Instructions

Upload your own solutions (handwritten or typed) to the corresponding Canvas Assignment on or before the due date: 03/12/2021.

Grading scheme

a	b	c	d	e	f	g	h	Total
3	2	3	3	2	2	3	2	20

---

a. A small sample confidence interval for the mean μ of a normal population when the **variance is unknown** (only the **sample** standard deviation was given) are given as

$$\left(\bar{x} - t_{23,1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + t_{23,1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right) = \left(26.676 - 2.069\tfrac{12.839}{\sqrt{24}}, 6.676 + 2.069\tfrac{12.839}{\sqrt{24}}\right) = (21.254, 32.098).$$

b. Provided the normality assumption is reasonable, from part **a.** we can conclude that we can be 95% confident that the true average task time as well as the median task time are contained in the interval $(21.254, 32.098)$. Since the interval also contains numbers larger than 30, a true average task time/median task time of 32 minutes (say) is plausible and cannot be ruled out based on this confidence interval. Therefore, we **should not** make the claim that the task under investigation can, on average or in median, be performed under 30 minutes using this new app.

c. The histogram shows a right skewed distribution, The QQ-Plot shows that both the left and the right tail are too heavy to match a normal distribution and the boxplot again highlights the skewness (to the right) of the data. Therefore, the normality assumption is not justified.

d. The histogram shows a distribution that matches a normal distribution in principle (more or less symmetric, bell-shape is plausible). The QQ-Plot supports this and shows in particular that both the left and the right tail seem to match a normal distribution. The boxplot also shows the symmetry of the data and does not highlight any outliers. Therefore, the normality assumption seems reasonable.

e. We should only apply a certain method if the assumptions seem to be satisfied. While it is obvious that the un-transformed task times come from a skewed distribution and should not be considered as normal sample, our analysis suggests that we can assume normality for the log-transformed times. To be precise, we can assume that $Y_1 = \ln(X_1), \dots, Y_{24} = \ln(X_{24})$ follow a normal distribution $\mathcal{N}(\mu_Y, \sigma_Y^2)$, if $X_1, \dots, X_{24}$ is a random sample of the task times in the population of all potential users of the app. μ_Y is the average log-task time. We can now construct a confidence interval for the mean with unknown variance for the log-times such that the following holds true:

$$\mathrm{P}\left(\mu_Y \in \left(\bar{Y} - t_{1-\frac{\alpha}{2},23}\frac{S_Y}{\sqrt{24}}, \bar{Y} + t_{1-\frac{\alpha}{2},23}\frac{S_Y}{\sqrt{24}}\right)\right) = 1 - \alpha.$$

To get from the log-transformed data to the task times again we need to apply the inverse function:

$$e^{Y_1} = X_1, \dots e^{Y_{24}} = X_{24}.$$

Let m_X denote the median task time. We have that

$$e^{\mu_Y} = m_X.$$

Since the exponential function ($x \mapsto e^x$) is the **strictly increasing** inverse of the natural logarithm, we obtain

$$\mathrm{P}\left(e^{\mu_Y} = m_X \in \left(e^{\bar{Y}-t_{1-\frac{\alpha}{2},23}\frac{S_Y}{\sqrt{24}}}, e^{\bar{Y}+t_{1-\frac{\alpha}{2},23}\frac{S_Y}{\sqrt{24}}}\right)\right) \tag{1}$$

$$= \mathrm{P}\left(\mu_Y \in \left(\bar{Y} - t_{1-\frac{\alpha}{2},23}\frac{S_Y}{\sqrt{24}}, \bar{Y} + t_{1-\frac{\alpha}{2},23}\frac{S_Y}{\sqrt{24}}\right)\right) = 1-\alpha. \tag{2}$$

Therefore, the interval $(e^{2.991}, e^{3.375}) = (19.906, 29.224)$ is a 95% confidence for the median task time.

f. Based on the analysis in the previous parts, we can assume that $Y_1 = \ln(X_1), \ldots, Y_{24} = \ln(X_{24})$ follow a normal distribution $\mathcal{N}(\mu_Y, \sigma_Y^2)$, if $X_1, \ldots, X_{24}$ is a random sample of the task times in the population of all potential users of the app. This gives us a 95% confidence interval for the median task times: $(19.906, 29.224)$ (see part **e**). Since this interval only contains numbers below 30, the claim of the company is supported by the confidence interval.

g. A 90% large sample confidence interval for the population proportion is given by

$$\left(\hat{p} - z_{1-\frac{0.1}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + z_{1-\frac{0.1}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right)$$

$$= \left(\frac{43}{50} - 1.645\frac{\sqrt{\frac{43}{50}(1-\frac{43}{50})}}{\sqrt{50}}, \frac{43}{50} + 1.645\frac{\sqrt{\frac{43}{50}(1-\frac{43}{50})}}{\sqrt{50}}\right) = (0.779, 0.941).$$

Assuming a random sample, we can use the large sample confidence interval since $n = 50 > 25$, $n\hat{p} = 43 > 5$ and $n(1-\hat{p}) = 7 > 5$.

h.

(i) We are 90% confident that the true proportion of users who can perform the task in less than 10 minutes is between 0.779 and 0.941.

(ii) Being "90% confident" means that if we could repeat the experiment many times, about 90% of the obtained intervals would contain the correct value of the population proportion.