Covid Depression Case Study - Summary Statistics

Natashia Dobbs

```
# Set working directory to the folder that contains your Excel file
setwd("C:/Users/Tashi/OneDrive/Documents/New folder")

# Load required package
library(readxl)

# Load the Excel file
x <- read_excel("Covid_Depression_Case_Studyxlsx.xlsx")
```

```
summary(x$Quality_score)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   6.000   6.000   7.000   6.667   7.000   7.000
```

```
sd(x$Quality_score)
```

```
## [1] 0.492366
```

```
# Standard deviation of the Quality Score
sd(x$Quality_Score)
```

```
## Warning: Unknown or uninitialised column: `Quality_Score`.
```

```
## [1] NA
```

```
colnames(x)
```

```
##  [1] "Study"                "Country"              "Sampling Method"
##  [4] "Sample_Size"          "Mean Age"             "Percent_Female"
##  [7] "Response_Rate"        "Depression_assessment" "Depression_Prevalence"
## [10] "Quality_score"
```

```
head(x)
```

```
## # A tibble: 6 × 10
##   Study        Country `Sampling Method`   Sample_Size `Mean Age` Percent_Female
##   <chr>        <chr>   <chr>                     <dbl>      <dbl>          <dbl>
## 1 Ahmed        China   Convenience Sampli…        1074       33.5           46.8
## 2 Gao et al    China   Convenience Sampli…        4872       32.2           67.7
## 3 Huang & Zhao China   Convenience Sampli…        7236       35.3           54.6
## 4 Kazmi et al. India   Randon Sampling            1000        0             62
## 5 Let et al.   China   Convenience Sampli…        1593       32.3           61.3
## 6 Mazza et al  Italy   Convenience Sampli…        2766       32.9           71.7
## # ℹ 4 more variables: Response_Rate <dbl>, Depression_assessment <chr>,
## #   Depression_Prevalence <dbl>, Quality_score <dbl>
```

Summary Interpretation

The summary statistics of the **Quality_score** variable show that the lowest score was **6**, and the highest was **7**. The **mean** was **6.67**, and the **median** was **7**, suggesting that most studies were rated highly in quality.

The **standard deviation** was **0.49**, which is relatively low — this means the quality scores across the studies were very consistent, with little variation. Overall, most studies had similar and strong quality scores.

Part Two: Written Responses

1. Describe an Experiment

To evaluate whether the Covid-19 pandemic led to a higher prevalence of depression, we could design a **longitudinal study**. This study would follow a group of individuals over time, collecting depression scores at multiple points — ideally **before**, **during**, and **after** the pandemic. Tools like the PHQ-9 or BDI-II could be used to measure depression levels consistently. By comparing changes in scores over time within the same group, we could assess the potential impact of the pandemic on mental health.

---

2. Why Is Sample Size Important?

Sample size is important because it affects the **accuracy** and **generalizability** of the results. A larger sample size increases the likelihood that the results represent the true population. It also reduces the effects of outliers and increases statistical power, which helps detect real differences or trends in the data. A small sample could lead to unreliable or misleading conclusions.

---

3. Why Is Causation Hard to Prove?

Causation is difficult to prove because **many other variables** can influence depression beyond Covid-19. While we might observe a correlation between the pandemic and higher depression rates, this doesn't mean one caused the other. Other factors — like **job loss, isolation, financial stress, or pre-existing mental health conditions —** can also play a role. To prove causation, researchers would need to isolate Covid-19 as the only influencing factor, which is challenging in real-world settings.