Does the Weather Affect Cycling's Popularity?

Introduction

Waka Kotahi NZTA has plans to grow the share of local travel by public transport, walking and cycling. The national benefits would span accessibility, economic, environmental, safety and health outcomes. The plan does not address cyclists' exposure to the weather during their journeys, which could affect their choice to cycle or use a different mode of transport. We set out to investigate the impact of the weather on the popularity of cycling as a transport option.

Data

The Bike Sharing Dataset was created by Hadi Fanaee-T in 2013 and published to the UC Irvine Machine Learning Repository. It combines daily bike hire data from a Washington DC bike-hire company, Capital Bike Share, with local weather and seasonal data. The dataset is cross-sectional and includes 731 days of records from 2010 and 2011.

Y Variable: PC (number of bikes hired daily, as proxy for popularity of cycling)

X Variables:

Seasons (dummies)	Weather Situations (dummies)	Other Factors
SP - spring	CW - clear weather	TM - temperature
SU - summer	MC - misty/cloudy	AT - adjusted temperature
AU - autumn	LR - light rain/snow	HU - humidity
WI - winter		WS - wind speed

The dataset also contains other variables irrelevant to our question, which we excluded from our analysis e.g. months, weekdays, registered user status. We also excluded the heavy rain/ snow category from the weather situations, as there is no data for this variable.

Two issues with the dataset are 1) the weather situation is by nature a subjective measure, and 2) the data source is American, so it could differ from New Zealand cyclists' experiences.

Hypotheses

Null: There is no relationship between the weather conditions and popularity of cycling.

Alternative: There is a relationship between the weather conditions and popularity of cycling.

Method

We will conduct three tests, all to an $\alpha = 0.05$ level of significance:

1. *Individual Tests* to find the effect of each weather condition on the popularity of cycling, individually. Test Statistic: P-values.

   $H_0: \beta = 0 \mid H_1: \beta \neq 0$

2. *Unrestricted Joint Test* to find the effect of all weather conditions on the popularity of cycling, jointly. Test Statistic: Significance F.

   $H_0: R^2 = 0 \mid H_1: R^2 \neq 0$

3. *Restricted Joint Test* to find the effect of weather conditions with significant explanatory power (restricted weather conditions) on the popularity of cycling, jointly; and check for Omitted Variable Bias in the restricted model. Test Statistic: Significance F.

   $H_0: \beta_{SU}=\beta_{AU}=\beta_{WI}=\beta_{CW}=\beta_{MC}=\beta_{TM}=\beta_{HU}=\beta_{WS} = 0$
   $H_1: \beta_{SU}=\beta_{AU}=\beta_{WI}=\beta_{CW}=\beta_{MC}=\beta_{TM}=\beta_{HU}=\beta_{WS} \neq 0$

Results

Individual Tests:

With the exception of WI, all weather conditions have p-values low enough (less than $\alpha = 0.05$) to reject the null hypothesis. We can conclude that all weather conditions, except WI, have statistically significant individual relationships with the popularity of cycling.

Variable	Coefficient	P-value	R^2
SP	-2525.5601	0.0000	0.3171
SU	652.1304	0.0001	0.0214
AU	1534.6347	0.0000	0.1201
WI	295.8555	0.0763	0.0043
CW	1015.8645	0.0000	0.0639
MC	-707.5695	0.0000	0.0299
LR	-2780.9537	0.0000	0.0576
TM	6640.7100	0.0000	0.3937
AT	7501.8339	0.0000	0.3982
HU	-1369.0807	0.0065	0.0101
WS	-5862.9128	0.0000	0.0550

Unrestricted Joint Test: The unrestricted joint model has a Significance F low enough (less than $\alpha = 0.05$) to reject the null hypothesis. We can conclude that all weather conditions, jointly, have a statistically significant relationship with the popularity of cycling.

PC^ = 1210.21 + 925.80(SU) + 476.51(AU) + 1494.49(WI) + 1897.93(CW) + 1680.51(MC)
(0.0272) (0.0000) (0.0445) (0.0000) (0.0000) (0.0000)

+ 5683.10(TM) + 584.96(AT) - 2642.62(HU) - 3301.08(WS)
(0.0100) (0.8079) (0.0000) (0.0000)

Restricted Joint Test: The restricted joint model has a Significance F low enough (less than $\alpha = 0.05$) to reject the null hypothesis. We can conclude that the weather conditions with significant explanatory power, jointly, have a statistically significant relationship with the popularity of cycling.

PC^ = 1226.09 + 926.97(SU) + 471.32(AU) + 1496.78(WI) + 1902.85(CW) + 1684.08(MC)
(0.0242) (0.0000) (0.0459) (0.0000) (0.0000) (0.0000)

+ 6205.41(TM) - 2635.17(HU) - 3330.79(WS)
(0.0000) (0.0000) (0.0000)

Omitted Variable Bias: The Adjusted R^2 values for our unrestricted and restricted models are incredibly similar (difference = 0.0006). This shows that the omitted variable AT does not affect the popularity of cycling with any statistical significance.

	Unrestricted	Restricted
R^2	0.5565	0.5565
Adjusted R^2	0.5508	0.5514
Significance F	0.000	0.000

Conclusion

All three tests produced significant results rejecting the null hypotheses. Therefore, we can accept our alternative hypothesis that there is a relationship between the weather conditions and popularity of cycling. We conclude that cyclists' exposure to the weather during their journeys will affect their choice to cycle or to use a different mode of transport.

The Summer and Winter seasons see the biggest increases in the popularity of cycling. We suggest that Waka Kotahi NZTA considers the balance of cycling and other modes of transport in Spring and Autumn, perhaps increasing public bus frequencies in those seasons.

Future research could expand on our findings by analysing similar datasets with New Zealand-specific data to assess the effect of our weather on the popularity of cycling here.

Contribution Statement

TG, CR and AS all contributed equally to this project.

References

Waka Kotahi NZ Transport Agency. (2019). *Keeping Cities Moving*. https://www.nzta.govt.nz/assets/resources/keeping-cities-moving/Keeping-cities-moving.pdf

Fanaee-T, Hadi. (2013). Bike Sharing Dataset. UCI Machine Learning Repository. https://doi.org/10.24432/C5W894