INFERENCES ON POPULATION PROPORTIONS WORKSHEET
MTH 3210 SPRING 2020

A population proportion can be thought of as a mean of a Bernoulli population (only 0's and 1's for "failure" or "success," respectively), but a more efficient, specialized confidence interval and hypothesis testing methods has been developed specifically for the case of population proportions. These methods are covered on the extra credit on Exam 3.

Please read Section 8.4 of our textbook carefully for a thorough development of these methods. The notes below provide the formulas for these methods and include some solved exercises useful for preparing for Exam 3.

1. Confidence interval for a population proportion, p

When estimating population proportions we use a different version of the confidence interval. Our assumptions are

- We have a simple random sample
- $n(\hat{p}) > 5$ and $n(1-\hat{p}) > 5$

A $100(1-\alpha)\%$ confidence interval for the population proportion, p, is given by,

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

Here n is the sample size, $\hat{p}$ is the sample proportion, and $z_{\alpha/2}$ is the z-value under the normal curve for which the area under the curve, to the right of z, is $\alpha/2$.

Our textbook does not give the formula above; rather, on page 548 (Procedure 12.1), they provide a simplified version of this formula that is valid if n is very large. The simplified version follows below.

<u>If n is very large</u>, then a $100(1-\alpha)\%$ confidence interval for the true population proportion p is given by,

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

All of the problems covered below and on Exam 3 will use the assumption that n is large so that we can use this simplified version of the formula.

2. Hypothesis tests for population proportions

Hypothesis test can be used to make inferences pertaining to a single population proportion, p, and can be used to compare two unknown population proportions; these methods are covered in Sections 12.2 and 12.3 in the book, respectively. These notes only cover inferences about a single population proportion.

The tests covered below assume that a simple random sample is taken and that the sample size is large enough that the number of “successes” and “failures” in the sample are both at least 5.

2.1. **Two-sided test for a population proportion p.** A two-sided test uses a double interval rejection region that includes both $-\infty$ and ∞.

- **Null Hypothesis** $H_0 : p = p_0$
- **Alternative Hypothesis** $H_A : p \neq p_0$
- **Test Statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- **Critical Values** $-z_{\alpha/2}$ and $z_{\alpha/2}$
- **Rejection Region** $(-\infty, -z_{\alpha/2}]$ or $[z_{\alpha/2}, \infty)$
- ***P*-value** $2P[Z > |z|]$ where Z is a standard normal distribution.

2.2. **The Right-sided paired t-test.** For this test, we use a single interval rejection region.

- **Null Hypothesis** $H_0 : p = p_0$
- **Alternative Hypothesis** $H_A : p > p_0$
- **Test Statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- **Critical Value** z_α
- **Rejection Region** $[z_\alpha, \infty)$
- ***P*-value** $P[Z > z]$ where Z is a standard normal distribution.

2.3. **The Left-sided paired t-test.** For this test, we use a single interval rejection region.

- **Null Hypothesis** $H_0 : p = p_0$
- **Alternative Hypothesis** $H_A : p < p_0$
- **Test Statistic**

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

- **Critical Value** $-z_\alpha$
- **Rejection Region** $(-\infty, -z_\alpha]$
- ***P*-value** $P[Z < z]$ where Z is a standard normal distribution.

3. Exercises (confidence intervals)

Find a $100(1-\alpha)\%$ confidence interval for the population proportion p in each case.

(1) $\hat{p} = .034, n = 2500, \alpha = .05$
(2) $\hat{p} = .8, n = 361, \alpha = .1$
(3) A simple random sample of 112 starting professional baseball pitchers includes 36 left-handed pitchers and 76 right-handed pitchers. Construct a confidence interval for the true proportion of professional baseball pitchers that are left-handed using a 90% confidence level.

Additional confidence interval exercises from the textbook:
Section 12.1: 49, 51, 53, 57

4. Exercises (hypothesis tests)

(1) Use the appropriate test to perform the required hypothesis test about the proportion, p, for the population from which the sample is drawn.
$\hat{p} = .766,\ n = 1250,\ H_0 : p = .75,\ H_a : p > .75, \alpha = 0.05.$

(2) Use the appropriate test to perform the required hypothesis test about the proportion, p, for the population from which the sample is drawn. $\hat{p} = .439,\ n = 278,\ H_0 : p = .5,\ H_a : p \neq .5, \alpha = 0.05.$

(3) A sample of 9000 honey bees from one honey bee colony was tested for mites. A total of 228 of the 9000 bees sampled were found to have mites. Test the null hypothesis that the proportion of bees in the entire colony with mites is 2% against the alternative hypothesis that more than 2% of the colony have mites using an $\alpha = 0.01$ significance level. State the conclusion of your hypothesis test in terms of the real world problem.

Additional hypothesis testing exercises from the textbook:
Section 12.2: 85, 87, 89, 91

Solutions to confidence interval exercises:

(1)

$$\left(.034 - z_{.025}\sqrt{\frac{.034(1-.034)}{2500}}, .034 + z_{.025}\sqrt{\frac{.034(1-.034)}{2500}}\right) = [.029, .041]$$

(2)

$$\left(.8 - z_{.05}\sqrt{\frac{.8(1-.8)}{361}}, .8 - z_{.05}\sqrt{\frac{.8(1-.8)}{361}}\right) = [.765, .835]$$

(3)

$$\frac{36}{112} \pm z_{.05}\sqrt{\frac{\frac{36}{112}(1-\frac{36}{112})}{112}} = [0.249, .394]$$

We are 90% confident that the population proportion of all professional pitchers that are left handed is captured by the interval $[0.249, 0.394]$.

Solutions to hypothesis testing exercises:

(1)
- **Null Hypothesis** $H_0 : p = .75$
- **Alternative Hypothesis** $H_A : p > .75$
- **Test Statistic**
$$z = \frac{.766 - .75}{\sqrt{\frac{.75(1-.75)}{1250}}} = 1.306$$
- **Critical Value** $z_{.05} = 1.645$
- **Rejection Region** $[1.645, \infty)$
- **Conclusion**: Do not reject the null hypothesis.

(2)
- **Null Hypothesis** $H_0 : p = .5$
- **Alternative Hypothesis** $H_A : p \neq .5$
- **Test Statistic**
$$z = \frac{.439 - .5}{\sqrt{\frac{.5(1-.5)}{278}}} = -2.034$$
- **Critical Value** $z_{.025} = 1.96$
- **Rejection Region** $(-\infty, -1.96]$ or $[1.96, \infty)$
- **Conclusion**: Reject the null hypothesis.

(3)
- **Null Hypothesis** $H_0 : p = .02$
- **Alternative Hypothesis** $H_A : p > .02$
- **Test Statistic**
$$z = \frac{\frac{228}{9000} - .02}{\sqrt{\frac{.02(1-.02)}{9000}}} = 3.614$$
- **Critical Value** $z_{.01} = 2.326$
- **Rejection Region** $[2.326, \infty)$
- **Conclusion**: At the 1% significance level, reject the null hypothesis that the proportion of the bees in this colony is 0.02 in favor of the alternative hypothesis that it is more than 0.02.