Practice 10

Elena Tuzhilina

March 21, 2023

Question

A study was conducted on 50 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	difference
1	75.25909	65.98412	9.274965
2	43.47533	46.69892	-3.223586
3	76.59599	100.00000	-23.404015
4	75.44859	34.13983	41.308760
5	58.29283	56.35368	1.939153
6	19.20100	65.50424	-46.303241
7	31.42866	76.54730	-45.118640
8	44.10559	52.82129	-8.715704
9	49.88466	0.00000	49.884656
10	98.09307	20.09157	78.001499
11	65.27187	68.76187	-3.489998
12	34.01982	57.66864	-23.648821
13	27.04686	29.78053	-2.733665
14	44.21077	54.52524	-10.314472
15	44.01570	33.55094	10.464759
16	41.76978	65.26790	-23.498121
17	55.04447	15.24705	39.797421
18	32.16158	68.97823	-36.816656
19	58.71367	65.45238	-6.738714
20	25.24923	59.95865	-34.709414
21	45.51464	58.57469	-13.060049
22	57.54791	65.72015	-8.172238
23	52.66673	38.52970	14.137030
24	66.08379	54.42494	11.658853
25	48.85786	77.92407	-29.066206
26	60.07216	91.02907	-30.956914
27	71.71539	62.31314	9.402243
28	36.18092	54.46739	-18.286469
29	24.30801	30.63795	-6.329936
30	50.93452	14.87241	36.062111

pair	twin1	twin2	difference
31	45.28587	34.08731	11.198555
32	39.14223	95.62249	-56.480258
33	41.33379	81.16427	-39.830472
34	37.01057	51.41453	-14.403964
35	64.53501	45.25569	19.279323
36	73.03824	45.43060	27.607638
37	69.84321	87.90961	-18.066398
38	41.40974	49.72666	-8.316921
39	74.76608	95.68056	-20.914482
40	44.41307	77.40023	-32.987157
41	85.15806	96.97937	-11.821307
42	61.21492	31.80214	29.412785
43	40.94432	58.25113	-17.306808
44	33.35913	31.57385	1.785286
45	26.66859	75.88777	-49.219181
46	28.68819	61.59153	-32.903341
47	18.72436	49.53478	-30.810425
48	73.13074	100.00000	-26.869260
49	66.64094	64.87059	1.770355
50	45.45343	87.89632	-42.442892

The difference variable in the data set represents stress level of twin 1 minus stress levels of twin 2. Here are the sample mean and standard deviation for the difference.

```
mean(difference)
```

```
## [1] -7.679487
```

```
sd(difference)
```

```
## [1] 27.98065
```

1. You want to test if sport influences average stress level. State null and alternative hypotheses in terms of the population average of the difference μ_d.

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

$H_a : \mu_d \neq 0$, i.e. twin 1 and twin 2 have different stress levels

2. Compute the 90% confidence interval for the difference.

$[\bar{d} - 1.68\frac{s_d}{\sqrt{n}}, \bar{d} + 1.68\frac{s_d}{\sqrt{n}}] = [-14.31, -1.05]$

3. Does the confidence interval cover zero? What conclusion can be drawn?

It does not. Thus we can reject H_0 in favor of H_a and we are 90% confident that sport impacts stress level.

4. Now find the test statistics and estimate the p-value.

$t_{obs} = \frac{\bar{d}-0}{s_d/\sqrt{n}} = -1.94$

$p\text{-}value = P(|T| > |t_{obs}|) = 0.06$

    5. Draw a conclusion using p-vlaue from 4. What significance level we need to use to make the results consisted with the conclusion from 3?

Significance level $\alpha = 0.1$. As $p\text{-}value < 0.1$ we can reject H_0 in favor of H_a.

    6. Perform statistical testing to check if sport reduces the stress level, use significance level $\alpha = 0.01$.

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

$H_a : \mu_d < 0$, i.e. twin 1 has lower stress level than twin 2

$p\text{-}value = P(T < t_{obs}) = 0.03$

$p\text{-}value > 0.01$, we cannot reject H_0 in favor of H_a. We do not have enough evidence to say that sport reduces the stress level with 99% confidence.

    7. Suppose we know that sport has no influence on stress. What type of error did we make in 3?

type I error = reject null if null is true

    8. Provide the reason why we may made this error.

Significance level $\alpha = 0.1$ implies that type I error will appear in 10% of the studies.

    9. Suppose we know that sport reduces the stress level. What type of error did we make in 6?

type II error = fail to reject null if alternative is true

    10. How can we reduce the probability to make this error?

Increasing α will increase the power of statistical test and decrease the chance to make type II error.