Sampling Distributions and One Sample Tests

So far, we have only talked about hypothesis testing in a very limited set of situations. We will now expand our discussion to cover a much broader array of cases. We begin with single sample tests: Hypotheses which involve tests of only one population. Later, we will talk about tests which involve multiple populations, e.g. comparisons of men and women.

In the first part of this handout, we will discuss three different types of situations involving single sample tests. We will show how our usual 5-step hypothesis testing procedure can be applied. The second part of the handout will show how, when the alternative hypothesis is two-tailed, confidence intervals can also be used for hypothesis testing.

I. SINGLE SAMPLE TESTS FOR $\bar{X}$

A. Case I: Sampling distribution of $\bar{X}$, Normal parent population (i.e. X is normally distributed), σ is known.

1. Suppose that, by some miracle, σ is known but μ isn't. Suppose further that the variable of interest, X, has a normal distribution; or, even if X is not normally distributed, N is large enough so that $\bar{X}$ is normally distributed.

2. Hypothesis testing. When doing hypothesis testing in this case, an appropriate test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{N}}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{N}}}$$

where μ_0 is the value for E(X) specified in the null hypothesis. If H_0 is true, then the test statistic will have a N(0,1) distribution. As we discussed earlier, recall that $\sigma/\sqrt{N}$ is the True Standard Error of the Mean.

EXAMPLES.

1. Suppose Shoe sizes are ~ N(9, 1), and a sample of size n = 16 is drawn.

a. What is the distribution of the sample mean, $\bar{X}$?

b. What is $P(X \geq 11)$?

c. What is $P(\bar{X} \geq 11)$?

SOLUTION.

a. $\bar{X}$ (i.e., $\hat{\mu}$) ~ $N(\mu, \sigma^2/16)$, i.e. $N(9, 1/4^2)$.

b. $P(X \geq 11) = P(Z \geq (11 - 9)/1 = 2) = 1 - F(2) = .0228$

c. $P(\bar{X} \geq 11) = P(Z \geq (11 - 9)/(1/4)) = 1 - F(8) =$ almost 0

Comment. Note that any one shoe has a small chance of being a size 11. However, the likelihood that all 16 shoes could average a size 11 is virtually zero.

2. A manufacture of steel rods considers that the manufacturing process is working properly if the mean length of the rods is 8.6. The standard deviation of these rods always runs about 0.3 inches. Suppose a random sample of size n = 36 yields an average length of 8.7 inches. Should the manufacturer conclude the process is working properly or improperly?

SOLUTION.

Step 1. The null and alternative hypotheses are:

H_0:	$E(X) = 8.6$	(or, $\mu_X = 8.6$)
H_A:	$E(X) <> 8.6$	(or, $\mu_X <> 8.6$)

A two tailed test is called for, since rods are defective if they are too long or too short. Note that we are not told whether or not X is normally distributed. Nevertheless, because of the central limit theorem, we know that, if H_0 is true, the sample mean $\overline{X}$ is probably ~ $N(8.6, .3^2/36)$.

Step 2. The appropriate test statistic is

$$z = \frac{\overline{x} - \mu_0}{\sqrt{\frac{\sigma^2}{N}}} = \frac{\overline{x} - 8.6}{\sqrt{\frac{.3^2}{36}}} = \frac{\overline{x} - 8.6}{.05}$$

Note also that

$$\overline{x} = z * \sqrt{\frac{\sigma^2}{N}} + E(X\ )_0 = z * .05 + 8.6$$

Step 3. Critical region. For $\alpha = .05$, accept H_0 if $-1.96 \le Z \le 1.96$; values outside this range fall in the critical region and mean that H_0 should be rejected. Converting these z-values into corresponding values for $\overline{X}$, this means we accept H_0 if $8.502 \le \overline{X} \le 8.698$ (since $-1.96 * .05 + 8.6 = 8.502$, and $1.96 * .05 + 8.6 = 8.698$)

Step 4. The computed value of the test statistic is

$$z = \frac{\overline{x} - \mu_0}{\sqrt{\frac{\sigma^2}{N}}} = \frac{8.7 - 8.6}{.05} = 2.0$$

Step 5. Decision. At the .05 level of significance, reject H_0 (i.e. conclude that the process is not functioning correctly). A sample result this far away from the hypothesized mean would occur only 4.56% of the time if H_0 were true. (Note that we would not reject H_0 if we were using the .01 level of significance.)

B. Case II: Sampling distribution for the binomial parameter p, using the normal approximation.

1. We have already dealt with this case at length. Let X_1 - X_N be independent and identically distributed Bernoulli trials. In the past, we have let X = the number of successes out of N trials. Instead, we will now say that X_1 - X_N have the same distribution as X. Thus, $E(\bar{X}) = \mu_X = p$, and $V(\bar{X}) = \sigma^2_X = pq$. Thus, for N sufficiently large, $\bar{X} \sim N(\mu, \sigma^2/N)$, i.e. $\bar{X} \sim N(p, pq/N)$.

2. <u>Hypothesis testing</u>. As noted before, the <u>appropriate test statistic</u> is

$$z = \frac{\hat{p} \pm CC/N - p_0}{\frac{\sqrt{p_0 q_0}}{\sqrt{N}}} = \frac{\hat{p} \pm CC/N - p_0}{\sqrt{\frac{p_0 q_0}{N}}}$$

where CC = the correction for continuity. To make the correction for continuity, add .5/N to the sample value of $\hat{p}$ when $\hat{p} < p_0$, subtract .5/N from $\hat{p}$ when $\hat{p} > p_0$, and do nothing if $\hat{p} = p_0$. If the null hypothesis is true, then the test statistic will be distributed N(0,1). Also, if the null hypothesis is true, $\sqrt{\frac{p_0 q_0}{N}}$ is the <u>True Standard Error of the Mean</u>.

3. Rather than give additional examples, simply refer back to the previous handout on hypothesis testing. All the examples where we expressed hypotheses and solved problems using the binomial parameter p fall under this case.

4. Note that the Central Limit Theorem provides us with the justification for using the Normal approximation to the binomial. As N becomes large, the sampling distribution for $\bar{X}$ more and more approximates a normal distribution.

5. Finally, note that, for normally distributed variables, the mean and variance are independent of each other. However, with a binomially distributed variable, the mean and variance are not independent, since $\mu = p$ and $\sigma^2 = pq$, that is, both the mean and variance depend on the parameter p.

C. Case III: Sampling distribution of $\bar{X}$, normal parent population, σ unknown.

1. In Case I, we assumed that σ was known, even though μ was not. In reality, such a situation would be extremely rare. Much more common is the situation where both μ and σ are unknown. What happens if we do not want to (or cannot) assume a value for σ (i.e. σ is unknown)? When σ is unknown, we substitute the sample variance s^2; and instead of doing a Z transformation which produces a N(0,1) variable, we do a T transformation which

produces a variable with a T distribution. Other than always having to look up values in the table, working with the T distribution is pretty much the same as working with the N(0, 1) distribution.

2. Hypothesis testing: When sampling from a normal population, σ unknown, the appropriate test statistic is

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{N}}} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}}$$

If the null hypothesis is true, then the test statistic will have a T_{N-1} distribution. Also, recall that $s/\sqrt{N}$ is the Estimated Standard Error of the Mean.

EXAMPLES:

1. A Bowler claims that she has a 215 average. In her latest performance, she scores 188, 214, and 204.

a. Calculate the sample mean, variance, and standard deviation

b. What is the probability the sample mean would be this low if the bowler is right? (Assume that her bowling scores are normally distributed.)

c. Would you conclude the bowler is "off her game?"

SOLUTION.

a. Sample mean, variance, and standard deviation:

$$\bar{X} = \frac{1}{N}\sum X_i = \hat{\mu} = \frac{188 + 214 + 204}{3} = 202,$$

$$s^2 = \frac{1}{N-1}\sum(x_i - \bar{x})^2 = \hat{\sigma}^2 = \frac{(188-202)^2 + (214-202)^2 + (204-202)^2}{2} = 172,$$

$$s = \hat{\sigma} = \sqrt{172} = 13.11$$

b. Note that

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{N}}} = \frac{202 - 215}{\frac{13.11}{\sqrt{3}}} = \frac{-13}{7.57} = -1.72$$

Hence, if $\mu = 215$, $P(\bar{X} \le 202) = P(T_2 \le -1.72)$ = somewhere between .1 and .25. (Look at v = 2; $P(T_2 \le -1.886) = .1$, $P(T_2 \le -0.816) = .25$, so, since -1.72 is between -.816 and -1.886, $.1 \le P(T_2 \le -1.72) \le .25$.

c. The bowler may be off her game. At least, we would not reject her claim that she has a 215 average if we were using the .05 level of significance.

2. The University contends that the average graduate student makes at least $8,000 a year. A random sample of 6 students has an average income (measured in thousands of dollars) of 6.5 and a sample variance of 2. Test the university's claim at the .05 and .01 levels of significance.

Step 1: The null and alternative hypotheses are:

H_0: E(X) = 8 (or $\mu_X = 8$)
H_A: E(X) < 8 (or $\mu_X < 8$)

Step 2: Since the population variance is unknown, an appropriate test statistic is

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}} = \frac{\bar{x} - 8}{\sqrt{\frac{s^2}{6}}}$$

Step 3. At the .05 level of significance, accept H_0 if $T_5 \ge -2.015$. At the .01 level, accept H_0 if $T_5 \ge -3.365$.

Step 4. The test statistic equals

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}} = \frac{\bar{x} - 8}{\sqrt{\frac{s^2}{6}}} = \frac{6.5 - 8}{\sqrt{\frac{2}{6}}} = -2.598$$

Step 5. Decision. For $\alpha = .05$, we reject H_0, because -2.598 falls outside the acceptance region. But, for $\alpha = .01$, we do not reject.

Comment: Note that, at Step 2, I did not fill in the value for s^2; and, at step 3, I did not tell what values of $\bar{x}$ would lead to acceptance or rejection of the null hypothesis. This is because information about s^2 would not be available until after the data had been analyzed. By way of contrast, with the other 2 cases σ is either known (Case I) or else implied by the null hypothesis (Case II). Remember, you should ideally go through steps 1, 2, and 3 before you know anything about the results from the data.

II. HYPOTHESIS TESTING USING CONFIDENCE INTERVALS.

A. Review of confidence intervals. Recall that the particular value chosen as most likely for a population parameter is called the point estimate. Because of sampling error, we know the point estimate probably is not identical to the population parameter. Hence, we usually specify a range of values in which the population parameter is likely to be.

B. Hypothesis testing. A nice feature of confidence intervals is that they can be easily used for hypothesis testing. One can think of the confidence interval as containing all the hypotheses about E(X) that could not be rejected at the α level of significance (two-tailed) in light of the evidence $\bar{x}$. That is, if H_A is a *two-tailed alternative* to H_0, then H_0 will not be rejected at the α level of significance if the value it specifies falls within the $100(1-\alpha)\%$ confidence interval. If the value specified by H_0 does not fall within the $100(1-\alpha)\%$ confidence interval, H_0 will be rejected.

EXAMPLE. Suppose you are testing

H_0: $\mu = 10$.
H_A: $\mu <> 10$.

Suppose the 95% confidence interval ranges from 8 to 12. Since μ_0 falls within the confidence interval, you will not reject the null hypothesis (at the .05 level of significance).

Suppose instead that the confidence interval ranges from 11 to 15. Since μ_0 falls outside this range, you will reject the null hypothesis.

C. Examples:

1. Suppose $\bar{x} = 100$, $\sigma = 36$, $n = 9$. This falls under Case I above, so the 95% c.i. for E(X) is

$$100 \pm (1.96 * 36/\sqrt{9}), \text{ or } 76.48 \le E(X) \le 123.52$$

Now, suppose we want to test, at the .05 level,

H_0: $E(X) = 125$
H_A: $E(X) <> 125$

Since the hypothesized mean of 125 does NOT fall in the 95% c.i. for E(X), we will reject H_0 at the .05 level of significance.

To double-check, let us solve this problem the same way we have done it in the past.

Since we know σ, the appropriate test statistic is

$$Z = (\overline{X} - E(X)_0)/(\sigma/\sqrt{n}) = (\overline{X} - 125)/12$$

and the acceptance region is

$$-1.96 \le z \le 1.96; \text{ equivalently, converting the z's back to x's:}$$

$$\mu_0 - (z_{\alpha/2} * \sigma/\sqrt{N}) \le \overline{x} \le \mu_0 + (z_{\alpha/2} * \sigma/\sqrt{N}), \text{ i.e.,}$$

$$125 - 1.96 * 36/3 \le \overline{x} \le 125 + 1.96 * 36/3, \text{ i.e.,}$$

$$101.48 \le \overline{x} \le 148.52$$

Since $\overline{X} = 100$, it does not fall in the acceptance region, and H_0 should be rejected. (Also, z_c = -2.08, which is outside the acceptance region.)

2. Let us once again consider the following problem: A manufacture of steel rods considers that the manufacturing process is working properly if the mean length of the rods is 8.6. The standard deviation of these rods always runs about 0.3 inches. Suppose a random sample of size n = 36 yields an average length of 8.7 inches. Should the manufacturer conclude the process is working properly or improperly? This time, use confidence intervals to solve the problem.

Solution: This again falls under Case I. Hence, the 95% c.i. is

$$8.7 \pm (1.96 * .3/\sqrt{36}), \text{ i.e. } 8.602 \le E(X) \le 8.798.$$

The hypothesized mean of 8.6 is outside the c.i., so reject H_0. (Note that this is the same conclusion we reached when we examined this problem before.)

3. Let us again consider a slightly modified version of the case of Notre Dame and the underpaid graduate students: The Deans contend that the average graduate student makes $8,000 a year. Zealous administration budget cutters contend that the students are being paid more than that, while the Graduate Student Union contends that the figure is less. A random sample of 6 students has an average income (measured in thousands of dollars) of 6.5 and a sample variance of 2. Use confidence intervals to test the Deans' claim at the .10 and .02 levels of significance.

Solution. This falls under Case III, since σ is unknown. $t_{.10/2,5}$ = 2.015, $t_{.02/2,5}$ = 3.365. Hence, for α = .10, the 90% confidence interval is

$$6.5 \pm (2.015 * \text{sqrt}(2/6)), \text{ i.e. } 5.34 \le E(X) \le 7.66.$$

Hence, using $\alpha = .10$, we reject the Dean's claim. For $\alpha = .02$, the 98% c.i. is

$$6.5 \pm (3.365 * \text{sqrt}(2/6)), \text{ i.e. } 4.56 \le E(X) \le 8.44.$$

Hence, using $\alpha = .02$, we do not reject the Deans' claim.

4. The mayor contends that 25% of the city's employees are black. Various left-wing and right-wing critics have claimed that the mayor is either exaggerating or understating the number of black employees. A random sample of 120 employees contains 18 blacks. Use confidence intervals to test the mayor's claim at the .01 level of significance.

Solution. This falls under case II, binomial parameter p. $\hat{p} = .15$, n = 120. Using the formula for the approximate c.i., we get

$$.15 \pm 2.58 * \text{sqrt}((.15*.85)/120), \text{ i.e. } .0659 \le p \le .2341.$$

Hence, according to the approximate c.i., we should reject H_0. Let us now confirm our results using our standard hypothesis testing procedure:

Step 1.

H_0: p = .25
H_A: p <> .25

Step 2. The appropriate test statistic is

$$z = \frac{\hat{p} \pm CC/N - p_0}{\sqrt{\frac{p_0 q_0}{N}}} = \frac{\hat{p} \pm CC/120 - .25}{.03953}$$

Step 3. For $\alpha = .01$, accept H_0 if $-2.58 \le z \le 2.58$. Equivalently, accept H_0 if

$$p_0 - CC/N - z_{\alpha/2}\sqrt{\frac{p_0 q_0}{N}} \le \hat{p} \le p_0 + CC/N + z_{\alpha/2}\sqrt{\frac{p_0 q_0}{N}}, i.e.$$

$$.25 - .5/120 - 2.58\sqrt{\frac{.25 * .75}{120}} \le \hat{p} \le .25 + .5/120 + 2.58\sqrt{\frac{.25 * .75}{120}}, i.e.$$

$$.1438 \le \hat{p} \le .3562$$

Step 4. The test statistic equals

$$z = \frac{\hat{p} \pm CC/N - p_0}{\sqrt{\frac{p_0 q_0}{N}}} = \frac{.15 + .5/120 - .25}{.03953} = -2.42$$

Step 5. Do not reject the null hypothesis. The computed test statistic falls within the acceptance region.

Whoops! We have an inconsistency. When we use the approximate confidence interval, we reject H_0, but when we use our usual hypothesis testing procedure we do not reject. Looks like we'll have to break down and compute the more precise Wilson 99% confidence interval.

$$\frac{N}{N + z^2_{\alpha/2}} \left[\hat{p} + \frac{z^2_{\alpha/2}}{2N} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{N} + \frac{z^2_{\alpha/2}}{4N^2}} \right], i.e.$$

$$\frac{120}{120 + 2.58^2} \left[.15 + \frac{2.58^2}{240} - 2.58 \sqrt{\frac{.15 * .85}{120} + \frac{2.58^2}{57{,}600}} \right] \le p \le$$

$$\frac{120}{120 + 2.58^2} \left[.15 + \frac{2.58^2}{240} + 2.58 \sqrt{\frac{.15 * .85}{120} + \frac{2.58^2}{57{,}600}} \right], i.e.$$

$$.0845 \le p \le .2523$$

We should not reject H_0. Hence, while the approximation usually works ok, it can sometimes lead us astray. In this case, while N was large, it wasn't quite large enough, because p substantially differs from .5.

D. Confidence intervals compared to 2-tailed acceptance regions: We have previously specified acceptance regions when deciding whether to accept or reject 2-tailed alternative hypotheses. We are now saying that confidence intervals can be used for the same purpose. In fact, the two approaches, though not identical, share many things in common.

The following table lists the confidence intervals and the 2-tailed acceptance regions (expressed in the metric of the original variable rather than in standardized form). Note how similar the formulas are.

Case	Confidence Interval	2-tailed acceptance region (non-standardized)
I: Population normal, σ known:	$\bar{x} \pm (z_{\alpha/2} * \sigma/\sqrt{N})$, i.e., $\bar{x} - (z_{\alpha/2} * \sigma/\sqrt{N}) \le \mu \le \bar{x} + (z_{\alpha/2} * \sigma/\sqrt{N})$	$\mu_0 \pm (z_{\alpha/2} * \sigma/\sqrt{N})$, i.e., $\mu_0 - (z_{\alpha/2} * \sigma/\sqrt{N}) \le \bar{x} \le \mu_0 + (z_{\alpha/2} * \sigma/\sqrt{N})$
II: Binomial parameter p (approximate CI)	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{N}}$, i.e. $\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{N}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{N}}$	$p_0 \pm CC/N \pm z_{\alpha/2}\sqrt{\frac{p_0 q_0}{N}}$, i.e. $p_0 - CC/N - z_{\alpha/2}\sqrt{\frac{p_0 q_0}{N}} \le \hat{p} \le p_0 + CC/N + z_{\alpha/2}\sqrt{\frac{p_0 q_0}{N}}$
III: Population normal, σ unknown	$\bar{x} \pm (t_{\alpha/2,v} * s/\sqrt{N})$, i.e. $\bar{x} - (t_{\alpha/2,v} * s/\sqrt{N}) \le \mu \le \bar{x} + (t_{\alpha/2,v} * s/\sqrt{N})$	$\mu_0 \pm (t_{\alpha/2,v} * s/\sqrt{N})$, i.e. $\mu_0 - (t_{\alpha/2,v} * s/\sqrt{N}) \le \bar{x} \le \mu_0 + (t_{\alpha/2,v} * s/\sqrt{N})$

Hence, with an acceptance region, you start with a hypothesized value for μ_0, and then determine which sample values might reasonably occur if μ_0 is correct. You then see whether the actual sample mean falls within this range. Confidence intervals work in just the opposite direction. With a confidence interval, you first compute the sample mean, and then determine which values for μ could reasonably account for the observed results. If μ_0 falls within this range, you accept the null hypothesis. *The use of either confidence intervals or acceptance regions should lead to the same conclusion when testing a two-tailed alternative.*

Warning: Remember, you only use confidence intervals for hypothesis testing if the alternative hypothesis is 2-tailed.