Analysis of Scientific Data – Summer Semester, 2022

Week 5 Statistics Half of the Tutorial Solutions

Part A

Dextroamphetamine is a drug that has been commonly used to treat hyperkinetic children. A paper in the *Journal of Nervous and Mental Disorders* (1968, vol. 146, pp. 136–146) reported the following data on the amount of dextroamphetamine (%) excreted in a trial over 7 hours by a sample of children having organically related disorders and a sample of children with non-organic disorders:

Organic	17.53	20.60	17.62	28.93	27.10
NonOrganic	15.59	14.76	13.32	12.45	12.79

The mean amount excreted for the organic group is 22.36 % with a standard deviation of 5.35 %, while for the non-organic group the mean is 13.78 % with a standard deviation of 1.34 %. Does this give evidence of a difference in mean dextroamphetamine excretion between the two groups?

a) State the null and alternative hypotheses of interest in terms of μ_1 and μ_2, the underlying mean dextroamphetamine excretion for the two populations.

Organic group – Group1; NonOrganic group – Group 2

$H_0: \mu_1 = \mu_2$ Vs $H_1: \mu_1 \neq \mu_2$

b) Does a pooled two-sample t-test seem appropriate here? Why or why not?

Pooled two sample t-test is not appropriate.

Pooled two sample t-test assumes that the unknown population standard deviations are equal. The sample standard deviations are very different to each other. Therefore, it is not reasonable to assume that unknown population standard deviations are equal.

c) Calculate the t statistic to test this null hypothesis.

$$t_{\text{stat}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(22.36 - 13.78) - (0)}{\sqrt{\frac{5.35^2}{5} + \frac{1.34^2}{5}}} = \frac{8.58 - 0}{2.4655} = 3.4786$$

d) Considering the concept of using the minimum of two sample sizes, what is the corresponding P-value? What do you conclude?

p-value = $2 * P(t_{stat} \geq 3.4786) = 2 * 0.011269213 = 0.0254$

In R: 1-pt(3.4786, df=4) = 0.01269213

Moderate evidence ($p < 0.05$) to conclude that mean dextroamphetamine excretion is different for two groups.

e) Calculate the 90% confidence interval for the difference in the mean dextroamphetamine excretion between two groups. (Hint: use the degrees of freedom for t-distribution from part d)).

90% CI for $(\mu_1 - \mu_2)$ takes the following form.

$(\bar{x}_1 - \bar{x}_2) \pm t^* se((\bar{x}_1 - \bar{x}_2)$

Using qt(0.95, df=4) in RStudio, $t^* = 2.131847$

```
qt(0.95, 4)
[1] 2.131847
```

$(22.36 - 13.78) \pm 2.131847 \times 2.4655$

8.58 ± 5.256068

$(3.323932, 13.83607)$

$3.32\% \leq (\mu_1 - \mu_2) \leq 13.84\%$

f) Create a "csv" file using the data and perform Welch t-test in RStudio to test the hypothesis in part a).

g) Using the Welch t-test results from part f), what is the margin of error in a 95% confidence interval for the difference in the mean dextroamphetamine excretion between two groups?

The results from Welch t-test:

```
        Welch Two Sample t-test

data:  Dextroamp by Group
t = -3.4755, df = 4.5008, p-value = 0.02104
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -15.133208  -2.014792
sample estimates:
mean in group NonOrganic    mean in group Organic
                  13.782                   22.356
```

Please note that RStudio has treated NonOrganic group as the group 1 and Organic group as the group 2 following the alphabetical order of the first letter in the categorical variable levels (or groups). Thus 95% CI has been calculated for the difference in the mean dextroamphetamine excretion between NonOrganic and Organic groups.

$$MOE = \frac{-2.014792 - (-15.133208)}{2}$$

$$MOE = \frac{13.11842}{2} = 6.5592$$

Part B

A study reported on the survival rate among patients suffering cardiac arrest both when resuscitation was started by (trained) lay people and when it was delayed until the arrival of an ambulance crew. A total of 27 patients survived out of the 75 attended by lay people, while 130 survived out of the 556 attended by an ambulance crew.

a) Give a 95% confidence interval for the difference between p_1 and p_2, where p_1 denotes the proportion of successful resuscitations in the lay-trained class and p_2 the proportion of delayed successful resuscitations by ambulance crews.

$\hat{p}_1 = \frac{27}{75} = 0.36$ $\hat{p}_2 = \frac{130}{556} = 0.2338$

95% CI for $(p_1 - p_2)$ is: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$$(0.36 - 0.2338) \pm 1.959964\sqrt{\frac{0.36(1-0.36)}{75} + \frac{0.2338(1-0.2338)}{556}}$$

$0.1262 \pm 1.959964*0.05826$

0.1262 ± 0.114189, or $(0.0120, 0.2404)$

b) Is there any evidence of a difference in the proportion of successful resuscitations in the lay-trained class compared to the proportion of delayed successful resuscitations by ambulance crews?

95% CI for $(p_1 - p_2)$ does not include zero (0). We have evidence to conclude that there is a difference in the two population proportions.

c) How would the analysis need to change if the researchers were aiming to show that a program to train lay people to perform resuscitation would be beneficial?

We need to perform a hypothesis test as follows (i.e. an upper tail test)

$H_0: p_1 = p_2$ Vs $H_1: p_1 > p_2$

d) What is the conclusion from the analysis in part c)?

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = \frac{(0.36-0.2338)-0}{0.05826} = 2.166152$$

p-value = 0.015149 ← in R: 1-pnorm(2.16652) or pnorm(-2.16652)

Moderate evidence to conclude that a program to train lay people to perform resuscitation would be beneficial.

PART C

About Paper Review which is due on 11 Jan 2023 at 3:00 pm.