Assignment 2

Tim But, Mehmet Bedirhan Gursoy, Vincentas Ryliskis, group 035

20 November 2023

Setting a Seed

In order to ensure that the results in this report are reproducible, seed (123) was used when generating random numbers.

```
set.seed(123)
```

Task 1

a)

In order to achieve a 95% confidence interval, we will use a Z-value of $Z = 1.96$, and in order to achieve a margin of error of 2%, we will use a value of $E = 0.02$. Due to the fact that we are working with a proportion, our calculation for the number (n) of surveyed flights should look as follows:

$n = (Z\frac{\sqrt{p(1-p)}}{E})^2$

Because we do not have a value for p, we can assume that p is equal to 0.5 - the value at which $p(1-p)$ will have it's highest value. Therefore, after accounting for all values, we get:

$n = (1.96\frac{\sqrt{0.5(1-0.5)}}{0.02})^2 = (1.96\frac{\sqrt{0.25}}{0.02})^2 = 2401$

Therefore, we need 2401 samples in order to get the desired margin of error over the desired confidence interval

b)

As the desired confidence interval and Z score remain the same as in part a), we can keep the previously used values of $E = 0.02$ and $Z = 1.96$. Additionally, the formula for calculating n stays the same:

$n = (Z\frac{\sqrt{p(1-p)}}{E})^2$

However, we are given a value for p in this part: 0.9. For this value of p, the calculation for n looks as follows:

$n = (1.96\frac{\sqrt{0.9(1-0.9)}}{0.02})^2 = (1.96\frac{\sqrt{0.09}}{0.02})^2 = 864.36$

Therefore, we need 865 samples in order to get the desired margin of error over the desired confidence interval

Task 2

In task 2, we are given the following values:

$\bar{x}_1 = 1124.3$

$\bar{x}_2 = 1118.1$

$s_d = 57.8$

$n = 30$

In order to find a confidence interval for $\bar{x}_1 - \bar{x}_2$, we will use the formula for two dependent samples, as the two sets are of first and second-born twins:

$[\bar{d} - t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}, \bar{d} + t_{n-1,\alpha/2}\frac{s_d}{\sqrt{n}}]$

Where $\bar{d} = \bar{x}_1 - \bar{x}_2 = 6.2$. Therefore, when we substitute the numbers, we get:

$[6.2 - t_{29,0.05}\frac{57.8}{\sqrt{30}}, 6.2 + t_{29,0.05}\frac{57.8}{\sqrt{30}}]$

Further evaluating the numbers (and critical t values) gives us:

$[6.2 - 1.699\frac{57.8}{\sqrt{30}}, 6.2 + 1.699\frac{57.8}{\sqrt{30}}] = [6.2 - 17.929, 6.2 + 17.929]$

Resulting in a confidence interval of:

$[-11.729, 6.2 + 24.129]$

Task 3

Test Claim: Whether there is a difference in the population mean volumes of first- and second-born babies.

$H_0 : \mu1 = \mu2, H_1 : \mu1 \neq \mu2, \alpha = 0.05$

In task 3, we are given the following values:

$\bar{x}_1 = 1131.3$

$\bar{x}_2 = 1123.8$

$s_1 = 129.0$

$s_2 = 127.2$

$n_1 = 25$

$n_2 = 20$

We will be using the following test statistic:

$t_2 = \frac{x_1 - x_2}{\sqrt{\frac{s_1^2}{n1} + \frac{s_2^2}{n2}}}$

When all values are accounted for, it looks as follows:

$t_2 = \frac{1131.3 - 1123.8}{\sqrt{\frac{129^2}{25} + \frac{127.2^2}{20}}} = \frac{7.5}{\sqrt{\frac{16641}{25} + \frac{16179.84}{20}}} = \frac{7.5}{\sqrt{1474.632}} = 0.195$

Under the null hypothesis, the test statistic has a t-distribution with $\tilde{n}$ degrees of freedom. Without a calculator, we can make the assumption that $\tilde{n} = min(n_1 - 1, n_2 - 1) = min(24, 19) = 19$.

$t_{19,0.05/2} = 2.093$

Since $0.195 > -2.093$ and $0.195 < 2.093$, we fail to reject H_0. There is not sufficient evidence to reject the claim that first- and second-born children have the same mean brain volume.

Assumptions Made:
- $\tilde{n} = min(n_1 - 1, n_2 - 1)$ - This is an estimate of the true value of $\tilde{n}$
- $H_0 : \mu 1 = \mu 2$ - This assumption was made in order to evaluate the data.
- Additionally, a t-distribution is assumed for the collected data.

Task 4

Task 5