Federal State Autonomous Educational Institution of Higher Education

"NATIONAL RESEARCH UNIVERSITY

"HIGH SCHOOL OF ECONOMICS"

Faculty of Software Engineering

HW №1

duration_3

MATHEMATICAL STATISTICS

Completed by:

Student group 197, software engineering

Yanal Yahya

Full Name

Moscow 2021

NOTE: ALL OPERSTIONS WAS CALCULATED FIRSTLY IN THIS EXCEL FILE

sample	24	22	25	41	19	94	24	17	20	16	23	25	15	22	21	16	24	56	9	20	22	24

a)

as an input we have only this data and we need to calculate the 90% confidence interval for the medium duration. Because of the distribution of the trait is normal and unknown variance, so we can use this confidence interval for medium with confidence level (1−α):

$$\bar{X} - t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}} < \mu < \bar{X} + t_{n-1,\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n}}$$

As we know in this case calculate $\bar{X}$ *and* $\hat{\sigma}$ is easy, and we can do it using these formals:

$$\bar{X} = \frac{X_1 + X_{2^+} \cdots + x_n}{n}$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=0}^{n} (X_i - \bar{X})^2$$

And it was calculated in excel file, and the results are:

$$\bar{X} = 26.31818$$
$$\hat{\sigma} = 17.80686$$

And the value of $t_{n-1,\frac{\alpha}{2}}$ we can get it from Student's t-distribution table or using Excel:

$$t_{21,0.05} = 1.721$$

As a result, we get all the unknown variables, which are necessary to get the interval, so let's compensate variables into inequality:

$$26.31818 - 1.721 \frac{17.80686}{\sqrt{22}} < \mu < 26.31818 + 1.721 \frac{17.80686}{\sqrt{22}}$$

Answer: $19.78452 < \mu < 32.85185$

---

b) For plotting a quantile-quantile plot we do these 4 steps:

1. We estimate the parameters of the normal distribution for the sample:

   $$\bar{X} \; and \; \hat{\sigma}^2$$

   We already have its values.
2. Calculating sample order quantiles $\frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}$

   These will just be ascending order of base data $X_1, \dots, X_n$

   $$\hat{Q}\left(\frac{1}{n+1}\right) = X_1, \dots, \hat{Q}\left(\frac{n}{n+1}\right) = X_n$$

3. calculate the quantiles of the normal distribution with the parameters

$$\mu = \bar{X} \; and \; \sigma^2 = \hat{\sigma}^2$$

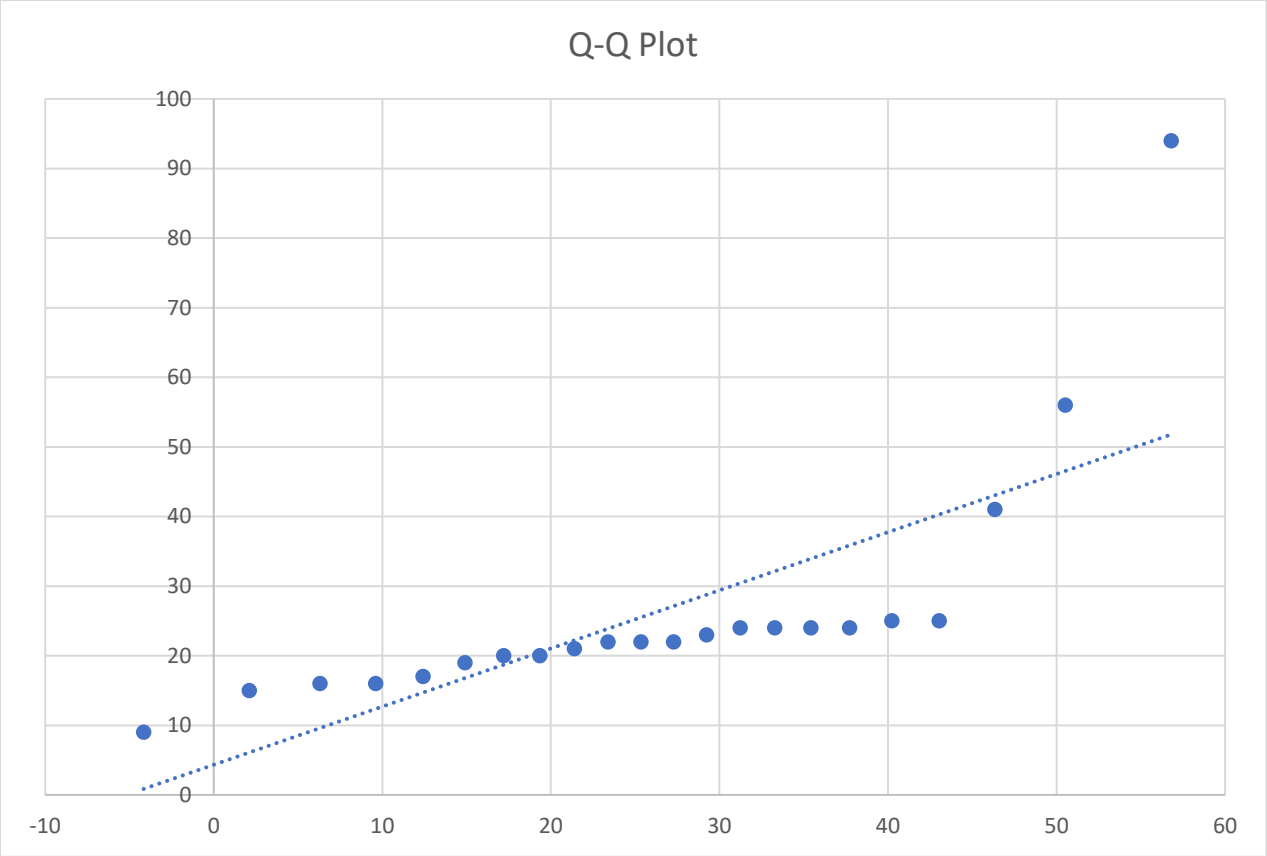
$$Q\left(\frac{1}{n+1}\right) = \bar{X} + \hat{\sigma}\Phi^{-1}\left(\frac{1}{n+1}\right), \dots, Q\left(\frac{n}{n+1}\right) = \bar{X} + \hat{\sigma}\Phi^{-1}\left(\frac{n}{n+1}\right),$$

4. build a graph in axes $(Q, \hat{Q})$- for sample from the normal distribution should be $Q(p) \approx \hat{Q}(p)$ (the graph is built along the bisector of the angle formed by the axes).

By applying all these steps and putting results into a table we get:

№	p	$\hat{Q}(p)$	$\Phi^{-1}(p)$,	$Q(p)$	$\hat{Q}(p)$
1	**0.0434783**	**9**	**-1.711675**	**-4.161388**	**9**
2	**0.0869565**	**15**	**-1.359737**	**2.1055228**	**15**
3	**0.1304348**	**16**	**-1.124338**	**6.2972435**	**16**
4	**0.173913**	**16**	**-0.938814**	**9.6008427**	**16**
5	**0.2173913**	**17**	**-0.781034**	**12.410419**	**17**
6	**0.2608696**	**19**	**-0.640667**	**14.909913**	**19**
7	**0.3043478**	**20**	**-0.511936**	**17.202203**	**20**
8	**0.3478261**	**20**	**-0.391196**	**19.352203**	**20**
9	**0.3913043**	**21**	**-0.275921**	**21.404893**	**21**
10	**0.4347826**	**22**	**-0.164211**	**23.394103**	**22**
11	**0.4782609**	**22**	**-0.054519**	**25.347371**	**22**
12	**0.5217391**	**22**	**0.0545189**	**27.288993**	**22**
13	**0.5652174**	**23**	**0.1642108**	**29.242261**	**23**
14	**0.6086957**	**24**	**0.2759211**	**31.231471**	**24**
15	**0.6521739**	**24**	**0.3911963**	**33.28416**	**24**
16	**0.6956522**	**24**	**0.5119362**	**35.434161**	**24**
17	**0.7391304**	**24**	**0.6406669**	**37.72645**	**24**
18	**0.7826087**	**25**	**0.7810338**	**40.225945**	**25**
19	**0.826087**	**25**	**0.9388143**	**43.035521**	**25**
20	**0.8695652**	**41**	**1.1243382**	**46.33912**	**41**
21	**0.9130435**	**56**	**1.3597374**	**50.530841**	**56**
22	**0.9565217**	**94**	**1.7116753**	**56.797752**	**94**

And using Excel we can plot a Q-Q plot for the last two columns, so we get:



We can see in our Q-Q plot above that the data values tend to deviate from the 45-degree line, especially on the tail ends, which could be an indication that the data set is not normally distributed.

---

c) For this purpose a C# script was written, you can take a look on it by going through this link, and to see the result of this script, you can go through this link.

This script gives 1000 averages of 1000 re-selections, each containing 22 random selections from the random sample.
And in Excel file, these 1000 averages were ascending sorted, then was found:

$$\hat{Q}(\alpha/2) = \hat{Q}(0{,}1/2) = \hat{Q}(0.05)$$
$$= number\ located\ on\ the\ 50th\ place\ after\ sorting = 21$$
$$\hat{Q}(1-\alpha/2) = \hat{Q}(1-0.1/2) = \hat{Q}(0.95)$$
$$= number\ located\ on\ the\ 950th\ place\ after\ sorting = 32.45455$$

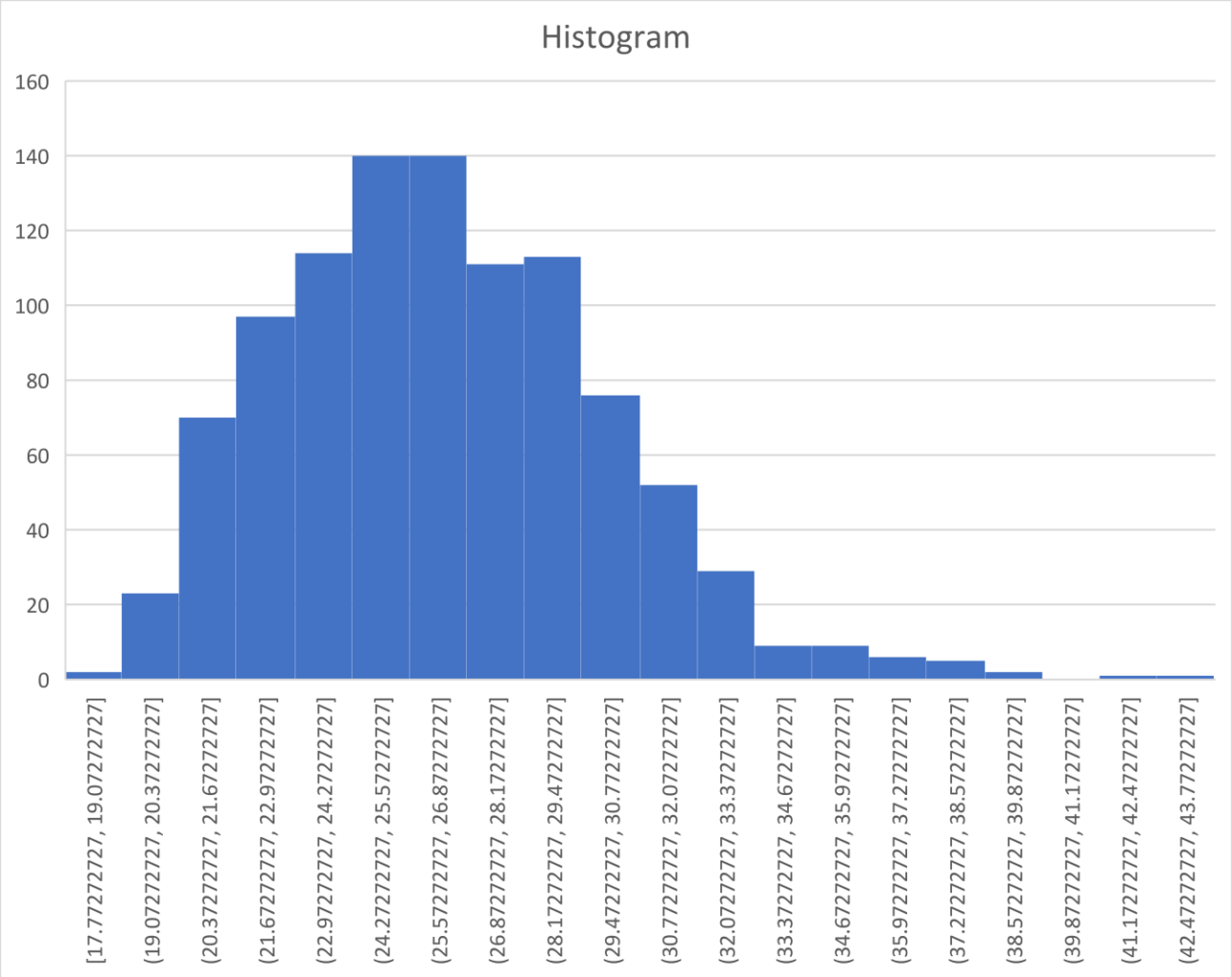
⇨

Answer: $21 < \mu < 32.45455$

We can note that, finding confidence interval for medium duration using this method gives more accurate interval then the method in the first bunk.

---

d) Using Excel and these 1000 averages we can build this histogram, so we get:



Actually, it's closer to exponential distribution, and by taking bigger sample it will become close to normal distribution.