Math 3081 (Probability and Statistics)

Lecture #20 of 27 $\sim$ August 9th, 2021

Unknown Proportion and Errors in Hypothesis Testing

- More z Tests for Unknown Proportion
- Type I and Type II Errors

This material represents §4.2.3-4.3.1 from the course notes, and problems 11-15 from WeBWorK 6.

Review, I

Suppose we have a binomially distributed test statistic $B_{n,p}$ counting the number of 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 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 $\sigma = \sqrt{np(1-p)}$. (Note that we have incorporated the continuity correction in our estimate.)
- 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.

Review, II

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.

Another Binomial Example, I

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?
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.

Another Binomial Example, II

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?

Another Binomial Example, II

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 $520 \cdot 5/52 = 50$.

Another Binomial Example, III

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.

Another Binomial Example, III

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 $50 \pm 1.9600 \cdot 6.7225 = (36.8, 63.2)$.

Another Binomial Example, IV

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.

Another Binomial Example, IV

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!

Another Binomial Example, V

Example: A coin is flipped 3000 times, yielding 1600 heads. Test at the 4% and the 0.1% significance levels whether the coin is fair.

Another Binomial Example, V

<u>Example</u>: A coin is flipped 3000 times, yielding 1600 heads. Test at the 4% and the 0.1% significance levels whether the coin is fair.

- Our hypotheses are $H_0 : p = 1/2$ and $H_a : p \neq 1/2$.
- Since $k = 1600$ and $n - k = 1400$ are large, we can use the normal approximation.
- The number of heads is approximately normally distributed with hypothesized mean $np = 1500$ and standard deviation $\sqrt{np(1-p)} \approx 27.386$.
- Thus, the p-value is $2P(B_{3000,0.5} \geq 1600) \approx 2P(N_{1500,27.386} > 1599.5) = 2P(N_{0,1} > 3.6332) \approx 0.028\%$.
- Since the p-value is below both significance levels, we reject the null hypothesis in both cases.

Another Binomial Example, VI

What if 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.

Another Binomial Example, VII

Example: A 6-sided die is rolled 18 times, yielding six 4s. Test at the 15%, 4%, and 1% significance levels the hypothesis that the true probability of rolling a 4 is equal to $1/6$.

Another Binomial Example, VII

Example: A 6-sided die is rolled 18 times, yielding six 4s. Test at the 15%, 4%, and 1% significance levels the hypothesis that the true probability of rolling a 4 is equal to $1/6$.

- Our hypotheses are $H_0 : p = 1/6$ and $H_a : p \neq 1/6$.
- Under the conditions of the null hypothesis, the total number of 4s rolled is binomially distributed with parameters $n = 18$ and $p = 16$. Here, $np = 3$ is too small for us to apply the normal approximation to the binomial distribution, so we will work directly with the binomial distribution itself.

Another Binomial Example, VIII

Example: A 6-sided die is rolled 18 times, yielding six 4s. Test at the 15%, 4%, and 1% significance levels the hypothesis that the true probability of rolling a 4 is equal to $1/6$.

Another Binomial Example, VIII

Example: A 6-sided die is rolled 18 times, yielding six 4s. Test at the 15%, 4%, and 1% significance levels the hypothesis that the true probability of rolling a 4 is equal to 1/6.

- Our hypotheses are $H_0 : p = 1/6$ and $H_a : p \neq 1/6$.
- The desired p-value is $P(\left|B_{18,1/6} - 3\right| \geq |6 - 3|) = P(B_{18,1/6} \geq 6) + P(B_{18,1/6} \leq 0) = 0.1028$.
- The result is statistically significant at the 15% significance level, and we accordingly reject the null hypothesis.
- However, it is not statistically significant at the 4% or 1% significance levels, and so we fail to reject the null hypothesis in these cases.
- We interpret this result as saying that there is moderate evidence against the hypothesis that the probability of rolling a 4 is equal to 1/6.

Comparing Two Proportions, I

If we have two independent, binomially-distributed quantities each of which is well approximated by a normal distribution, we can use the method for a two-sample z test to set up a hypothesis test for the difference of these quantities: we refer to this as a two-sample z-test for unknown proportion.

Comparing Two Proportions, II

Suppose the two populations are A and B.

- We use the null hypothesis $H_0 : p_A - p_B = 0$ to test whether $p_A = p_B$, and take our test statistic to be the difference between the proportions.
- By hypothesis, A is approximately normally distributed with mean p_A and standard deviation $\sigma_A = \sqrt{p_A(1 - p_A)/n_A}$ while B is approximately normally distributed with mean p_B and standard deviation $\sigma_B = \sqrt{p_B(1 - p_B)/n_B}$.
- Under the assumption that H_0 is true, the test statistic $p_A - p_B$ is normally distributed with mean 0 (the true mean postulated by the null hypothesis).

Comparing Two Proportions, II

However, the null hypothesis $H_0 : p_A - p_B = 0$ does not actually tell us the standard deviations of p_A and p_B (that would only be the case if the null hypothesis were to state a specific value for p_A and for p_B).

- What we must do instead is estimate the two standard deviations using the sample data.
- Here, under the null hypothesis assumption that the two proportions are actually equal, we can calculate a <u>pooled estimate</u> for the true proportion p by putting the two samples together.

Comparing Two Proportions, III

Suppose sample A has k_A successes in n_A trials and sample B has k_B successes in n_B trials.

- Then, together, there were $k_A + k_B$ successes in $n_A + n_B$ trials, so our pooled estimate for both p_A and p_B is $p_{\text{pool}} = \dfrac{k_A + k_B}{n_A + n_B}$.
- Then the standard deviation of A is $\sigma_A = \sqrt{\dfrac{p_{\text{pool}}(1 - p_{\text{pool}})}{n_A}}$ and the standard deviation of B is $\sigma_B = \sqrt{\dfrac{p_{\text{pool}}(1 - p_{\text{pool}})}{n_B}}$.
- That means the standard deviation of $A - B$ is $\sigma_{A-B} = \sqrt{\sigma_A^2 + \sigma_B^2} = \sqrt{p_{\text{pool}}(1 - p_{\text{pool}})\left(\dfrac{1}{n_A} + \dfrac{1}{n_B}\right)}$.

Comparing Two Proportions, IV

Once we have the standard deviation, we then perform our hypothesis test as normal.

- The desired p-value is then the probability that the normally-distributed random variable $N_{\mu_{A-B},\sigma_{A-B}}$ will take a value further from the hypothesized value 0 (in the direction of the alternative hypothesis, as applicable) than the observed test statistic $z = \hat{p}_A - \hat{p}_B$.

To Pool or Not To Pool? That Is The Question

We will note that there is another way to estimate the standard deviation:

- Specifically, we could estimate the two standard deviations from their sample proportions separately as $\sigma_A = \sqrt{\hat{p}_A(1-\hat{p}_A)/n_A}$ and $\sigma_B = \sqrt{\hat{p}_B(1-\hat{p}_B)/n_B}$.
- These are called the unpooled standard deviations, and they give a slightly different estimate $\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$ for the standard deviation of $A - B$.

To Pool or Not To Pool? Is There An Answer?

There is not universal consensus on the usage of the pooled versus unpooled standard deviations.

- Ultimately, our choice of test statistic and parameters is up to us: we can do hypothesis tests however we want.
- The only real issue is whether one approach yields more reliable results than the other.
- As a practical matter, if the sample proportions $\hat{p}_A$ and $\hat{p}_B$ are actually close to each other, these values will both also be close to p_{pool}, and thus the two estimates for σ_{A-B} will also be very close.
- One should use the unpooled standard deviations to perform more complicated tests on the observed proportions (e.g., if we wanted to test whether the proportion for A exceeded the proportion for B by 2% or more).

Two Proportions Examples, I

Example: In a sample from a statistics class taught with a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas in a sample from a statistics class taught with a revised curriculum, 86 students out of 284 received an A (30.3%). If p_t is the proportion of students getting an A with the traditional curriculum and p_r is the proportion of students getting an A with the revised curriculum, test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels

1. with alternative hypothesis $p_t > p_r$.
2. with alternative hypothesis $p_t < p_r$.
3. with alternative hypothesis $p_t \neq p_r$.

Two Proportions Examples, I

Example: In a sample from a statistics class taught with a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas in a sample from a statistics class taught with a revised curriculum, 86 students out of 284 received an A (30.3%). If p_t is the proportion of students getting an A with the traditional curriculum and p_r is the proportion of students getting an A with the revised curriculum, test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels

1. with alternative hypothesis $p_t > p_r$.
2. with alternative hypothesis $p_t < p_r$.
3. with alternative hypothesis $p_t \neq p_r$.

- The proportion of students getting an A in each of the samples will be binomially distributed, and the parameters are all in the range where the normal approximation is applicable.

Two Proportions Examples, II

Example: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

1. Test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels with alternative hypothesis $p_t > p_r$.

Two Proportions Examples, II

Example: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

1. Test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels with alternative hypothesis $p_t > p_r$.

- The null hypothesis is $H_0 : p_t - p_r = 0$ with alternative hypothesis $p_t - p_r > 0$.
- Here, we have $n_t = 311$, $n_r = 284$, $\hat{p}_t = 125/311 = 0.4019$, and $\hat{p}_r = 86/284 = 0.3028$, so that $\hat{p}_{t-r} = 0.0991$.
- To find the pooled standard deviation, we have $p_{\text{pool}} = (125 + 86)/(311 + 284) = 0.3546$. Then

$$\sigma_{t-r,\text{pool}} = \sqrt{p_{\text{pool}}(1 - p_{\text{pool}})\left[\frac{1}{n_A} + \frac{1}{n_B}\right]} = 0.03927.$$

Two Proportions Examples, III

Example: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

1. Test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels with alternative hypothesis $p_t > p_r$.

- We have $\sigma_{t-r,\text{pool}} = 0.03927$, so the p-value is $P(N_{0,0.03927} \geq 0.0991) = P(N_{0,1} \geq 2.5242) \approx 0.00580$.
- Thus, the result is statistically significant at the 10%, 5%, and 1% significance levels, and we accordingly reject the null hypothesis in these cases, but it is not statistically significant at the 0.1% significance level.
- We interpret this result as saying that there is strong evidence for the hypothesis that the students with the traditional curriculum had a higher proportion of As than the students with the revised curriculum.

Two Proportions Examples, IV

<u>Example</u>: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

2. Test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels with alternative hypothesis $p_t < p_r$.

Two Proportions Examples, IV

Example: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

2. Test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels with alternative hypothesis $p_t < p_r$.

- The null hypothesis is $H_0 : p_t - p_r = 0$ with alternative hypothesis $p_t - p_r < 0$. The parameters are the same as before: the only difference is that the p-value is now $P(N_{0,0.03927} \leq 0.0991) = P(N_{0,1} \leq 2.5242) \approx 0.99420$.
- Thus, the result is (extremely!) not statistically significant at any of the indicated significance levels, and we fail to reject the null hypothesis in all cases.
- We interpret this result as saying that there is essentially zero evidence for the hypothesis that the students with the revised curriculum had the higher proportion of As.

Two Proportions Examples, V

Example: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

3. Test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels with alternative hypothesis $p_t \neq p_r$.

Two Proportions Examples, V

Example: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

3. Test the null hypothesis $p_t = p_r$ at the 10%, 5%, 1%, and 0.1% significance levels with alternative hypothesis $p_t \neq p_r$.

- The alternative hypothesis is now $p_t - p_r \neq 0$. The parameters are the same, so the p-value is $P(|N_{0,0.03927} - 0| \geq |0.0991 - 0|) = 2P(N_{0,0.03927} \geq 0.0991) = 2P(N_{0,1} \geq 2.5242) \approx 0.01160$.
- Thus, the result is statistically significant at the 10% and 5% significance levels, so we accordingly reject the null hypothesis there, but not at the 1% or 0.1% significance levels.
- We interpret this result as saying that there is relatively strong evidence for the hypothesis that the students with the two curricula had different proportions of As.

Pooled Versus Unpooled, Again

Example: With a traditional curriculum, 125 students out of 311 received an A (40.2%), whereas with a revised curriculum, 86 students out of 284 received an A (30.3%).

- If we wanted to use the unpooled standard deviations, we would have $\sigma_t = \sqrt{\hat{p}_t(1-\hat{p}_t)/n_t} = 0.02780$, $\sigma_r = \sqrt{\hat{p}_r(1-\hat{p}_r)/n_r} = 0.02726$.
- Then $\sigma_{t-r,\text{unpool}} = \sqrt{\hat{p}_t(1-\hat{p}_t)/n_t + \hat{p}_r(1-\hat{p}_r)/n_r} = 0.03894$.
- This is almost exactly the same as the pooled standard deviation $\sigma_{t-r,\text{pool}} = 0.03927$.
- So as a practical matter here, it makes essentially no difference whether we used the pooled or unpooled standard deviation.

Two Proportions Examples, VI

Example: A pollster conducts a poll on the favorability of Propositions ♣ and ♡. They poll 2,571 people and find that 1,218 of them favor Proposition ♣ (47.4%) and 1,344 of them favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

1. Proposition ♣ has at least 50% support.
2. Proposition ♣ has exactly 50% support.
3. Proposition ♡ has at least 50% support.
4. Proposition ♡ has exactly 55% support.
5. Proposition ♡ has more support than Proposition ♣.

Two Proportions Examples, VII

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

1. Proposition ♣ has at least 50% support.

Two Proportions Examples, VII

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

1. Proposition ♣ has at least 50% support.

- Our hypotheses are $H_0 : p_{\clubsuit} = 0.50$ and $H_a : p_{\clubsuit} < 0.50$, since Proposition ♣ actually does have under 50% support.
- Here, we have $np = n(1-p) = 1285.5$ so we can use the normal approximation. Note that $np = 1285.5$ and $\sqrt{np(1-p)} = 25.35$.
- We compute the p-value as $P(B_{2571,0.5} \leq 1218) \approx P(N_{1285.5,25.35} < 1218.5) = P(N_{0,1} < -2.6427) = 0.00411$.
- Thus, the result is statistically significant at both the 8% and 1% significance levels.
- Interpretation: there is strong evidence against the hypothesis that the support for Proposition ♣ is 50% or above.

Two Proportions Examples, VIII

<u>Example</u>: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

2. Proposition ♣ has exactly 50% support.

Two Proportions Examples, VIII

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

2. Proposition ♣ has exactly 50% support.

- Our hypotheses are $H_0 : p_\clubsuit = 0.50$ and $H_a : p_\clubsuit \neq 0.50$.
- We have the same parameters as above ($np = 1285.5$ and $\sqrt{np(1-p)} = 25.35$), so the p-value is $\approx 2P(N_{1285.5,25.35} < 1218.5) = 2P(N_{0,1} < -2.6427) = 0.00822$.
- Thus, the result is statistically significant at both the 8% and 1% significance levels.
- Interpretation: there is strong evidence against the hypothesis that the support for Proposition ♣ is equal to 50%.

Two Proportions Examples, IX

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

3. Proposition ♡ has at least 50% support.

Two Proportions Examples, IX

<u>Example</u>: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

3. Proposition ♡ has at least 50% support.

- Our hypotheses are $H_0 : p_\heartsuit = 0.50$ and $H_a : p_\heartsuit > 0.50$, since the sample suggests Proposition ♡ has at least 50% support.
- We still have the same parameters (only the actual test statistic value will differ), so the p-value is $P(B_{2571,0.5} \geq 1344) \approx P(N_{1285.5,25.35} > 1343.5) = P(N_{0,1} > 2.2877) = 0.01107$.
- Thus, the result is statistically significant at the 8% significance level, but not statistically significant at the 1% significance level.
- Interpretation: there is moderately strong evidence against the hypothesis that the support for Proposition ♡ is 50% or below.

Two Proportions Examples, X

<u>Example</u>: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

4. Proposition ♡ has exactly 55% support.

Two Proportions Examples, X

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

4. Proposition ♡ has exactly 55% support.

- Our hypotheses are $H_0 : p_\heartsuit = 0.55$ and $H_a : p_\heartsuit \neq 0.55$.
- Here, $n = 2571$ and $p = 0.55$ so $np = 1414.05$ and $\sqrt{np(1-p)} = 25.225$.
- Then the p-value is $\approx 2P(N_{1414.05,25.225} < 1344.5) = P(N_{0,1} < -2.7571) = 0.00291$.
- Thus, the result is statistically significant at both the 8% and 1% significance levels, and we accordingly reject the null hypothesis.
- Interpretation: there is strong evidence against the hypothesis that the support for Proposition ♡ is 55%.

Two Proportions Examples, XI

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

5. Proposition ♡ has more support than Proposition ♣.

Two Proportions Examples, XI

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

5. Proposition ♡ has more support than Proposition ♣.

- This is a two-sample test so our null hypothesis is $H_0 : p_{\clubsuit} - p_{\heartsuit} = 0$.
- We want to test whether or not Proposition ♡ has more support than Proposition ♣: this requires a one-sided alternative hypothesis.
- Because the sampling suggests that Proposition ♡ does actually have more support than Proposition ♣, we want the alternative in that direction.
- Thus, we take our hypotheses as $H_0 : p_{\clubsuit} - p_{\heartsuit} = 0$ and $H_a : p_{\clubsuit} - p_{\heartsuit} < 0$.

Two Proportions Examples, XII

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

5. Proposition ♡ has more support than Proposition ♣.

Two Proportions Examples, XII

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

5. Proposition ♡ has more support than Proposition ♣.

- Here, we have $n_♣ = n_♡ = 2571$, $\hat{p}_♣ = 1218/2571 = 0.4737$, and $\hat{p}_♡ = 1344/2571 = 0.5228$, so that $\hat{p}_{♣-♡} = -0.04901$.
- To find the pooled standard deviation, we have $p_{\text{pool}} = (1218 + 1344)/(2571 + 2571) = 0.4982$, so then $\sigma_{♣-♡,\text{pool}} = \sqrt{p_{\text{pool}}(1 - p_{\text{pool}})\left[\frac{1}{n_♣} + \frac{1}{n_♡}\right]} = 0.01395$.

Two Proportions Examples, XIII

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

5. Proposition ♡ has more support than Proposition ♣.

Two Proportions Examples, XIII

Example: 2,571 people are polled: 1,218 favor Proposition ♣ (47.4%) and 1,344 favor Proposition ♡ (52.3%). Perform hypothesis tests at the 8% and 1% significance levels that

5. Proposition ♡ has more support than Proposition ♣.

- The pooled standard deviation is $\sigma_{\clubsuit-\heartsuit,\text{pool}} = 0.01395$.
- Then the desired p-value is $P(N_{0,0.01395} < -0.04901) = P(N_{0,1} < -3.5143) = 0.000220$.
- Thus, the result is statistically significant at both the 8% and 1% significance levels, and we accordingly reject the null hypothesis.
- Interpretation: there is very strong evidence that Proposition ♡ has more support than Proposition ♣.

Errors in Hypothesis Testing, I

Now that we have some concrete methods for testing hypotheses, and have worked through enough examples to get a sense of how hypothesis testing works (and doesn't work!), we can discuss some of the finer points.

- Specifically, we will discuss errors in hypothesis testing, and a number of related misinterpretations and misuses of hypothesis testing.

Errors in Hypothesis Testing, II

When we perform a hypothesis test, there are two possible outcomes (reject H_0 or fail to reject H_0).

- The correctness of the result depends on the actual truth of H_0: if H_0 is false then it is correct to reject it, while if H_0 is true than it is correct not to reject it.
- The other two situations, namely "rejecting a correct null hypothesis" and "failing to reject an incorrect null hypothesis" are refered to as hypothesis testing errors.

Errors in Hypothesis Testing, III

Since these two errors are very different, we give them very different names:

Definition

If we are testing a null hypothesis H_0, we commit a type I error if we reject H_0 when H_0 was actually true. We commit a type II error if we fail to reject H_0 when H_0 was actually false.

We usually summarize these errors with a small table:

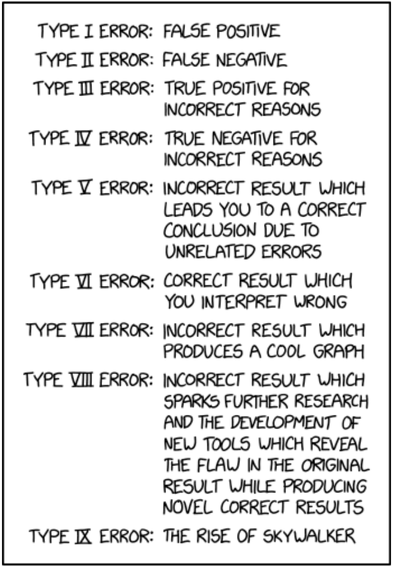
$H_0 \setminus$ Result	Fail to Reject H_0	Reject H_0
H_0 is true	Correct Decision	Type I Error
H_0 is false	Type II Error	Correct Decision

Errors in Hypothesis Testing, IV

The names for these two errors are very unintuitive, and it must simply be memorized which one is which.

- If we view a positive result as one in which we reject the null hypothesis, which in most cases is the practical interpretation, then a type I error corresponds to a false positive (a positive test on an actual negative sample) while a type II error corresponds to a false negative (a negative test on an actual positive sample).
- For example, if the purpose of the hypothesis test is to determine whether or not to mark an email as spam (with H_0 being that the email is not spam), a type I error would be marking a normal email as spam, while a type II error would be marking a spam email as normal.

Not Actually Real Statistical Errors



Citation:
Randall Munroe
("Error Types",
xkcd #2303,
xkcd.com/2303)

Errors in Hypothesis Testing, V

We would like, in general, to minimize the probabilities of making a type I or type II error.

- The probability of committing a type I error is the significance level α of the test, since by definition this is the probability of rejecting the null hypothesis when it is actually true.
- The probability of committing a type II error is denoted by β. This value is more difficult to calculate, since it will depend on the actual nature in which H_0 is false.
- If we postulate the actual value of the test statistic, we can calculate the probability of committing a type II error.

Errors in Hypothesis Testing, VI

Example: A new mathematics curriculum is being tested in schools to see if students score more highly on standardized tests. The scores for students using the old curriculum are normally distributed with mean 200 and standard deviation 20. It is assumed that scores using the new curriculum are also normally distributed with mean μ and standard deviation 20. The hypothesis $H_0 : \mu = 200$ is tested against the alternative $H_a : \mu > 200$ using a sample of 400 students using the new curriculum. The null hypothesis will be rejected if the sample mean $\hat{\mu} > 202$.

1. Find the probability of making a type I error.

2-5. Find the probability of making a type II error if the true mean is actually 201 / 202 / 203 / 204 / 205.

Errors in Hypothesis Testing, VII

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

1. Find the probability of making a type I error.

Errors in Hypothesis Testing, VII

<u>Example</u>: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

1. Find the probability of making a type I error.

- We want to calculate the probability of rejecting the null hypothesis when it is true.
- If the null hypothesis is true, then the sample mean $\hat{\mu}$ will be normally distributed with mean 200 and standard deviation $20/\sqrt{400} = 1$.
- Then, the probability of rejecting the null hypothesis is $P(N_{200,1} > 202) = P(N_{0,1} > 2) = 0.02275$. (Note that this value is the significance level α for this hypothesis test.)

Errors in Hypothesis Testing, VIII

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

2. Find the probability of making a type II error if the true mean is actually 201.

Errors in Hypothesis Testing, VIII

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

2. Find the probability of making a type II error if the true mean is actually 201.

- We want to calculate the probability of failing to reject the null hypothesis when it is false.
- Under the assumption given, the sample mean $\hat{\mu}$ will be normally distributed with mean 201 and standard deviation $20/\sqrt{400} = 1$.
- Then, the probability of failing to reject the null hypothesis is $P(N_{201,1} \leq 202) = P(N_{0,1} \leq 1) = 0.8413$: quite large.

Errors in Hypothesis Testing, IX

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

3. Find the probability of making a type II error if the true mean is actually 202.

Errors in Hypothesis Testing, IX

<u>Example</u>: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

3. Find the probability of making a type II error if the true mean is actually 202.

- Now $\hat{\mu}$ is normally distributed with mean 202 and standard deviation 1, so the probability of failing to reject the null hypothesis is $P(N_{202,1} \leq 202) = P(N_{0,1} \leq 0) = 0.5$.

4. Find the probability of making a type II error if the true mean is actually 203.

Errors in Hypothesis Testing, IX

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

3. Find the probability of making a type II error if the true mean is actually 202.

- Now $\hat{\mu}$ is normally distributed with mean 202 and standard deviation 1, so the probability of failing to reject the null hypothesis is $P(N_{202,1} \leq 202) = P(N_{0,1} \leq 0) = 0.5$.

4. Find the probability of making a type II error if the true mean is actually 203.

- Now $\hat{\mu}$ is normally distributed with mean 203 and standard deviation 1, so the probability of failing to reject the null hypothesis is $P(N_{203,1} \leq 202) = P(N_{0,1} \leq -1) = 0.1587$.

Errors in Hypothesis Testing, X

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

5. Find the probability of making a type II error if the true mean is actually 204.

Errors in Hypothesis Testing, X

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

5. Find the probability of making a type II error if the true mean is actually 204.
   - Now $\hat{\mu}$ is normally distributed with mean 204 and standard deviation 1, so the probability of failing to reject the null hypothesis is $P(N_{204,1} \leq 202) = P(N_{0,1} \leq -2) = 0.02275$.
6. Find the probability of making a type II error if the true mean is actually 205.

Errors in Hypothesis Testing, X

Example: Old curriculum scores are normally distributed with mean 200 and standard deviation 20. New curriculum scores are normally distributed with mean μ and standard deviation 20. The hypotheses are $H_0 : \mu = 200$ and $H_a : \mu > 200$ with a sample of 400 students. H_0 will be rejected if the sample mean $\hat{\mu} > 202$.

5. Find the probability of making a type II error if the true mean is actually 204.
   - Now $\hat{\mu}$ is normally distributed with mean 204 and standard deviation 1, so the probability of failing to reject the null hypothesis is $P(N_{204,1} \leq 202) = P(N_{0,1} \leq -2) = 0.02275$.
6. Find the probability of making a type II error if the true mean is actually 205.
   - Now $\hat{\mu}$ is normally distributed with mean 205 and standard deviation 1, so the probability of failing to reject the null hypothesis is $P(N_{205,1} \leq 202) = P(N_{0,1} \leq -3) = 0.00135$.

Errors in Hypothesis Testing, XI

We can see that as the true mean gets further away from the mean predicted by the null hypothesis, the probability of making a type II error drops.

- The idea here is quite intuitive: the bigger the distance between the true mean and the predicted mean, the better our hypothesis test will be better at picking up the difference between them.

If we use the same rejection rule, but instead vary the sample size, the probability of making either type of error will change.

Errors in Hypothesis Testing, XII

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

1. Find the probability of a type I error if $n = 100$.
2. Find the probability of a type II error if $n = 100$.
3. Find the probability of a type I error if $n = 400$.
4. Find the probability of a type II error if $n = 400$.
5. Find the probability of a type I error if $n = 1600$.
6. Find the probability of a type II error if $n = 1600$.

Errors in Hypothesis Testing, XIII

<u>Example</u>: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

1. Find the probability of a type I error if $n = 100$.

Errors in Hypothesis Testing, XIII

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

1. Find the probability of a type I error if $n = 100$.

- To find the probability of a type I error, we assume the null hypothesis is correct, so that $\mu = 200$.
- Then the sample mean $\hat{\mu}$ is normally distributed with mean 200 and standard deviation $\sigma = 20/\sqrt{100} = 2$
- Thus, the probability of a type I error is $P(N_{200,2} > 202) = P(N_{0,1} > 1) = 0.1587$.

Errors in Hypothesis Testing, XIV

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

2. Find the probability of a type II error if $n = 100$.

Errors in Hypothesis Testing, XIV

<u>Example</u>: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

2. Find the probability of a type II error if $n = 100$.

- For a type II error, we assume the given value $\mu = 203$ is correct.
- Then the sample mean $\hat{\mu}$ is normally distributed with mean 203 and standard deviation $\sigma = 20/\sqrt{100} = 2$.
- That means the probability of a type II error is $P(N_{203,2} \leq 202) = P(N_{0,1} \leq -0.5) = 0.3085$.

Errors in Hypothesis Testing, XV

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

3. Find the probability of a type I error if $n = 400$.

Errors in Hypothesis Testing, XV

<u>Example</u>: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

3. Find the probability of a type I error if $n = 400$.

- The sample mean $\hat{\mu}$ is normally distributed with mean 200 and standard deviation $\sigma = 20/\sqrt{400} = 1$.
- Thus, the probability of a type I error is $P(N_{200,1} > 202) = P(N_{0,1} > 2) = 0.02275$.

Errors in Hypothesis Testing, XVI

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

3. Find the probability of a type II error if $n = 400$.

Errors in Hypothesis Testing, XVI

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

3. Find the probability of a type II error if $n = 400$.

- The sample mean $\hat{\mu}$ is normally distributed with mean 203 and standard deviation $\sigma = 20/\sqrt{400} = 1$.
- Thus, the probability of a type II error is $P(N_{203,1} \leq 202) = P(N_{0,1} \leq -1) = 0.1587$.

Errors in Hypothesis Testing, XVII

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

5. Find the probability of a type I error if $n = 1600$.

Errors in Hypothesis Testing, XVII

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

5. Find the probability of a type I error if $n = 1600$.

- The sample mean $\hat{\mu}$ is normally distributed with mean 200 and standard deviation $\sigma = 20/\sqrt{1600} = 0.5$.
- Thus, the probability of a type I error is $P(N_{200,0.5} > 202) = P(N_{0,1} > 4) = 0.0000316 = 3.16 \cdot 10^{-5}$.

Errors in Hypothesis Testing, XVIII

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

6. Find the probability of a type II error if $n = 1600$.

Errors in Hypothesis Testing, XVIII

Example: The school wants to gather more data on the new curriculum. Assume as before the scores with the old curriculum are normally distributed with mean 200 and standard deviation 20, and the new curriculum scores also have standard deviation 20. We again test $H_0 : \mu = 200$ against $H_a : \mu > 200$ and reject H_0 if $\hat{\mu} > 202$. Suppose the true mean is $\mu = 203$.

6. Find the probability of a type II error if $n = 1600$.

- The sample mean $\hat{\mu}$ is normally distributed with mean 203 and standard deviation $\sigma = 20/\sqrt{1600} = 0.5$.
- Thus, the probability of a type II error is $P(N_{203,0.5} \leq 202) = P(N_{0,1} \leq -2) = 0.02275$.

Summary

We discussed more examples of z tests for unknown proportion.

We discussed two-sample z tests for unknown proportion.

We introduced type I and type II errors, and calculated probabilities in some examples.

Next lecture: More about errors in hypothesis testing