Math 3081 (Probability and Statistics)

Lecture #19 of 27 ∼ August 5th, 2021

More z Tests

- One-Sample z Tests
- z Tests for Unknown Proportion

This material represents §4.1.2-4.1.3 from the course notes, and problems 8-11 from WeBWorK 6.

Hypothesis Testing Procedure, I

Recall our general framework for performing hypothesis tests:

1. Identify the null and alternative hypotheses for the given problem, and select a significance level α.
2. Identify the most appropriate test statistic and its distribution according to the null hypothesis (usually, this is an average or occasionally a sum of the given data values) including all relevant parameters.
3. Calculate the p-value: the probability that a value of the test statistic would have a value at least as extreme as the value observed.
4. Determine whether the p-value is less than the significance level α (reject the null hypothesis) or greater than or equal to the significance level α (fail to reject the null hypothesis).

Hypothesis Testing Procedure, II

Here is the more detailed procedure for a one-sample z-test:

- First, we must identify the appropriate null and alternative hypotheses and select a significance level α.
- We will use the test statistic $\hat{\mu}$, the sample mean, since this is the minimum-variance unbiased estimator for the true population mean μ.
- Then the null hypothesis will be of the form $H_0 : \mu = c$, for some specific value of c.
- Under the assumption that H_0 is true, the test statistic is normally distributed with mean μ (the true mean postulated by the null hypothesis) and standard deviation σ (which we must be given).

Hypothesis Testing Procedure, III

Once we have written down the test statistic, we can compute the p-value:

- If the hypotheses are $H_0 : \mu = c$ and $H_a : \mu > c$, then the p-value is $P(N_{\mu,\sigma} \geq z)$.
- If the hypotheses are $H_0 : \mu = c$ and $H_a : \mu < c$, then the p-value is $P(N_{\mu,\sigma} \leq z)$.
- If the hypotheses are $H_0 : \mu = c$ and $H_a : \mu \neq c$, it is
$$P(|N_{\mu,\sigma} - \mu| \geq |z - \mu|) = \begin{cases} 2P(N_{\mu,\sigma} \geq z) & \text{if } z \geq \mu \\ 2P(N_{\mu,\sigma} \leq z) & \text{if } z < \mu \end{cases}.$$
- In each case, we are simply calculating the probability that the normally-distributed random variable $N_{\mu,\sigma}$ will take a value further from the hypothesized mean μ (in the direction of the alternative hypothesis, as applicable) than the observed test statistic z.

Hypothesis Testing Procedure, IV

Finally, once we compute the p-value, we compare it to the significance level α.

- If $p < \alpha$, we reject the null hypothesis. Our interpretation is that the test statistic is so far away from the prediction that it could not reasonably have happened by chance (for "reasonable" as defined by the significance level α).
- If $p \geq \alpha$, we fail to reject the null hypothesis. Our interpretation is that the test statistic is close enough from the prediction that it could reasonably have happened by chance (again, for "reasonable" as defined by the significance level α).

One-Sample z Tests, I

Example: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.76 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

1. Whether the average in Class B is greater than 81 points.
2. Whether the average in Class A is greater than 81 points.

One-Sample z Tests, VI

<u>Example</u>: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.76 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

1. Whether the average in Class B is greater than 81 points.

- Like before, we want $H_0 : \mu_B = 81$. The question is which alternative hypothesis we want.

One-Sample z Tests, VI

Example: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.76 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

1. Whether the average in Class B is greater than 81 points.

- Like before, we want $H_0 : \mu_B = 81$. The question is which alternative hypothesis we want.
- Here, because the actual average is 81.76 points (greater than 81), we take the alternative hypothesis $H_a : \mu_B > 81$.
- The test statistic will be normally distributed with mean 81 (per H_0) and standard deviation $5/\sqrt{100} = 0.5$.
- Thus, the p-value is $P(N_{81,0.5} \geq 81.76) = 0.0643$.
- This is statistically significant at the 10% significance level (so we reject the null there) but not at the 3% significance level (so we fail to reject the null in this case).

One-Sample z Tests, II

Example: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.76 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

2. Whether the average in Class A is greater than 81 points.

One-Sample z Tests, II

Example: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.76 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

2. Whether the average in Class A is greater than 81 points.

- We want to test the hypothesis that $\mu_A \geq 81$. However, we cannot take this as our null hypothesis: it needs to specify a value for μ_A, so the only sensible choice is $H_0 : \mu_A = 81$.
- For the alternative hypothesis, we clearly want it to be one-sided; the only question is which direction.
- We will try both possibilities $H_a : \mu_A > 81$ and $H_a : \mu_A < 81$ to try to understand which one of them we actually want.

One-Sample z Tests, III

Example: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.16 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

2. Whether the average in Class A is greater than 81 points.

- For $H_0 : \mu_A = 81$, $H_a : \mu_A > 81$, our test statistic is again the average score in class A.
- The test statistic will be normally distributed with mean 81 (per H_0) and standard deviation $5/\sqrt{64} = 0.625$.
- Thus, the p-value is $P(N_{81,0.625} \geq 80.05) = 0.9357$.
- This is (extremely!) not statistically significant at the 10% and 3% significance levels, so we fail to reject the null hypothesis in this case.

One-Sample z Tests, IV

<u>Example</u>: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.76 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

2. Whether the average in Class A is greater than 81 points.

- For $H_0 : \mu_A = 81$, $H_a : \mu_A < 81$, everything is the same as what we just did, except now the tail of the distribution is the "other side".
- The p-value is now $P(N_{81,0.625} \leq 80.05) = 0.0643$.
- This is statistically significant at the 10% significance levels (so we reject the null hypothesis in this case) but not at the 3% significance level (so we fail to reject the null hypothesis there).

One-Sample z Tests, V

<u>Example</u>: A sample from Class A has 64 students and average score 80.05 points. A sample from Class B has 100 students and average 81.76 points. Assume the standard deviation is known to be 5 points. Test at the 10% and 3% significance levels

2. Whether the average in Class A is greater than 81 points.

- For $H_0 : \mu_A = 81$, $H_a : \mu_A > 81$, we got a p-value of 0.9357. We interpret this result as providing essentially zero evidence to disprove the statement that the true mean of A is 81 or below.
- For $H_0 : \mu_A = 81$, $H_a : \mu_A < 81$, we got a p-value of 0.0643. We interpret this result as providing relatively strong evidence against the statement that the true mean of A is 81 or above.
- The second statement has more content, since we are explicitly rejecting the null hypothesis there.

Interlude: Which Side Is The Right Side?, I

In this last example, the question of which one-sided alternative hypothesis we wanted to test depended on the actual result of doing the tests.

- Either alternative hypothesis $H_a : \mu_A > 81$ and $H_a : \mu_A < 81$, depending on whether we reject or fail to reject their corresponding null hypotheses, could potentially be the one that carries more useful information.
- Although it might seem obvious that when we fail to reject the null hypothesis with $p = 0.9357$ for $H_a : \mu_A > 81$ that it is really saying that the true average should be less than 81, that is *not* how we can interpret the result of the test.
- All we can say from failing to reject the null hypothesis is that we have weak or no evidence suggesting that the true average is not 81.

Interlude: Which Side Is The Right Side?, II

In general, we interpret "rejecting the null hypothesis" as a much stronger statement than "failing to reject the null hypothesis".

- This is because rejecting the null hypothesis takes substantial evidence, since the p-value must be less than the significance level α (usually a stringent requirement).
- On the other hand, failing to reject the null hypothesis does not take substantial evidence.

Thus, the version of the alternative hypothesis in which we reject the null hypothesis (if there is one) is usually the one we will want to discuss.

- For our one-sample z tests, that will be the one-sided version that is on "the same side" as our sample statistic, relative to the hypothesized value.

Interlude: Which Side Is The Right Side?, III

This business of which alternative to use is a bit subtle, so I want to reiterate: one should **never, ever, ever** try to fit a hypothesis to the data in order to get a good result. But that is not what is being done here.

- In this specific situation, both of the one-sided alternative hypothesis tests contain the same information, and doing one test is equivalent to doing the other (since the p-values for the two one-sided alternatives will always sum to 1).
- In order to report the results in the most useful way possible, we would want to quote the alternative hypothesis with the smaller p-value, since that is the side that tells us the most information.

I will mention that different people have different conventions, so if you prefer to select your alternative hypothesis in some other way, that is fine: that's why we always write down the hypotheses!

Two-Sample z Tests, I

In some situations, we want to compare two quantities to decide whether one of them is larger than the other.

- In situations where both quantities are normally distributed and independent, we can make this decision by analyzing the difference between the two quantities, which will also be normally distributed.
- We can then apply the same decision procedures described before to test the appropriate null hypothesis about the value of the difference of the quantities.
- Because there are now two samples involved and we are studying the properties of a normally distributed test statistic z, this method is referred to as a <u>two-sample z-test</u>.

Two-Sample z Tests, II

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

1. whether the two class averages are equal.
2. whether the average score in class A is lower than the score in class B.

Two-Sample z Tests, II

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

1. whether the two class averages are equal.
2. whether the average score in class A is lower than the score in class B.

- This is a two-sample z test because we are comparing the averages of two normally-distributed quantities.
- We first identify the null and alternative hypotheses, then calculate the test statistic (the difference in the average scores) and its hypothesized distribution.
- Then we find the p-value and compare it to the significance level.

Two-Sample z Tests, III

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

1. whether the two class averages are equal.

Two-Sample z Tests, III

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

1. whether the two class averages are equal.

- Let μ_A and μ_B be the respective class averages.
- Since our testing procedure requires testing the distribution of a specific quantity, we write our hypotheses as $H_0 : \mu_A - \mu_B = 0$ and $H_a : \mu_A - \mu_B \neq 0$.
- Our test statistic is $z = 80.25 - 81.16 = -0.91$ points, the difference in the two class averages.
- Under the null hypothesis, $\mu_A - \mu_B$ is normal with mean 0 points and standard deviation $\sqrt{\sigma_A^2 + \sigma_B^2} = \sqrt{0.625^2 + 0.5^2} = 0.8004$ points. (Remember we found σ_A and σ_B earlier.)

Two-Sample z Tests, IV

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

1. whether the two class averages are equal.

Two-Sample z Tests, IV

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

1. whether the two class averages are equal.

- Our test statistic is normal with mean 0 and standard deviation 0.8004, and the observed value is -0.91.
- Thus, because our alternative hypothesis is $H_a : \mu_A - \mu_B \neq 0$ (which is two-sided), the p-value is $P(|N_{0,0.8004}| \geq 0.91) = 2 \cdot P(N_{0,0.8004} \leq -0.91) = 2 \cdot P(N_{0,1} \leq -1.1369) = 0.2556$.
- Since the p-value is relatively large, it is not significant at either the 10% or 3% significance level, and we accordingly fail to reject the null hypothesis in both cases.

Two-Sample z Tests, V

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

2. whether the average score in class A is lower than the score in class B.

Two-Sample z Tests, V

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

2. whether the average score in class A is lower than the score in class B.

- In this case the null hypothesis is the same, but because the sampled average score in class A is actually lower than the sampled average in class B, we want the alternative hypothesis to agree.
- Thus, we take $H_0 : \mu_A - \mu_B = 0$ and $H_a : \mu_A - \mu_B < 0$.

Two-Sample z Tests, VI

Example: A sample from Class A has 64 students (avg score 80.25 pts) and a sample from Class B has 100 students (avg score 81.16 pts). The standard deviation for any given student's score is 5 points. Test at the 10% and 3% significance levels

2. whether the average score in class A is lower than the score in class B.

- As before, the test statistic is normal with mean 0 and standard deviation 0.8004, and the observed value is -0.91.
- Per the alternative hypothesis, the p-value is now $P(N_{0,0.8004} < -0.91) = P(N_{0,1} \leq -1.1369) = 0.1278$.
- Since the p-value is still larger than the significance levels, we still fail to reject the null hypothesis in both cases.

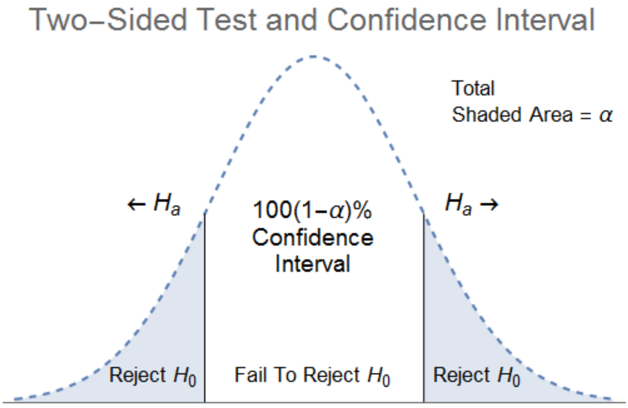
z Tests and Confidence Intervals, I

We will also mention that the results of a z test can also be interpreted in terms of confidence intervals.

- For a two-sided alternative hypothesis, if we give a $100(1-\alpha)\%$ confidence interval around the mean of a distribution under the conditions of the null hypothesis, then we will reject the null hypothesis with significance level α precisely when the sample statistic lies outside the confidence interval.
- The $100(1-\alpha)\%$ confidence interval is precisely giving the range of values around the null hypothesis sample statistic that we would believe are likely to have occurred by chance, in the sense that if we repeated the experiment many times, then we would expect a proportion $1-\alpha$ of the results to land inside the confidence interval.

z Tests and Confidence Intervals, II

See the diagram:



z Tests and Confidence Intervals, III

- Intuitively, the $100(1-\alpha)\%$ confidence interval is giving the range of values around the null hypothesis sample statistic that we would believe are likely to have occurred by chance, in the sense that if we repeated the experiment many times, then we would expect a proportion $1-\alpha$ of the results to land inside the confidence interval.
- If we interpret this probability as an area, what this means is that we would expect to see a test statistic "far away" from the null hypothesis value only with probability α: if we do obtain such an extreme value as our test statistic, we should take this as strong evidence (at the significance level α) that the true test statistic does not align with the prediction from the null hypothesis.

z Tests and Confidence Intervals, IV

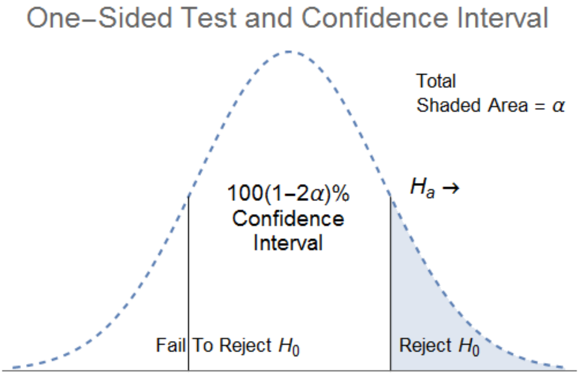
One minor caveat....

- Instead of quoting a confidence interval around the null-hypothesis prediction, we usually quote a confidence interval around the test statistic instead, and then check whether the null-hypothesis prediction lies within the confidence interval around the test statistic.

We can also make a similar interpretation for a one-sided alternative hypothesis, but because of the lack of symmetry in the rejection region, we instead need to use a $100(1 - 2\alpha)\%$ confidence interval to get the correct area.

z Tests and Confidence Intervals, V

Here is the picture for a one-sided alternative hypothesis:



The shaded region has area α, and there is a second region also of area α on the other side of the confidence interval, so the total area inside the confidence interval is $1 - 2\alpha$, meaning it is a $100(1 - 2\alpha)\%$ confidence interval.

z Tests and Confidence Intervals, VI

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points:

1. Find 90% and 97% CIs for the true average of Class A.
2. Find 90% and 97% CIs for the true average of Class B.
3. Find 90% and 97% CIs for the difference between the averages of the two classes.

Then test at the 10% and 3% significance levels

4. that the average of Class A is 80 points.
5. that the average of Class A is 79 points.
6. that the average of Class B is 80 points.
7. that the average of Class B is 82 points.
8. that the average scores in the classes are equal.
9. that the average score in Class A is 1 point greater than the average in Class B.

z Tests and Confidence Intervals, VI

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points:

1. Find 90% and 97% CIs for the true average of Class A.

[Note that the number of standard deviations for a 90% CI is $c = 1.6449$ and for a 97% CI it is $c = 2.1701$.]

z Tests and Confidence Intervals, VI

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points:

1. Find 90% and 97% CIs for the true average of Class A.

[Note that the number of standard deviations for a 90% CI is $c = 1.6449$ and for a 97% CI it is $c = 2.1701$.]

- As we calculated before, the estimator for the mean of Class A has $\hat{\mu}_A = 80.25$ and $\sigma_A = 5/\sqrt{64} = 0.625$.
- Thus, the 90% confidence interval for the mean is $80.25 \pm 1.6449 \cdot 0.625 = (79.22, 81.28)$,
- Also, the 97% confidence interval is $80.25 \pm 2.1701 \cdot 0.625 = (78.89, 81.61)$.

z Tests and Confidence Intervals, VII

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points:

2. Find 90% and 97% CIs for the true average of Class B.

[Note that the number of standard deviations for a 90% CI is $c = 1.6449$ and for a 97% CI it is $c = 2.1701$.]

z Tests and Confidence Intervals, VII

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points:

2. Find 90% and 97% CIs for the true average of Class B.

[Note that the number of standard deviations for a 90% CI is $c = 1.6449$ and for a 97% CI it is $c = 2.1701$.]

- As we calculated before, the estimator for the mean of Class A has $\hat{\mu}_B = 81.16$ and $\sigma_B = 5/\sqrt{100} = 0.5$.
- Thus, the 90% confidence interval for the mean is $81.16 \pm 1.6449 \cdot 0.5 = (80.34, 81.98)$, and the 97% confidence interval is $81.16 \pm 2.1701 \cdot 0.5 = (80.07, 82.25)$.

z Tests and Confidence Intervals, VIII

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points:

3. Find 90% and 97% CIs for the difference between the averages of the two classes.

[Note that the number of standard deviations for a 90% CI is $c = 1.6449$ and for a 97% CI it is $c = 2.1701$.]

z Tests and Confidence Intervals, VIII

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points:

3. Find 90% and 97% CIs for the difference between the averages of the two classes.

[Note that the number of standard deviations for a 90% CI is $c = 1.6449$ and for a 97% CI it is $c = 2.1701$.]

- As we calculated before, the estimator for the difference in the means has $\hat{\mu}_{A-B} = -0.91$ and $\sigma_{A-B} = \sqrt{0.625^2 + 0.5^2} = 0.8004$.
- Thus, the 90% confidence interval for the difference in the means is $-0.91 \pm 1.6449 \cdot 0.8004 = (-2.23, 0.41)$, and the 97% confidence interval is $-0.91 \pm 2.1701 \cdot 0.8004 = (-2.65, 0.83)$.

z Tests and Confidence Intervals, IX

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

4. that the average of Class A is 80 points.

- The 90% CI was $(79.22, 81.28)$, while the 97% CI was $(78.89, 81.61)$.

z Tests and Confidence Intervals, IX

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

4. that the average of Class A is 80 points.

- The 90% CI was $(79.22, 81.28)$, while the 97% CI was $(78.89, 81.61)$.
- Since 80 lies inside both confidence intervals, the result is not significant at either the 10% or 3% significance levels: we fail to reject the null hypothesis that the true mean is 80 points.

z Tests and Confidence Intervals, X

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

5. that the average of Class A is 79 points.

- The 90% CI was $(79.22, 81.28)$, while the 97% CI was $(78.89, 81.61)$.

z Tests and Confidence Intervals, X

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

5. that the average of Class A is 79 points.

- The 90% CI was $(79.22, 81.28)$, while the 97% CI was $(78.89, 81.61)$.
- Since 79 lies outside the first interval, the result is significant at the 10% level (we reject the null hypothesis that the true mean is 79 points).
- However, since 79 lies inside the second interval, the result is not significant at the 3% level (we fail to reject the null hypothesis with this more stringent significance level).

z Tests and Confidence Intervals, XI

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

6. that the average of Class B is 80 points.

- The 90% CI was $(80.34, 81.98)$, while the 97% CI was $(80.07, 82.25)$.

z Tests and Confidence Intervals, XI

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

6. that the average of Class B is 80 points.

- The 90% CI was $(80.34, 81.98)$, while the 97% CI was $(80.07, 82.25)$.
- Since 80 lies outside the first interval, the result is significant at the 10% level (we reject the null hypothesis that the true mean is 80 points).
- However, since 80 lies inside the second interval, the result is not significant at the 3% level (we fail to reject the null hypothesis with this more stringent significance level).

z Tests and Confidence Intervals, XI

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

7. that the average of Class B is 82 points.

- The 90% CI was $(80.34, 81.98)$, while the 97% CI was $(80.07, 82.25)$.

z Tests and Confidence Intervals, XI

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

7. that the average of Class B is 82 points.

- The 90% CI was $(80.34, 81.98)$, while the 97% CI was $(80.07, 82.25)$.
- Since 82 lies outside the first interval (barely!), the result is significant at the 10% level (we reject the null hypothesis that the average is 82)
- But it lies inside the second interval, so the result is not significant at the 3% level (we fail to reject the null hypothesis).

z Tests and Confidence Intervals, XII

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

8. that the average scores in the classes are equal.

- The 90% CI was $(-2.23, 0.41)$, while the 97% confidence interval is $(-2.65, 0.83)$.

z Tests and Confidence Intervals, XII

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

8. that the average scores in the classes are equal.

- The 90% CI was $(-2.23, 0.41)$, while the 97% confidence interval is $(-2.65, 0.83)$.
- Since 0 lies inside both intervals, the result is not significant at either the 10% or 3% significance levels: we fail to reject the null hypothesis that the means are equal.

z Tests and Confidence Intervals, XIII

Example: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

9. that the average score in Class A is 1 point greater than the average in Class B.

- The 90% CI was $(-2.23, 0.41)$, while the 97% confidence interval is $(-2.65, 0.83)$.

z Tests and Confidence Intervals, XIII

<u>Example</u>: Using the Class A (64 students, average 80.25) and Class B (100 students, average 81.16) data, with individual score standard deviation 5 points. Test at the 10% and 3% significance levels

9. that the average score in Class A is 1 point greater than the average in Class B.

- The 90% CI was $(-2.23, 0.41)$, while the 97% confidence interval is $(-2.65, 0.83)$.
- Be careful! This is actually a one-sided hypothesis test, so we can't use these confidence intervals to determine the results. (I know, I tricked you. But it's very important to be careful with this!)
- We would need to find the 80% and 94% CIs to answer this question.

z Tests for Unknown Proportion, I

We now discuss how to use z tests to handle situations with approximately normally distributed test statistics.

- Presently, we will discuss the scenario of approximating the binomial distribution by the normal distribution.
- We can then adapt our procedure for using a z test to handle hypothesis testing with a binomially-distributed test statistic.

z Tests for Unknown Proportion, II

Thus, suppose we have a binomially distributed test statistic $B_{n,p}$ counting total successes in n trials with success probability p.

- If np (the number of successes) and $n(1-p)$ (the number of failures) are both larger than 5, we are in the situation where the normal approximation to the binomial is good: then $P(a \leq B_{n,p} \leq b)$ will be well approximated[1] by $P(a - 0.5 < N_{np,\sqrt{np(1-p)}} < b + 0.5)$, where N is normally distributed with mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$.
- We can then test the null hypothesis $H_0 : p = c$ by equivalently testing the equivalent hypothesis $H_0 : np = nc$ using the normal approximation via a one-sample z test, where our test statistic is the number of observed successes k.

[1] Note that we have included a continuity correction here. Some software does not make this correction, which will lead to slightly different p-values.

z Tests for Unknown Proportion, III

Using the mean $\mu = np$ and standard deviation $\sigma = \sqrt{np(1-p)}$, we can compute the p-value.

- If the hypotheses are $H_0 : p = c$ and $H_a : p > c$, the associated p-value is $P(B_{n,p} \geq k) \approx P(N_{np,\sqrt{np(1-p)}} > k - 0.5)$.
- If the hypotheses are $H_0 : p = c$ and $H_a : p < c$, the associated p-value is $P(B_{n,p} \leq k) \approx P(N_{np,\sqrt{np(1-p)}} < k + 0.5)$.
- Finally, if the hypotheses are $H_0 : p = c$ and $H_a : p \neq c$, the associated p-value is $P(|B_{n,p} - c| \geq |k - c|) \approx$
$$\begin{cases} 2P(N_{np,\sqrt{np(1-p)}} > k - 0.5) & \text{if } k > c \\ 2P(N_{np,\sqrt{np(1-p)}} < k + 0.5) & \text{if } k < c \end{cases}.$$

We then compare the p-value to the significance level α.

z Tests for Unknown Proportion, IV

All of this still leaves open the question of what we can do in situations where the binomial distribution is not well approximated by the normal distribution.

- In such cases, we can work directly with the binomial distribution explicitly, or (in the event n is large but np or $n(1-p)$ is small) we could use a Poisson approximation.
- Of course, in principle, we could always choose to work with the exact distribution, but when n is large computing the necessary probabilities becomes cumbersome, which is why we usually use the normal approximation instead.

z Tests for Unknown Proportion, V

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypotheses

1. that the coin is fair.
2. that the coin is more likely to land heads than tails.
3. that the heads probability is $2/3$.
4. that the heads probability is less than $3/4$.

z Tests for Unknown Proportion, VI

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

1. that the coin is fair.

z Tests for Unknown Proportion, VI

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

1. that the coin is fair.

- Our hypotheses are $H_0 : p = 1/2$ and $H_a : p \neq 1/2$, since we only want to know whether or not the coin is fair.
- Here, we have $np = n(1-p) = 50$ so we can use the normal approximation. Note that $np = 50$ and $\sqrt{np(1-p)} = 5$.
- We compute the p-value as $P(|B_{100,0.5} - 50| \geq |64 - 50|) \approx 2P(N_{50,5} > 63.5) = 0.00693$.

z Tests for Unknown Proportion, VII

<u>Example</u>: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

1. that the coin is fair.

- The p-value is 0.00693.
- Thus, the result is statistically significant at the 11% and 4% levels and we reject the null hypothesis in these cases.
- But the result is not statistically significant at the 0.5% level, so we fail to reject the null hypothesis there.
- Overall, we interpret the result as giving fairly strong evidence that the heads probability is not $1/2$.

z Tests for Unknown Proportion, VIII

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

2. that the coin is more likely to land heads than tails.

z Tests for Unknown Proportion, VIII

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

2. that the coin is more likely to land heads than tails.

- It is easy to see that we want a one-sided alternative hypothesis; the only question is the appropriate direction.
- The actual heads proportion is 64%. Since this is above 50%, we want the alternative hypothesis $H_a : p > 1/2$.
- Thus, we take the null hypothesis as $H_0 : p = 1/2$ and the alternative hypothesis as $H_a : p > 1/2$.

z Tests for Unknown Proportion, IX

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

2. that the coin is more likely to land heads than tails.

- We have the same parameters as before, so $np = 50$ and $\sqrt{np(1-p)} = 5$, and then the p-value is $P(B_{100,0.5} \geq 64) \approx P(N_{50,5} > 63.5) = 0.00347$.
- Thus, the result is statistically significant at the 11%, 4%, and 0.5% significance levels, and we reject the null hypothesis in each case.
- We interpret this as giving strong evidence that the heads probability is greater than $1/2$.

z Tests for Unknown Proportion, X

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

3. that the heads probability is $2/3$.

z Tests for Unknown Proportion, X

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

3. that the heads probability is $2/3$.

- Our hypotheses are $H_0 : p = 2/3$ and $H_a : p \neq 2/3$, since we want to know whether or not the heads probability is $2/3$.
- Our parameter values are now $n = 100$, $p = 2/3$ so that $np = 66.667$ and $\sqrt{np(1-p)} = 4.714$.
- Since $n(1-p) = 33.333$ the normal approximation is still appropriate, so we compute the p-value as $P(|B_{100,2/3} - 66.667| \geq 2.667) \approx 2P(N_{66.667,4.714} < 64.5) = 0.6457$.

z Tests for Unknown Proportion, XI

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

3. that the heads probability is $2/3$.

- The p-value is 0.6457.
- Thus, the result is not statistically significant at the 11%, 4%, or 0.5% levels, and we accordingly fail to reject the null hypothesis in each case.
- We interpret this as giving minimal evidence against the hypothesis that the heads probability is $2/3$. (Quite sensible, since the observed frequency of heads was 64%!)

z Tests for Unknown Proportion, XII

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

3. that the heads probability is less than $3/4$.

z Tests for Unknown Proportion, XII

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

3. that the heads probability is less than $3/4$.

- It is easy to see that we want a one-sided alternative hypothesis; the only question is the appropriate direction.
- The actual heads proportion is 64%. Since this is below 75%, we try the alternative hypothesis $H_a : p < 3/4$.
- Thus, we have the null hypothesis as $H_0 : p = 3/4$ and the alternative hypothesis as $H_a : p < 3/4$.

z Tests for Unknown Proportion, XII

Example: A coin with unknown heads probability p is flipped $n = 100$ times, yielding 64 heads. Test at the 11%, 4%, and 0.5% significance levels the hypothesis

3. that the heads probability is less than 3/4.

- The parameter values now are $n = 100$ and $p = 3/4$ so that $np = 75$ and $\sqrt{np(1-p)} = 4.330$.
- Since $np = 75$ and $n(1-p) = 25$ the normal approximation is still appropriate.
- We compute the p-value as $P(B_{100,3/4} \leq 64) \approx P(N_{75,4.330} < 64.5) = 0.0077$.
- The result is statistically significant at the 11% and 4% levels, but not statistically significant at the 0.5% level.
- Interpretation: we have fairly strong evidence for the hypothesis that the true heads probability is less than 3/4.

z Tests for Unknown Proportion, XIII

<u>Example</u>: You think your poker opponent is cheating, so you tally the proportion of times his 5-card hand contains the ace of spades. In 520 hands, he has the ace of spades 78 times.

1. How many times would you expect him to have the ace of spades?
2. Find 80% and 95% confidence intervals for the number of times you would expect him to have the ace of spades.
3. Test at the 10%, 1%, and 0.1% significance levels the hypothesis that your opponent has the ace of spades more often than he should.

z Tests for Unknown Proportion, XIV

Example: You think your poker opponent is cheating, so you tally the proportion of times his 5-card hand contains the ace of spades. In 520 hands, he has the ace of spades 78 times.

1. How many times would you expect him to have the ace of spades?

z Tests for Unknown Proportion, XIV

Example: You think your poker opponent is cheating, so you tally the proportion of times his 5-card hand contains the ace of spades. In 520 hands, he has the ace of spades 78 times.

1. How many times would you expect him to have the ace of spades?

- Chapter 1 Review: the probability p of getting the ace of spades in a 5-card hand will be $5/52$ if the deck is fair.
- Chapter 2 Review: Since the probability of getting the ace of spades is $5/52$, the expected number of times to get the ace of spades should be $520 \cdot 5/52 = 50$.

z Tests for Unknown Proportion, XV

Example: You think your poker opponent is cheating, so you tally the proportion of times his 5-card hand contains the ace of spades. In 520 hands, he has the ace of spades 78 times.

2. Find 80% and 95% confidence intervals for the number of times you would expect him to have the ace of spades.

z Tests for Unknown Proportion, XV

Example: You think your poker opponent is cheating, so you tally the proportion of times his 5-card hand contains the ace of spades. In 520 hands, he has the ace of spades 78 times.

2. Find 80% and 95% confidence intervals for the number of times you would expect him to have the ace of spades.

- Chapter 3 Review: The actual proportion should be $p = 5/52$.
- Using the normal approximation to the binomial (appropriate since np and $n(1-p)$ are large), the true number of successes should be approximately normal with mean $np = 50$ and standard deviation $\sqrt{np(1-p)} = 6.7225$.
- Thus, the 80% confidence interval is $50 \pm 1.2816 \cdot 6.7225 = (41.4, 58.6)$.
- The 95% confidence interval is $50 \pm 1.9600 \cdot 6.7225 = (36.8, 63.2)$.

z Tests for Unknown Proportion, XVI

Example: You think your poker opponent is cheating, so you tally the proportion of times his 5-card hand contains the ace of spades. In 520 hands, he has the ace of spades 78 times.

3. Test at the 10%, 1%, and 0.1% significance levels the hypothesis that your opponent has the ace of spades more often than he should.

z Tests for Unknown Proportion, XVI

Example: You think your poker opponent is cheating, so you tally the proportion of times his 5-card hand contains the ace of spades. In 520 hands, he has the ace of spades 78 times.

3. Test at the 10%, 1%, and 0.1% significance levels the hypothesis that your opponent has the ace of spades more often than he should.

- Chapter 4 Review: Our hypotheses are H_0: $p = 5/52$ and H_a: $p > 5/52$.
- Our test statistic is the number of hands with an ace of spades, which is binomially distributed with mean $np = 50$ and standard deviation $\sqrt{np(1-p)} = 6.7225$.
- Our p-value is then $P(B_{520,5/52} \geq 78) \approx P(N_{50,6.7225} > 77.5) = 0.0000215$.
- This p-value is minuscule, so we reject the null hypothesis in all cases: our opponent does appear to be cheating!

Summary

We continued our discussion one-sample z tests with more examples.

We introduced z tests for unknown proportion and gave some examples.

Next lecture: Additional examples, errors in hypothesis testing