AP Statistics
Solutions to Packet 10

Introduction to Inference
Estimating with Confidence
Tests of Significance
Making Statistical Sense of Significance
Inference as a Decision
10.1 **POLLING WOMEN** A *New York Times* poll on women’s issues interviewed 1025 women randomly selected from the US, 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? 44% to 50%

(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.

We do not have information about the whole population; we only know about a small sample. We expect our sample to give us a good estimate of the population value, but it will not be exactly correct.

(c) Then explain clearly what “95% confidence” means.

The procedure used gives an estimate within 3 percentage points of the true value in 95% of all samples.

10.2 **NAEP SCORES** Young people have a better chance of full-time employment and good wages if they are good with numbers. How strong are the quantitative skills of young Americans of working age? One source of data is the National Assessment of Educational Progress (NAEP) Young Adult Literacy Assessment Survey, which is based on a nationwide probability sample of households. The NAEP survey includes a short test of quantitative skills, covering mainly basic arithmetic and the ability to apply it to realistic problems. Scores on the test range from 0 to 500. For example, a person who scores 233 can add the amounts of two checks appearing on a bank deposit slip; someone scoring 325 can determine the price of a meal from a menu; a person scoring 375 can transform a price in cents per ounce into dollars per pound.

Suppose that you give the NAEP test to a SRS of 840 people from a large population in which the scores have mean 280 and standard deviation \( \sigma = 60 \). The mean \( \bar{x} \) of the 840 scores will vary if you take repeated samples.

(a) Describe the shape, center, and spread of the sampling distribution of \( \bar{x} \). What guarantees this?

\[ N \left( 280, \frac{60}{\sqrt{840}} \right) \]

- 840 \( < 10\% \) population of young Americans of working age \( \Rightarrow \) standard deviation formula works
- We can assume the sampling distribution is approximately normal because \( n = 840 \) is large enough that we can use CLT
(b) Sketch the normal curve that describes how $\bar{x}$ varies in many samples from this population. Mark its mean and the values 1, 2, and 3 standard deviations on either side of the mean.

(c) According to the 68-95-99.7% rule, about 95% of all the values of $\bar{x}$ fall within $\pm 2\sigma = 4.2$ of the mean of the curve. What is the missing number? Call it $m$ for “margin of error”. Sketch the region from the mean minus $m$ to the mean plus $m$ on the axis of your sketch.

(d) Whenever $\bar{x}$ falls in the region you shaded, the true value of the population mean, $\mu = 280$, lies in the confidence interval between $\bar{x} - m$ and $\bar{x} + m$. Draw the confidence interval below your sketch for one value of $\bar{x}$ inside the shaded region and one value of $\bar{x}$ outside the shaded region.

(e) In what percent of all samples will the true mean $\mu = 280$ be covered the confidence interval $\bar{x} \pm m$? 95%

10.3 **EXPLAINING CONFIDENCE**  A student reads that a 95% confidence interval for the mean NAEP quantitative score for men of ages 21 to 25 is 267.8 to 276.2. Asked to explain the meaning of this interval, the student says, “95% of all young men have scores between 267.8 and 276.2.” Explain why this student is wrong.

- Individual scores may vary widely
- This is a statement about the mean NAEP scores for all young men
- We are only attempting to estimate the center of the population distribution
- 95% is not a probability, it is a confidence level
10.5 ANALYZING PHARMACEUTICALS  A manufacturer of pharmaceutical products analyzes a specimen from each batch of a product to verify the concentration of active ingredient. The chemical analysis is not perfectly precise. Repeated measurements on the same specimen give slightly different results. The results of repeated measurements follow a normal distribution quite closely. The analysis procedure has no bias, so the mean $\mu$ of all measurements is the true concentration in the specimen. The standard deviation of this distribution is known to be $\sigma = 0.0068$ grams per liter. The laboratory analyzes each specimen three times and reports the mean result. Three analyses of one specimen give concentrations \{0.8403, 0.8363, 0.8447\}.

Construct a 99% confidence interval for the true concentration $\mu$

**P**  – name the parameter and the population
$\mu$ = true mean of measurements of concentration of active ingredient in the specimen

**A**  – verify all assumptions
SRS

$n < 10\%$ of population
parent population is normal

**N**  – name the interval

**I**  – calculate the interval (Show your work)

$$
\bar{x} \pm z \frac{\sigma}{\sqrt{n}} = 0.84 \pm 2.576 \left( \frac{0.0068}{\sqrt{3}} \right) = (0.830, 0.850) \text{ grams/liter}
$$

**C**  – state the conclusion in context.

We are 99% confident that the true mean of measurements of concentration of active ingredient in the specimen is between 0.83 and 0.85 grams/liter.
10.6 **SURVEYING HOTEL MANAGERS** A study of the career paths of hotel general managers sent questionnaires to a 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. Find a 99% confidence interval for the mean number of years general managers of major-hotel chains 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.)

\[ P \] name the parameter and the population
\[ \mu = \text{true mean number of years general managers of major hotel chains have spent with their current company} \]

\[ A \] verify all assumptions

SRS
114 < 10% of population
\[ n = 114 > 30, \text{ so we can assume normality by CLT} \]

\[ N \] name the interval

\[ z \]-interval

\[ I \] calculate the interval (Show your work)

\[ \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 11.78 \pm 2.576 \left( \frac{3.2}{\sqrt{114}} \right) = (11.008, 12.552) \text{ years} \]

\[ C \] state the conclusion in context.
We are 99% confident that the true mean number of years general managers of major-hotel chains have spent time with their current companies is between 11.01 and 12.55 years.
Here are measurements (in millimeters) of a critical dimension on a sample of auto engine crankshafts:

\[
\begin{align*}
224.120 & \quad 224.001 & \quad 224.017 & \quad 223.982 & \quad 223.989 & \quad 223.961 & \quad 223.960 & \quad 224.089 \\
223.987 & \quad 223.902 & \quad 223.980 & \quad 224.098 & \quad 224.057 & \quad 223.913 & \quad 223.999 & \quad 223.976
\end{align*}
\]

The data come from a production process that is known to have standard deviation \( \sigma = 0.060 \text{mm} \). The process mean is supposed to be \( \mu = 224 \text{ mm} \) but can drift away from this target during production.

(a) Do we have enough observations to be able to assume normality based on the Central Limit Theorem? NO Make a histogram of the data on your calculator, sketch below, and describe the shape of the distribution. Make a normal probability plot of the data and sketch below. What does this plot indicate about the shape of the distribution? not too far from normal

In general how can we use normal probability plots to determine if the sample is approximately normal? If NPP is close to linear \( \Rightarrow \) data is close to normal

(b) Use your calculator to find the 95% confidence interval for the process mean at the time these crankshafts were produced. (YOU DO NOT HAVE TO PANIC.)

\( (223.97, 224.03) \text{ mm} \)
10.8  **CORN YIELD**  Corn researchers plant 15 plots with a new variety of corn. The yields in bushels per acre are

<table>
<thead>
<tr>
<th>Yield (bushels per acre)</th>
</tr>
</thead>
<tbody>
<tr>
<td>128.0</td>
</tr>
<tr>
<td>139.1</td>
</tr>
<tr>
<td>113.0</td>
</tr>
<tr>
<td>132.5</td>
</tr>
<tr>
<td>140.7</td>
</tr>
<tr>
<td>109.7</td>
</tr>
<tr>
<td>118.9</td>
</tr>
<tr>
<td>134.8</td>
</tr>
<tr>
<td>109.6</td>
</tr>
<tr>
<td>127.3</td>
</tr>
<tr>
<td>115.6</td>
</tr>
<tr>
<td>130.4</td>
</tr>
<tr>
<td>130.2</td>
</tr>
<tr>
<td>111.7</td>
</tr>
<tr>
<td>105.5</td>
</tr>
</tbody>
</table>

Assume that \( \sigma = 10 \) bushels per acre.

(a) Find the 90% confidence interval for the mean yield \( \mu \) for this variety of corn.

**P** – name the parameter and the population

\[ \mu = \text{true mean yield of corn in bushels per acre} \]

**A** – verify all assumptions

SRS – not given, proceed with caution

15 < 10% of population

15 is too small to use CLT, look at NPP of data

fairly linear, so assume sample is close to normal

**N** – name the interval

\( \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \)

\[ \bar{x} \pm z^* \frac{10}{\sqrt{15}} = 123.1 \pm 1.645 \left( \frac{10}{\sqrt{15}} \right) = (118.89, 127.38) \text{ bushels per acre.} \]

**C** – state the conclusion in context.

We are 90% confident that the true mean yield for this variety of corn is between 118.9 and 127.4 bushels per acre.

(b) Use your calculator to find the 95% confidence interval for the mean yield \( \mu \) for this variety of corn. What is the critical value \( z^* \)? 1.96

\( (118.7, 128.9) \text{ bu/acre} \)

(c) Use your calculator to find the 99% confidence interval for the mean yield \( \mu \) for this variety of corn. What is the critical value \( z^* \)? 2.576

\( (117.1, 130.5) \text{ bu/acre} \)

(d) How do the margins of error in (a), (b), and (c) change as the confidence level increases.

As the confidence level increases, the margin of error gets larger
10.9  **MORE CORN**  Suppose that the crop researchers in Exercise 10.8 obtained the same value of \( \bar{x} \) from a sample of 60 plots rather than 15.

(a)  Use your calculator to find the 95% confidence interval for the mean yield \( \mu \).

What is the critical value \( z^* \)? 1.96

\((120.6, 125.66)\)

(b) What is the margin of error?  \(1.96 \cdot \frac{10}{\sqrt{60}} = 2.53\)  Is the margin of error larger or smaller than the margin of error found for the sample of 15 plots in Exercise 10.8?  smaller  Explain in plain language why the change occurs.

With a larger sample, there is more information \( \Rightarrow \) there is less uncertainty about the value of \( \mu \).

(c) Will the 90% and 99% confidence intervals for a sample of size 60 be wider or narrower than those for \( n = 15 \)?  narrower  (You need not actually calculate these intervals.)

10.10  **CONFIDENCE LEVEL AND INTERVAL LENGTH**  Examples 10.5 and 10.6 give confidence intervals for the screen tension \( \mu \) based on 20 measurements with \( \bar{x} = 306.3 \) and \( \sigma = 43 \).  The 99% confidence interval is 281.6 to 331.1  and the 90% confidence interval is 290.5 to 322.1.

(a)  Use your calculator to find the 80% confidence interval for \( \mu \).

What is the critical value \( z^* \)?  1.282

\((293.98, 318.62)\)

(b)  Use your calculator to find the 99.9% confidence interval for \( \mu \).

What is the critical value \( z^* \)?  3.291

\((274.66, 337.94)\)

(c) Make a sketch to compare all four intervals.  How does increasing the confidence level affect the length of the confidence interval?  **Increasing confidence \( \Rightarrow \) increases width of interval**
10.12  **A BALANCED SCALE?**  To assess the accuracy of a laboratory scale, a standard weight known to weigh 10 grams is weighed repeatedly. The scale readings are normally distributed with unknown mean (this mean is 10 grams if the scale has no bias). The standard deviation of the scale readings is known to be 0.0002 gram.

(a) The weight is weighed 5 times. The mean result is 10.0023 grams. Give a 98% CI for the mean of repeated measurements of the weight.

\[ \mu = \text{true mean scale reading for 10 gram weight} \]

**A** – verify all assumptions

- SRS – not given, so proceed with caution
- 5 < 10% of all possible readings
- parent population is normally distributed

**N** – name the interval

- z-interval

\[ \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} = 10.0023 \pm 2.326 \left( \frac{0.0002}{\sqrt{5}} \right) = (10.0021, 10.0025) \text{ grams.} \]

\[ \bar{x} \pm z^* \frac{\sigma}{\sqrt{n}} \]

**C** – state the conclusion in context.

We are 98% confident that the true mean scale reading for 10 gram weight is between 10.0021 and 10.0025.

(b) How many measurements must be averaged to get a margin of error of ± 0.0001 with 98% confidence?  **SHOW YOUR WORK!!!**

\[ z^* \frac{\sigma}{\sqrt{n}} < 0.0001 \]

\[ 2.326 \left( \frac{0.0002}{\sqrt{n}} \right) < 0.0001 \]

\[ \left( \frac{2.326(0.0002)}{0.0001} \right)^2 < n \]

\[ n > 22 \]
10.13 **SURVEYING HOTEL MANAGERS, II** In an earlier problem we looked at surveys of hotel managers. The average time these 114 general managers had spent with their current company was 11.78 years \((\mu = 3.2\text{ years})\) How large a sample of the hotel managers would be needed to estimate the mean \(\mu\) within \(\pm 1\) year with 99% confidence? SHOW YOUR WORK!!!

\[
z^* \frac{\sigma}{\sqrt{n}} < 1
\]

\[
2.576 \frac{(3.2)}{\sqrt{n}} < 1
\]

\[
(2.576 (3.2))^2 < n
\]

\[n > 68\]

10.14 **ENGINE CRANKSHAFTS, II** In an earlier problem we looked at measurements of a critical dimension of auto engine crankshafts. The process mean is supposed to be \(\mu = 224\text{ mm}\) and \(\sigma = 0.060\text{ mm}\). How large a sample of crankshafts would be needed to estimate the mean \(\mu\) within \(\pm 0.020\text{ mm}\) with 95% confidence? SHOW YOUR WORK!!!

\[
z^* \frac{\sigma}{\sqrt{n}} < 0.020
\]

\[
1.96 \frac{(0.060)}{\sqrt{n}} < 0.020
\]

\[
\left(1.96 (0.060)\right)^2 < n
\]

\[n > 35\]

10.15 **A TALK SHOW OPINION POLL** A radio talk show invites listeners to enter a dispute about a proposed pay increase for city council members. “What yearly pay do you think council members should get? Call us with your number.” In all, 958 people call. The mean pay they suggest is \(\bar{x} = 8740\text{ per year}\), and the standard deviation of the responses \(s = 1125\). For a large sample such as this, \(s\) is very close to the unknown population \(\sigma\). The station calculates the 95% CI for the mean pay \(\mu\) that all citizens would propose for council members to be $8669 to $8811.

(a) Is the station’s calculation correct?
   Yes, the computations are correct.

(b) Does their conclusion describe the population of all the city’s citizens? Explain your answer.
   No, these numbers are based on a voluntary response sample, not a SRS \(\Rightarrow\) the interval does not apply to the whole population.
10.16 **THE 2000 PRESIDENTIAL ELECTION** A closely contested presidential election pitted George W. Bush against Al Gore in 2000. A poll taken immediately before the 2000 election showed that 50% of the sample intended to vote for Gore. The polling organization announced that they were 95% confident that the sample result was with ±2 points of the true percent of all voters who favored Gore.

(a) Explain in plain language to someone who knows no statistics what “95% confident” means in this announcement.

The interval was based on a method that gives correct results 95% of the time.

(b) The poll showed Gore leading. Yet the polling organization said the election was too close to call. Explain why.

Since the margin of error was 2%, the true value of $p$ could be as low as 49%. The confidence interval thus contains some values of $p$, which give the election to Bush.

(c) On hearing of the poll, a nervous politician asked, “What is the probability that over half the voters prefer Gore?” A statistician replied that this question can’t be answered from the poll results, and that it doesn’t even make sense to talk about such a probability. Explain why.

The proportion of voters that favor Gore is not random—either a majority favors Gore, or they don’t. Discussing probabilities about this proportion has little meaning: the “probability” the politician asked about is either 1 or 0 (respectively).

10.18 **95% CONFIDENCE** A student reads that a 95% CI for the mean SAT Math score of California high school seniors is 452 to 470. Asked to explain the meaning of this interval, the student says, “95% of California high school seniors have SAT Math scores between 452 and 470.” Is the student right? Justify your answer.

No: The interval refers to the mean math score, not to individual scores, which will be much more variable and have a much wider spread (indeed, if more than 95% of students score below 470, they are not doing very well.)
10.27 STUDENT ATTITUDES  The survey of Study Habits and Attitudes (SSHA) is a psychological test that measures the attitude toward school and study habits of students. Scores range from 0 to 200. The mean scores for U.S. college students is about 115, and the standard deviation is about 30. A teacher suspects that older students have better attitudes toward school. She gives the SSHA to 25 students who are at least 30 years of age. Assume that scores in the population of older students are normally distributed with standard deviation $\sigma = 30$. The teacher wants to test the hypotheses $H_0 : \mu = 115$ $H_a : \mu > 115$

(a) What is the sampling distribution of the mean score $\bar{x}$ of a sample of 25 older students if the null hypothesis is true? $N \left( 115, \frac{30}{\sqrt{25}} \right) = N(115, 6)$

Sketch the density curve of this distribution.  see below

(b) Suppose that the sample data give $\bar{x} = 118.6$. Mark this point on the axis of your sketch. In fact, the result was $\bar{x} = 125.7$. Mark this point on your sketch. Using your sketch, explain in simple language why one result is good evidence that the mean score of all older students is greater than 114 and why the other outcome is not.

$P(\bar{x} \geq 118.6) = 0.274$, not surprising, FAIL to reject $H_0$

$P(\bar{x} \geq 125.7) = 0.037$, unlikely to occur by chance alone, reject $H_0$

(c) Shade the area under the curve that is the $p$-value for the sample result $\bar{x} = 118.6$
10.28 SPENDING ON HOUSING  The Census Bureau reports that households spend an average of 31% of their total spending on housing. A homebuilders association in Cleveland believes that the average is lower in their area. They interview a sample of 40 households in the Cleveland metropolitan area to learn what % of their spending goes toward housing. Take \( \mu \) to be the mean percent of spending devoted to housing among all Cleveland households. We want to test the hypotheses

\[
H_0 : \mu = 31\% \\
H_a : \mu < 31\% .
\]

The population standard deviation is 9.6%.

(a) What is the sampling distribution of the mean percent \( \bar{x} \) that the sample spends on housing if the null hypothesis is true? \( N(31\%, 1.518\%) \). Sketch the density curve of the sampling distribution.

(b) Suppose that the study finds \( \bar{x} = 30.2\% \) for the 40 households in the sample. Mark this point on the axis of your sketch. Then suppose that the study result is \( \bar{x} = 27.6\% \). Mark this point on your sketch. Referring to your sketch, explain in simple language why one result is good evidence that average Cleveland spending on housing is less than 31%, whereas the other result is not.

\[
P(\bar{x} < 30.2) = 0.299, \text{ fairly likely, FAIL to reject } H_0
\]

\[
P(\bar{x} < 27.6) = 0.0126, \text{ unlikely to occur by chance alone, reject } H_0
\]

(c) Shade the area under the curve that gives the \( p \)-value for the result \( \bar{x} = 30.2\% \). (Note that we are looking for evidence that spending is less than the null hypothesis states.)
Each of the following situations calls for a significance test for a population mean $\mu$. State the null hypothesis $H_0$ and the alternative hypothesis $H_a$ in each case.

10.29 MOTORS  The diameter of a spindle in a small motor is supposed to be 5 mm. If the spindle is either too small or too large, the motor will not work properly. The manufacturer measures the diameter in a sample of motors to determine whether the mean diameter has moved away from the target.

$H_0 : \mu = 5 \text{ mm}$
$H_a : \mu \neq 5 \text{ mm}$

10.30 HOUSEHOLD INCOME  Census Bureau data show that the mean household income in the area served by a shopping mall is $42,500 per year. A market research firm question shoppers at the mall. The researchers suspect the mean household income of mall shoppers is higher than that of the general population.

$H_0 : \mu = 42,500$
$H_a : \mu > 42,500$

10.31 BAD TEACHING?  The examinations in a large accounting class are scaled after grading so that the mean score is 50. The professor thinks that one teaching assistant is a poor teacher and suspects that his students have a lower mean score than the class as a whole. The TA’s students this semester can be considered a sample from the population of all students in the course, so the professor compares their mean scores with 50.

$H_0 : \mu = 50 \text{ points}$
$H_a : \mu < 50 \text{ points}$

10.32 SERVICE TECHNICIANS  Last year, your company’s service technicians took an average of 2.6 hours to respond to trouble calls from business customers who had purchased service contracts. Do this year’s data show a different average response time?

$H_0 : \mu = 2.6 \text{ hours}$
$H_a : \mu \neq 2.6 \text{ hours}$
HW #16  33 - 39

10.33 Return to 10.27 (HW #19). (a) Starting from the picture you drew there, calculate the $p$-values for both $x = 118.6$ and $x = 125.7$. The two $p$-values express in numbers the comparison you made informally.

$$p\text{-value} = 0.274 \quad p\text{-value} = 0.037$$

(b) Which of the two observed values of $x$ are statistically significant at the

$\alpha = 0.05$ level?  No  Yes  $\alpha = 0.01$ level?  No  No

10.34 Return to 10.28 (HW #19). (a) Starting from the picture you drew there, calculate the $p$-values for both $x = 30.2\%$ and $x = 27.6\%$. The two $p$-values express in numbers the comparison you made informally.

$$p\text{-value} = 0.299 \quad p\text{-value} = 0.0126$$

(b) Is the result $x = 27.6$ statistically significant at the $\alpha = 0.05$ level?  Yes  $\alpha = 0.01$ level?  No

\[p\text{-value} \leq \alpha \Rightarrow \text{result is statistically significant}\]
\[p\text{-value} > \alpha \Rightarrow \text{result is not significant}\]
10.35 COFFEE SALES  Weekly sales of regular ground coffee at a supermarket have in the recent past varied according to a normal distribution with mean $\mu = 354$ units per week and standard deviation $\sigma = 33$ units. The store reduces the price by 5%. Sales in the next three weeks are 405, 378, and 411 units. Is this good evidence that average sales are now higher? The hypotheses are

$H_0 : \mu = 354$

$H_a : \mu > 354$

(a) Find the value of the test statistic $x$. (show your work)

$z = \frac{\bar{x} - \mu}{\sigma} = \frac{398 - 354}{\frac{33}{\sqrt{3}}} = 2.3094$

(b) Sketch the normal curve for the sampling distribution of $\bar{x}$ when $H_0$ is true. Why is the sampling distribution normal?  It is normal because the population distribution is normal

(c) What is the $p$-value for the observed outcome? 0.0105 Shade the area that represents this $p$-value.

(d) Is the result statistically significant at the $\alpha = 0.05$ level? Yes Is it significant at the $\alpha = 0.01$ level? No Do you think there is convincing evidence that mean sales are higher? Yes There is convincing evidence that the mean sales are higher, Reject $H_0$. 


10.36 CEO PAY A study of the pay of corporate chief executive officers (CEOs) examined the increase in cash compensation of the CEOs of 104 companies, adjusted for inflation, in a recent year. The mean increase in real compensation was \( x = 6.9\% \) and the standard deviation of the increases was \( s = 55\% \). Is this good evidence that the mean real compensation \( \mu \) of all CEOs increased that year? The hypotheses are

\[
H_0 : \mu = 0 \quad \text{(no increase)} \\
H_a : \mu > 0 \quad \text{(an increase)}.
\]

Because the sample size is large, the sample \( s \) is close to the population \( \sigma \).

(a) Sketch the normal curve for the sampling distribution of \( \bar{x} \) when \( H_0 \) is true. Why is the sampling distribution approximately normal?

\[
z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{6.9 - 0}{55} = 1.279
\]

\( n = 104 \), approximately normal by CLT

(b) What is the \( p \)-value for the observed outcome \( \bar{x} = 6.9\% \)? \( 0.1004 \) Shade the area that represents the \( p \)-value.

(c) Is the result statistically significant at the \( \alpha = 0.05 \) level? No Do you think the study gives strong evidence that the mean compensations of all CEOs went up? No, we would expect this result to occur by chance 10% of the time.

10.37 STUDENT EARNINGS The financial aid office of a university asks a random sample of students about their employment and earnings. The report says that “for academic year earnings, a significant difference \( (p = 0.038) \) was found between the sexes, with men earning more on the average. No difference \( (p = 0.476) \) was found between the earnings of black and white students.” Explain both of these conclusions for the effects of sex and race on mean earnings in language understandable to someone who knows no statistics.

Comparing men’s and women’s earnings for our sample, we observe a difference so large that it would only occur in 3.8% of all samples if men and women actually earned the same amount. Based on this, we conclude that men earn more.

While there is almost certainly some difference between earnings of black and white students in our sample, it is relatively small—if blacks and whites actually earn the same amount, we would still observe a difference as big as what we saw almost half (47.6%) of the time.
A pharmaceutical manufacturer forms tablets by compressing a granular material that contains the active ingredient and various fillers. The hardness of a sample from each lot of tablets produced is measured in order to control the compression process. The target values for the hardness are $\mu = 11.5$ and $\sigma = 0.2$. The hardness data for a sample of 20 tablets are

\[
\begin{align*}
11.627 & \quad 11.613 & \quad 11.493 & \quad 11.602 & \quad 11.360 & \quad 11.374 & \quad 11.592 & \quad 11.458 & \quad 11.552 & \quad 11.463 \\
11.383 & \quad 11.715 & \quad 11.485 & \quad 11.509 & \quad 11.429 & \quad 11.477 & \quad 11.570 & \quad 11.623 & \quad 11.472 & \quad 11.531 \\
\end{align*}
\]

Is there significant evidence at the 5% level that the mean hardness of the tablets is different from the target value?

**P** – name the parameter and the population

$\mu = \text{true mean hardness of tablets.}$

**H** – state the hypotheses

$H_0 : \mu = 11.5$

$H_a : \mu \neq 11.5$

**A** – verify all assumptions

SRS

$20 < 10\%$ of population of tablets

NPP looks approximately linear $\Rightarrow$

**N** – name the test

z-test for means

**T** – calculate the test statistic (Show your work)

$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{11.5164 - 11.5}{\frac{0.2}{\sqrt{20}}} = 0.3667$

**O** – obtain the $p$-value

$p$-value $= 2P(Z \geq 0.3667) = 2(0.357) = 0.714$

**M** – make decision

$p$-value is very large, so it does NOT provide evidence against $H_0$.  FAIL to reject $H_0$.

**S** – state the conclusion in context.

The sample does not provide evidence against $H_0$.  The sample does not show evidence against the mean hardness being equal to 11.5 units.
10.39 **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 6 bottles. The results are 299.4 297.7 301.0 298.9 300.2 297.0

Is this convincing evidence at the 5% level that the mean contents of cola bottles is less than the advertised 300 ml?

- **P** – name the parameter and the population
  $\mu =$ true mean volume of soda in bottles filled from machine

- **H** – state the hypotheses
  $H_0 : \mu = 300 \text{ ml}$
  $H_A : \mu < 300 \text{ ml}$

- **A** – verify all assumptions
  - **SRS**
    6 < 10% of number of bottles filled by machine
  - **Parent population is normal**

- **N** – name the test
  z-test for means

- **T** – calculate the test statistic (Show your work)
  
  $$z = \frac{\bar{x} - \mu}{\sigma} = \frac{299.03 - 30}{\frac{3}{\sqrt{n}}} = -0.7893$$

- **O** – obtain the $p$-value
  
  $$p\text{-value} = P(Z < -0.7893) = 0.215$$

- **M** – make decision
  
  $p\text{-value} > \alpha = 0.05 \Rightarrow \text{Fail to reject } H_0$

- **S** – state the conclusion in context.
  We would expect to see this result by chance 21% of the time. This does not provide evidence against $H_0$. 
10.40 TESTING A RANDOM NUMBER GENERATOR A random number generator is supposed to produce random numbers that are uniformly distributed on the interval from 0 to 1. If this is true, the numbers generated come from a population with \( \mu = 0.5 \) and \( \sigma = 0.2887 \). A command to generate 100 random numbers gives outcomes with mean \( \bar{x} = 0.4365 \). Assume that the population \( \sigma \) remains fixed.

We want to test

\[
H_0 : \mu = 0.5 \\
H_a : \mu \neq 0.5
\]

(a) Calculate the value of the \( z \) test statistic. (show your work)

\[
z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.4365 - 0.5}{\frac{0.2887}{\sqrt{100}}} = -2.20
\]

\[p\text{-value} = 2(0.0139) = 0.0278\]

(b) Is the result significant at the 5% level (\( \alpha = 0.05 \))? Yes

(c) Is the result significant at the 1% level (\( \alpha = 0.01 \))? No

(d) Between which 2 normal critical values in the bottom row of Table C does \( z \) lie?

\[
\text{Invnorm}(0.02/1) = 2.054 \quad \text{Invnorm}(0.04/2) = 2.326
\]

Between what two numbers does the \( p\) -value lie? 0.02, 0.04

10.42 NICOTINE IN CIGARETTES To determine whether the mean nicotine content of a brand of cigarettes is greater than the advertised value of 1.4 milligrams, a health advocacy group tests

\[
H_0 : \mu = 1.4 \\
H_a : \mu > 1.4
\]

The calculated value of the test statistic is \( z = 2.42 \).

\[p\text{-value} = 0.0078\]

(a) Is the result significant at the 5% level? Yes

(b) Is the result significant at the 1% level? Yes

(c) By comparing \( z \) with the critical values in the bottom row of Table C, give 2 numbers that catch the \( p\)-value between them. 0.005 and 0.01
IQ TEST SCORES Here are the IQ test scores of 31 7th grade girls in a Midwest school district:

<table>
<thead>
<tr>
<th>114</th>
<th>100</th>
<th>104</th>
<th>89</th>
<th>102</th>
<th>91</th>
<th>114</th>
<th>114</th>
<th>103</th>
<th>108</th>
<th>130</th>
<th>120</th>
<th>132</th>
<th>111</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>119</td>
<td>86</td>
<td>72</td>
<td>111</td>
<td>103</td>
<td>74</td>
<td>112</td>
<td>107</td>
<td>103</td>
<td>98</td>
<td>96</td>
<td>112</td>
<td>112</td>
<td>93</td>
</tr>
</tbody>
</table>

Treat the 31 girls as a SRS of all 7th grade girls in the school district. Suppose that the standard deviation of IQ scores in this population is known to be $\sigma = 15$.

(a) Give a 95% CI for the mean IQ score $\mu$ for the population

- **P** – name the parameter and the population
  $\mu =$ true mean IQ score of 7th grade girls in a Midwest school district

- **A** – verify all assumptions
  SRS
  72 < 10% of all students in district
  $n = 72 > 30$ is large enough to assume normality by CLT

- **N** – name the interval
  $z$-interval

- **I** – calculate the interval (Show your work)

  $\bar{x} \pm z \frac{\sigma}{\sqrt{n}} = 105.87 \pm 1.96 \left( \frac{15}{\sqrt{72}} \right) = (100.59, 111.15)$

- **C** – state the conclusion in context.
  We are 95% confident that the true mean IQ score of 7th grade girls in this district is between 100.59 and 111.15.

(b) Is there significant evidence at the 5% level that the mean IQ score in the population differs from 100? Give appropriate statistical evidence to support your conclusion.

  Our hypotheses are $H_0: \mu = 100$ versus $H_a: \mu \neq 100$. Since 100 does not fall in the 95% confidence interval, we reject $H_0$ at the 5% level in favor of the two-sided alternative.

(c) In fact, the scores are those of all 7th grade girls in one of the several schools in the district. Explain carefully why your results from (a) and (b) cannot be trusted.

  Since the scores all came from a single school, the sample may not be representative of the school district as a whole.
10.57 IS IT SIGNIFICANT? Suppose that in the absence of special preparation SAT Math scores vary normally with mean \( \mu = 475 \) and \( \sigma = 100 \). 100 students go through a rigorous training program designed to raise their SAT Math scores by improving their mathematics skills. Carry out a test of

\[
H_0 : \mu = 475 \quad \text{vs} \quad H_a : \mu > 475
\]

in each of the following situations:

(a) The students’ average score is \( \bar{x} = 491.4 \). Is this result significant at the 5% level?

\[
z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{491.4 - 475}{100} = 1.64 \quad \text{not significant at 5% level (} p = 0.0505 \).
\]

(b) The average score is \( \bar{x} = 491.5 \). Is this result significant at the 5% level?

\[
z = 1.65; \quad \text{significant at 5% level (} p = 0.0495 \).
\]

The difference between the two outcomes in (a) and (b) is of no importance. Beware of attempts to treat \( \alpha = 0.05 \) as sacred.

10.58 COACHING AND THE SAT Let us suppose that SAT Math scores in the absence of coaching vary normally with mean \( \mu = 475 \) and \( \sigma = 100 \). Suppose also that coaching may change \( \mu \) but not \( \sigma \). An increase in the SAT Math score from 475 to 478 is of no importance in seeking admission to college, but this unimportant change can be statistically very significant. To see this, calculate the \( p \)-value for the test of \( H_0 : \mu = 475 \) in each of the following situations:

(a) A coaching service coaches 100 students. Their SAT Math scores average \( \bar{x} = 478 \).

\[
z = 0.3, \quad \text{p-value} = 0.382 \Rightarrow \text{fail to reject } H_0
\]

(b) By the next year, the service has coached 1000 students. Their SAT Math scores average \( \bar{x} = 478 \).

\[
z = 0.9949, \quad \text{p-value} = 0.171 \Rightarrow \text{fail to reject } H_0
\]

(c) An advertising campaign brings the number of students coached to 10,000. Their average score is still \( \bar{x} = 478 \).

\[
z = 3, \quad \text{p-value} = 0.0013 \Rightarrow \text{reject } H_0
\]
10.59 **COACHING AND THE SAT, II** Use your calculator to find a 99% confidence interval for the mean SAT Math score \( \mu \) after coaching in each part of the previous exercise. For large samples, the confidence interval tells us, “Yes, the mean score is higher after coaching, but only by a small amount.”

\[
\begin{align*}
  n = 100 & \quad (452.24, 503.76) \\
  n = 1000 & \quad (469.85, 486.15) \\
  n = 10,000 & \quad (475.42, 480.58)
\end{align*}
\]

10.60 **A CALL-IN OPINION POLL** A local TV station announces a question for a call-in opinion poll on the six o’clock news, and then gives the response on the eleven o’clock news. Today’s question concerns a proposed gun-control ordinance. Of the 2373 calls received, 1921 oppose the new law. The station, following standard statistical practice, makes a confidence statement: “81% of the Channel 13 Pulse Poll sample oppose gun control. We can be 95% confident that the proportion of all viewers who oppose the law is within 1.6% of the sample result.” Is the station’s conclusion justified? Explain your answer.

No—the percentage was based on a voluntary response sample, and so cannot be assumed to be a fair representation of the population. Such a poll is likely to draw a higher-than-actual proportion of people with a strong opinion, especially a strong negative opinion.

10.61 **DO YOU HAVE ESP?** A researcher looking for evidence of extrasensory perception (ESP) tests 500 subjects. Four of these subjects do significantly better \((p < 0.01)\) than random guessing.

(a) Is it proper to conclude that these four people have ESP? Explain your answer.

No—in a sample of size 500, we expect to see about 5 people who have a “P-value” of 0.01 or less. These four might have ESP, or they may simply be among the “lucky” ones we expect to see.

(b) What should the researcher now do to test whether any of these four subjects have ESP?

The researcher should repeat the procedure on these four to see if they again perform well.

10.62 Which of the following questions does a test of significance answer?

(a) Is the sample or experiment properly designed? **NO**

(b) Is the observed effect due to chance? **YES**

(c) Is the observed effect important? **NO**
10.66 MEDICAL SCREENING  Your company markets a computerized medical diagnostic program. The program scans the results of routine medical tests (pulse rate, blood tests, etc.) and either clears the patient or refers the case to a doctor. The program is used to screen thousands of people who do not have specific medical complaints. The program makes a decision about each person.

(a) What are the two hypotheses and the two types of error that the program can make? Describe the two types of error in terms of “false positive” and “false negative” test results. Include a chart.

- **H₀**: the patient is ill (or “the patient should see a doctor”)
- **Hₐ**: the patient is healthy (or “the patient should not see a doctor”).

A Type I error means a false negative—telling an ill person that he is healthy
A Type II error is a false positive—telling a healthy person he is ill and should see a doctor

<table>
<thead>
<tr>
<th>Reject H₀</th>
<th>H₀ true</th>
<th>Ha true</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type I error</td>
<td>correct</td>
<td></td>
</tr>
<tr>
<td>correct</td>
<td>Type II error</td>
<td></td>
</tr>
</tbody>
</table>

(b) The program can be adjusted to decrease one error probability at the cost of an increase in the other error probability. Which error probability would you choose to make smaller, and why? (This is a matter of judgment. There is no single correct answer.)

One might wish to lower the probability of a false negative so that most ill patients are treated. On the other hand, if money is an issue, or there is concern about sending too many patients to see the doctor, lowering the probability of false positives might be desirable.
You have the NAEP quantitative scores for a SRS of 840 young Americans. You plan to test hypotheses about the population mean score,

\[ H_0 : \mu = 275 \]
\[ H_a : \mu < 275 \]

at the 1% level of significance. The population standard deviation is known to be \( \sigma = 60 \). The z test statistic is

\[ z = \frac{\bar{x} - 275}{\frac{60}{\sqrt{840}}} \].

(a) What is the rule for rejecting \( H_0 \) in terms of \( z \)?

Reject \( H_0 \) if p-value < 0.01, this occurs if \( z < -2.326 \).

(b) What is the probability of a Type I error?

Probability of a Type I error = \( \alpha = 0.01 \) (the significance level)

(c) You want to know whether this test will usually reject \( H_0 \) when the true population mean is 270, 5 points lower than the null hypothesis claims. Answer this question by calculating the probability of a Type II error when \( \mu = 270 \).

Reject \( H_0 \) if \( \bar{x} \leq 270.185 \), \( \text{invnorm} \ (0.01, 275, 60/\sqrt{840}) \) so when

\[ \mu = 270, P(\text{Type II error}) = P(\bar{x} \geq 270.185) = P\left( \frac{\bar{x} - 270}{60/\sqrt{840}} \geq \frac{270.815 - 270}{60/\sqrt{840}} \right) = 0.4644 \]
10.69 OPENING A RESTAURANT You are thinking about opening a restaurant and are searching for a good location. From research you have done, you know that the mean income of those living near the restaurant must be over $45,000 to support the type of upscale restaurant you wish to open. You decide to take a SRS of 50 people living near one potential location. Based on the mean income of this sample, you will decide whether to open a restaurant there. A number of similar studies have shown that $\sigma = 5000$.

(a) Describe the two types of errors that you might make. Identify which is a Type I error and which is a Type II error.

Type I error: concluding that the local mean income exceeds $45,000 when in fact it does not (and therefore, opening your restaurant in a locale that will not support it).
Type II error: concluding that the local mean income does not exceed $45,000 when in fact it does (and therefore, deciding not to open your restaurant in a locale that would support it).

(b) Which of the two types of error is most serious? Explain.

Type I error is more serious. If you opened your restaurant in an inappropriate area, then you would sustain a financial loss before you recognized the mistake. If you failed to open your restaurant in an appropriate area, then you would miss out on an opportunity to earn a profit there, but you would not necessarily lose money (e.g., if you chose another appropriate location in its place).

(c) State the null and alternative hypotheses.

$H_0 : \mu = 45,000$
$H_A : \mu > 45,000$

(d) If you had to choose one of the “standard” significance levels for your significance test, would you choose $\alpha = 0.01, 0.05, 0.10$? ___________ Justify your choice.

$\alpha = 0.01$ would be most appropriate, because it would minimize your probability of committing a Type I error.

(e) Based on your choice in (d), how high will the sample mean need to be before you decide to open a restaurant in that area.

In order to reject $H_0$ at level $\alpha = 0.01$, we must have $z \geq 2.326$, or

$\bar{x} \geq 45,000 + 2.326 \frac{5000}{\sqrt{50}} = 46645.$

The sample mean would have to be at least $46,645.$

(f) If the mean income in a certain area is $47,000, how likely are you not to open a restaurant in that area? ______________ What is the probability of a Type II error? __________

$P(\text{fail to reject } H_0 \text{ when } \mu = 47,000) = P(\bar{x} < 46,645 \text{ when } \mu = 47,000)$

$= P \left( Z < \frac{46,645 - 47,000}{5000/\sqrt{50}} \right) = P(Z < -0.5) = 0.3085$

When $\mu = 47,000$, the probability of committing a Type II error is 0.3085, or about 31%.
10.78 **STATING HYPOTHESES** Each of the following situations require a significance test about a population mean $\mu$. State the appropriate null hypothesis $H_0$ and alternative hypothesis $H_a$ in each case.

(a) Larry’s car averages 32 miles per gallon on the highway. He now switches to a new motor oil that is advertised as increasing gas mileage. After driving 3000 highway miles with the new oil, he wants to determine if his gas mileage actually has increased.

$H_0 : \mu = 32 \text{ mpg}$

$H_A : \mu > 32 \text{ mpg}$

(b) A university gives credit in French language courses to students who pass a placement test. The language department wants to see if students who get credit in this way differ in their understanding of spoken French from students who actually take the French courses. Some faculty think the students who test out of the courses are better, but others argue that they are weaker in oral comprehension. Experience has shown that the mean score of students in the courses on a standard listening test is 24. The language department gives the same listening test to a sample of 40 students who passed the credit examination to see if their performance is different.

$H_0 : \mu = 24$

$H_A : \mu \neq 24$

10.82 **WHY ARE LARGE SAMPLES BETTER?** Statisticians prefer large samples. Describe briefly the effect of increasing the size of a sample (or the number of subjects in an experiment) on each of the following:

(a) The margin of error of a 95% confidence interval.

Increasing $n \rightarrow$ decreases margin of error

(b) The $p$-value of a test, when $H_0$ is false and all facts about the population remain unchanged as $n$ increases. Increasing $n \rightarrow$ decreases $P$-value (the evidence against $H_0$ becomes stronger).

(c) The power of a fixed level $\alpha$ test, when $\alpha$, the alternative hypothesis, and all facts about the population remain unchanged.

Increasing $n \rightarrow$ increases the power of the test (the test becomes better at distinguishing between the null and alternative hypotheses).
Here are the Degree of Reading Power (DRP) scores for a SRS of 44 3rd grade students from a suburban school district.

<table>
<thead>
<tr>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>42</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>43</td>
</tr>
<tr>
<td>46</td>
</tr>
<tr>
<td>27</td>
</tr>
<tr>
<td>29</td>
</tr>
<tr>
<td>47</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>44</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>38</td>
</tr>
<tr>
<td>31</td>
</tr>
<tr>
<td>46</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>25</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>33</td>
</tr>
<tr>
<td>29</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>49</td>
</tr>
<tr>
<td>28</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>47</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>48</td>
</tr>
<tr>
<td>22</td>
</tr>
<tr>
<td>33</td>
</tr>
<tr>
<td>41</td>
</tr>
<tr>
<td>51</td>
</tr>
<tr>
<td>27</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>54</td>
</tr>
<tr>
<td>45</td>
</tr>
</tbody>
</table>

DRP scores are approximately normal. Suppose that the standard deviation of scores in this school district is known to be $\sigma = 11$. The researcher believes that the mean score $\mu$ of all 3rd graders in this district is higher than the national mean, which is 32. Carry out the test at the 5% confidence level.

**P** – name the parameter and the population

$\mu = \text{true mean score of all 3rd graders in district}$

**H** – state the hypotheses

$H_0 : \mu = 32$

$H_A : \mu > 32$

**A** – verify all assumptions

SRS

$44 < 10\%$ of all 3rd graders in district

Parent population is normal

**N** – name the test

z-test for means

**T** – calculate the test statistic (Show your work)

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{35.318 - 32}{\frac{11}{\sqrt{44}}} = 2$$

**O** – obtain the $p$-value

$p$-value $= P(Z > 2) = 0.0227$

**M** – make decision

$p$-value $< \alpha = 0.05 \Rightarrow \text{Reject } H_0$

**S** – state the conclusion in context.

Because the $p$-value $< 0.05$ this provides strong evidence against $H_0$. Observations this extreme would only occur about 2.2% of the time. We do not believe the true mean scores of all 3rd graders in district equals 32.
10.85 **WELFARE PROGRAMS** A study compare two groups of mothers with young children who were on welfare two years ago. One group attended a voluntary training program offered free of charge at a local vocational school and advertised in the local news media. The other group did not choose to attend the training program. The study finds a significant difference ($p < 0.01$) between the two proportions of the mothers in the two groups who are still on welfare. The difference is not only significant but quite large. The reports says that with 95% confidence the percent of the nonattending group still on welfare is $21\% \pm 4\%$ higher than that of the group who attended the program. You are on the staff of a member of Congress who is interested in the plight of welfare mothers, and who asks you about the report.

(a) Explain in simple language what “a significant difference ($p < 0.01$)” means.
   The difference observed in the study would occur in less than 1% of all samples if the two populations actually have the same proportion.

(b) Explain clearly and briefly what “95%” confidence means.
   The interval is constructed using a method that is correct (i.e., contains the actual proportion) 95% of the time.

(c) Is this study good evidence that requiring job training of all welfare mothers would greatly reduce the percent who remain on welfare for several years?
   No—treatments were not randomly assigned, but instead were chosen by the mothers. Mothers who choose to attend a job training program may be more inclined to get themselves out of welfare.

10.87 **CORN YIELD** The mean yield of corn in the U.S. is about 120 bushels per acre. A survey of 40 farmers this year gives a sample mean yield of $\bar{x} = 123.8$ bushels per acre. We want to know whether this is good evidence that the national mean this year is not 120 bushels per acre. Assume that the farmers surveyed are a SRS from the population of all commercial corn growers and that the standard deviation of the yield in this population is $\sigma = 10$ bushels per acre. Give the $p$-value for the test of $H_0 : \mu = 120$

$H_a : \mu \neq 120$.

$z = \frac{123.8 - 120}{\frac{10}{\sqrt{40}}} = 2.40$, which gives $P = 0.0164$

Are you convinced that the population mean is not 120 bushels per acres? This is strong evidence that this year’s mean is different. Is your conclusion correct if the distribution of corn yields is somewhat nonnormal? Slight nonnormality will not be a problem since we have a reasonably large sample size (greater than 30).