Practice 11

Elena Tuzhilina

March 28, 2023

Question 1

A study was conducted on 8 pairs on twins. In each pair:

- twin 1 regularly exercised
- twin 2 was not involved in any sport activities.

The stress level for each study participant was recorded as a score from 0 to 100.

pair	twin1	twin2
1	75.25909	57.82698
2	43.47533	100.00000
3	76.59599	80.90780
4	75.44859	34.02972
5	58.29283	23.57029
6	19.20100	49.31615
7	31.42866	49.02355
8	44.10559	45.65467

1. You want to test if sport decreases the average stress level. State null and alternative hypotheses. What type of test is appropriate in this scenario?

$H_0 : \mu_d = 0$, i.e. twin 1 and twin 2 have the same stress levels

$H_a : \mu_d < 0$, i.e. twin 1 is less stressed than twin 2

We will use signed test as n is small and we cannot apply normal approximation.

2. Restate the hypotheses in terms of p, the probability to observe positive difference between stress levels of twins (twin 1 - twin 2).

$H_0 : p = 0.5$, i.e. we have equal chance to observe positive and negative differences.

$H_a : p < 0.5$, i.e. we will observe positive difference less often than the negative one.

3. What would be the test statistic for this test?

The test statistic is the number of positive differences N.

4. What would be the null distribution? Draw the null distribution.

Under the null, $N \sim Binomial(8, 0.5)$. We use the binomial table for the second part.

4. What is the observed value of test statistic?

We have $n_{obs} = 3$ positive differences (twin pairs 1, 4 and 5).

5. Find the p-value.

As this alternative is one-sided

$p\text{-}value = P(N \leq n_{obs}) = P(N = 0) + P(N = 1) + P(N = 2) + P(N = 3) = 0.004 + 0.031 + 0.109 + 0.219 = 0.363$

6. What conclusion can we draw at significance level 0.1?

As $p\text{-}value > 0.1$ we do not have enough evidence to conclude that sport decreases stress levels.

Question 2

A study was conducted on 50 male and 50 female first-year students at U of T.

The stress level for each study participant was recorded as a score from 0 to 100 and the summary statistics were computed.

```
mean(male)
```

```
## [1] 50.47862
```

```
sd(male)
```

```
## [1] 18.32589
```

```
mean(female)
```

```
## [1] 58.15811
```

```
sd(female)
```

```
## [1] 23.72332
```

1. You want to test if average stress level is different for male and female students. What test will you use? State null and alternative hypotheses in terms of the male and female population averages.

We use t-test for non-matching pairs.

If x and y correspond to male and female samples, respectively, then

$H_0 : \mu_x = \mu_y$, i.e. male and female students have the same stress levels

$H_a : \mu_x \neq \mu_y$, i.e. male and female students have different stress levels

2. Well, compute degrees of freedom for this test :(

We use "pooled" degrees of freedom formula and get $df = 92.124$ (we can approximate it by $df = 92$).

3. Compute the $t_{df}^{\alpha/2}$ quantile for 90% confidence interval.

We don't have $df = 92$ in the table, so we use $df = 90$ instead and approximate $t_{92}^{0.05} \approx 1.66$.

4. Compute 90% confidence interval for the difference in population means.

$$[\bar{x} - \bar{y} - 1.66\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}, \bar{x} - \bar{y} + 1.66\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}] = [-14.72, -0.64]$$

5. What conclusion can we draw from the confidence interval?

As it doesn't cover zero we can reject the null hypothesis and say with 90% confidence that there is a difference in stress levels between male and female students.

6. Now find the upper 90% CI for the difference in population means.

First we find $t_{92}^{0.1} = 1.29$ (again use $df = 90$ from the table).

$$[\bar{x} - \bar{y} - 1.29\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}, +\infty) = [-13.15, +\infty)$$

7. What alternative hypothesis corresponds to this CI? What conclusion can we draw from this CI?

One-sided alternative

$H_a : \mu_x > \mu_y$, i.e. male students are more stressed than female students

It covers zero, thus we do not have enough evidence to conclude that male students are more stressed than female students.

8. Suppose that we know that the population variances for male and female stress levels are equal, i.e. $\sigma_{male}^2 = \sigma_{female}^2$. How can you use the summary statistics to approximate the values of the population variances?

Use "pooled" variance formula:

$$\sigma_{male}^2 = \sigma_{female}^2 \approx s^2 = 449.32$$

9. Find test statistic t_{obs} for the case when $\sigma_{male}^2 = \sigma_{female}^2$.

$$t_{obs} = \frac{\bar{x} - \bar{y}}{\sqrt{s^2(1/n + 1/m)}} = -1.81$$

10. Suppose you want to check that female students are more stressed than male students for the case when $\sigma_{male}^2 = \sigma_{female}^2$. What would be the p-value?

We use one-sided alternative

$H_a : \mu_x < \mu_y$, i.e. male students are less stressed than female students

The $pvalue = P(T < t_{obs})$ where T is a random variable with $df = n + m - 2 = 98$.

We use $df = 100$ in the table and conclude that $pvalue$ is between 0.025 and 0.05.

11. What conclusion can you make for the hypothesis from 10 at significance level $\alpha = 0.05$?

We can reject the null hypothesis and conclude that female student are more stressed with 95% confidence.

Question 3

A study was conducted on 50 male and 50 female first-year students at U of T.

Each study participant was asked if they feel stressed. The following results were received:

- 30 out of 50 female students are stressed

- 25 out of 50 male students are stressed

1. You want to test if proportions of stressed male and female students are different. What test will you use? State null and alternative hypotheses.

We use t-test for proportions for non-matching pairs.

If x and y correspond to male and female samples, respectively, then

$H_0 : p_x = p_y$, i.e. male and female categories have the same proportions of stressed students

$H_a : p_x \neq p_y$, i.e. male and female categories have different proportions of stressed students

2. Find the value of observed statistic.

As $p_x = p_y$ under the null, we can use "pooled" estimate for these proportions.

$p_x = p_y \approx \frac{30+25}{100} = 0.55$

Then $z_{obs} = \frac{0.6-0.5}{\sqrt{0.55(1-0.55)(1/50+1/50)}} \approx 1$

3. Find the p-value.

For two-sided alternative

$p\text{-}value = P(|Z| > |z_{obs}|) = 2 \cdot 0.159 = 0.318$

4. Can we conclude that female students stress out more often at significance level 0.05?

No, as $p\text{-}value > 0.05$ we do not have enough evidence to reject the null hypothesis.