Unit 25: Tests of Significance

PREREQUISITES

In this unit, we use inference about the mean μ of a normal distribution to illustrate the reasoning of significance tests. Hence, students should be familiar with normal distributions (Units 7, 8, and 9). Inference about μ is based on the sample mean $\bar{x}$. Hence, students should be familiar with facts about the sampling distribution of $\bar{x}$ presented in Unit 22, Sampling Distributions. Unit 24, Confidence Intervals, should be covered before this unit since several of the exercises ask students to compute a confidence interval either before or after performing a significance test.

ADDITIONAL TOPIC COVERAGE

An introduction into significance tests can be found in *The Basic Practice of Statistics*, Chapter 15, Tests of Significance: The Basics.

ACTIVITY DESCRIPTION

In this activity, students will check whether the mean number of chips per cookie in Nabisco's Chips Ahoy chocolate chip cookies has changed from what it was advertised to be in the 1980s. In Unit 27's activity, students will determine whether the mean number of chips in regular Chips Ahoy Chocolate Chip Cookies differs significantly from the mean number of chips in reduced fat Chips Ahoy Chocolate Chip Cookies. So, you may want to collect the data from both types of cookies now.

MATERIALS

One or two bags of Nabisco's Chips Ahoy chocolate chip cookies; paper plates or paper towels. (Include a bag or two of Nabisco's reduced fat Chips Ahoy chocolate chip cookies if you want to also collect the data needed for Unit 27's activity.)

This activity should be done in groups with 2 to 4 students. If you need to have a group of 3, students in the group can trade off being chip counters.

For question 1, students will need to collect data. Before students begin counting the chips in the cookies, hold a discussion to establish rules that students will follow when counting chips. To start the discussion, hand a cookie to one student, and ask him to count the chips but not to reveal his result. Next, the cookie should be passed to two other students who do the same. After all three have counted the chips in the same cookie, they should report on the number of chips they counted. Often the chip counts for the same cookie are very different, which is an indication that the variability due to the counting procedure needs to be controlled. At this point, students should discuss in their groups how they think the counting should be done so that the variability due to the counting process is reduced. Give groups a chance to present their counting plans. After groups have presented their plans, the class should decide on a set of rules. Here is a sample set of rules:

- Count chips that look larger than a half chip and ignore anything smaller.
- Count chips appearing on the top and bottom of the cookie as separate chips. Chips on the side of the cookie only get counted once, even though they might appear both from the top and bottom.
- Two independent counts will be taken on each cookie and the cookie count will be the average of the two independent counts. (This will reduce the variability due to the counting procedure.)
- Give each group a paper plate or paper towel. Place a bunch of cookies on each plate.

Group members will count the number of chips in each cookie – labeling them cookie #1, cookie #2, etc. so that the independent counts get recorded and matched to the same cookie. If students are also collecting data on the reduced fat cookies, make sure they label the cookie type as well. When students have finished counting the number of chips in their bunch of cookies, they should hand in their data so that it can be consolidated into a single data sheet (or spreadsheet). (Once the counting is completed, students can eat the cookies!)

Students will need a copy of the class data for questions 2 – 4. If you decide not to collect the data (a task students really enjoy), use the sample data (see sample answer to question 1), which was collected in two statistics classes (each class got a bag of cookies).

THE VIDEO SOLUTIONS

1. After entering a poem, the program could tell you how many new words there are in the poem that Shakespeare did not use in any of his other writings.

2. The null hypothesis was that Shakespeare was the author of the poem. The alternative hypothesis was that someone else authored the poem.

3. The number of unique words per poem was approximately normally distributed with mean $\mu = 7$ and standard deviation $\sigma = 2.6$.

4. No. Thisted could expect to find a value at least as extreme as 10 unique words about 25% of the time when the poems were Shakespeare's.

5. A small *p*-value.

UNIT ACTIVITY SOLUTIONS

1. The sample data below are from the non-broken cookies in two bags of Nabisco's Chips Ahoy chocolate chip cookies. Two students counted the chips in each cookie independently and the results were averaged. Sample answers to questions 2 and 3 are based on these data.

18.5	17.0	14.0	14.5	15.0	13.5	16.5	15.5	19.0	16.0
20.5	23.5	17.5	18.5	21.5	18.5	22.5	15.0	16.0	17.5
12.0	14.5	12.5	12.0	20.5	21.5	22.5	24.0	18.5	16.5
22.5	22.0	18.5	18.5	21.5	20.0	17.5	16.5	17.5	17.5
19.5	21.5	24.5	18.0	23.0	20.5	19.5	25.0	19.0	20.0
22.0	19.0	21.5	18.0	14.5	17.0	21.0	10.5	18.0	18.0
20.0	13.5	23.5	16.5	19.5					

2. Sample answer: $\bar{x} = 18.462$ and s = 3.308

3. a. Let μ be the mean number of chips per cookie in Nabisco Chips Ahoy chocolate chip cookies. The null and alternative hypotheses are:

$H_0 : \mu = 16$
$H_a : \mu \neq 16$

b. $z = \dfrac{18.462 - 16}{3.308/\sqrt{65}} \approx 6.00$

c. The p-value is essentially 0. Hence, we conclude that the mean number of chips in Chips Ahoy chocolate chip cookies has changed since the 1980s.

4. The confidence interval is $18.462 \pm 1.96\left(\frac{3.308}{\sqrt{65}}\right) = 18.462 \pm 0.804$ or (17.7, 19.3).

Since all the numbers in the interval are larger than 16, it appears that the average number of chips per cookie is greater than 16.

EXERCISE SOLUTIONS

1. a. Let μ be Larry's average miles per gallon when using the new oil. He hopes to show that μ is greater than the 32 mpg he got before switching to the new oil. The hypotheses are:

$H_0 : \mu = 32$
$H_a : \mu > 32$

The null hypothesis says "no change" and the alternative says that mileage has increased.

b. The question asked is "Do students who get credit by the placement exam differ from the usual level, in either direction?" So the alternative hypothesis is two-sided. Take μ to be the average listening score for all students who get credit by the placement exam.
The hypotheses are:

$H_0 : \mu = 24$
$H_a : \mu \neq 24$

c. The student hopes to show that the mice take less time when responding to a loud noise. Let μ be the mean time for mice to run the maze after hearing a loud noise.
The hypotheses are:

$H_0 : \mu = 18$
$H_a : \mu < 18$

2. The null hypothesis states that μ = 115. Because the standard deviation of the population is σ = 30 the standard deviation of the sample mean for n = 25 older students is

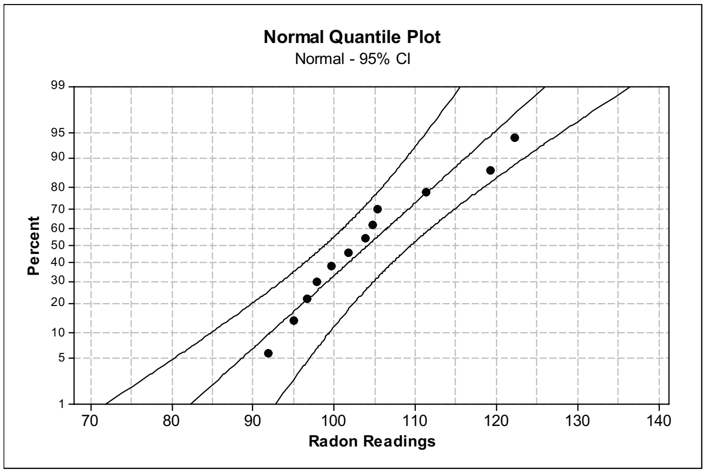
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{25}} = 6 .$$

The value of the z-test statistic is $z = \frac{125.2 - 115}{6} = 1.7$.

The alternative is one-sided on the high side. So, the p-value is the area under a standard normal curve to the right of 1.7, which gives a p-value ≈ 1 – 0.9554 = 0.0446.

A sample of older students would have an average score at least as high as 125.2 in less than 5% of all samples assuming the null hypothesis is true. Because a sample result this high is unlikely to occur just by chance, it is evidence that the mean for all older students is really higher than 115.

3. a. In the normal quantile plot below, all dots lie between the curved bands. So, the normality assumption is reasonably satisfied.



b. The null and alternative hypotheses are:

$H_0 : \mu = 105$
$H_a : \mu \neq 105$

The observed sample mean is $\bar{x} = 104.13$. The value of the test statistic is

$$z = \frac{104.13 - 105}{9/\sqrt{12}} \approx -0.335$$

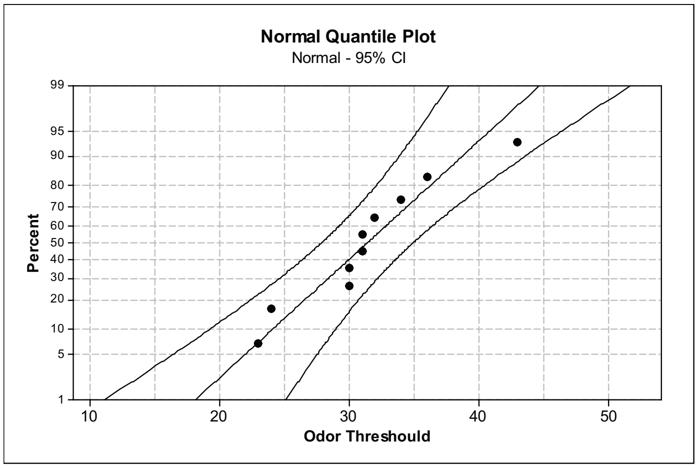
This is a two-sided alternative. Hence the p-value is 2(0.3688) ≈ 0.738. That tells us that if the true mean reading of all radon detectors of this type is 105, we would expect to see an observed *z*-test value at least as extreme as the one we observed about 74% of the time. Based on these data, we have no evidence that the mean reading of these detectors differs from 105.

4. a. Let μ = the mean BMI for 6-year-old girls. The null and alternative hypotheses are $H_0 : \mu = 15.2$ versus $H_a : \mu > 15.2$.

b. $\bar{x} = 16.173$ kg/m^2 ; $s = 2.669$ kg/m^2

c. $z = \frac{16.173 - 15.2}{2.669/\sqrt{30}} \approx 2.00$; $p \approx 0.023$

d. Because $p \approx 0.023 < 0.05$, there is sufficient evidence to reject the null hypothesis and conclude that the mean BMI for 6-year-old girls has increased since the time the CDC collected data to construct its BMI charts.



REVIEW QUESTIONS SOLUTIONS

1. a. The dots in the normal quantile plot are within the curved bands. In addition, the pattern is roughly linear. So, it is reasonable to assume these data are approximately normal.

b. The null and alternative hypotheses are:

$$H_0 : \mu = 25$$
$$H_a : \mu > 25$$

c. The observed sample mean is $\bar{x} = 30.4$. The value of the test statistic is

$z = \dfrac{30.4 - 25}{7/\sqrt{10}} \approx 2.44$; the p-value is 1 – 0.9927 or 0.0073.

That is, an observed average result at least as high as 30.4 would happen only 7 times in 1000 samples if the true mean were really 25. This is unlikely; hence, we have strong evidence that the mean odor threshold for students is higher than 25.

2. a. $\bar{x} = 516.2$ and s = 80.7

b. The null and alternative hypotheses are:

$$H_0 : \mu = 514$$
$$H_a : \mu \neq 514$$

The value of the test statistic is $z = \dfrac{516.2 - 514}{80.7/\sqrt{50}} \approx 0.19$. Since this is a two-sided alternative, the p-value is 2(0.4247) ≈ 0.85. There is insufficient evidence to reject the null hypothesis.

c. The confidence interval is $516.2 \pm 1.96(80.7/\sqrt{50})$ or (493.8, 538.6). Since 514 is in the confidence interval, it is one of the plausible values for μ.

3. a. The null and alternative hypotheses are:

$$H_0 : \mu = 90$$
$$H_a : \mu > 90$$

b. The sample mean and standard deviation are $\bar{x} = 118.4$ and s = 186.5;

$$z = \frac{118.4 - 90}{186.5/\sqrt{100}} \approx 1.52$$

the *p*-value is 0.064. The *p*-value is not sufficiently small to reject the null hypothesis.

c. Yes. Since the *p*-value is less than 0.10, we would conclude that there is sufficient evidence that the mean length of calls coming into the center have increased.

4. a. $H_0 : \mu = 5$ versus $H_a : \mu > 5$. You would want to gather evidence that the mean mercury concentration was above the recommended safe limit before closing the lake to fishing.

b. Sample answer #1: Significance level preference – 0.01. Choosing the smaller significance level means that there must be very strong evidence that the mercury level is unsafe before Lake Natoma is closed to fishing. Closing the lake would anger fishermen and could depress the part of the local economy that depends on fishing.

Sample answer #2: Significance level preference – 0.1. Since eating fish with high mercury concentrations is a serious health risk, it is better to err on the side of caution.