Does Temperature Affect The Number Of Pedestrians Outside?

Abstract

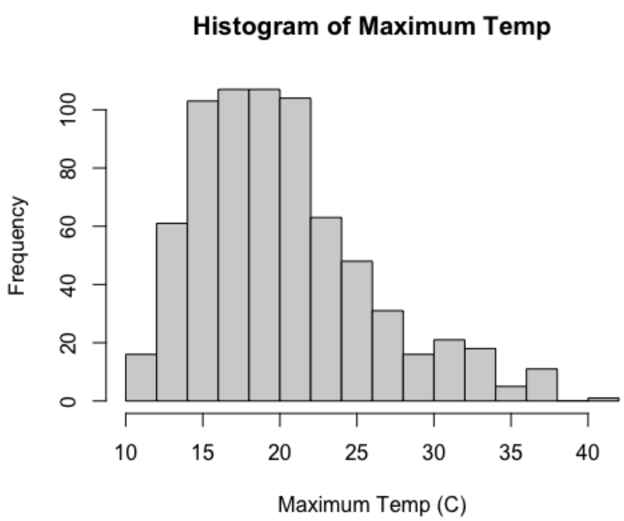
Does the temperature affect if people choose to go outside and how long they spend outside? Using Melbourne, Australia as a test case, we find and model the relationship between the temperature and the number of pedestrians recorded by sensors within Melbourne. First, we performed a simple linear regression, finding that the temperature is a significant variable in predicting the number of pedestrians on a given day; yet, a simple linear model explains only a little of the variance of the number of pedestrians. Thus, we next explored a polynomial regression model, using cross-validation to choose the degree of the polynomial. As such, we fit a quadratic regression model, which better aligns with the scatterplot of the two variables than a linear model. This resulted in a slightly more promising model, yet still explaining a small percentage of the variance in the number of pedestrians recorded. While temperature is a factor, more is needed to explain more fully the number of pedestrians per day.

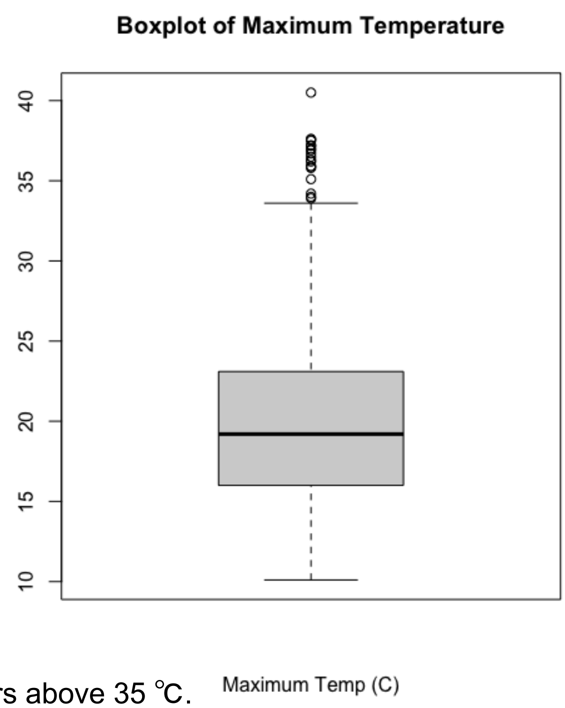
Introduction

Getting outside and being physically active, whether in the form of participating in an activity or simply walking to the store, is important in maintaining good health—mental, physical, and spiritual. Accordingly, knowing what can prevent people from spending time outside is critical to understanding when an individual's health may have declined due to a lack of physical activity and what can be done to help overcome these reasons for not going outside. In this paper, we are analyzing one possible reason a person may choose to stay inside or go outside: the temperature. Do extreme temperatures cause fewer people to go outside and, of those who do go outside, spend a shorter duration outside? Using daily data for 2023 and 2024 in Melbourne, Australia, we examine whether the daily maximum temperature affects the total number of pedestrians seen on the sidewalks of the city per day. Melbourne has pedestrian sensors on various sidewalks in the city which track the number of pedestrians who pass underneath the sensor per day.

Descriptive Analysis

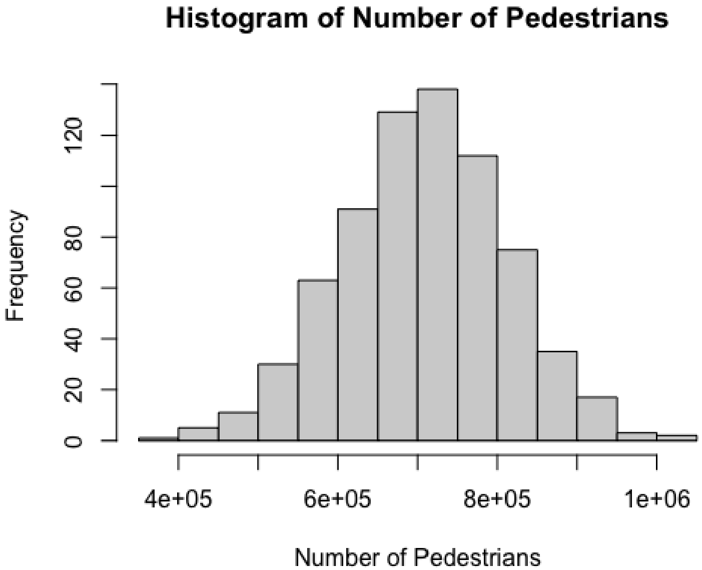
We now examine our two variables in greater detail. Our first variable, the predictor variable, is the daily temperature in Melbourne, a busy Australian city. The mean temperature was found to be 20.23 ℃, with a median of 19.20℃, an interquartile range of 7.10℃, and a standard deviation of 5.64℃. Below are the graphical representations of the temperature.

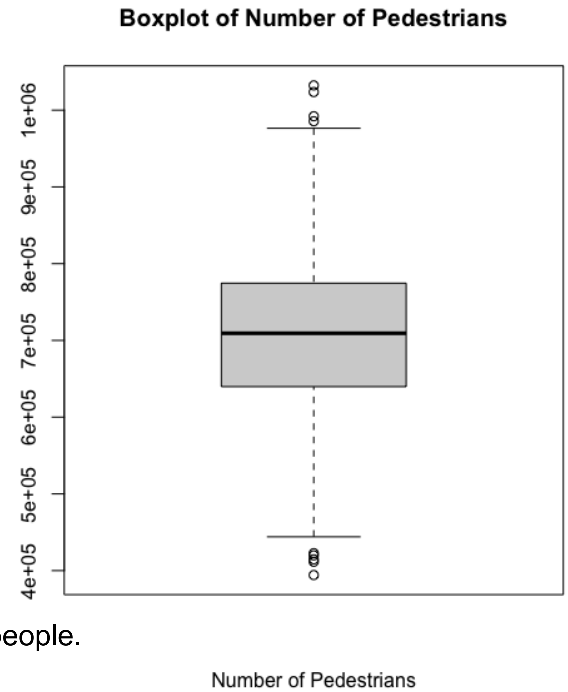




- The most common temperatures are 16-20 ℃.
- The temperatures seem to be skewed right, with outliers above 35 ℃.

The response variable is the number of pedestrians measured on variance sensors each day. The mean number of pedestrians was 707828 pedestrians, with a median of 709322 pedestrians in a day, an interquartile range of 134642 pedestrians, and a standard deviation of 104119.3 pedestrians.

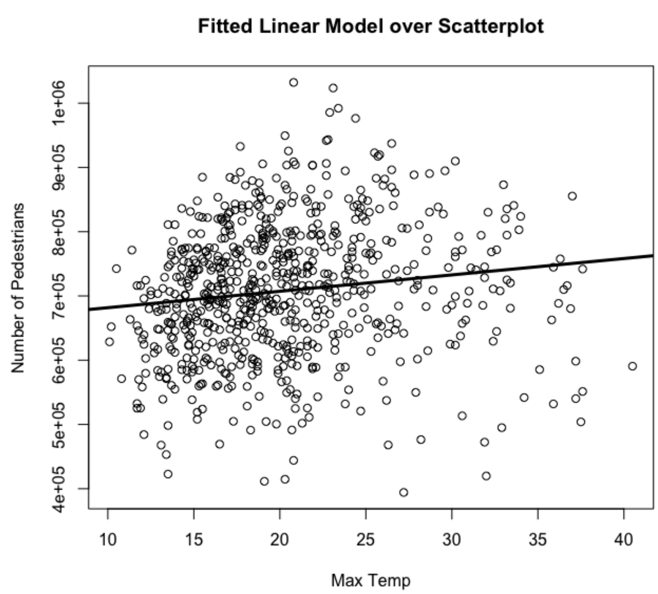




- The most common number of pedestrians is 700,000 people.
- The number of pedestrians seems to be normally distributed.

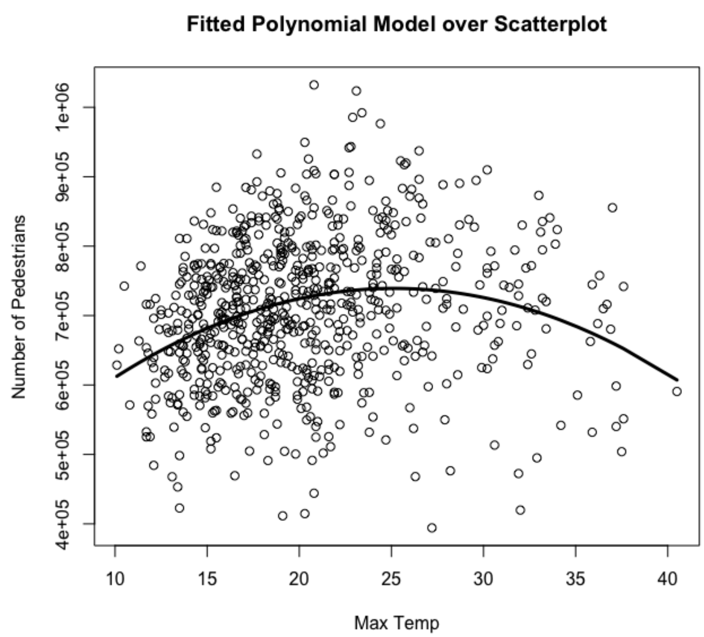
We aim to find the relationship between these two variables. We speculate that there does exist a relationship between the temperature and the number of pedestrians. Intuitively, we know that people will remain indoors during extreme weather and go outside during nicer weather conditions.

Statistical Analysis



We performed simple linear regression on the two variables to find a linear relationship, which resulted in an estimated intercept of 656170.7, slope of 2553.8 and standard error of 686.2. Both the intercept and slope were significant variables. We tested the null hypothesis that there existed no relationship between the two variables. The resulting t-value was 3.722, which yielded a p-value of 0.000213. This is below our significance level of 5%, which means that we reject the null hypothesis and conclude that there exists a relationship between the temperature and the number of pedestrians. This model, however, only had an R-squared value of 0.01914. This model fails to capture a significant amount of the variance.

In viewing the scatterplot of the number of pedestrians and the max temperature that day, we noticed a slight non-linear relationship between the two variables; we believed this looked like a quadratic relationship, where the number of pedestrians peaks with a temperature of around 25℃. As such, we chose to do a polynomial regression model after the completion of the linear model. To validate our hypothesis of a quadratic relationship, we used cross-validation in R to determine the best degree of polynomial for a polynomial regression model without overfitting to our data. The results of the cross-validation confirmed our hypothesis that a quadratic regression model was the best model for the data.



In the fitted polynomial model, each estimator is statistically significant. The estimate for the linear coefficient changed from 2553.8 pedestrians to 384059 pedestrians, a large increase due to the addition of the quadratic term. The estimate for the quadratic coefficient was -628478 pedestrians; moreover, this affirms our hypothesis that the number of pedestrians peaks, on average, around 25℃ and decreases in extreme cold and extreme heat. While this model improves the simple linear model, it is still a poor model overall, explaining only 7% of the variance of the number of pedestrians.

Conclusion

While our results indicated that temperature is a statistically significant variable, our model left much to be desired. We struggled to find a model that could capture much of the variability. Perhaps this is due to how volatile the number of pedestrians on any given day can be. Many other variables could artificially inflate or deflate the data points, such as a large social event or weather hazards that do not change the temperature. There was also an issue with data collection, as we could not utilize records before 2023 due to the impact of the COVID-19 lockdown. We would likely need to remodel the data to factor in other variables to find a better-fitting model.

Data Sources

Melbourne Pedestrian Data

Latest Weather Observations for the Melbourne Area