Student Information

Name : Onat Özdemir

ID : 2310399

Answer 1

a)

We can divide the UK into two different populations: people with age 40 and above and people with age under 40. Let X and Y represents our independently chosen samples which are belong to the people with age 40 and above and people with age under 40, respectively. Additionally, let μ_X and μ_Y represent the population means of the people with age 40 and above and the people with age under 40, respectively and $\bar{X}$ and $\bar{Y}$ be sample means that estimate μ_X and μ_Y, respectively.

As can be seen, population standard deviations are unknown (not given) in this question. If we would have sufficiently large samples where the both size of X and the size of Y are above 30 then sample standard deviations would be accurate estimators for population standard deviations and then we could find the confidence interval by using Z-statistic. However, since we do not have sufficiently large samples we have to use T statistic to calculate the confidence interval.

Firstly, let's assume that the two population have equal variances, $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. With this assumption, we can use the formula given in "Probability and Statistics for Computer Scientists, page 262, Confidence interval for the difference of means; equal, unknown standard deviations" to calculate the confidence interval;

$$\bar{X} - \bar{Y} \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} \tag{1}$$

where $t_{\alpha/2}$ is a critical value from T-distribution with $(n+m-2)$ degrees of freedom and α is the significance level. And s_p is the pooled standard deviation where

$$s_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} \tag{2}$$

where s_X and s_Y are sample standard deviations, n and m are sizes of samples.

From the question we have, n(sample size of the first group) = 19, m(sample size of the second group) = 15, $s_X = 0.96$, $s_Y = 1.12$, $\bar{X} = 3.375$ and $\bar{Y} = 2.05$. Also, since question asks for the 95% confidence interval then α $(significance\ level) = 1 - 0.95 = 0.05$. (from $(1-\alpha)100\% = 95\%$)

To find $t_{\alpha/2} = t_{0.025}$ we can use the Student's T-Distribution Table (A5) given in "Probability and Statistics for Computer Scientists, page 419" by taking degrees of freedom $= (n+m-2) = 19+15-2 = 32$. Since according to the table, $P\{t > 2.037\} = 0.025$ then $t_{0.025} = 2.037$.

Then, let us calculate the s_p using the formula give in (2);

$$s_p = \sqrt{\frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}} = \sqrt{\frac{(18)*(0.96)^2 + (14)*(1.12)^2}{32}} = 1.03305 \tag{3}$$

To calculate 95% confidence interval on the difference between the means, put the parameters we have found to (1),

$$\bar{X} - \bar{Y} \pm t_{0.025} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} = 3.375 - 2.05 \pm 2.037 * 1.03305 * \sqrt{\frac{1}{19} + \frac{1}{15}} \tag{4}$$

Thus, 95% confidence interval on the difference between the means equals to $[0.598176, 2.05182]$

b)

In 1-a, the definitions and values of $\bar{X}$, $\bar{Y}$, n, m, s_p, s_X, s_Y are given and since the values of those parameters do not depend on the confidence level, I will use those parameters with the same definitions and values as given in 1-a.

To calculate 90% confidence interval on the difference between the means we can use the same formula given in (1);

$$\bar{X} - \bar{Y} \pm t_{\alpha/2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} \tag{5}$$

Although other parameters don't depend on the confidence level, $t_{\alpha/2}$ depends on it so we should recalculate it. Since question asks for the 90% confidence interval then α (*significance level*) $= 1 - 0.90 = 0.10$ (from $(1-\alpha)100\% = 90\%$).

To find $t_{\alpha/2} = t_{0.05}$ we can use the Student's T-Distribution Table (A5) given in "Probability and Statistics for Computer Scientists, page 419" by taking degrees of freedom $= n+m-2 = 19+15-2 = 32$. Since according to the table, $P\{t > 1.694\} = 0.05$ then $t_{0.05} = 1.694$.

To calculate 90% confidence interval on the difference between the means, put the parameters we found in 1-a and $t_{0.05}$ to (5);

$$\bar{X} - \bar{Y} \pm t_{0.05} s_p \sqrt{\frac{1}{n} + \frac{1}{m}} = 3.375 - 2.05 \pm 1.694 * 1.03305 * \sqrt{\frac{1}{19} + \frac{1}{15}} \tag{6}$$

Thus, 90% confidence interval on the difference between the means equals to $[0.720562, 1.92944]$

c)

We are interested in if the age 40 and above supports BREXIT or equivalently if the population mean of the people with age 40 and above (μ_X) is higher than 3 with 95% confidence level. Therefore, we will test H_0 against a one sided right tail alternative H_A using t-test (we will use t-test since the population standard deviation is unknown and our sample size is not sufficiently large to estimate population standard deviation accurately($n = 19 < 30$)) where;

$$H_0 : \mu_X = 3 \qquad H_A : \mu_X > 3 \tag{7}$$

We are given that $s_X = 0.96$, $\bar{X} = 3.375$, n = 19 and $\alpha(level\ of\ significance) = 1 - 0.95 = 0.05$. (from $(1 - \alpha)100\% = 95\%$)

Since we have one sided right tail alternative, if $t \geq t_\alpha$ we will reject H_0 and if $t < t_\alpha$ then we will say that there is no significant evidence to reject H_0 where t is our T-statistic and $t_\alpha = t_{0.05}$ represents the critical value from T-distribution with $n - 1 = 19 - 1 = 18$ degrees of freedom (the reason behind using t_α as our critical value rather than $t_{\alpha/2}$ is that as stated before we have one sided right tail alternative). Therefore our rejection region $R = [t_{0.05}, \infty)$. To find $t_\alpha = t_{0.05}$ we can use Student's T-Distribution Table (A5) placed in "Probability and Statistics for Computer Scientists", by taking $df(degrees\ of\ freedom) = n - 1 = 19 - 1 = 18$ and $\alpha = 0.05$, hence we have found that $t_{0.05} = 1.734$.

Then, let us calculate the T-statistic by using the formula given in "Probability and Statistics for Computer Scientists, page 276, table 9.2", since the population standard deviation is not given we will estimate it with s_X;

$$t = \frac{\bar{X} - 3}{s_X/\sqrt{n}} = \frac{3.375 - 3}{0.96/\sqrt{19}} = 1.70269 \tag{8}$$

Since $t = 1.70269 < t_{0.05} = 1.734$, $t \notin R$. Hence, we cannot reject H_0 since we do not have significant evidence. Thus, we cannot say people with age 40 and above supports BREXIT with 95% confidence level.

Answer 2

a)

$$H_0 : \mu = \mu_0$$

where μ represents the population mean of weights (in kg) of the olympic bars in the line while μ_0 represents the average weight of the olympic bars produced by the company where $\mu_0 = 20\ kg$ as given in the text.

b)

$$H_A : \mu \neq \mu_0$$

where μ represents the population mean of weights (in kg) of the olympic bars in the line while μ_0 represents the average weight of the olympic bars produced by the company where $\mu_0 = 20\ kg$ as given in the text.

c)

We will test H_0 against a two sided alternative H_A using one sample t-test.
In the question, we are given;

$$\begin{aligned} \mu_0\ (population\ mean) &= 20 \\ \bar{X}\ (sample\ mean) &= 20.07 \\ s\ (sample\ standart\ deviation) &= 0.07 \\ \alpha\ (level\ of\ significance) &= 0.01 \\ n\ (sample\ size) &= 11 \end{aligned} \tag{9}$$

Since we have two sided alternative, if $|t| \geq t_{\alpha/2}$ we will reject H_0 and if $|t| < t_{\alpha/2}$ then we will say that there is no significant evidence to reject H_0 where t is our T-statistic and $t_{\alpha/2} = t_{0.005}$ represents the critical value from T-distribution with $n - 1 = 11 - 1 = 10$ degrees of freedom. Therefore our rejection region $R = (-\infty, -t_{0.005}] \cup [t_{0.005}, \infty)$. To find $t_{\alpha/2} = t_{0.005}$ we can use Student's T-Distribution Table (A5) placed in "Probability and Statistics for Computer Scientists", by taking $df(degrees\ of\ freedom) = n - 1 = 10$ and $\alpha = 0.005$, hence we have found that $t_{0.005} = 3.169$.

t-test diagram is given below where the coloured area represents rejection regions:

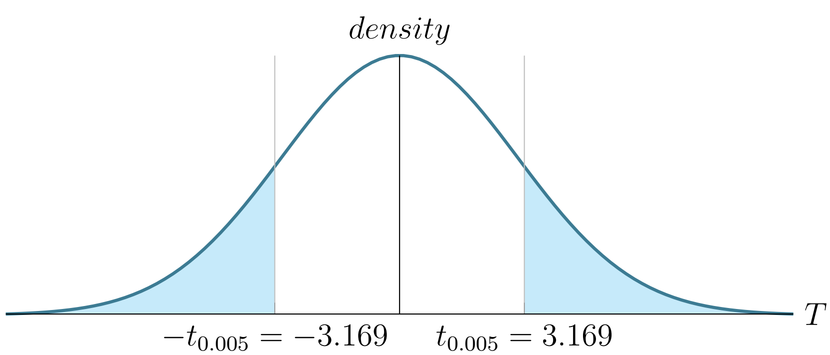

Figure 1: t-test diagram

Then, let us calculate the T-statistic by using the formula given in "Probability and Statistics for Computer Scientists, page 276, table 9.2" and the values given in (9);

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{20.07 - 20}{0.07/\sqrt{11}} = 3.31662 \tag{10}$$

Since $|t| = 3.31662 \geq t_{0.005} = 3.169$, $t \in R$. Thus, we reject the null hypothesis H_0 and conclude that there is a significant evidence of an undesired production line. In conclusion, they should stop production and check the line.

Answer 3

a)

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

where μ_X represents the population mean of the time (in minute) that the new painkiller reduces headache in while μ_Y represents the population mean of the time (in minute) that the painkillers existing in the market reduce headache in.

b)

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

where μ_X represents the population mean of the time (in minute) that the new painkiller reduces headache in while μ_Y represents the population mean of the time (in minute) that the painkillers existing in the market reduce headache in.

c)

We will test the H_0 against a one sided left tail alternative H_A by using two sample Z-test. In the question, we are given (Let X and Y represents our independently chosen samples which are belong to the people who use the new painkiller and the people who use existing painkillers, respectively.);

$$\begin{aligned} \bar{X} &= 2.8\ (in\ minute) \\ s_X &= 1.7 \\ n\ (size\ of\ X) &= 68 \\ \bar{Y} &= 3\ (in\ minute) \\ s_Y &= 1.4 \\ m\ (size\ of\ Y) &= 68 \\ \alpha\ (significance\ level) &= 0.05 \end{aligned} \tag{11}$$

Since we have one sided left tail alternative, if $Z \le -z_\alpha$ we will reject H_0 and if $Z > -z_\alpha$ then we will say that there is no significant evidence to reject H_0 where Z represents our Z-statistic $z_\alpha = z_{0.05}$ represents the critical value from Standard Normal Distribution. Therefore our rejection region $R = (-\infty, -z_{0.05}]$. To find $z_\alpha = z_{0.05}$ we can use the Standard Normal Distribution Table(A4) given in "Probability and Statistics for Computer Scientists, page 417". Since according to the table, $\Phi(1.645) = (1 - 0.05) = 0.950$ then $z_{0.05} = \Phi^{-1}(0.950) = 1.645$. So that $R = (-\infty, -1.645]$

z-test diagram is given below where the coloured area represents rejection region:

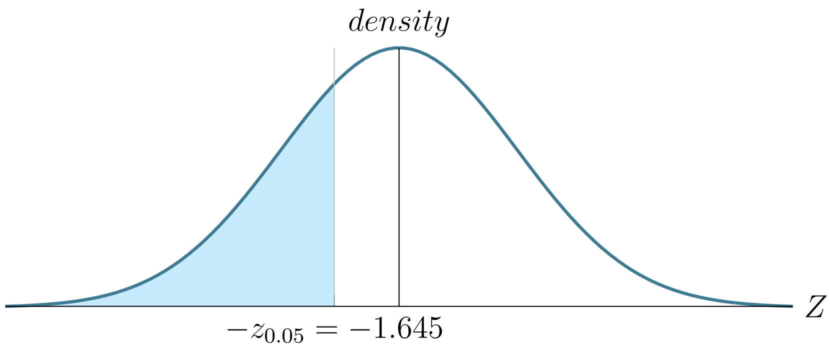

Figure 2: z-test diagram

Then, let us calculate test statistic using the formula given in "Probability and Statistics for Computer Scientists, page 273, table 9.1" and the values in (11);

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \tag{12}$$

where $n = m = 68$, σ_X and σ_Y are population standard deviations of new painkiller and other painkillers, respectively. Since both sample sizes are sufficiently large ($68 > 30$), then we can use s_X and s_Y to estimate σ_X and σ_Y, respectively (in other words s_X and s_Y will be accurate estimators since sample sizes are sufficiently large). Then;

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} = \frac{2.8 - 3}{\sqrt{\frac{(1.7)^2}{68} + \frac{(1.4)^2}{68}}} = -0.748882 \tag{13}$$

Since $Z = -0.748882 > -z_{0.05} = -1.645$, $Z \notin R$. Hence, we cannot reject the null hypothesis H_0 since there is no significant evidence. Thus, we can not state the new painkiller really produce better results with 5% level of significance.