Appendix 1

Distribution of Respondents' Choice of Responses to the Degree of Satisfaction with the Results of Mentoring Activities

Degree of satisfaction	The number of respondents, in units	Frequency of selection, in %
10.0	–	–
9.5	3	0,90
9.0	8	2,40
8.5	17	5,20
8.0	59	18,30
7.5	39	12,10
7.0	37	11,50
6.5	37	11,50
6.0	34	10,50
5.5	22	6,80
5.0	26	8,00
4.5	17	5,20
4.0	8	2,40
3.5	4	0,01
3.0	5	1,50
2.5	2	0,60
2.0	1	0,30
1.5	1	0,30
1.0	1	0,30
0.5	–	–
–	–	–

Source: Compiled by the authors.

Methodology for Calculating Correlation and Reliability

Data Analysis. To analyze the dataset and detect the dependencies of the parameter “Quality to improve” from parameters “Age” and “Working years”, we followed these steps:

1. Load the dataset (dataset.csv) and inspect it to understand its structure and contents.
2. Check for any missing values and handle them appropriately.

The provided dataset has the following structure and contents (Table 1).

Table 1. **Overview of the dataset structure and variables**

Column Name	Description
Age	Age range of the respondent
Working years	Number of years the respondent has worked
Needs	Specific needs expressed by the respondent
Mentor	Whether the respondent had a mentor
Qualities to improve	Aspects the respondent wants to improve

Source: Compiled by the authors.

Qualitative data, by its nature, cannot have a normal distribution because it represents categorical or discrete values, not continuous numerical values like quantitative data. The normal distribution is a probability distribution that describes the distribution of continuous variables, not categorical variables.

Qualitative data, also known as categorical data, can take on a finite number of distinct categories or levels. These variables are often analyzed using methods such as contingency tables, chi-square tests, or logistic regression, which are appropriate for categorical data analysis.

When working with qualitative data, we are typically interested in frequencies or proportions of each category, as well as relationships between different categories. It’s essential to choose appropriate statistical methods and visualizations that are suited to the nature of qualitative data and the specific research questions being addressed.

Hypothesis Formulation. Hypothesis about the relationship between “Quality to improve”, “Age”, and “Working years” is as follows: Parameter “Qualities to improve” doesn’t depend of “Age” and “Working years”.

Statistical Analysis. To test the hypothesis that “Qualities to improve” doesn’t depend on “Age” and “Working years”, we can use a chi-square test of independence. This test determines whether there is a significant association between two categorical variables. If the p-value from the chi-square test is less than a chosen significance level (commonly 0.05), we reject the null hypothesis, suggesting that there is a significant association between the variables.

Let’s conduct the chi-square test using Python:

```
import pandas as pd
from scipy.stats import chi2_contingency

# Create the dataset
data = [ data.array]

columns = ["Age", "Working years", "Needs", "Mentor", "Qualities to
improve"]
df = pd.DataFrame(data, columns=columns)

# Create contingency table
contingency_table = pd.crosstab(df['Age'], df['Qualities to im-
prove'])
```

```
# Perform chi-square test of independence
chi2, p_value, dof, expected = chi2_contingency(contingency_table)

# Print the results
print("Chi-square Test of Independence:")
print(f"Chi-square statistic: {chi2}")
print(f"P-value: {p_value}")
print(f"Degrees of freedom: {dof}")
print("Expected frequencies:")
print(expected)
if p_value > 0.05:
    print("There is no significant association between Age and Qual-
ities to improve (p > 0.05)")
else:
    print("There is a significant association between Age and Quali-
ties to improve (p <= 0.05)")
```

This code will conduct a chi-square test of independence between "Age" and "Qualities to improve". It calculates the chi-square statistic, p-value, degrees of freedom, and expected frequencies. Based on the p-value, we determine whether there is a significant association between the variables.

We performed a similar test for "Working years" as well by replacing "Age" with "Working years" in the code.

Coefficient Interpretation. Dependency between "*Age*" *and* "*Qualities*" *to Improve.*

```
Chi-square statistic: 186.79950351744827
P-value: 1.4027012430147167e-28
Degrees of freedom: 21
Expected frequencies: [[5.78034682e-03 1.27167630e-01 1.73410405e-
02  1.21387283e-01  3.41040462e-01  8.09248555e-02  2.31213873e-
01  7.51445087e-02]  [3.12138728e-01  6.86705202e+00  9.36416185e-
01  6.55491329e+00  1.84161850e+01  4.36994220e+00  1.24855491e+01
4.05780347e+00]   [2.08092486e-01   4.57803468e+00   6.24277457e-
01  4.36994220e+00  1.22774566e+01  2.91329480e+00  8.32369942e+00
2.70520231e+00]  [4.73988439e-01  1.04277457e+01  1.42196532e+00
9.95375723e+00   2.79653179e+01   6.63583815e+00   1.89595376e+01
6.16184971e+00]]
```

There is a significant association between "Age" and "Qualities to improve" ($p \leq 0.05$).

Dependency between "Working years" and "Qualities to improve".

```
Chi-square   Test   of   Independence:   Chi-square   statistic:
184.75561967028597
P-value: 3.515019507355743e-28
Degrees of freedom: 21
Expected frequencies: [[5.78034682e-03 1.27167630e-01 1.73410405e-
02  1.21387283e-01  3.41040462e-01  8.09248555e-02  2.31213873e-
01  7.51445087e-02]  [2.02312139e-01  4.45086705e+00  6.06936416e-
01  4.24855491e+00  1.19364162e+01  2.83236994e+00  8.09248555e+00
2.63005780e+00]  [5.26011561e-01  1.15722543e+01  1.57803468e+00
1.10462428e+01   3.10346821e+01   7.36416185e+00   2.10404624e+01
6.83815029e+00]   [2.65895954e-01   5.84971098e+00   7.97687861e-
01  5.58381503e+00  1.56878613e+01  3.72254335e+00  1.06358382e+01
3.45664740e+00]]
```

There is a significant association between "Working years" and "Qualities to improve" ($p \leq 0.05$).

Results Visualisation. To visualize the results of the chi-square test of independence between "Age" and "Qualities to improve", you can create a heatmap of the contingency table, showing the observed frequencies. Additionally, you can highlight cells with significant differences (e.g., using colors) based on the expected frequencies.

Here’s how we can visualize the results using Python with Matplotlib and Seaborn:

```
import seaborn as sns
import matplotlib.pyplot as plt

# Create heatmap of observed frequencies
plt.figure(figsize=(10, 6))
sns.heatmap(contingency_table, annot=True, fmt="d", cmap="YlGnBu")
plt.title(‘Contingency Table: Observed Frequencies’)
plt.xlabel(‘Qualities to improve’)
plt.ylabel(‘Age’)
plt.show()

# Highlight significant differences based on expected frequencies
plt.figure(figsize=(10, 6))
sns.heatmap(contingency_table, annot=True, fmt="d", cmap="YlGnBu",
mask=(contingency_table < expected))
plt.title(‘Contingency Table: Observed vs Expected Frequencies’)
plt.xlabel(‘Qualities to improve’)
plt.ylabel(‘Age’)
plt.show()
```

These visualizations will provide insights into the relationship between “Age” and “Qualities to improve”. The first heatmap shows the observed frequencies, while the second heatmap highlights cells with significant differences between observed and expected frequencies based on the chi-square test results.

Conclusion. Based on our calculations we detected significant association between “Age” and “Qualities to improve” ($p \leq 0.05$) and “Working years” and “Qualities to improve” ($p \leq 0.05$).

So, our hypothesis is false, and “Qualities to improve” strongly depends of “Age” and “Working years”.