Math 218, Mathematical Statistics
D Joyce, Spring 2016

Chap. 6, page 231: 17–20, 22; and Chap. 7, page 262, exercises 1, 2, 9, 11, 12

Selected answers.

17. In each of the following cases, state the null hypothesis H_0 and the alternative hypothesis H_1 about the population proportion p under test. Interpret p in each instance.

a. The winner in the last congressional election received 54% of the vote. One year before the next election, the congressman hired a polling agency to conduct a sample survey of voters to determine if this proportion has changed.

Since the question is whether or not the proportion p of voters who will vote for the congressman has changed, and not which direction, this is a two-sided hypothesis test with $H_0 : p = 0.54$ and $H_1 : p \neq 0.54$.

b. The proportion of overdue books in a library is 5%. It is proposed to raise the fine for overdue books from 5 cents to 10 cents per day. It is felt that this would reduce the proportion of overdue books.

Here, the question is whether p, the proportion of overdue books, will stay the same or go down, so a one-sided test is called for, in particular, a lower one-sided test with $H_0 : p = 0.05$ and $H_1 : p < 0.05$. You could also state $H_0 : p \geq 0.05$, but H_1 has to be $p < 0.05$.

c. p is the mean scrap rate for the new process. $H_0 : p = 0.40$ and $H_1 : p < 0.40$.

d. p is the mean propportion that prefer Gatorade. Since it is not known which of the two drinks might be favored, this should be a two-sided test, not a one-sided test. $H_0 : p = \frac{1}{2}$ and $H_1 : p \neq \frac{1}{2}$.

18. In each of the following cases, interpret the population mean μ and state H_0 and H_1.

a. You could state your answer either in terms of percents or in terms of grams. Here it is in terms of grams. The population mean μ is the mean grams of fat per cup. $H_0 : \mu = 3.4$ and $H_1 : \mu > 3.4$.

b. The population mean μ is the mean shear strength for a fastener. In both cases mentioned $H_0 : \mu = 10000$.

In the first case, the vendor is new, so, according to an example in the text, the burden of proof is on the vendor to show that the quality standard is met, meaning the shear strength of the vendor's fasteners is high enough. Therefore, $H_1 : \mu > 10000$.

In the second case, the vendor is old with a past history of quality, so $H_1 : \mu < 10000$.

c. The population mean μ is the mean commuting time for the new route. $H_0 : \mu = 25$ and $H_1 : \mu < 25$.

d. μ is the mean difference between the two scores. $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$.

19. In each of the following cases, explain which hypothesis should be set up as H_0 and which as H_1 by deciding which hypothesis, if incorrectly rejected, would result in a more serious error, and make that H_0. State any assumptions you make.

a. Make H_0 (ii) the compound is not safe in the amount normally consumed. You wouldn't want to conclude it is safe when it isn't, because a dangerous compound shouldn't go to market.

b. If (i), the new analgesic is effective, is rejected when it is effective, then a useful medicine wouldn't go to market. If (ii), the new analgesic is not effective, is rejected when it is in fact not effective, then a useless medicine would go to market. But we already have effective analgesics, so make (ii) H_0.

c. If (i), the generic drug is not the same as the original drug, is rejected, then it would go to market and people would use it even though it's not the same. If (ii), it is the same, is rejected, then it wouldn't go to market even though it's just as good, and people would have to continue to pay more for the prescription drug. Better make (i) H_0.

d. If (i), cloud seeding increases precipitation, is rejected, we'd use a way to make it rain. If (ii), cloud seeding doesn't increase precipitation, is rejected, then a useless technique would be put into practice. Is there an alternative to cloud seeding to increase precipitation? How much does that cost? Without more information, it's not clear whether rejecting (i) or rejecting (ii) is a more serious error.

20. Let X be a Bernoulli random variable with probability p of success ($X = 1$) and $q = 1 - p$ of failure ($X = 0$). We want to test $H_0 : p = \frac{1}{4}$ versus $H_1 : p = \frac{3}{4}$.

a. Suppose that based on a single observation X the decision rule is: do not reject H_0 if $X = 0$; reject H_0 if $X = 1$. Find the probabilities of type I and type II errors for this rule.

The probability of a type I error is

$$P(H_0 \text{ is rejected} \mid H_0 \text{ is true}) = P(X = 1 \mid p = \frac{1}{4}) = \frac{1}{4}.$$

The probability of a type II error is

$$P(H_0 \text{ is not rejected} \mid H_0 \text{ is false}) = P(X = 0 \mid p = \frac{3}{4}) = \frac{1}{4}.$$

b. Now suppose that there are two observations instead of one and the decision rule rule is: reject H_0 if both are successes. Find the probabilities of type I and type II errors for this rule.

The probability of a type I error is

$P(H_0 \text{ is rejected}|H_0 \text{ is true})$
$= P(2 \text{ successes} \,|\, p = \frac{1}{4}) = (\frac{1}{4})^2 = \frac{1}{16}$.

The probability of a type II error is

$P(H_0 \text{ is not rejected}|H_0 \text{ is false})$
$= P(\text{not 2 successes} \,|\, p = \frac{3}{4}) = 1 - (\frac{3}{4})^2 = \frac{7}{16}$.

22. Refer to exercise 6.17d. Let p be the proportion of people in the population who prefer Gatorade over All Sport. We want to decide if more people prefer Gatorade. The hypotheses are set up as $H_0 : p \leq \frac{1}{2}$ versus $H_1 : p > \frac{1}{2}$. Fifteen tasters participated in the taste-testing experiment.

a. Suppose that 11 of the 15 tasters preferred Gatorade over All Sport. What is the P-value? Can you reject H_0 at $\alpha = 0.10$?

Let's let X denote the number of tasters who preferred Gatorade over All Sport.

The P-value, also called the observed level of significance, is the smallest α-level at which the observed test result is significant. The smaller the P-value, the more significant the test result. A test at any α greater than P will reject H_0, but an α less than P will not reject H_0. Thus, P is probability under H_0 of obtaining a result at least as extreme as the observed one.

Under H_0, p is $\frac{1}{2}$. What is the probability that 11 or more tasters would prefer Gatorade if $p = \frac{1}{2}$, that is, what is

$$P(X \geq 11)?$$

We could work that out using binomial coefficients, but table A.1 has the information. When $n = 15$ and $p = 0.50$, the table says $P(X \leq 10) = 0.941$. Therefore $P(X \geq 11) = 0.059$. So, $P = 0.059$.

Since $P < 0.10$, H_0 is rejected at $\alpha = 0.10$, that is, the 90% one-sided hypothesis test concludes that p is not less than $\frac{1}{2}$, so Gatorade is preferred over All Sport. (However, a 95% test couldn't make that conclusion.)

b. Suppose that a priori there is no reason to claim that one drink is preferred over the other. We want the taste-testing experiment to tell us if there is a significant difference, and if so, in which direction. Therefore the alternative hypothesis is now two-sided. What is the P-value? Can you reject H_0 at $\alpha = 0.10$?

"At least extreme as" means something different now. What is the probability that 11 or more tasters would prefer one of the two drinks over the other if $p = \frac{1}{2}$? We need to find

$$P(X \geq 11 \text{ or } X \leq 4).$$

With $p = \frac{1}{2}$ we can use symmetry and note that $P(X \geq 11)$ is the same as $P(X \leq 4)$. Therefore, $P = 2 \cdot 0.059 = 0.118$.

Since $P > 0.10$, H_0 is not rejected at $\alpha = 0.10$, that is, the 90% two-sided hypotheses can't conclude that there's a difference between the preferences for the two drinks. (But if one more taster preferred Gatorade, the test would.)

Chapter 7.

2. A textile engineer wants to know how many fabric fibers to test to obtain a 90% confidence interval for the mean tensile strength having a margin of error of no more than 0.5 psi. From past experience it is known that the range of measurements is approximately ± 5 psi around the mean.

a. Calculate the necessary sample size. Use a rough estimate of σ obtained from the range of measurements.

The range of measurements is about ± 5 psi. In an earlier exercise we used such a range to estimate 2σ since about 95% of a normal distribution is within 2σ of its mean. We have no idea with what confidence ± 5 psi was generated, but using this rule of thumb to estimate σ is the best we can do. Thus, we can take $\sigma \approx 2.5$.

Now, the margin of error E for a z-interval is

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

and we want $E = 0.5$. Therefore, we need to solve for n in that equation, and that gives

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2.$$

Since $z_{\alpha/2} = z_{0.05} = 1.645$, therefore

$$n = \left(\frac{1.645 \cdot 2.5}{0.5}\right)^2 = 67.65,$$

so $n = 68$ will do.

9. A consumer watchdog group suspects that a yogurt that is advertised to be 98% fat free has in fact a higher mean content. The group will take action against the company if it can substantiate its suspicion with factual data. For this purpose, the group takes a sample of 25 yogurt cups (each containing 170 grams) and measures the fat contnets. If the company's claim is correct, then the mean fat content should be no more than 2%, i.e., 3.4 grams.

a. Set up the hypotheses to be tested. Explain why you set up the hypotheses the way you did.

Assume the fat content is normally distributed with an unknown mean μ and variance. The null hypothesis is $H_0 : \mu = 3.4$, and since we want to know if μ is actually greater than 3.4, the alternative hypothesis should be $H_1 : \mu > 3.4$.

b. Suppose that the mean fat content for 25 sample cups was 3.6 grams. Also suppose that σ of fat contents is 0.5 grams. [As we see in the next section, we should be using a t-test and we wouldn't need this assumption, which, in fact, is as hard or harder to verify than the value of μ we're looking for.] Do a 0.01-level test of the hypotheses. Is there enough statistical evidence to support the consumer group's suspicion?

This is an upper one-sided z-test, so we reject H_0 if

$$\overline{x} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}.$$

Since $\overline{x} = 3.6$, and

$$\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} = 3.4 + 2.33 \frac{0.5}{\sqrt{25}} = 3.63,$$

we don't reject H_0 at the $\alpha = 0.01$ level.

Alternatively, you could find the P-value and compare it to $\alpha = 0.01$. The P-value, from the table on page 241 is $1 - \Phi(z)$ where

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.6 - 3.4}{0.5/\sqrt{25}} = 2$$

so the P-value is $1 - \Phi(2) = 1 - 0.9772 = 0.0228$, and since that's bigger than 0.1, we can't reject H_0 at the 0.01 level. (But we could at the 0.05 level.)

Whichever way you do it, there's not enough evidence to support the consumer group's suspicion.

c. If the true mean fat content per cup is 3.7 grams, what is the probability that this test will detect it? How many yogurt cups should be tested if this probability must be at least 0.95?

The value $\pi(3.7)$ will give the first probability. From the table on page 247,

$$\begin{aligned} \pi(\mu) &= \Phi\left(-z_\alpha + \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right) \\ &= \Phi\left(-z_{0.01} + \frac{(3.7 - 3.4)\sqrt{25}}{0.5}\right) \\ &= \Phi(-2.33 + 3) = \Phi(0.67) = 0.7486 \end{aligned}$$

So the test will detect $\mu = 3.7$ about 75% of the time.

For the next question, we need to find n so that $\pi(\mu) = 0.95$. That means we need

$$0.95 = \Phi\left(-2.33 + \frac{(3.7 - 3.4)\sqrt{n}}{0.5}\right)$$

which equals $\Phi(-2.33 + 0.6\sqrt{n})$. Now, $\Phi(1.645) = 0.95$, so that means we need $1.645 = -2.33 + 0.6\sqrt{n}$, so $\sqrt{n} = 6.625$ and $n = 43.89$. Therefore, 44 yogurt cups need to be tested.

11. Suppose that 100 random samples of size 25 are drawn from a normal distribution with $\mu = 12$ and $\sigma = 2$.

a. If a 95% z-interval is calculated for each sample, how many of the intervals would you expect to contain the true $\mu = 12$?

The expected number is 95. It doesn't matter what kind of interval or hypothesis test is used, if they're at the 95% confidence level, then over the long run you expect about 95% of the intervals to contain the parameter or 95% of the hypotheses that you reject to be correct rejections.

b. If a 95% t-interval is calculated for each sample, will the answer be different from part a? Why or why not?

Still 95 for the same reason. The intervals will usually be bigger ones than for the z-test, but that's because you know more for the z-test (namely the correct value of σ).

12. A random sample of size 16 is drawn from a normal distribution with $\mu = 70$ and $\sigma = 3$. The mean of the sample is $\overline{x} = 68.45$ and its standard deviation is $s = 1.73$.

a. Calculate a 90% z-interval for μ assuming that you know $\sigma = 3$.

$\alpha/2$ is 5%., so $z_{\alpha/2} = 1.645$. Therefore, the endpoints are

$$\overline{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 68.45 \pm 1.645 \frac{3}{\sqrt{16}} = 68.45 \pm 1.23,$$

so the interval is $[67.22, 69.68]$.

b. Calculate a 90% t-interval for μ assuming that you do not know σ.

Since $t_{15,.05}$ is 1.753 (from the t-table), therefore the endpoints are

$$\overline{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} = 68.45 \pm 1.753 \frac{2.73}{\sqrt{16}} = 68.45 \pm 1.20,$$

so the interval is $[67.25, 69.65]$.

c. Which interval is shorter for this sample? Which interval would be shorter on the average if a large number of samples are drawn form this normal distribution, and z and t intervals are calculated for each sample? Explain.

The t-interval is slightly shorter for $n = 16$. As n grows the two intervals become almost the same.

Math 218 Home Page at
`http://math.clarku.edu/~djoyce/ma218/`