CENG222

Statistical Methods for Computer Engineering

Spring 2021-2022

Homework 3

Due: May 29th, 2022, Sunday 23:59

Question 1

a) $\bar{X} = 16.96$, $\sigma = 3$. Then the asked confidence intervals are

$$\bar{X} \pm z_{0.05}\, \sigma/\sqrt{n} \approx \boxed{16.96 \pm 1.56} \quad \bar{X} \pm z_{0.005}\, \sigma/\sqrt{n} \approx \boxed{16.96 \pm 2.44}$$

b) The expression for the margin is margin $\leq \Delta = z_{\alpha/2}\, \sigma/\sqrt{n}$. Then for a desired Δ we use $n \geq \left(\frac{z_{\alpha/2}\,\sigma}{\Delta}\right)^2$. Plugging in the values $z_{\alpha/2} = 2.326, \sigma = 3, \Delta = 1.55$ we get $n = 20.27$. Then the smallest sample size is $\boxed{21}$.

Question 2

a) Open-ended question and as such there is no unique answer. To be able to perform Z-test we need the additional information on σ or the distribution of ratings provided n is large. To be able to perform t-test again we need the distribution of ratings to get an expression for $s(\hat{\theta})$. For these, sample mean and size are not enough. On the other hand, one might not be interested in performing a statistical test to get a more accurate estimation and choose to stay on the safe-zone as in Question 2 part d.

It is also possible that the question is interpreted in a more general sense, and answers have a wide spectrum. As a result, the answers will be graded in terms of the consistency of provided explanations. If an explanation is not given, the grade will be zero.

b) Here $H_0 : \mu = 7.5, H_A : \mu < 7.5$. Hence we perform a one-sided left-tail test and we accept H_0 if $Z \geq -z_{0.05}$ and reject it otherwise. In our case $Z = \dfrac{7.4 - 7.5}{0.8/\sqrt{256}} = -2 < -1.645$. Thus, we reject H_0 meaning that the data provides sufficient evidence that the restaurant is significantly lower than 7.5 at a 5% level of significance.

c) Our calculated Z changes to $Z = \dfrac{7.4 - 7.5}{1.0/\sqrt{256}} = -1.6 \geq -1.645$. The data does not provide sufficient evidence to reject H_0, hence we must accept it. So we have a new candidate restaurant.

d) If the sample mean is already larger than 7.5 we know for sure that H_0 cannot be rejected, sample mean is always on the safe-zone, *i.e.*, Z can never be less than 0.

Question 3

The null hypothesis is $H_0 : \mu_X - \mu_Y = D$ and the alternative is $H_A : \mu_X - \mu_Y < D$.
We use one-sided left-tail t-tests where $\bar{X} = 211$, $\bar{Y} = 133$, $s_X = 5.2$, $s_Y = 22.8$, $D = 90$, $n = 20$, $m = 32$. The null hypothesis is rejected if $t \leq -t_{0.01}$ and accepted otherwise.

a) First we calculate the pooled sample variance $s_p^2 = \dfrac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2} = 332.58$ so that $s_p = 18.24$. Then our t variable is $t = \dfrac{\bar{X} - \bar{Y} - D}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} = \dfrac{211 - 133 - 90}{18.24\sqrt{\frac{1}{20} + \frac{1}{32}}} = -2.309$. Now we need to find the value of $t_{0.01}$ from the table. Since we assume the population standard deviations are the same $\nu = 20 + 32 - 2 = 50$. Hence, $t_{0.01} = 2.403$. Since $t > -t_{0.01}$, we accept H_0, *i.e.*, a 90-minute or better improvement is achieved at the specified significance level.

b) Since population variances are different we use Satterthwaite approximation to estimate the degree of freedom which gives us $\nu = 35.97$. From Table A5 we use $t_{0.01} = 2.434$.

The parameter t is calculated as $t = \dfrac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} = -2.861$.

Since $t \leq -t_{0.01}$ we reject the null hypothesis, *i.e.*, enough evidence is present to refuse H_0.