**Feifei Li**

**Applied Statistics 1210**

**Instructor: David Liew**

**Does Onsite Work Enhance Productivity More Than Remote Work?**

**Introduction**

This project investigates whether onsite work enhances productivity more than remote work. The debate has gained attention due to contrasting opinions, such as Elon Musk's argument that remote work is "not moral" and that onsite work fosters productivity. Conversely, many employees favor remote work, citing mental health benefits, improved work-life balance, and better focus. **Does onsite work enhance productivity more than remote work?** Addressing this question is critical for shaping future workplace policies and practices, offering valuable insights for HR professionals, business leaders, and employees alike.

**Rationale**

After the pandemic, many companies required employees to return to office-based work, with public figures like Elon Musk arguing that onsite work leads to higher productivity. However, numerous studies and surveys highlight the significant benefits of remote work. One key advantage is reduced commuting stress, as employees can reallocate the time previously spent commuting to more productive tasks or personal well-being. Additionally, remote work offers improved work-life balance, allowing individuals to manage personal and professional responsibilities more effectively. Furthermore, the flexibility of remote work enables employees to customize their workspaces, fostering better focus and increased productivity.

These findings suggest that the relationship between work environment and productivity is nuanced and has significant implications for HR and policymaking. For instance, organizations might explore hybrid work models or invest in technology that supports remote collaboration. Personally, this topic resonates with my career aspirations in remote work, making it both societally and professionally relevant.

**Research Population**

All employees worldwide working onsite or remotely across various industries, roles, and locations comprise the targeted population.

**Variables**

The variables in this study are the two groups (onsite and remote employees) and their respective productivity scores. The first variable (the two groups) is categorical. The second variable (productivity scores) is quantitative, with a unit of measurement as "scores." The respective productivity scores are treated as discrete quantitative variables.

**Data Collection Process**

The dataset, *Remote Work Productivity*, was sourced from Kaggle and contains 1,000 records from employees across diverse industries and roles. It is a reliable and appropriate source because Kaggle datasets are often curated by professionals, accompanied by metadata, and cover representative populations.

This study used a cross-sectional design, capturing data at a single point in time to focus on employees' productivity within either remote or onsite work environments.

**Sample Selection and Bias Prevention**

To ensure unbiased representation, the data preparation process involved several key steps. First, random sampling was used to select 50 records each for onsite and remote workers, ensuring a balanced dataset and minimizing selection bias through an online random number generator. The original work environment classification, stored under the employee\_type column, was then split into two distinct groups: 510 onsite and 490 remote workers, using Python for better data organization. The primary variable, productivity\_score, was categorized into these two groups, creating a new dataset while maintaining data integrity and representativeness. Finally, the cleaned dataset was imported into RStudio for comprehensive statistical analysis and visualization.

**Descriptive Statistics**

Using RStudio, I analyzed the productivity scores for both groups, focusing on measures of center, variation, and visual representation through boxplots.

**Onsite Workers:**

**Measures of Central Tendency**  
The mean productivity score for onsite workers is 62.2, and the median is 62.5. The mode is 49, as it appears most frequently. The midrange is 66.5, calculated by adding the maximum value (90) and the minimum value (43), then dividing by two.

**Measures of Variation**  
The range of productivity scores is 47, calculated by subtracting the minimum value (43) from the maximum value (90). The interquartile range (IQR) is 12.25, determined by subtracting the first quartile (Q1 = 55.25) from the third quartile (Q3 = 67.50). The standard deviation is 10.42 (to 2 d.p.), and the variance is 108.58 (to 2 d.p.).

**Measures of Relative Standing**  
The maximum productivity score is 90, and the minimum is 43. The first quartile is 55.25, the second quartile (median) is 62.5, and the third quartile is 67.50. A boxplot constructed in R-Studio revealed an outlier, as the upper fence is 85.875, and the lower fence is 36.875. One score exceeded the upper fence, confirming the presence of an outlier.



**Interpretation:**

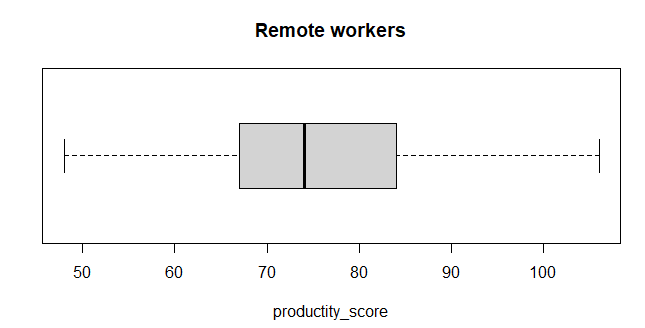
Based on the boxplot, the distribution of onsite workers' productivity scores is right-skewed, with the median score of 62.5 serving as a better measure of central tendency due to the outlier. The IQR of 12.25 suggests moderate dispersion, while the outlier indicates that some workers may have significantly higher productivity.

**Remote Workers:**

**Measures of Central Tendency**  
The mean productivity score for remote workers is 74.84, and the median is 74. The mode is 79, as it appears most frequently. The midrange is 77, calculated by adding the maximum value (106) and the minimum value (48), then dividing by two.

**Measures of Variation**  
The range of productivity scores is 58, calculated by subtracting the minimum value (48) from the maximum value (106). The interquartile range (IQR) is 16.75, determined by subtracting the first quartile (Q1 = 67) from the third quartile (Q3 = 83.75). The standard deviation is 12.97 (to 2 d.p.), and the variance is 168.10 (to 2 d.p.).

**Measures of Relative Standing**  
The maximum productivity score is 106, and the minimum is 48. The first quartile is 67, the second quartile (median) is 74, and the third quartile is 83.75. A boxplot constructed in R-Studio confirmed no outliers, indicating a consistent dataset.



**Interpretation:**

Based on the boxplot, the distribution of productivity scores for remote workers appears symmetrical, suggesting that the data is evenly spread around the center. With a mean of 74.84 and a standard deviation of 12.97, demonstrating moderate variability, indicating that most workers' productivity levels are relatively close to the average.

**Hypothesis Testing for Two Means (Independent Samples)**

**Test Details:**

Test Type: Welch’s Two Sample t Test (independent t test)

Significance Level (α): 0.05

**Step-by-step Procedure:**

1. **Check Requirements:**

Both samples were selected randomly to ensure unbiased representation, meeting the requirement for simple random sampling. Additionally, the two groups, onsite and remote workers, are independent of each other, fulfilling the condition for independent samples. Each group consists of 50 participants, which satisfies the sample size requirement of (n > 30) for normal approximation. Therefore, all conditions necessary for conducting a two-sample t-test are met.

2. **Hypothesis Statement:**

Null Hypothesis (H₀): There is no significant difference in productivity between onsite and remote workers.

H₀: μ₁ = μ₂ (onsite productivity = remote productivity)

Alternative Hypothesis (H₁): Onsite workers have higher productivity than remote workers.

H₁: μ₁ > μ₂ (onsite productivity > remote productivity)

3. **Sampling Distribution:** Independent t-test.

4. **Analysis in RStudio:**

After we used R-studio, we found that the test statistics is -5.3734 and p\_value is 1.00

**5. Decision:**

Since the p\_value (1.00) is greater than the significance level (0.05), we fail to reject the null hypothesis (H₀).

**6. Conclusion:**

There is no sufficient evidence to support the claim that onsite workers have higher productivity than remote workers.

**7. Answer to Research Question:**

Based on this analysis, onsite work does not enhance productivity more than remote work.

**Test the claim by constructing an appropriate confidence interval**

Since the test is right tailed with a significance level (alpha) of 0.05, the corresponding confidence level is: 1-2(alpha) which is 1-0.1=0.90 as 90%.

We use this 90% confidence level to test out our claim by using the R-studio.

**The 90% confidence interval for the difference in means is:**

Confidence Interval:(−16.55, −8.73)

**Confidence Interval Interpretation**

We are 90% confident that the true difference in mean productivity scores between onsite and remote workers lies between -16.55 and -8.73.

**Conclusion:**   
Since the interval contains only negative values, then we can conclude thatu1 < u2, indicating that remote workers consistently outperform onsite workers in productivity.

**Final Conclusion**

Evidence suggesting that remote workers may have higher productivity by the confidence interval test. While the hypothesis that onsite work enhances productivity was not supported, this finding highlights the importance of other factors—such as reduced commuting stress, improved work-life balance, and better mental health—in shaping employee performance. Future research could explore additional variables, such as job satisfaction and collaboration, to provide a more comprehensive understanding of workplace productivity dynamics.

**Reference:**

**Original dataset:**

**<https://www.kaggle.com/code/alaaabdelstar/remote-work-productivity/notebook>**

**DATASETS AFTER CLEAN as productivity\_score:**

[productivity\_score.xlsx](https://1drv.ms/x/c/4ae7d083cfc9aeab/Ea9ylyoLHuRMqZY1W14s30IBaX9AZSks8UhK7W49Gx1HsQ?e=PjfLpd)

> View(p)

> o=p$Onsite

> r=p$Remote

> mean(o)

[1] 62.2

> median(o)

[1] 62.5

> sd(o)

[1] 10.41976

> var(o)

[1] 108.5714

> summary(o)

Min. 1st Qu. Median Mean 3rd Qu. Max.

43.00 55.25 62.50 62.20 67.50 90.00

> sort(table(o))

o

43 44 47 51 55 58 61 63 68 70 74 76 79 85 90 57 62 71 72 77 53 56 60 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 3 3 3 64 65 66 49

3 4 4 5

> boxplot(o,horizontal = TRUE,main="onsite workers",xlab="productity\_score")

> mean(r)

[1] 74.84

> median(r)

[1] 74

> sort(table(r))

r

48 56 57 60 61 62 64 65 66 71 73 77 83 88 89 96 97 98 99 106

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

50 68 70 81 84 75 67 72 86 79

2 2 2 2 2 3 4 4 4 5

> summary(r)

Min. 1st Qu. Median Mean 3rd Qu. Max.

48.00 67.00 74.00 74.84 83.75 106.00

> sd(r)

[1] 12.9652

> var(r)

[1] 168.0963

> 83.75-67

[1] 16.75

> 106-48

[1] 58

> 106+48

[1] 154

> 154/2

[1] 77

> boxplot(r,horizontal = TRUE,main="Remote workers",xlab="productity\_score")

> t.test(o, r, mu=0, alternative="greater", var.equal=FALSE)

**Output:**

Welch Two Sample t-test

data: o and r

t = -5.3734, df = 93.664,

p-value = 1

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-16.54786 Inf

sample estimates:

mean of x mean of y

62.20 74.84

> x=p$Onsite

> y=p$Remote

> t.test(x,y,conf.level=0.90,var.equal=FALSE)

Welch Two Sample t-test

data: x and y

t = -5.3734, df = 93.664, p-value = 5.624e-07

alternative hypothesis: true difference in means is not equal to 0

90 percent confidence interval:

-16.547855 -8.732145

sample estimates:

mean of x mean of y

62.20 74.84