Student Information

Name : Batuhan Karaca

ID : 2310191

Answer 1

NOTE: First group stands for the sample of 19 people with age 40 and above, while second group stands for the sample of 15 people aged under 40, in the sample of 34 people in the UK.

a)

We are asked to find the *Confidence interval*. We cannot use the *Normal Distribution* since the population Standard Deviations σ_x, σ_y are unknown. Furthermore, we cannot approximate our confidence interval using the *Normal Distribution*, since we have relatively small sample sizes for the first and second group ($19 < 30$ and $15 < 30$). Therefore, we should use ***Student's T-distribution***. We are asked to find the *Confidence interval* *on the difference between two means*. We have two cases, for equal and unequal variances. First group's standard variation is $s_x = 0.96$, and second group's standard variation is $s_y = 1.12$. Since standard variations s_x, s_y are not equal, variances s_x^2, s_y^2 are not equal either. We use the formula of the *confidence interval for the difference of means* using these unequal variances s_x^2, s_y^2, in page 263 of the textbook.

$$\bar{X} - \bar{Y} \pm t_{\frac{\alpha}{2}} \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$$

where $\bar{X} = 3.375$ is the sample mean of the first group, whereas $\bar{Y} = 2.05$ is the sample mean of the second group. $t_{\frac{\alpha}{2}}$ is a critical value from *T-distribution* with ν degrees of freedom given by the formula below (which is also in page 263)

$$\nu = \frac{\left(\frac{s_x^2}{n} + \frac{s_y^2}{m}\right)^2}{\frac{s_x^4}{n^2(n-1)} + \frac{s_y^4}{m^2(m-1)}}$$

which is also known as *Satterthwaite approximation*. In both formulas, $n = 19$ is the sample size of the first group, while $m = 15$ is the sample size of the second group. Subsituting the values in *Satterthwaite approximation*

$$\nu = \frac{(\frac{(0.96)^2}{(19)} + \frac{(1.12)^2}{(15)})^2}{\frac{(0.96)^4}{(19)^2((19)-1)} + \frac{(1.12)^4}{(15)^2((15)-1)}}$$

$$\nu \sim 27.7$$

We take the closest integer $\nu = 28$, since the table of ***Student's T-distribution***, in page 419, contains integers not exceeding 200. We need to find $(1-\alpha)100\% = 95\%$ confidence interval. Therefore $\alpha = 0.05$. Looking at the table, at the column where $\alpha/2 = 0.025$ and the row where $\nu = 28$, we find $t_{\frac{\alpha}{2}} = 2.048$. Subsituting values in the confidence interval formula, the *Confidence interval (I)* is

$$I = 3.375 - 2.05 \pm 2.048\sqrt{\tfrac{0.96^2}{19} + \tfrac{1.12^2}{15}}$$

$$I = 1.325 \pm 2.048\sqrt{\tfrac{0.96^2}{19} + \tfrac{1.12^2}{15}}$$

$$I \sim [0.580, 2.069]$$

Note that both ends of the interval (I) is approximated

b)

We are asked a question similar to **part a**. Similar to **part a**, we will use the *Confidence interval* formula. However, in this case $(1-\alpha)100\% = 90\%$, therefore $\alpha = 0.1$, hence $\frac{\alpha}{2} = 0.05$. From the table of ***Student's T-distribution*** in page 419, at the column where $\alpha/2 = 0.05$ and the row where $\nu = 28$,(ν does not change) we find $t_{\frac{\alpha}{2}} = 1.701$. Subsituting values in the confidence interval formula,(other values do not change) the *Confidence interval (I)* is

$$I = 3.375 - 2.05 \pm 1.701\sqrt{\tfrac{0.96^2}{19} + \tfrac{1.12^2}{15}}$$

$$I = 1.325 \pm 1.701\sqrt{\tfrac{0.96^2}{19} + \tfrac{1.12^2}{15}}$$

$$I \sim [0.707, 1.943]$$

Note that both ends of the interval (I) is approximated

c)

We shall calculate the 95%*confidence interval* of the first group *using Student's T-distribution*, since as stated in **part a**, sample size of the first group $n = 19$ is small $(19 < 30)$. In textbook page 259, formula of the *confidence interval for mean* is given

$$\bar{X} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

In this case, degrees of freedom $\nu = n - 1 = 18$ ($n = 19$ is introduced in **part a**). Looking at the table of ***Student's T-distribution***, in page 419, at the column where $\alpha/2 = 0.025$ (from **part a**, since the confidence level is the same at 95%) and the row where $\nu = 18$ we find $t_{\frac{\alpha}{2}} = 2.101$. Subsituting the values, $\bar{X} = 3.375,\ s = s_x = 0.96,\ n = 19$ (please refer to **part a**) ,and $t_{\frac{\alpha}{2}} = 2.101$, the *Confidence interval (I)* is

$$I = 3.375 \pm (2.101)\frac{0.96}{\sqrt{19}}$$
$$I \sim [2.912, 3.838]$$

The sample mean $\bar{X} = 3.375$ is greater than 3, does not indicate that the population mean is greater than 3. We may have a sampling error. This data does not guarantee that people with age 40 and above supports BREXIT with 95% confidence level.

Answer 2

The company claims that they are producing 20.00kg olympic bars. We can test whether the average weight of the bars is 20.00kg or not. Then, the *null hypothesis*

a)

$$H_0 : \mu = 20.00$$

, and the *alternate hypothesis*

b)

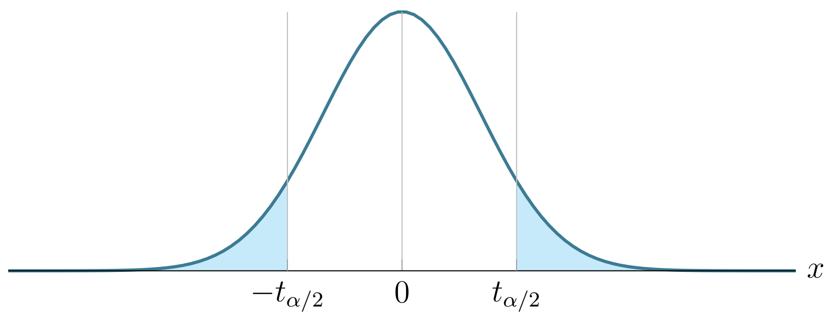
$$H_A : \mu \neq 20.00$$

which is a **two sided alternate hypothesis**. It is stated that

if the statistical significance is above 1%, they stop production (statement 1)

which means for significance level $\alpha = 0.01$, the staff stops producing when the statistic $|T| \geq t_{\alpha/2}$. Absolute value of T comes from the fact that we stated a **two sided alternate hypothesis**. Since our sample size is small, and the *population standard deviation* σ is unknown, we are conducting a *t-test*, *not a z-test*. That is the reason we use $t_{\alpha/2}$ instead of $z_{\alpha/2}$ as the critical value. The areas under the curve where the staff stops producing are the rejection regions. The diagram is below

c)



t-test diagram for the experiment. The blue color indicates the rejection regions.

In textbook page 276, formulas for different type of t statistics are given in a table. Since our *null hypothesis* is of the form $\mu = \mu_0$, the formula for t

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

with degrees of freedom $n - 1 = 10$, since we have 11 samples.

First, we find the critical value $t_{\frac{\alpha}{2}}$. Looking at the *(statement 1)*, we see the level of significance $\alpha = 0.01$, the area of the non-colored region under the curve in the diagram , between $\pm t_{\frac{\alpha}{2}}$. . Looking at the table of ***Student's T-distribution***, in page 419, at the column where $\alpha/2 = 0.005$ and the row where degrees of freedom is 10, we find $t_{\frac{\alpha}{2}} = 3.169$.

We can calculate t. Subsituting values ($\bar{X} = 20.07$, $s = 0.07$, $n = 11$ are the data from the sample in the question; and $\mu_0 = 20.00$ is the population mean), we find

$$t = \frac{20.07 - 20.00}{0.07/\sqrt{11}}$$

$$t = \frac{0.07}{0.07/\sqrt{11}}$$

$$t = \sqrt{11}$$
$$t \sim 3.317$$

We reach $|t| = 3.317 \geq t_{\frac{\alpha}{2}} = 3.169$. The statistic t is within the field of rejection, hence we reject H_0. Therefore, we conclude there is significant evidence in favor of H_A, and the staff should stop production.

Answer 3

We have one sample of $n = 68$ people that are experimented. The company claims the headache allows for smaller duration of headache, compared to other painkillers in the market which show 3 minutes on average on the group. We are comparing in our analysis. Therefore, the *null hypothesis*

a)

$$H_0 : \mu_X - \mu_Y = 0$$

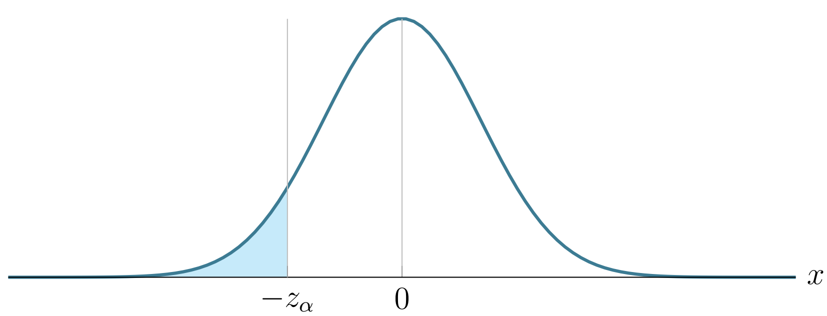
, and the *alternate hypothesis*

b)

$$H_A : \mu_X - \mu_Y < 0$$

Which is a **one-sided, left-tail alternate hypothesis**. μ_X is the mean of the experiment with the new painkiller whereas μ_Y is the mean of the experiment with other painkillers in the market. Since our sample size is large, we can approximate $s \sim \sigma$, using a *z-test*. We use $-z_\alpha$ for the critical value, since the *alternate hypothesis* is *left-tailed* (We are not dividing by 2). Below is the diagram

c)



z-test diagram for the experiment. The blue color indicates the rejection region.

We know the area of the rejection region is α that we can observe the *cdf* at $-z_\alpha$ (the distribution is *Normal* of course)

$\Phi(-z_\alpha) = \alpha$, then
$z_\alpha = -\Phi^{-1}(\alpha)$

We find the critical value $-z_\alpha = -z_{0.05}$ (5% level of significance is given). We need to find $-z_{0.05}$.

Looking at the table of ***Standard Normal distribution***, in page 417, we see that the row with number -1.6 and and the column with number -0.05 intersects at 0.0495, then

$\Phi(-1.65) = 0.0495$
$-1.65 = \Phi^{-1}(0.0495)$
$1.65 = -\Phi^{-1}(0.0495) \sim -\Phi^{-1}(0.05)$
$-z_{0.05} \sim 1.65$

We can calculate Z, our statistic for the *Standard Normal null distribution*, that we symbolically represented as the diagram above. In textbook page 273, formulas for different type of Z statistics are given in a table. Since our *null hypothesis* is of the form $\mu_X - \mu_Y < D$, the formula for Z

$$Z = \frac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$$

Subsituting values ($\bar{X} = 2.8, \bar{Y} = 3, \sigma_x \sim s_x = 1.7, \sigma_y \sim s_y = 1.4, n = m = 68$ are the data from the sample in the question; and mean difference $D = 0$,we find

$$Z = Z = \frac{2.8 - 3 - 0}{\sqrt{\frac{(1.7)^2}{68} + \frac{(1.4)^2}{68}}}$$

$$Z = \frac{-0.2}{\sqrt{4.85/68}}$$

$$Z \sim -0.749$$

We reach $Z = -0.749 \leq -z_\alpha = 1.65$. The statistic Z is within the field of rejection, hence we reject H_0. Therefore, we conclude there is significant evidence in favor of H_A, that is $\mu_X - \mu_Y < 0$, $\mu_X < \mu_Y$. we can state the new painkiller produces better results