

This analysis was conducted using the Employee Survey results from Lainguyen123. (2023). *Employee Survey* [Data set]. Kaggle. <https://www.kaggle.com/datasets/lainguyen123/employee-survey>

RQ6: To what extent do demographic, organizational, and lifestyle factors predict employee stress levels?

Null Hypothesis (H_{06}):

Demographic (e.g., gender, age), organizational (e.g., job level, department, overtime status), and lifestyle factors (e.g., sleep hours, physical activity, commute distance) do not significantly predict employee stress levels.

Alternative Hypothesis (H_{16}):

Demographic, organizational, and lifestyle factors significantly predict employee stress levels.

Results

Linear Regression

Model Fit Measures

Model	R	R ²
1	0.0807	0.00651

Note. Models estimated using sample size of N=3025

Model Summary

- **R² = 0.0065**
The model explains **less than 1%** of the variance in stress scores. This is a very small effect size, indicating that the included predictors — while theoretically relevant — do not meaningfully account for differences in reported stress levels in this dataset

Model Coefficients - Stress

Predictor	Estimate	SE	t	p
Intercept ^a	1.79711	0.19802	9.0754	<.001

Model Coefficients - Stress

Predictor	Estimate	SE	t	p
Gender:				
2 – 1	0.03674	0.04306	0.8531	0.394
3 – 1	0.01074	0.07333	0.1464	0.884
Age	-0.00633	0.00305	-2.0744	0.038
JobLevel:				
Junior – Intern/Fresher	-0.07910	0.08723	-0.9068	0.365
Lead – Intern/Fresher	0.17923	0.12564	1.4266	0.154
Mid – Intern/Fresher	0.02317	0.09312	0.2488	0.804
Senior – Intern/Fresher	0.05938	0.09603	0.6184	0.536
Dept:				
Finance – Customer Service	0.12842	0.09749	1.3172	0.188
HR – Customer Service	0.07824	0.12519	0.6250	0.532
IT – Customer Service	-0.08624	0.10194	-0.8460	0.398
Legal – Customer Service	0.09085	0.10904	0.8332	0.405
Marketing – Customer Service	0.06501	0.10652	0.6103	0.542
Operations – Customer Service	0.00512	0.10107	0.0506	0.960
Sales – Customer Service	0.03018	0.10843	0.2783	0.781
haveOT:				
1 – 0	0.10013	0.05162	1.9396	0.053

Model Coefficients - Stress

Predictor	Estimate	SE	t	p
SleepHours	0.01334	0.01936	0.6890	0.491
PhysicalActivityHours	-0.01483	0.02008	-0.7386	0.460
CommuteDistance	-6.91e-4	0.00231	-0.2991	0.765

^a Represents reference level

Significant Predictors

Age

- **Estimate = -0.00633, p = .038**
- Interpretation: Older employees report slightly lower stress levels. While the effect is small, it's statistically significant. This may reflect increased emotional regulation, coping strategies, or role stability with age.

Overtime Status (haveOT)

- **Estimate = 0.10013, p = .053**
- Interpretation: Employees who report working overtime tend to have slightly higher stress scores. This is **marginally significant**, suggesting a practical effect worth exploring further, even if it narrowly misses the conventional $p < .05$ threshold.

Non-Significant Predictors

These variables did **not** significantly predict stress:

- **Gender** (Female and Other vs. Male)
- **JobLevel** (all levels vs. Intern/Fresher)
- **Department** (all comparisons vs. Customer Service)
- **EmpType** (Full-Time and Part-Time vs. Contract)
- **SleepHours**
- **PhysicalActivityHours**

- **CommuteDistance**

Interpretation:

Despite their theoretical relevance, these predictors didn't show statistical influence. This could be due to:

- Low variability in stress scores
- Measurement limitations (e.g., Likert scale compression)
- Unmeasured confounding factors (e.g., team dynamics, leadership style, emotional intelligence)

Assumption Checks

Collinearity Statistics

	VIF	Tolerance
Gender	1.03	0.970
Age	1.61	0.622
JobLevel	1.13	0.887
Dept	1.04	0.964
haveOT	1.23	0.811
SleepHours	1.00	0.997
PhysicalActivityHours	1.00	0.997
CommuteDistance	1.00	0.997

Normality Test (Shapiro-Wilk)

Statistic	p
0.773	<.001

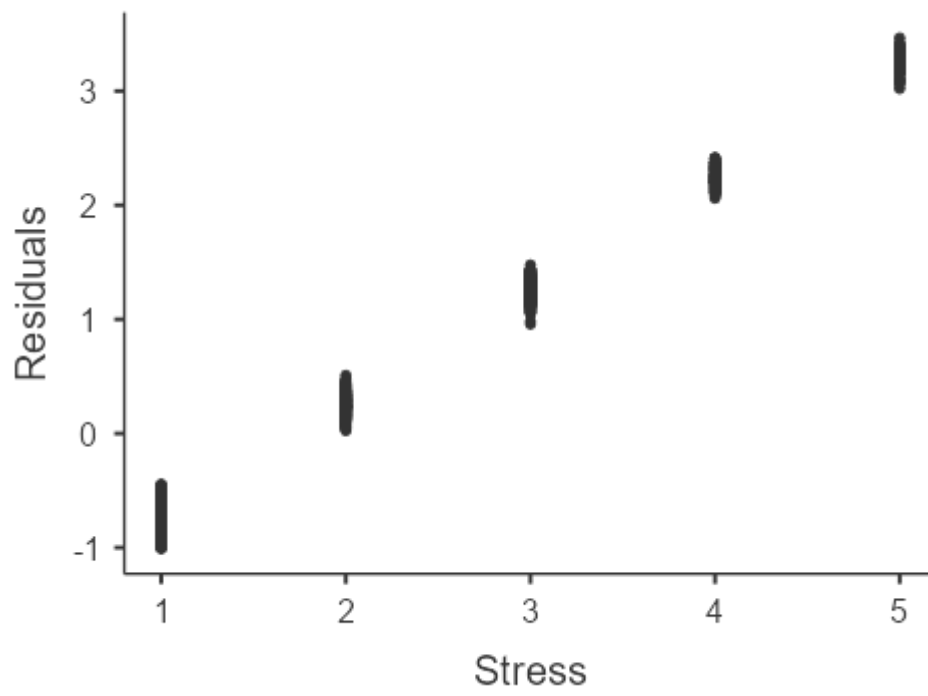
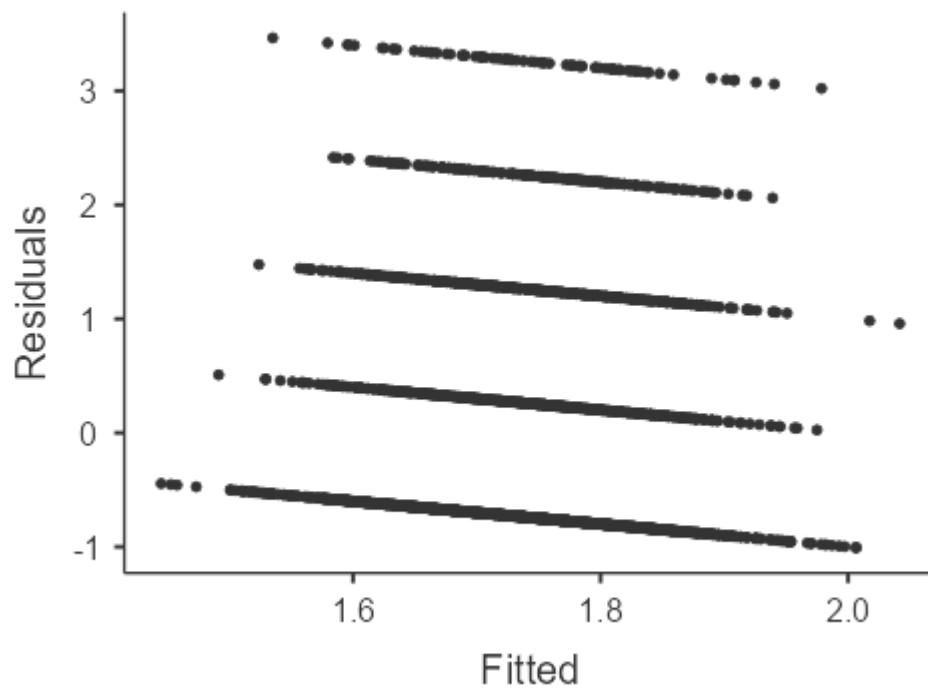
Assumption Checks

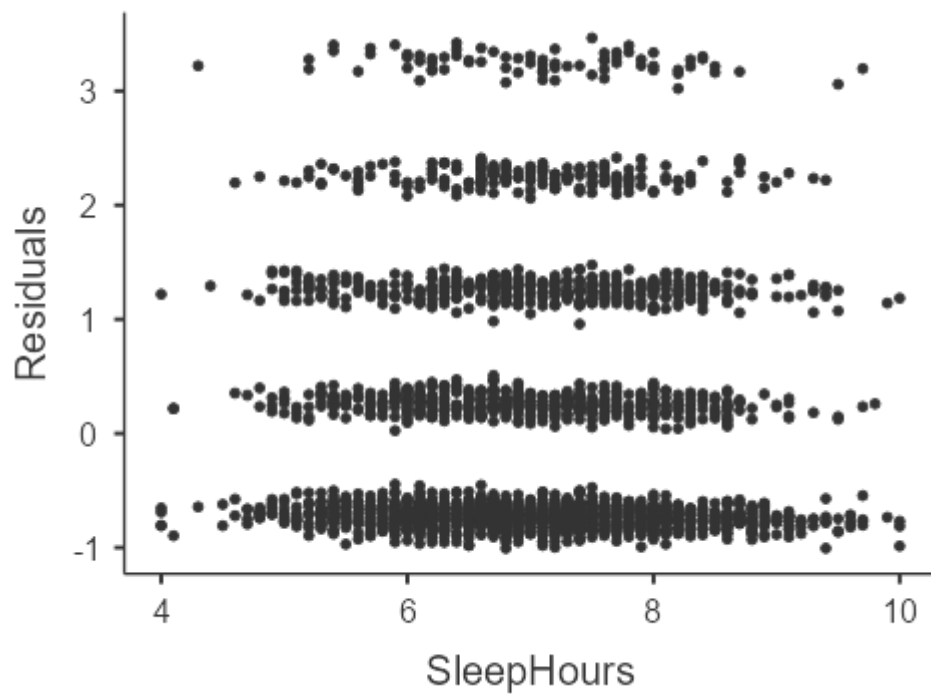
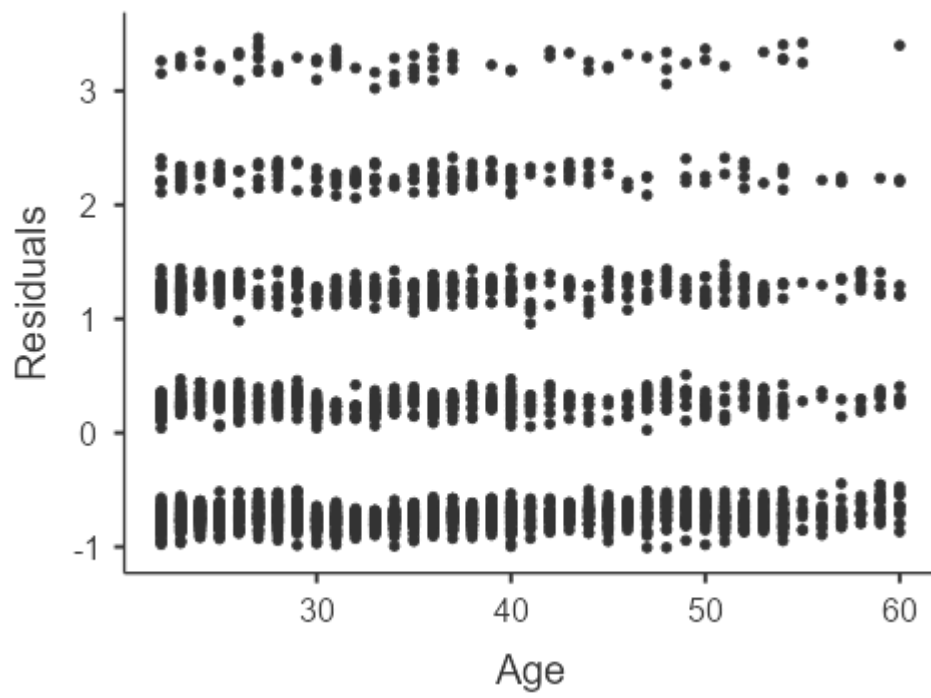
- **VIF scores:** All below 2 — no multicollinearity concerns.
- **Shapiro-Wilk test:** $p < .001 \rightarrow$ residuals are **not normally distributed**.
- **Residual plots:** Should be reviewed visually, but statistical tests suggest deviation from normality.

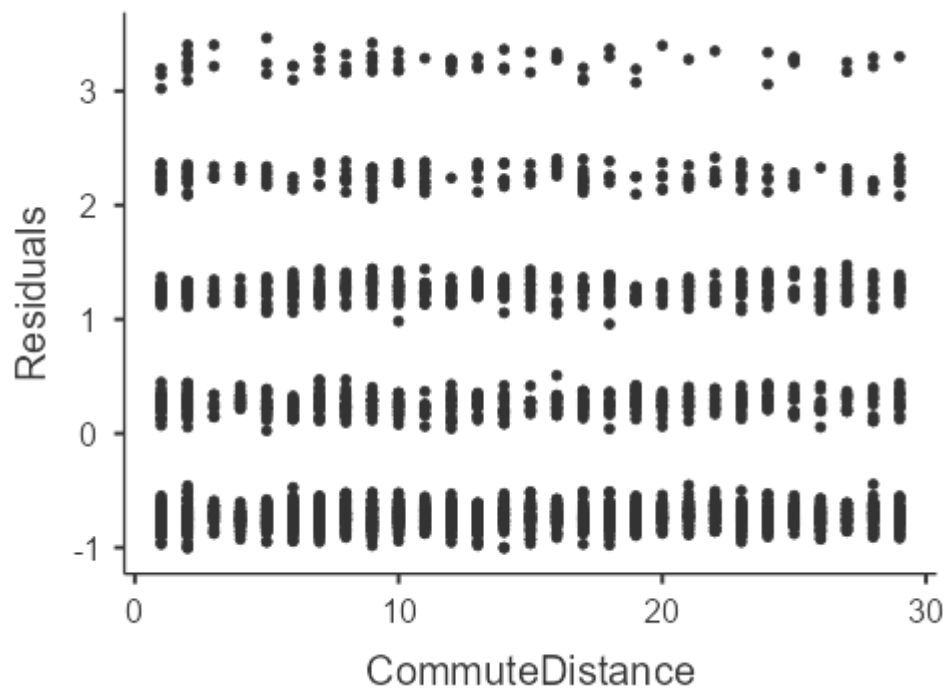
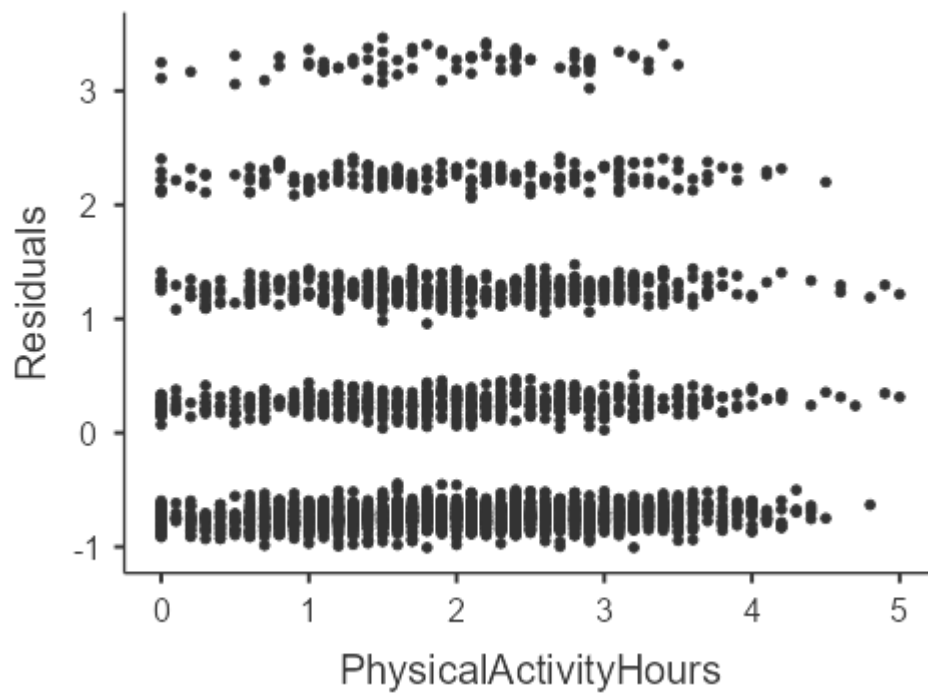
Interpretation:

The model is statistically valid, but the residuals violate normality. Since stress is measured on a Likert scale, this isn't surprising. You might note this as a limitation and consider robust or non-parametric alternatives in future modeling.

Residuals Plots







Summary Analysis

The purpose of this analysis was to examine the extent to which demographic, organizational, and lifestyle factors predict employee stress levels. A multiple linear regression was conducted

using stress scores as the dependent variable and a set of predictors including gender, age, job level, department, employment type, overtime status, sleep hours, physical activity hours, and commute distance.

The overall model was statistically significant; however, it accounted for only a small proportion of the variance in stress scores ($R^2 = 0.0065$). Among the predictors, age was found to be statistically significant ($p = .038$), with older employees reporting slightly lower stress levels. Overtime status approached significance ($p = .053$), suggesting that employees who work overtime may experience marginally higher stress. All other predictors, including gender, job level, department, and lifestyle factors, did not reach statistical significance.

Given these results, the null hypothesis (H_{06}), that demographic, organizational, and lifestyle factors do not significantly predict employee stress levels, cannot be rejected. While age and overtime status had a minor effect, the model's overall explanatory power was minimal. These findings suggest that stress may be influenced by factors not captured in this model, such as interpersonal dynamics, emotional intelligence competencies, or organizational culture.

Future research may benefit from incorporating qualitative data or mixed-methods approaches to explore the nuanced and context-dependent nature of workplace stress.

What Does This All Mean?

Failing to reject the null doesn't mean the null is true; it means we didn't find enough evidence to support the alternative. In this case, demographic, organizational, and lifestyle factors do not significantly predict stress levels in this dataset.

References

The jamovi project (2025). *jamovi*. (Version 2.7) [Computer Software]. Retrieved from <https://www.jamovi.org>.

R Core Team (2025). *R: A Language and environment for statistical computing*. (Version 4.5) [Computer software]. Retrieved from <https://cran.r-project.org>. (R packages retrieved from CRAN snapshot 2025-05-25).

Fox, J., & Weisberg, S. (2024). *car: Companion to Applied Regression*. [R package]. Retrieved from <https://cran.r-project.org/package=car>.

Lainguyn123. (2023). *Employee Survey* [Data set]. Kaggle. <https://www.kaggle.com/datasets/lainguyn123/employee-survey>

