AP Stats Topic 6: Means Inferencing

One of the two fire stations in a certain town responds to calls in the northern half of the town, and the other fire station responds to calls in the southern half of the town. One of the town council members believes that the two fire stations have different mean response times. Response time is measured by the difference between the time an emergency call comes into the fire station and the time the first fire truck arrives at the scene of the fire.

Data were collected to investigate whether the council member's belief is correct. A random sample of 50 calls selected from the northern fire station had a mean response time of 4.3 minutes with a standard deviation of 3.7 minutes. A random sample of 50 calls selected from the southern fire station had a mean response time of 5.3 minutes with a standard deviation of 3.2 minutes.

(a) Construct and interpret a 95 percent confidence interval for the difference in mean response times between the two fire stations.

(b) Does the confidence interval in part (a) support the council member's belief that the two fire stations have different mean response times? Explain.

Solution

Part (a):

Step 1: Identify the appropriate confidence interval by name or formula and check for appropriate conditions.

The two-sample *t* interval for $\mu_N - \mu_S$, the difference in population mean response times, is

$$\left(\bar{x}_N - \bar{x}_S\right) \pm t^* \sqrt{\frac{s_N^2}{n_N} + \frac{s_S^2}{n_S}}$$

where μ_N denotes the mean response for calls from the northern fire station and μ_S denotes the mean response for calls from the southern fire station.

Conditions: 1. Independent random samples
2. Large samples or normal population distributions

A random sample of 50 calls was selected from the northern fire station, independent of the random sample of 50 calls selected from the southern fire station.

The use of the two-sample *t* interval is reasonable because both sample sizes are large ($n_N = 50 > 30$ and $n_S = 50 > 30$), and by the central limit theorem, the sampling distributions for the two sample means are approximately normal. Therefore the sampling distribution of the difference of the sample means $\bar{x}_N - \bar{x}_S$ is approximately normal.

Step 2: Correct Mechanics

Unequal variances: Degrees of freedom = 96.

$$(4.3 - 5.3) \pm 1.985\sqrt{\frac{3.7^2}{50} + \frac{3.2^2}{50}}$$

$$-1.0 \pm 1.985 \times .6918$$

$$(-2.37, 0.37)$$

Step 3: Interpretation

Based on these samples, one can be 95 percent confident that the difference in the population mean response times (northern - southern) is between -2.37 minutes and 0.37 minutes.

Part (b):

Zero is within the 95 percent confidence interval of plausible values for the difference in population means. Therefore this confidence interval does not support the council member's belief that there is a difference in mean response times for the two fire stations.

Patients with heart-attack symptoms arrive at an emergency room either by ambulance or self-transportation provided by themselves, family, or friends. When a patient arrives at the emergency room, the time of arrival is recorded. The time when the patient's diagnostic treatment begins is also recorded.

An administrator of a large hospital wanted to determine whether the mean wait time (time between arrival and diagnostic treatment) for patients with heart-attack symptoms differs according to the mode of transportation. A random sample of 150 patients with heart-attack symptoms who had reported to the emergency room was selected. For each patient, the mode of transportation and wait time were recorded. Summary statistics for each mode of transportation are shown in the table below.

Mode of Transportation	Sample Size	Mean Wait Time (in minutes)	Standard Deviation of Wait Times (in minutes)
Ambulance	77	6.04	4.30
Self	73	8.30	5.16

(a) Use a 99 percent confidence interval to estimate the difference between the mean wait times for ambulance-transported patients and self-transported patients at this emergency room.

(b) Based only on this confidence interval, do you think the difference in the mean wait times is statistically significant? Justify your answer.

Solution

Part (a):

Step 1: Identifies the appropriate confidence interval by name or formula and checks appropriate conditions.

Two sample *t* interval for $\mu_A - \mu_S$, the difference in mean waiting times, or

$(\overline{x}_A - \overline{x}_S) \pm t^*_{df}\sqrt{\frac{s_A^2}{n_A} + \frac{s_S^2}{n_S}}$ [See the next page for possible values of *df*.]

Conditions: 1. Independent random samples
2. Large samples or normal population distributions

One sample of 150 patients divided into two groups after sampling does not meet the condition of two independent random samples with fixed sample sizes. Nevertheless, it is reasonable to assume that mode of transportation splits the patients into two independent groups. Secondly, use of the two sample t interval is reasonable because each sample size is large (e.g., $n_A = 77 > 30$ and $n_S = 73 > 30$).

Alternatively, we could assume that the waiting times are (at least approximately) normally distributed, but we have no way to check this assumption with the information provided.

Step 2: Correct Mechanics

Degrees of freedom = $\min\{(77-1),(73-1)\} = 72$.

$(6.04 - 8.30) \pm 2.6459\sqrt{\frac{4.30^2}{77} + \frac{5.16^2}{73}}$

$-2.26 \pm 2.6459 \cdot (0.7777)$

-2.26 ± 2.0577

$(-4.3177, -0.2023)$

Step 3: Interpretation

Based on this sample, we are 99 percent confident that the true difference in the populations' mean waiting times (ambulance – self) is between –4.3177 minutes and –0.2023 minutes.

Equivalently,

With 99 percent confidence, the true mean wait time for those who arrive by ambulance is shorter than those who are self transported by somewhere between 0.2 and 4.3 minutes.

Part (b):

Since zero is not in the 99 percent confidence interval of plausible values for the true difference in means, we can reject $H_0: \mu_A - \mu_S = 0$ in favor of the alternative $H_a: \mu_A - \mu_S \neq 0$ at the $\alpha = .01$ significance level.

Thus, we have statistically significant evidence that there is a difference in the mean wait times for the two groups.

Baby walkers are seats hanging from frames that allow babies to sit upright with their legs dangling and feet touching the floor. Walkers have wheels on their legs that allow the infant to propel the walker around the house long before he or she can walk or even crawl. Typically, babies use walkers between the ages of 4 months and 11 months.

Because most walkers have tray tables in front that block babies' views of their feet, child psychologists have begun to question whether walkers affect infants' cognitive development. One study compared mental skills of a random sample of those who used walkers with a random sample of those who never used walkers. Mental skill scores averaged 113 for 54 babies who used walkers (standard deviation of 12) and 123 for 55 babies who did not use walkers (standard deviation of 15).

(a) Is there evidence that the mean mental skill score of babies who use walkers is different from the mean mental skill score of babies who do not use walkers? Explain your answer.

(b) Suppose that a study using this design found a statistically significant result. Would it be reasonable to conclude that using a walker causes a change in mean mental skill score? Explain your answer.

a. **part 1**: States a correct pair of hypotheses

H_o: $\mu_W = \mu_N$ OR H_o: $\mu_W - \mu_N = 0$

H_a: $\mu_W \neq \mu_N$ H_a: $\mu_W - \mu_N \neq 0$

where μ_W is the mean mental skill score for babies who used walkers and μ_N is the mean for those who did not. Nonstandard notation must be explained. Hypotheses about statistics (e.g. $\bar{x}$ or $\hat{p}$) are unacceptable.

part 2: Identifies a correct test (by name or by formula), and **<u>checks</u>** appropriate assumptions.

Note: Problem states that samples are random samples, so this does not need to be addressed in the assumptions.

Independent samples t test. Assumptions: large sample or normal population distributions. Check: OK, because, for example, n_1&n_2>30.

part 3: Correct mechanics, including value of test statistic, df (if appropriate), and P-value or rejection region (except for minor arithmetic errors)

- For independent samples t test:

$$t = \frac{\bar{x}_W - \bar{x}_N}{\sqrt{\frac{s_W^2}{n_W} + \frac{s_N^2}{n_N}}} = \frac{113 - 123}{\sqrt{\frac{12^2}{54} + \frac{15^2}{55}}} = \frac{-10}{\sqrt{6.7576}} = -3.8468$$

(Calculator: t = -3.846843677)
df = 102.828 (OK to use 102), P-value = .0002

OR conservative df = 54 - 1 = 53, P-value = 2(.00016)=.00032
OR using tables (for either df) P-value < 2(.0005) = .001

- For pooled t test: s_p = 13.597, t = -3.839, df = 107, P-value = .0002 (or < .001 from tables)

- For independent samples z test, z = -3.8468, P-value = .0001 (or <2(.0002) = .0004 from tables)

part 4: Stating a correct conclusion in the context of the problem, using the result of the statistical test (i.e., **linking the conclusion to the result of the hypothesis test**).

Reject the null hypothesis because P-value is less than stated α (or because P-value is very small, or because test statistic falls in the rejection region). There is convincing evidence that the mean mental score of babies who used walkers is different from the mean score for babies who did not use walkers.

If both an α and a P-value are given, the linkage is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the P-value or explaining how the conclusion follows from the P-value.

If the P-value in part 3 is incorrect but the conclusion is consistent with the computed P-value, part 4 can be considered as correct.

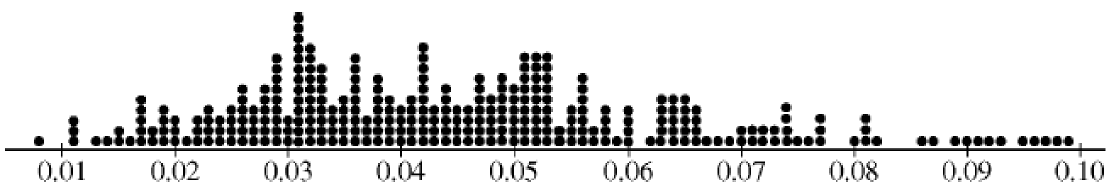
part (b): No. This was an observational study, and a causal relationship can not be inferred from an observational study.

A bottle-filling machine is set to dispense 12.1 fluid ounces into juice bottles. To ensure that the machine is filling accurately, every hour a worker randomly selects four bottles filled by the machine during the past hour and measures the contents. If there is convincing evidence that the mean amount of juice dispensed is different from 12.1 ounces or if there is convincing evidence that the standard deviation is greater than 0.05 ounce, the machine is shut down for recalibration. It can be assumed that the amount of juice that is dispensed into bottles is normally distributed.

During one hour, the mean number of fluid ounces of four randomly selected bottles was 12.05 and the standard deviation was 0.085 ounce.

(a) Perform a test of significance to determine whether the mean amount of juice dispensed is different from 12.1 fluid ounces. Assume the conditions for inference are met.

(b) To determine whether this sample of four bottles provides convincing evidence that the standard deviation of the amount of juice dispensed is greater than 0.05 ounce, a simulation study was performed. In the simulation study, 300 samples, each of size 4, were randomly generated from a normal population with a mean of 12.1 and a standard deviation of 0.05. The sample standard deviation was computed for each of the 300 samples. The dotplot below displays the values of the sample standard deviations.



Use the results of this simulation study to explain why you think the sample provides or does not provide evidence that the standard deviation of the juice dispensed exceeds 0.05 fluid ounce.

Part (a):

Step 1: State a correct pair of hypotheses.

H_0: $\mu = 12.1$
H_a: $\mu \neq 12.1$, where μ is the mean number of fluid ounces dispensed into all juice bottles filled in the past hour

Step 2: Identify a correct test (by name or by formula) and check appropriate conditions. (Stem of the question said to assume that conditions for inference are met.)

One-sample t test for a mean *OR* $t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

Step 3: Correct mechanics, including the value of the test statistic, df, and p-value (or rejection region).

test statistic: $t = \frac{12.05 - 12.1}{\frac{0.085}{\sqrt{4}}} = \frac{-0.05}{0.0425} = -1.176$

p-value: $2 \cdot P(T_{3df} < -1.176) = 0.324$

Step 4: State a correct conclusion in the context of the problem, using the result of the computations.

Because the p-value of 0.324 is larger than any reasonable significance level, such as $\alpha = .05$, do not reject the null hypothesis that the mean number of fluid ounces being dispensed is 12.1 fluid ounces. There is not sufficient evidence to conclude that the machine is filling the juice bottles with an average amount different from 12.1 fluid ounces.

Part (b):

In 300 simulated sample standard deviations, the value of the computed standard deviation (0.085) from our sample in part (a) or a value larger than 0.085 occurred only 12 times. This is a simulated p–value of $\frac{12}{300}$ or 0.04. If the actual population standard deviation is 0.05, then we estimate that the chance of observing a sample standard deviation of 0.085 or larger is 4 percent.

Because this simulated p-value is less than a significance level of 5 percent, the sample from part (a) provides strong evidence that the standard deviation of the juice being dispensed exceeds 0.05 ounces.

High cholesterol levels in people can be reduced by exercise, diet, and medication. Twenty middle-aged males with cholesterol readings between 220 and 240 milligrams per deciliter (mg/dL) of blood were randomly selected from the population of such male patients at a large local hospital. Ten of the 20 males were randomly assigned to group A, advised on appropriate exercise and diet, and also received a placebo. The other 10 males were assigned to group B, received the same advice on appropriate exercise and diet, but received a drug intended to reduce cholesterol instead of a placebo. After three months, posttreatment cholesterol readings were taken for all 20 males and compared to pretreatment cholesterol readings. The tables below give the reduction in cholesterol level (pretreatment reading minus posttreatment reading) for each male in the study.

Group A (placebo)

Reduction (in mg/dL)	2	19	8	4	12	8	17	7	24	1

Mean Reduction: 10.20 Standard Deviation of Reductions: 7.66

Group B (cholesterol drug)

Reduction (in mg/dL)	30	19	18	17	20	−4	23	10	9	22

Mean Reduction: 16.40 Standard Deviation of Reductions: 9.40

Do the data provide convincing evidence, at the $\alpha = 0.01$ level, that the cholesterol drug is effective in producing a reduction in mean cholesterol level beyond that produced by exercise and diet?

Solution

Step 1: States a correct pair of hypotheses.

Let μ_A represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet and also receive a placebo.

Let μ_B represent the mean cholesterol reduction if all such male patients at this hospital are advised on appropriate exercise and diet but receive the drug instead of a placebo.

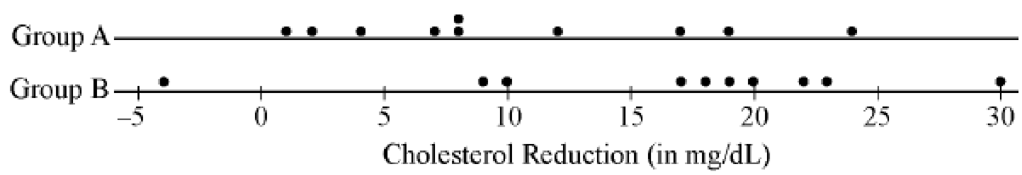
The hypotheses to be tested are $H_0: \mu_A = \mu_B$ versus $H_a: \mu_A < \mu_B$.

Step 2: Identifies a correct test procedure (by name or by formula) and checks appropriate conditions.

The appropriate procedure is a two-sample *t*-test.

When comparing two experimental treatments using a two-sample *t*-test, the subjects must be randomly assigned to the treatments. This condition is stated in the question (10 men were randomly assigned to group A and the remaining 10 men to group B).

The second condition is that the two populations are approximately normally distributed or the sample sizes are sufficiently large. Because of the small sample sizes (10 in each treatment group), we need to check whether it is reasonable to assume that the samples came from populations that are normally distributed. The following dotplots reveal slight skewness and a possible outlier for group B, but it appears reasonable to proceed with the two-sample *t*-test.



Step 3: Demonstrates correct mechanics, including the value of the test statistic and *p*-value (or the rejection region).

The test statistic is: $t = \dfrac{\overline{x}_A - \overline{x}_B}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} = \dfrac{10.20 - 16.40}{\sqrt{\dfrac{7.66^2}{10} + \dfrac{9.40^2}{10}}} \approx -1.62$

With df = 17.3, *p*-value $\approx$ 0.062.

Step 4: States a correct conclusion in the context of the problem, using the result of the statistical test.

Because the *p*-value is greater than the significance level of $\alpha = 0.01$, we fail to reject H_0. The data do not provide enough evidence at the 0.01 level of significance to conclude that the drug is effective in producing a mean cholesterol reduction beyond that provided by exercise and dietary advice.

The developers of a training program designed to improve manual dexterity claim that people who complete the 6-week program will increase their manual dexterity. A random sample of 12 people enrolled in the training program was selected. A measure of each person's dexterity on a scale from 1 (lowest) to 9 (highest) was recorded just before the start of and just after the completion of the 6-week program. The data are shown in the table below.

Person	Before Program	After Program
A	6.7	7.8
B	5.4	5.9
C	7.0	7.6
D	6.6	6.6
E	6.9	7.6
F	7.2	7.7
G	5.5	6.0
H	7.1	7.0
I	7.9	7.8
J	5.9	6.4
K	8.4	8.7
L	6.5	6.5
Total	81.1	85.6

Can one conclude that the mean manual dexterity for people who have completed the 6-week training program has significantly increased? Support your conclusion with appropriate statistical evidence.

Solution

Step 1: States a correct pair of hypotheses.

Let μ_D denote the mean difference (after – before) in dexterity scores for the population of individuals enrolled in the program.
$H_0 : \mu_D = 0$ versus $H_a : \mu_D > 0$

Step 2: Identifies a correct test (by name or formula) and checks appropriate conditions.

One sample t-test or paired t-test or $t = \frac{\overline{x}_D}{s_D/\sqrt{n}}$.

We are told that the 12 people are a random sample. Assume that the differences (after – before) are approximately normal. This check may be done with a histogram, dotplot, stem-and-leaf display, or normal probability plot. The student should note that the normal assumption is not unreasonable because the plot displays no obvious skewness or outliers.

Step 3: Correct mechanics, including the value of the test statistic and the p-value (or rejection region).

$\overline{x}_D = 0.375$, $s_D = 0.367$
Degrees of freedom $= 12 - 1 = 11$

$$t = \frac{0.375}{\frac{0.367}{\sqrt{12}}} = 3.54$$

p-value $= 0.002$

Step 4: States a correct conclusion in the context of the problem.

Since the p-value is less than 0.05, we can reject the null hypothesis of no difference in favor of the alternative and conclude that, on average, people who completed the program have significantly increased manual dexterity.

Regulations require that product labels on containers of food that are available for sale to the public accurately state the amount of food in those containers. Specifically, if milk containers are labeled to have 128 fluid ounces and the mean number of fluid ounces of milk in the containers is at least 128, the milk processor is considered to be in compliance with the regulations. The filling machines can be set to the labeled amount. Variability in the filling process causes the actual contents of milk containers to be normally distributed. A random sample of 12 containers of milk was drawn from the milk processing line in a plant, and the amount of milk in each container was recorded.

(a) The sample mean and standard deviation of this sample of 12 containers of milk were 127.2 ounces and 2.1 ounces, respectively. Is there sufficient evidence to conclude that the packaging plant is not in compliance with the regulations? Provide statistical justification for your answer.

Inspectors decide to study a particular filling machine within this plant further. For this machine, the amount of milk in the containers has a mean of 128.0 fluid ounces and a standard deviation of 2.0 fluid ounces.

(b) What is the probability that a randomly selected container filled by this machine contains at least 125 fluid ounces?

(c) An inspector will randomly select 12 containers filled by this machine and record the amount of milk in each. What is the probability that the minimum (smallest amount of milk) recorded in the 12 containers will be at least 125 fluid ounces? (Note: In order for the minimum to be at least 125 fluid ounces, each of the 12 containers must contain at least 125 fluid ounces.)

An analyst wants to use simulation to investigate the sampling distribution of the minimum. This analyst randomly generates 150 samples, each consisting of 12 observations, from a normal distribution with mean 128 and standard deviation 2 and finds the minimum for each sample. The 150 minimums (sorted from smallest to largest) are shown on the next page.

Part (a):

Step 1: State a correct pair of hypotheses.

$H_0 : \mu = 128$ fluid ounces versus $H_a : \mu < 128$ fluid ounces

Step 2: Identify a correct test (by name or by formula) and checks appropriate conditions.

One sample *t*-test for a mean

OR $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$

Condition: The random sample is taken from a normal population. (This information is stated in the stem so it does not need to be repeated here.)

Step 3: Use correct mechanics, including the value of the test statistic, degrees of freedom, and *p*-value (or rejection region)

Test Statistic: $t = \dfrac{127.2 - 128}{2.1/\sqrt{12}} = \dfrac{-0.8}{0.6062} = -1.3192$

p-value: $P(T_{11 d.f.} < -1.3192) = 0.1070$

Step 4: Using the result of the statistical test, state a correct conclusion in the context of the problem.

Since the *p*-value = 0.1070 is greater than any reasonable significance level, say $\alpha = .05$, we do not have statistically significant evidence to refute the claim that the company is in compliance with the regulations. That is, we cannot reject the null hypothesis that the mean quantity of milk in 12 containers is at least 128 fluid ounces.

If both an α and a *p*-value are given, the linkage is implied. If no α is given, the solution must be explicit about the linkage by giving a correct interpretation of the *p*-value or explaining how the conclusion follows from the *p*-value.

If the *p*-value in step 3 is incorrect but the conclusion is consistent with the computed *p*-value, step 4 can be considered as correct.

Part (b)

Let X denote the amount of fluid in a randomly selected container from this group. The probability that a randomly selected container from this group contains at least 125 fluid ounces is:

$$P(X \geq 125) = P\left(Z \geq \frac{125 - 128}{2}\right) = P(Z \geq -1.5) = 1 - 0.0668 = 0.9332$$

Part (c):

Let $X_{(1)} = \min\{X_1, X_2, \ldots, X_{12}\}$. The probability that the smallest amount of milk recorded in the 12 randomly selected containers will be at least 125 fluid ounces is:

$$\begin{aligned} P\left(X_{(1)} \geq 125\right) &= P(X_1 \geq 125, X_2 \geq 125, \ldots, X_{12} \geq 125) \\ &= [P(X_1 \geq 125)]^{12} \\ &= (0.9332)^{12} \\ &= 0.4362 \end{aligned}$$

Part (d):

Looking at the sorted list of 150 minimums, we notice that 66 of the minimums are at least 125 and 84 of the minimums are less than 125. Thus, the probability in part (c) is approximated by $P\left(X_{(1)} \geq 125\right) \approx \frac{66}{150} = 0.44$ or $P\left(X_{(1)} \geq 125\right) \approx 1 - \frac{84}{150} = 0.44$. The difference between the simulated value and the theoretical value, $0.44 - 0.4362 = 0.0038$, is very small. In other words, the simulation provides a very good approximation to the theoretical value.

A pharmaceutical company has developed a new drug to reduce cholesterol. A regulatory agency will recommend the new drug for use if there is convincing evidence that the mean reduction in cholesterol level after one month of use is more than 20 milligrams/deciliter (mg/dl), because a mean reduction of this magnitude would be greater than the mean reduction for the current most widely used drug.

The pharmaceutical company collected data by giving the new drug to a random sample of 50 people from the population of people with high cholesterol. The reduction in cholesterol level after one month of use was recorded for each individual in the sample, resulting in a sample mean reduction and standard deviation of 24 mg/dl and 15 mg/dl, respectively.

(a) The regulatory agency decides to use an interval estimate for the population mean reduction in cholesterol level for the new drug. Provide this 95 percent confidence interval. Be sure to interpret this interval.

(b) Because the 95 percent confidence interval includes 20, the regulatory agency is not convinced that the new drug is better than the current best-seller. The pharmaceutical company tested the following hypotheses.

$$H_0: \mu = 20 \text{ versus } H_a: \mu > 20,$$

where μ represents the population mean reduction in cholesterol level for the new drug.

The test procedure resulted in a t-value of 1.89 and a p-value of 0.033. Because the p-value was less than 0.05, the company believes that there is convincing evidence that the mean reduction in cholesterol level for the new drug is more than 20. Explain why the confidence interval and the hypothesis test led to different conclusions.

(c) The company would like to determine a value L that would allow them to make the following statement.

We are 95 percent confident that the true mean reduction in cholesterol level is greater than L.

A statement of this form is called a one-sided confidence interval. The value of L can be found using the following formula.

$$L = \bar{x} - t^* \frac{s}{\sqrt{n}}$$

This has the same form as the lower endpoint of the confidence interval in part (a), but requires a different critical value, t^*. What value should be used for t^*?

Recall that the sample mean reduction in cholesterol level and standard deviation are 24 mg/dl and 15 mg/dl, respectively. Compute the value of L.

(d) If the regulatory agency had used the one-sided confidence interval in part (c) rather than the interval constructed in part (a), would it have reached a different conclusion? Explain.

Part (a):

Step 1: States and checks appropriate conditions.

We are told that the sample was randomly selected. Since the sample size is large (e.g., $n = 50 > 30$), the one sample t interval should be valid. Alternatively, we could assume that the reduction in cholesterol level after one month is (at least approximately) normally distributed, but we have no way to check this assumption with the information provided.

Step 2: Identifies the appropriate confidence interval by name or formula.

One sample t interval for μ, the mean reduction in cholesterol for the new drug or $\bar{x} \pm t^*_{n-1}\frac{s}{\sqrt{n}}$.

Step 3: Correct mechanics.

Degrees of freedom = n-1 = 49.

$$\bar{x} \pm t^*_{n-1}\frac{s}{\sqrt{n}} = 24 \pm 2.0096\frac{15}{\sqrt{50}} = 24 \pm 4.2631 = (19.7369, 28.2631).$$

Step 4: Interprets the confidence interval in context.

We are 95% confident that the mean reduction in cholesterol for the new drug in the population of people with high cholesterol is between 19.74 and 28.26 mg/dl.

Part (b):

The decision based on a 95% confidence interval only corresponds to the two-sided test of significance at the 5% level, not necessarily the one-sided test. The confidence interval in (a) is equivalent to testing H_0: $\mu = 20$ against H_a: $\mu \neq 20$. In this test, the tail probability would be doubled, and this two-sided p-value, .066, is larger than .05, failing to reject the null hypothesis. However, in testing H_0: $\mu = 20$ against H_a: $\mu > 20$, the one-sided p-value of .033 is small enough to reject H_0 at the 5% level.

Alternatively, if we had compared the p-value of .033 to an alpha level of .025, the conclusions would match.

Part (c):

The critical value for the lower confidence bound is the 95^{th} percentile (instead of the 97.5^{th} percentile) of the t distribution with 49 degrees of freedom.

$$t^* = 1.676$$

and

$$L = 24 - 1.676\frac{15}{\sqrt{50}} = 24 - 3.5553 = 20.4447$$

Part (d):

Yes, the decision would change. Since the lower bound L is more than 20, the agency would now be convinced that μ is greater than 20 and the new drug is statistically significantly better than the current drug.

Researchers want to see whether training increases the capability of people to correctly predict outcomes of coin tosses. Each of twenty people is asked to predict the outcome (heads or tails) of 100 independent tosses of a fair coin. After training, they are retested with a new set of 100 tosses. (All 40 sets of 100 tosses are independently generated.) Since the coin is fair, the probability of a correct guess by chance is 0.5 on each toss. The numbers correct for each of the 20 people were as follows.

Score Before Training (number correct)	Score After Training (number correct)
46	61
48	62
50	53
54	46
54	50
54	52
54	53
54	59
54	60
54	61
55	55
56	59
57	55
58	50
58	56
61	58
61	64
63	57
64	61
65	54
Sum 1,120	Sum 1,126

To answer the following questions, you may want to enter these data into your calculator. As a check that you have entered the data correctly, the sum of the first column is 1,120 and the sum of the second column is 1,126.

a. Do the data suggest that after training people can correctly predict coin toss outcomes better than the 50 percent expected by chance guessing alone?

   Give appropriate statistical evidence to support your conclusion.

b. Does the statistical test that you completed in part (a) provide evidence that this training is effective in improving a person's ability to predict coin toss outcomes?

   If yes, justify your answer. If no, conduct an appropriate analysis that would allow you to determine whether or not the training is effective.

c. Would knowing a person's score before training be helpful in predicting his or her score after training?

   Justify your answer.

a. Let MU_A = mean number of correct responses after training

Ho: $MU_A = 50$ (or <=) Ha: $MU_A > 50$

One-sample t test Test statistic: $$t = \frac{\bar{x}_A - \mu_0}{s_A / \sqrt{n}}$$

Assumption: Normal population distribution.

Checking Assumption for t test: Boxplot, dot plot, stem and leaf plot or histogram does not show any outliers or extreme skewness. Or normal probability plot looks OK. (See attached plots.)

$\bar{x}_A = 56.3$

$s_A = 4.725$

n = 20

$$t = \frac{56.3 - 50}{4.725 / \sqrt{20}} = 5.963$$

df = 19

P-value = .0000049

For any reasonable choice of ALPHA, reject H_o. There is convincing evidence that the mean number of correct responses after training is higher than 50, the value expected by chance alone.

(For a rejection region approach, Rejection Region boundaries are at 1.33 for ALPHA = .1, 1.73 for ALPHA = .05, and 2.54 for ALPHA= .01)

b. The analysis in part (a) does not provide evidence that the training is effective, because it does not compare before and after scores. Data is paired.

H_o: $MU_D = 0$ Where MUD is the mean difference in the score before training and the score after training

H_a: $MU_D > 0$ (would be < if student defines differences as before-after)

OR

H_o: $MU_B - MU_A = 0$ H_a: $MU_B - MU_A > 0$ (< if student defines differences as before-after)

One-sample t test Test statistic: $$t = \frac{\bar{x}_D - \mu_0}{s_D / \sqrt{n}}$$

Assumption: Normal difference population

Checking Assumption for t test: Boxplot, dot plot, stem and leaf plot or histogram of sample differences does not show any outliers or extreme skewness. Or normal probability plot of differences looks OK. (See attached graphs.)

$\bar{x}_D = .3$

$s_D = 6.837$

n = 20

$$t = \frac{.3 - 0}{6.837/\sqrt{20}} = .1962$$

df = 19

P-value = .4233

For any reasonable choice of ALPHA, fail to reject Ho. There is not convincing evidence that training is effective in improving a person's ability to predict coin tossing outcomes.

(For a rejection region approach, Rejection Region boundaries are at 1.33 for ALPHA = .1, 1.73 for ALPHA= .05, and 2.54 for ALPHA = .01)

c. The response indicates that knowledge of a person's score before training would not be helpful in predicting after training score and justifies this response in at least one of the following ways:

- A scatter plot of the after scores vs. the before scores shows no pattern indicating a relationship between before and after scores
- The correlation coefficient for before and after scores is very close to 0 (r = .020 or $R^2 = .0396\%$)
- The test for slope in a regression of y = "after score" on x = "before score" does not indicate that the slope is significantly different from 0 (t = .08, P-value = .934).

6. Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(a) Describe the population for which it is appropriate for Emma to generalize the results from her sample.

The distribution of the 50 rental prices of the available apartments is shown in the following histogram.

The distribution of the 50 rental prices of the available apartments is shown in the following histogram.

(b) Emma wants to estimate the typical rental price of a one-bedroom apartment in the city. Based on the distribution shown, what is a disadvantage of using the mean rather than the median as an estimate of the typical rental price?

(c) Instead of using the sample median as the point estimate for the population median, Emma wants to use an interval estimate. However, computing an interval estimate requires knowing the sampling distribution of the sample median for samples of size 50. Emma has one point, her sample median, in that sampling distribution. Using information about rental prices that are available on the Web site, describe how someone could develop a theoretical sampling distribution of the sample median for samples of size 50.

Because Emma does not have the resources to develop the theoretical sampling distribution, she estimates the sampling distribution of the sample median using a process called bootstrapping. In the bootstrapping process, a computer program performs the following steps.

- Take a random sample, with replacement, of size 50 from the original sample.
- Calculate and record the median of the sample.
- Repeat the process to obtain a total of 15,000 medians.

Emma ran the bootstrap process, and the following frequency table is the bootstrap distribution showing her results of generating 15,000 medians.

Bootstrap Distribution of Medians					
Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65
2,497.5	136	2,735	35	2,985	12
2,500	1,899	2,747.5	619	2,987.5	1
2,522.5	2	2,750	2	2,995	6
2,525	945	2,795	278	3,000	2
2,550	1,673	2,812.5	16	3,062.5	3

The bootstrap distribution provides an approximation of the sampling distribution of the sample median. A confidence interval for the median can be constructed using a percentage of the values in the middle of the bootstrap distribution.

(d) Use the frequency table to find the following.
  (i) Value of the 5th percentile:
  (ii) Value of the 95th percentile:

(e) Find the percentage of bootstrap medians in the table that are equal to or between the values found in part (d)

(f) Use your values from parts (d) and (e) to construct and interpret a confidence interval for the median rental price.

a. The population is all one-bedroom apartments in this city that are voluntarily posted to the website.

b. Because the distribution of rental prices are skewed to the right, the mean price will be higher than the median. The median price will be a better "typical" value, so if Emma uses the mean as "typical", the disadvantage is Emma will over-estimate the typical rental price.

c. To create a theoretical sampling distribution of median rental prices for samples of size 50, one would have to randomly sample 50 one-bedroom apartments and record the median. This sampling and recording of sample medians would have to be repeated many times until all combinations of 50 have been sampled. Those medians of all the samples would be the sampling distribution of medians for samples of size 50.

d. The 5^{th} percentile would occur at the (0.05)(1500) or 75^{th} lowest median, which occurs at rental price of \$2500.

The 95^{th} percentile would be at the 75^{th} from the highest median, which is at \$2950.

e. The percentage of bootstrap medians equal to or between \$2500 to \$2950 is 14,404/15,000 or 96.0%

f. Based on the bootstrap process, we are 96% confident the true median of rental prices is between \$2500 and \$2950.