Lesson 24

Tuesday 4/30/24

Single-Sample Hypothesis Test Example 2

Step 1: State the hypothesis you are going to test (what your textbook in chapter 8 calls the null hypothesis); also decide whether the test will be directional (one-tailed) or non-directional (2-tailed).

We examine America's 200 largest counties and in each of these counties, we calculate a homicide clearance rate for both of the last 2 years. Next, in each of the counties, we determine whether the clearance rate increased from one year to the next. We find that the homicide clearance rate decreased in 93 of the counties (which is slightly less than half the counties). <u>Test the hypothesis that the homicide clearance rate was equally likely to go up or down across the counties</u>. In this study, we are concerned that homicide clearance rates might be declining so we conduct a 1-tailed test. This means we will interpret evidence of declining homicide clearance rates as evidence against our hypothesis.

Single-Sample Hypothesis Test Example 2 (Cont'd)

Step 2: Decide on the test statistic and sampling distribution that will be used.

For this problem, we can assume a normal (z) sampling distribution because we have a large sample of cases (n = 200) and a proportion estimate (93/200 = 0.465). The test statistic will be a one-tailed z-test.

Single-Sample Hypothesis Test Example 2 (Cont'd)

Step 3: Identify the significance level of the test and the evidence that would convince you the hypothesis is wrong (i.e., the critical region).

We will conduct our one-tailed test at the $p < 0.10$ significance level (you can also say the test will be conducted at the $\alpha = 0.10$ significance level). The critical region of the z-statistic for a one-tailed $p < .10$ significance level is (from Table B.3 on page 536):

Critical region: $z < -1.282$

Single-Sample Hypothesis Test Example 2 (Cont'd)

Step 4: Calculate the test statistic (i.e., look at the evidence).

$$z\text{-test} = \frac{\hat{p} - P}{\sqrt{P(1-P)/n}}$$

where p = the sample proportion; P = the hypothesized proportion; and n = the number of cases in the sample.

$$z\text{-test} = \frac{0.465 - 0.5}{\sqrt{0.5(1-0.5)/200}} = \frac{-0.035}{\sqrt{0.250/200}} = \frac{-0.035}{0.035} = -1$$

Single-Sample Hypothesis Test Example 2 (Cont'd)

Step 5: Draw your conclusion - decide whether the hypothesis should be rejected.

$$z\text{-test} = \frac{0.465 - 0.5}{\sqrt{0.5(1-0.5)/200}} = \frac{-0.035}{\sqrt{0.250/200}} = \frac{-0.035}{0.035} = -1$$

Decision: since the z-test we obtained is not in the critical region (z < -1.282), we fail to reject the hypothesis and conclude that the sample proportion is not significantly less than 0.5.

Note: substantive interpretation of the z-test we obtained is that the observed sample proportion is about 1 standard error below 0.5.

Single-Sample Hypothesis Test Example 3

Step 1: State the hypothesis you are going to test (what your textbook in chapter 8 calls the null hypothesis); also decide whether the test will be directional (one-tailed) or non-directional (2-tailed).

Suppose that, over the long run, people on death row have waited an average of 8.5 years before execution. Now, supposed we look at the 17 executions occurring in a U.S. state over the past 10 years. For each of these executions, we consider the waiting time until execution and we find that the average waiting time was 11.29 years with a standard deviation of 2.53 years. <u>Test the hypothesis that the mean waiting time for executions has not changed from the historical mean waiting time.</u> In this study, we are interested in whether execution waiting times have been increasing so we conduct a 1-tailed test. This means we will interpret evidence of increasing execution wait times as evidence against our hypothesis.

Single-Sample Hypothesis Test Example 3 (Cont'd)

Step 2: Decide on the test statistic and sampling distribution that will be used.

For this problem, we must assume a *t* sampling distribution because we have a small sample of cases ($n = 17$). The test statistic will be a one-tailed *t*-test.

Single-Sample Hypothesis Test Example 3 (Cont'd)

Step 3: Identify the significance level of the test and the evidence that would convince you the hypothesis is wrong (i.e., the critical region).

We will conduct our one-tailed test at the $p < 0.05$ significance level (you can also say the test will be conducted at the $\alpha = 0.05$ significance level). The critical region of the t-statistic for a one-tailed $p < .05$ significance level is (from Table B.3 on page 536):

Critical region: $t > 1.746$ (with 17-1=16 degrees of freedom)

Single-Sample Hypothesis Test Example 3 (Cont'd)

Step 4: Calculate the test statistic (i.e., look at the evidence).

$$t\text{-test} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

where $\overline{X}$ = the sample mean; μ = the historical population mean; s = the sample standard deviation and n = the number of cases in the sample.

$$t\text{-test} = \frac{11.29 - 8.5}{2.53/\sqrt{17}} = \frac{2.79}{2.53/4.123} = 2.79/0.614 = 4.544$$

Single-Sample Hypothesis Test Example 3 (Cont'd)

Step 5: Draw your conclusion - decide whether the hypothesis should be rejected.

$$t\text{-test} = \frac{11.29 - 8.5}{2.53/\sqrt{17}} = \frac{2.79}{2.53/4.123} = 2.79/0.614 = 4.544$$

Decision: since the t-test we obtained is in the critical region (t > 1.746), we reject the hypothesis and conclude that the sample mean of 11.29 is significantly greater than the historical average of 8.5.

Single-Sample Hypothesis Test Example 4

Step 1: State the hypothesis you are going to test (what your textbook in chapter 8 calls the null hypothesis); also decide whether the test will be directional (one-tailed) or non-directional (2-tailed).

Suppose we live in a small state with 8 counties. Historically, the average theft loss across these counties (in inflation adjusted dollars) is $367. When this year's data came in, the values for the 8 counties were: 358, 447, 401, 439, 366, 393, 447, and 473. Test the hypothesis that this year's mean is equal to the historical average theft loss. Because we do not have a prediction about whether the new mean is higher or lower than the historical mean, our test is non-directional (two-tailed).

Single-Sample Hypothesis Test Example 4 (Cont'd)

Step 2: Decide on the test statistic and sampling distribution that will be used.

For this problem, we must assume a *t* sampling distribution because we have a small sample of cases (n = 8). The test statistic will be a two-tailed *t*-test.

Single-Sample Hypothesis Test Example 4 (Cont'd)

Step 3: Identify the significance level of the test and the evidence that would convince you the hypothesis is wrong (i.e., the critical region).

We will conduct our two-tailed test at the $p < 0.10$ significance level (you can also say the test will be conducted at the $\alpha = 0.10$ significance level). The critical region of the *t*-statistic for a two-tailed $p < .10$ significance level is (from Table B.3 on page 536):

Critical region: $t > 1.895$ or $t < -1.895$ (with 8-1=7 degrees of freedom)

Single-Sample Hypothesis Test Example 4 (Cont'd)

Step 4: Calculate the test statistic (i.e., look at the evidence).

$$t\text{-test} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

where $\overline{X}$ = the sample mean; μ = the historical population mean; s = the sample standard deviation and n = the number of cases in the sample.

Because the sample mean and standard deviation are not given for this problem, we have to calculate them now.

Single-Sample Hypothesis Test Example 4 (Cont'd)

First, let's calculate the sample mean:

County	Mean Theft Loss
1	358
2	447
3	401
4	439
5	366
6	393
7	447
8	473

$$\overline{X} = \frac{358 + 447 + 401 + 439 + 366 + 393 + 447 + 473}{8} = 415.5$$

Single-Sample Hypothesis Test Example 4 (Cont'd)

Next, let's calculate the standard deviation:

	dev = score-mean	squared(dev)
358-415.5	-57.5	3306.25
447-415.5	31.5	992.25
401-415.5	-14.5	210.25
439-415.5	23.5	552.25
366-415.5	-49.5	2450.25
393-415.5	-22.5	506.25
447-415.5	31.5	992.25
473-415.5	57.5	3306.25
Sum	0	12316.00

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{X})^2} = \sqrt{\frac{1}{7} \times 12316} = \sqrt{1759.429} = 41.946$$

Single-Sample Hypothesis Test Example 4 (Cont'd)

Step 4: Calculate the test statistic (i.e., look at the evidence).

$$t\text{-test} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

sample mean = 415.5
population mean = 367
sample standard deviation = 41.946
sample size = 8
degrees of freedom = 8-1 = 7

$$\text{Calculated } t\text{-test} = \frac{415.5 - 367}{41.946/\sqrt{8}} = \frac{48.5}{41.946/2.828} = \frac{48.5}{14.832} = 3.270$$

Single-Sample Hypothesis Test Example 4 (Cont'd)

Step 5: Draw your conclusion - decide whether our hypothesis should be rejected.

$$\text{Calculated } t\text{-test} = \frac{415.5 - 367}{41.946/\sqrt{8}} = \frac{48.5}{41.946/2.828} = \frac{48.5}{14.832} = 3.270$$

Decision: since the *t*-test we obtained is in the critical region ($t > 1.895$ or $t < -1.895$), we reject our hypothesis and conclude that the sample mean of 415.5 is significantly different from the historical average of 367 (at the $p < .10$ significance level).

Chapter 9: Categorical Data

Marginal Distributions

Dependent Variable

Independent Variable

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	A	B	379
Incarcerated = Yes	C	D	403
Total	333	449	782

p(recidivism=yes) = 449/782 = 0.574

p(recidivism=no) = 333/782 = 0.426

p(incarcerated=no) = 379/782 = 0.485

p(incarcerated=yes) = 403/782 = 0.515

Chapter 9: Categorical Data

2 x 2 Contingency Table

	Recidivism=No	Recidivism=Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

Calculating Marginal Probabilities from a 2x2 Contingency Table (pp. 267-268)

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

p(recidivism=yes | incarcerated=no) = 203/379 = 0.536

p(recidivism=yes | incarcerated=yes) = 246/403 = 0.610

Difference Between 2 Conditional Probabilities (p. 268)

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

p(recidivism=yes | incarcerated=no) = 203/379 = 0.536

p(recidivism=yes | incarcerated=yes) = 246/403 = 0.610

Difference = 0.610 - 0.536 = 0.074

Interpretation: Difference between recidivism probability between the 2 groups is 0.610-0.536 = 0.074

Relative Risk Statistic (p. 268)

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	176	203	379
Incarcerated = Yes	157	246	403
Total	333	449	782

p(recidivism=yes | incarcerated=no) = 203/379 = 0.536

p(recidivism=yes | incarcerated=yes) = 246/403 = 0.610

Relative risk = 0.610/0.536 = 1.138

Interpretation #1: risk of recidivism is 13.8% greater in the incarcerated group compared to the non-incarcerated group

Interpretation #2: risk of recidivism is 1.138 times greater in the incarcerated group compared to the non-incarcerated group.

Yule's Q Statistic (p. 285, p. 287)

	Recidivism= No	Recidivism= Yes	Total
Incarcerated = No	A 176	B 203	379
Incarcerated = Yes	C 157	D 246	403
Total	333	449	782

Yule's Q = (AD-BC)/(AD+BC) = 11425/75167 = 0.152

- Interpreting a Measure of Association (p. 287)
- Using the chart on page 287 (figure 9.2), we would call this a "weak positive" relationship.