools lies in the interval $66\% \pm 3\%$.

The news article goes on to say: "The theoretical errors do not take into account a margin of additional error resulting from the various practical difficulties in taking any survey of public opinion." List some of the "practical difficulties" that may cause errors in addition to the $\pm 3\%$ margin of error. Pay particular attention to the news article's description of the sampling method.

IV. Exercises for Confidence Intervals and Hypothesis Tests for a Population Mean μ (σ Unknown)

1. **Measuring blood pressure.** A medical study finds that $\bar{x} = 114.9$ and $s = 9.3$ for the seated systolic blood pressure of the 27 members of one treatment group. What is the standard error of the mean?

2. **Measuring acculturation.** The Acculturation Rating Scale for Mexican Americans (ARSMA) measures the extent to which Mexican Americans have adopted Anglo/English culture. During the development of ARSMA, the test was given to a group of 17 Mexicans. Their scores, from a possible range of 1.00 to 5.00, had a symmetric distribution with $\bar{x} = 1.67$ and $s = 0.25$. Because low scores should indicate a Mexican cultural orientation, these results helped to establish the validity of the test.⁵⁷

- Give a 95% confidence interval for the mean ARSMA score of Mexicans.
- What assumptions does your confidence interval require? Which of these assumptions is most important in this case?

3. **Red wine is good for the heart.** Observational studies suggest that moderate use of alcohol reduces heart attacks, and that red wine may have special benefits. One reason may be that red wine contains polyphenols, substances that do good things to cholesterol in the blood and so may reduce the risk of heart attacks. In an experiment, healthy men were assigned at random to several groups. One group of 9 men drank half a bottle of red wine each day for two weeks. The level of polyphenols in their blood was measured before and after the two-week period. Here are the percent changes in level:

3.5 8.1 7.4 4.0 0.7 4.9 8.4 7.0 5.5

Make a stemplot of the data. It is difficult to assess Normality from just 9 observations. Give a 90% t confidence interval for the mean percent change in blood polyphenols among all healthy men if all drank this amount of red wine.

4. **DDT poisoning.** Poisoning by the pesticide DDT causes tremors and convulsions. In a study of DDT poisoning, researchers fed several rats a measured amount of DDT. They then made measurements on the rats' nervous systems that might explain how DDT poisoning causes tremors. One important variable was the "absolutely refractory period," the time required for a nerve to recover after a stimulus. This period varies Normally. Measurements on four rats gave the data below (in milliseconds):⁶¹

1.6 1.7 1.8 1.9

- Find the mean refractory period \bar{x} and the standard error of the mean.
- Give a 90% confidence interval for the mean absolutely refractory period for all rats of this strain when subjected to the same treatment.
- Suppose that the mean absolutely refractory period for unpoisoned rats is known to be 1.3 milliseconds. DDT poisoning should slow nerve recovery and so increase this period. Do the data give good evidence for this claim? State H_0 and H_a and do a t test. Between what levels from Table C does the P -value lie? What do you conclude from the test?

5. **Growing tomatoes.** An agricultural field trial compares the yield of two varieties of tomatoes for commercial use. The researchers divide in half each of 10 small plots of land in different locations and plant each tomato variety on one half of each plot. After harvest, they compare the yields in pounds per plant at each location. The 10 differences (Variety A - Variety B) give $\bar{x} = 0.34$ and $s = 0.83$. Is there convincing evidence that Variety A has the higher mean yield?

- (a) Describe in words what the parameter μ is in this setting.
- (b) Carry out the test.

6. **Mutual-fund performance.** Do "index funds" that simply buy and hold all the stocks in one of the stock market indexes, such as the Standard & Poor's 500 Index, perform better than actively managed mutual funds? Compare the percent total return (price change plus dividends) of a large actively managed fund with that of the Vanguard Index 500 fund for the 24 years from 1977 to 2000. Vanguard did better by an average of 2.83% per year, and the standard deviation of the 24 annual differences was 11.65%. Is there convincing evidence that the index fund does better?

- (a) Describe in words the parameter μ for this comparison.
- (b) State the hypotheses H_0 and H_a .
- (c) Find the matched pairs t statistic and its P -value. What do you conclude?

7. **ARSMA versus BI.** The ARSMA test (Exercise 17.7) was compared with a similar test, the Bicultural Inventory (BI), by administering both tests to 22 Mexican Americans. Both tests have the same range of scores (1.00 to 5.00) and are scaled to have similar means for the groups used to develop them. There was a high correlation between the two scores, giving evidence that both are measuring the same characteristics. The researchers wanted to know whether the population mean scores for the two tests are the same. The differences in scores (ARSMA - BI) for the 22 subjects had $\bar{x} = 0.2519$ and $s = 0.2767$.

- (a) Describe briefly how to arrange the administration of the two tests to the subjects, including randomization.
- (b) Carry out a significance test for the hypothesis that the two tests have the same population mean. Give the P -value and state your conclusion.
- (c) Give a 95% confidence interval for the difference between the two population mean scores.

V. Exercises for Confidence Intervals and Hypothesis Tests for Comparing Two Population Means μ_1 and μ_2 (σ_1 and σ_2 Unknown)

1. **Which data design?** Is each of these designs (1) single sample, (2) matched pairs, or (3) two independent samples? Explain your choices.

(a) An education researcher wants to learn whether it is more effective to put questions before or after introducing a new concept in an elementary school mathematics text. He prepares two text segments that teach the concept, one with motivating questions before and the other with review questions after. He uses each text segment to teach a separate group of children. The researcher compares the scores of the groups on a test over the material.

(b) Another researcher approaches the same issue differently. She prepares text segments on two unrelated topics. Each segment comes in two versions, one with questions before and the other with questions after. The subjects are a single group of children. Each child studies both topics, one (chosen at random) with questions before and the other with questions after. The researcher compares test scores for each child on the two topics to see which topic he or she learned better.

2. **Beetles in oats.** In a study of cereal leaf beetle damage on oats, researchers measured the number of beetle larvae per stem in small plots of oats after randomly applying one of two treatments: no pesticide or malathion at the rate of 0.25 pound per acre. The data appear roughly Normal. Here are the summary statistics:⁶⁶

Group	Treatment	n	\bar{x}	s
1	Control	13	3.47	1.21
2	Malathion	14	1.36	0.52

Is there significant evidence at the 1% level that malathion reduces the mean number of larvae per stem?

3. **Social insight among men and women.** The Chapin Social Insight Test is a psychological test designed to measure how accurately a person appraises other people. The possible scores on the test range from 0 to 41. During the development of the Chapin test, it was given to several different groups of people. Here are the results for male and female college students majoring in the liberal arts:⁶⁹

Group	Sex	n	\bar{x}	s
1	Male	133	25.34	5.05
2	Female	162	24.94	5.44

Do these data support the contention that female and male students differ in average social insight?

4. **Market research.** A market research firm supplies manufacturers with estimates of the retail sales of their products from samples of retail stores. Marketing managers are prone to look at the estimate and ignore sampling error. An SRS of 75 stores this month shows mean sales of 52 units of a small appliance, with standard deviation 13 units. During the same month last year, an SRS of 53 stores gave mean sales of 49 units, with standard deviation 11 units. An increase from 49 to 52 is a rise of 6%. The marketing manager is happy, because sales are up 6%.

(a) Use the two-sample t procedure to give a 95% confidence interval for the difference between this year and last year in the mean number of units sold at all retail stores.

(b) Explain in language that the manager can understand why he cannot be confident that sales rose by 6%, and that in fact sales may even have dropped.

5. **Depressed teens.** To study depression among adolescents, investigators administered the Children's Depression Inventory (CDI) to teenagers in rural Newfoundland, Canada. As is often the case in social science studies, there is some question about whether the subjects can be considered a random sample from an interesting population. We will ignore this issue. One finding was that "older adolescents scored significantly higher on the CDI." Higher scores indicate symptoms of depression. Here are summary data for two grades:⁷⁴

Group	n	\bar{x}	s
Grade 9	84	6.94	6.03
Grade 11	70	8.98	7.08

Do an analysis to verify the quoted conclusion.

VI. Exercises for Confidence Intervals and Hypothesis Tests for a Population Proportion p

1. **More online publishing.** The previous exercise describes a survey of authors of papers in a medical journal. Another question in the survey asked whether authors would accept a stronger move toward online publishing: "As an author, how acceptable would it be for us to publish only the abstract of papers in the paper journal and continue to put the full long version on our website?" Of the 104 authors in the sample, 65 said "Not at all acceptable." What proportion of all authors feel that abstract-only publishing is not acceptable? (Estimate with 95% confidence.) Do the data provide good evidence that more than half of all authors feel that abstract-only publishing is not acceptable? Answer both questions with appropriate inference methods.

2. **Unhappy HMO patients.** How likely are patients who file complaints with a health maintenance organization (HMO) to leave the HMO? In one recent year, 639 of the more than 400,000 members of a large New England HMO filed complaints. Fifty-four of the complainers left the HMO voluntarily. (That is, they were not forced to leave by a move or a job change.)⁸¹ Consider this year's complainers as an SRS of all patients who will complain in the future. Give a 90% confidence interval for the proportion of complainers who voluntarily leave the HMO. Can you use the large-sample method?

3. **School vouchers.** A national opinion poll found that 44% of all American adults agree that parents should be given vouchers good for education at any public or private school of their choice. The result was based on a small sample. How large an SRS is required to obtain a margin of error of 0.03 (that is, $\pm 3\%$) in a 95% confidence interval?

- (a) Answer this question using the previous poll's result as the guessed value p^* .
- (b) Do the problem again using the conservative guess $p^* = 0.5$. By how much do the two sample sizes differ?

4. **Attitudes toward nuclear power.** A Gallup Poll on energy use asked 512 randomly selected adults if they favored "increasing the use of nuclear power as a major source of energy." Gallup reported that 225 said "Yes." Does this poll give good evidence that fewer than half of all adults favor increased use of nuclear power?

5. **Do chemists have more girls?** Some people think that chemists are more likely than other parents to have female children. (Perhaps chemists are exposed to something in their laboratories that affects the sex of their children.) The Washington State Department of Health lists the parents' occupations on birth certificates. Between 1980 and 1990, 555 children were born to fathers who were chemists. Of these births, 273 were girls. During this period, 48.8% of all births in Washington State were girls.⁸⁵ Is there evidence that the proportion of girls born to chemists is higher than the state proportion?

6. **We want to be rich.** In a recent year, 73% of first-year college students responding to a national survey identified "being very well-off financially" as an important personal goal. A state university finds that 132 of an SRS of 200 of its first-year students say that this goal is important.

(a) Give a 95% confidence interval for the proportion of all first-year students at the university who would identify being well-off as an important personal goal.

(b) Is there good evidence that the proportion of first-year students at this university who think being very well-off is important differs from the national value, 73%? (Be sure to state hypotheses, give the P -value, and state your conclusion.)

(c) Check that you can safely use the methods of this chapter in both (a) and (b).

VII. Exercises for Chi-square Tests

1. **Treating cocaine addiction.** Cocaine addicts need the drug to feel pleasure. Perhaps giving them a medication that fights depression will help them stay off cocaine. A three-year study compared an antidepressant called desipramine with lithium (a standard treatment for cocaine addiction) and a placebo. The subjects were 72 chronic users of cocaine who wanted to break their drug habit. Twenty-four of the subjects were randomly assigned to each treatment. Here are the counts and proportions of the subjects who avoided relapse into cocaine use during the study:¹⁰⁶

Group	Treatment	Subjects	No relapse	Proportion
1	Desipramine	24	14	0.583
2	Lithium	24	6	0.250
3	Placebo	24	4	0.167

Does data analysis suggest that desipramine is more successful than the other two treatments? Are there significant differences among the outcomes for the treatments?

2. **Ebonics awareness.** Ebonics, often called "black English," is a variety of English common among blacks in the United States. How aware are college students of the existence of Ebonics? Here are data from a sample of students at a racially diverse college in the South:¹¹⁰

	Aware	Not aware
Black students	121	11
White students	159	21
Other students	75	28

Do both data analysis and a formal test to compare awareness in the three groups of students. Write a clear summary of your findings.

MA 116 Elements of Statistics
Solutions to Selected Review Exercises for Final Exam
 Spring 2010
 Organized by Dr. Paul Duty (MA 116 Course Chair)

I. Solutions to Exercises for the Normal Distribution

1. **Three great hitters.** The standardized values are found by subtracting the mean and then dividing by the standard deviation. For example, the standardized score for Cobbs is $z_{cobb} = \frac{.420 - .266}{.0371} = 4.15$. Based on the standardized scores, all three hitters were over 4 standard deviations larger than their respective peers.

	Decade	Hitter	Hitters Batting	Mean	Std. dev.	Std Score
1	1910s	Cobb	0.420	0.266	0.0371	4.15
2	1940s	Williams	0.406	0.267	0.0326	4.26
3	1970s	Bretts	0.390	0.261	0.0317	4.07

2. **How hard do locomotives pull?** Let A be the adhesive power. Then A has a $N(\mu = .37, \sigma = .04)$ distribution. (a) $P(A > .40) = P\left(\frac{A - .37}{.04} > \frac{.40 - .37}{.04}\right) = P(Z > .75) = .2266$. (b) $P(.40 < A < .50) = P\left(\frac{.40 - .37}{.04} < \frac{A - .37}{.04} < \frac{.50 - .37}{.04}\right) = P(.75 < Z < 3.25) = .9994 - .7734 = .2260$. This almost the same as (a) because the value of .50 is so high above the mean.

3. **Are we getting smarter?** (a) Let IQ be the IQ power. Then IQ has a $N(\mu = 100, \sigma = 15)$ distribution. $P(IQ > 130) = P\left(\frac{IQ - 100}{15} > \frac{130 - 100}{15}\right) = P(Z > 2) = .0227 = 2.27\%$ (b) Let IQ be the IQ power. Then IQ has a $N(\mu = 120, \sigma = 15)$ distribution. $P(IQ > 130) = P\left(\frac{IQ - 120}{15} > \frac{130 - 120}{15}\right) = P(Z > .67) = .2524 = 25.24\%$

4. **Deciles.** (a) The .1 and .9 deciles are -1.28 and 1.28 respectively for a standard normal distribution. (b) The .1 and .9 deciles are found as $65 \pm 1.28 \times 2.7$ or $(60.53, 67.46)$ inches respectively.

II. Solutions to Exercises for Confidence Intervals and Hypothesis Tests for a Population Mean (σ Known)

1. **Polling women.** (a) The 95% c.i. is $.47 \pm .03 = (.44, .50) = (44\%, 50\%)$.
 (b) Every poll gets a different result because it based on a sample taken from the population.

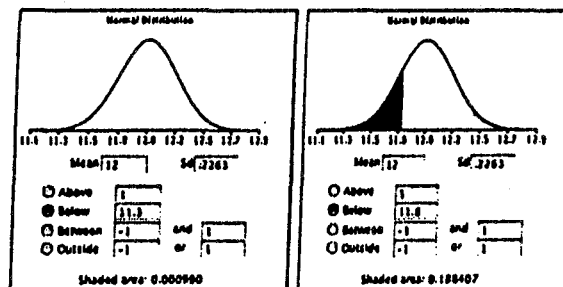
2. **Surveying hotel managers.** For a 99% c.i. Table C gives $z^* = 2.576$. The 99% c.i. is then: $\bar{X} \pm z^* \frac{\sigma}{\sqrt{n}} = 11.78 \pm 2.576 \frac{3.2}{\sqrt{114}} = (11.008, 12.552)$

3. **Cellulose in hay.** (a) For a 90% c.i., Table C gives $z^* = 1.645$. The 90% c.i. is: $\bar{X} \pm z^* \frac{\sigma}{\sqrt{n}} = 145 \pm 1.645 \frac{8}{\sqrt{15}} = (141.60, 148.40)$. (b) $H_0: \mu = 140$; $H_1: \mu > 140$. This is a one-sided test. The test-statistics is computed as $z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{145 - 140}{8/\sqrt{15}} = 2.42$. $P = P[Z > 2.42] = .0078$. There is evidence against the hypothesis that the mean is 140. (c) It assumes an SRS from a normal population with the correct value of σ .

4. **Filling cola bottles.** (a) $H_0: \mu = 300$; $H_1: \mu < 300$. This is a one-sided alternative because interest lies in seeing if the mean amount filled is less than 300 mL. This is a one-sided test. (b) The test-statistics is computed as $z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{299.033 - 300}{.21483/\sqrt{6}} = -7.983$. (c) $P = P[Z < -7.983] = .00000$. There is no evidence against the hypothesis that the mean amount filled is 300 mL, i.e. there is no evidence that the mean studying time is less than 300 mL.

z Test	
Hypothesized Value	300
Actual Estimate	299.033
df	5
Std Dev	1.50289
Sigma given	3
Test Statistic	-0.7893
Prob > z	0.4299
Prob > z	0.7850
Prob < z	0.2150

5. **Anemia.** (a) $\bar{X} \sim N(\mu_{\bar{X}} = 12, \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.6}{\sqrt{50}} = .2263)$. (b) Refer to sketches. The value of $\bar{X} = 11.3$ is much more convincing because the area to the left of this value is much smaller than the area to the left of $\bar{X} = 11.8$.



6. **What's the P-value?** (a) $P = P[Z > 1.8] = .0359$. (b) $P = P[Z < 1.8] = .9641$. (c) $P = 2P[Z > 1.8] = .0718$.

7. **P and significance.** (a) No, because the P-value is greater than 5%. (b) No, because the P-value is greater than 1%.

III. Solutions to Exercises for Thinking about Inference

1. **Confidence level and margin of error.** (a) (271.4, 278.6). (b) (272.0, 278.0), (270.3, 279.7). (c) ± 3 , ± 3.6 , ± 4.17 . A larger confidence level leads to a larger margin of error.

2. **Planning a sample.** We need to find n so that $z^* \frac{\sigma}{\sqrt{n}} = 2.576 \frac{3.2}{\sqrt{n}} = 1$. This gives $n = 68$.

3. **Explaining significance.** The difference in mean ethocentrism between church attenders and non-attenders was larger than could reasonably be expected by chance if the means were equal.

4. **Prayer in the schools?** Unable to contact people without phones; non-response; refusal to respond; etc.

IV. Solutions to Exercises for Confidence Intervals and Hypothesis Tests for a Population Mean (σ Unknown)

1. Measuring blood pressure. $se_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{9.3}{\sqrt{27}} = 5.2$.

2. Measuring acculturation. (a) $\bar{X} \pm t_{n-1} \frac{s}{\sqrt{n}} = 1.67 \pm 2.120 \frac{.25}{\sqrt{17}} = (1.54, 1.80)$.
 (b) SRS, normal distribution in small samples. The SRS is most important.

3. Red wine is good for the heart. See stem plot.
 $\bar{X} \pm t_{n-1} \frac{s}{\sqrt{n}} = 15.5 \pm 1.860 \frac{2.52}{\sqrt{9}} = (3.94, 7.6)$.

Stem	Leaf	Count
8	14	2
7		
7	04	2
6		
6		
5	5	1
5		
4	9	1
4	0	1
3	5	1
3		
2		
2		
1		
1		
0	7	1

0|7 represents 0.7

4. DDT poisoning. (a) $\bar{X} = 1.75$, $SEM = \frac{s}{\sqrt{n}} = \frac{.129}{\sqrt{4}} = .065$ ms. (b) $\bar{X} \pm t_{n-1} \frac{s}{\sqrt{n}} = 1.75 \pm 2.353 \frac{.129}{\sqrt{4}} = (1.69, 1.90)$ ms. (c) $H_0: \mu = 1.3$; $H_1: \mu > 1.3$. This is a one-sided alternative because interest lies in seeing if the mean recovery time has increased from 1.3 ms. This is a one-sided test. The test-statistic is computed as $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{1.75 - 1.3}{.129/\sqrt{4}} = 6.9714$. $P = P[t > 6.9714] = .0030$. There is strong evidence against the hypothesis that the mean is 1.3, i.e. there is strong evidence that the mean recovery time is greater than 1.3 ms. If you use Table C with 3 df, $.0025 < P < .005$.

Moments	
Mean	1.75
Std Dev	0.1290994
Std Err Mean	0.0645497

Test Mean Value	
Hypothesized Value	1.3
Actual Estimate	1.75
df	3
Std Dev	0.1291

t Test	
Test Statistic	6.9714
Prob > t	0.0061*
Prob > t	0.0030*
Prob < t	0.9970

5. **Growing tomatoes.** (a) μ is the DIFFERENCE in the mean yield between the two varieties. (b) $H_0: \mu = 0; H_1: \mu > 0$. This is a one-sided alternative because interest lies in seeing if the mean yield of Variety A is larger than the mean yield of Variety B. This is a one-sided test. The test-statistic is computed as $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{.34 - 0}{.83/\sqrt{10}} = 1.30$ with 9 *df*. Note that n is the number of plots. From Table C, $.10 < P < .15$. There is no evidence that the mean yield of Variety A is larger than the mean yield of Variety B.

6. **Mutual-fund performance.** (a) μ is the difference in the mean view of the Vanguard Index fund and the actively managed fund. (b) $H_0: \mu = 0; H_1: \mu > 0$. This is a one-sided alternative because interest lies in seeing if the mean yield of the Index fund is larger than the mean yield of the managed fund. This is a one-sided test. The test-statistic is computed as $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{2.83 - 0}{11.65/\sqrt{24}} = 1.19$ with 23 *df*. Note that n is the number of years. From Table C, $.10 < P < .15$. There is no evidence that the Index fund does better on average than the managed fund.

7. **ARSMA versus BI.** (a) Each subject should write both tests with half of the subjects writing each test first. (b) $H_0: \mu = 0; H_1: \mu \neq 0$. This is a two-sided alternative because interest lies in seeing if the mean score on both tests are the same or different. This is a two-sided test. The test-statistic is computed as $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{.2519 - 0}{.2767/\sqrt{22}} = 4.27$ with 21 *df*. Note that n is the number of people. From Table C, $P < .001$. There is strong evidence that the mean score on both tests is not the same. (c) $\bar{X} \pm t_{n-1} \frac{s}{\sqrt{n}} = .2519 \pm 2.080 \frac{.2767}{\sqrt{22}} = (.129, .375)$.

V. Solutions to Exercises for Confidence Intervals and Hypothesis Tests for Comparing Two Population Means μ_1 and μ_2 (σ_1 and σ_2 Unknown)

1. Which data design? (a) Two independent samples. Each child receives one of the treatments. (b) Matched pair design. Each child is measured twice.

2. Beetles in oats. $H_0: \mu_1 = \mu_2; H_1: \mu_1 > \mu_2$. $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{3.47 - 1.36}{\sqrt{\frac{1.21^2}{13} + \frac{.52^2}{14}}} = 5.81$

with $\min(n_1 - 1, n_2 - 1) = \min(14 - 1, 13 - 1) = 12$ *df*. From Table C, $P < .0005$. Yes the result is statistically significant at the 1% level.

3. Social insight among men and women. $H_0: \mu_1 = \mu_2; H_1: \mu_1 \neq \mu_2$.
 $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{25.34 - 24.94}{\sqrt{\frac{5.05^2}{133} + \frac{5.44^2}{162}}} = .65$ with $\min(n_1 - 1, n_2 - 1) = \min(133 - 1, 162 - 1) = 132$ *df*. From Table C, $P > .50$. There is no evidence that mean test score differs between males and females.

4. Market research. (a) $\bar{X}_1 - \bar{X}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 52 - 49 \pm 2.009 \sqrt{\frac{13^2}{75} + \frac{11^2}{53}} = (-1.3, 7.3)$ change in mean sales. (b) The c.i. for the difference in mean sales includes 0, or no change in the mean.

5. Depressed teens. $H_0: \mu_9 = \mu_{11}; H_1: \mu_9 < \mu_{11}$. $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{6.94 - 8.98}{\sqrt{\frac{6.03^2}{84} + \frac{7.03^2}{70}}} = -1.91$ with $\min(n_1 - 1, n_2 - 1) = \min(84 - 1, 70 - 1) = 69$ *df*. From Table C, $.025 < P < .05$. There is evidence that the mean score for older adolescents is higher.

VI. Solutions to Exercises for Confidence Intervals and Hypothesis Tests for a Population Proportion p

1. **More online publishing.** Both n and $n - x$ are ≥ 15 so the large sample methods can be used. $\hat{p} = .625$. The 95% c.i. is $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .625 \pm 1.96 \sqrt{\frac{.625(1-.625)}{104}} = (.532, .718)$. $H_0: p = .5; H_1: p > .5$. The test statistic is $z = \frac{\hat{p}-p_0}{p_0(1-p_0)/n} = \frac{.625-.5}{.5(1-.5)/104} = 2.55$. $P = P[Z > 2.55] = .0054$. There is strong evidence that the support is greater than .5.

2. **Unhappy HMO patients.** The large sample c.i. can be used because both n and $n - x$ are ≥ 15 . $\hat{p} = 54/639 = .0845$. The 90% c.i. is $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .0845 \pm 1.645 \sqrt{\frac{.0845(1-.0845)}{639}} = (.066, .102)$.

3. **School vouchers.** (a) $n = \left(\frac{z^*}{me}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{.03}\right)^2 .44(1-.44) = 1052$.
 (b) $n = \left(\frac{z^*}{me}\right)^2 p^*(1-p^*) = \left(\frac{1.96}{.03}\right)^2 .50(1-.50) = 1068$.

4. **Attitudes toward nuclear power.** $\hat{p} = 225/512 = .44$. $H_0: p = .5; H_1: p < .5$. The test statistic is $z = \frac{\hat{p}-p_0}{p_0(1-p_0)/n} = \frac{.44-.5}{.5(1-.5)/512} = -2.74$. $P = P[Z < -2.74] \approx .0031$. There is strong evidence that the proportion of people who favour the increased use of nuclear power is less than 50%.

5. **Do chemists have more girls?** $\hat{p} = 273/555 = .49$. $H_0: p = .488; H_1: p > .488$. The test statistic is $z = \frac{\hat{p}-p_0}{p_0(1-p_0)/n} = \frac{.49-.488}{.488(1-.488)/555} = .18$. $P = P[Z > .18] \approx .43$. There is no evidence that the proportion of females born to chemists is higher than expected.

6. **We want to be rich.** (a) $\hat{p} = 132/200 = .66$. The 95% c.i. is $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .66 \pm 1.96 \sqrt{\frac{.66(1-.66)}{200}} = (.59, .73)$. (b) $H_0: p = .73; H_1: p \neq .73$. The test statistic is $z = \frac{\hat{p}-p_0}{p_0(1-p_0)/n} = \frac{.66-.73}{.73(1-.73)/200} = -2.23$. $P = 2P[Z > 2.23] = .026$. There is some evidence the proportion of students who thing being well off is important is different than 73%. (c) Both x and $n - x$ are ≥ 15 so the large sample methods are appropriate.

VII. Solutions to Exercises for Chi-square Tests

1. Treating cocaine addiction. H_0 : no relationship between status (relapse/no relapse) and treatment. $\chi^2 = 10.5$, $P = .0052$. There is strong evidence of a difference among treatments in the relapse rate.

Treatment	Count	Status		Row %
		No relapse	Relapse	
Desipramine	14	10	24	58.33
Lithium	6	18	24	25.00
Placebo	4	20	24	16.67
	24	48	72	

	N	DF	-LogLike	RSquare (U)
	72	2	5.2188692	0.1139

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	10.438	0.0054*
Pearson	10.500	0.0052*

2. No title given. H_0 : Awareness of Ebonics does not depend upon race of student. $\chi^2 = 18.6$. $P < .0001$. There is strong evidence that awareness of Ebonics varies among the races.

Student	Count	Aware		Row %
		No	Yes	
Black	11	121	132	8.33
Other	28	75	103	27.18
White	21	159	180	11.67
	60	355	415	

	N	DF	-LogLike	RSquare (U)
	415	2	8.5055023	0.0496

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	17.011	0.0002*
Pearson	18.626	<.0001*