

# STAT452/652 Solution to HW04

Copyright 2020 William Ruth and Thomas Loughin  
Distribution without their permission is illegal and may be prosecuted

Due on Oct 09, 2020

## 1 Concepts

### 1.1 Question 1

- (a) If  $X_2 = c$  then the slope of  $X_1$  is  $\beta_1 + \beta_3/c$ .
- (b) If  $X_1 = d$  then the slope of  $X_2$  is  $\beta_2 - \beta_3 d/[x_2(x_2 + 1)]$

### 1.2 Question 2

The formula for the regression model being fit is  $f(X, Z) = \beta_0 + \beta_1 X + \beta_2 z_2 + \beta_3 X z_2$ .

## 2 Applications A

### 2.1 Question 1

The following table contains minima, maxima and means for our two constructed variables.

```
##      TWcp TWrat
## Min    216  3.03
## Max   1490 40.90
## Mean    757  9.42
```

### 2.2 Question 2

```
##      t value Pr(>|t|)
## CP      -3.99 0.000123
## Rat      4.01 0.000115
```

Both variables appear to be quite useful for predicting ozone. The cross-product looks a bit better (its p-value is half that of the ratio), but both effects are so strong that it's hard to say one the clear winner.

Slopes for temperature when wind speed is at its minimum and maximum values are given in the following display.

```
##      Min      Max
## 3.500 -0.497
```

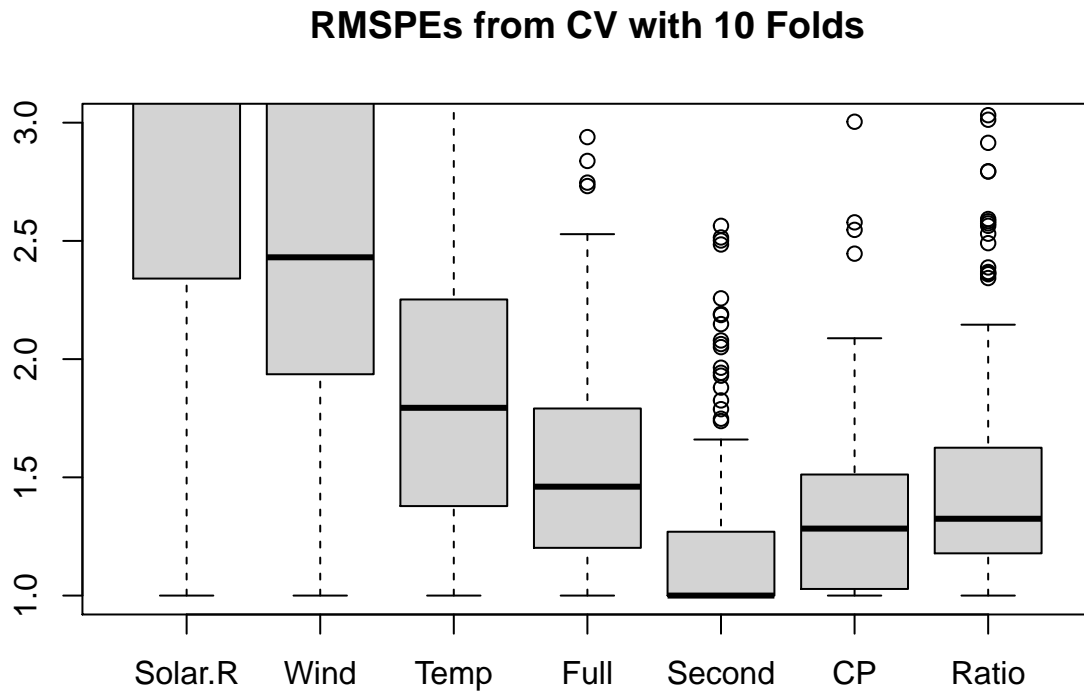
## 2.3 Question 3

Validation set MSPEs for our two models are:

```
##      CP Ratio
##    262   253
```

We see that the ratio model performs slightly better on this split than the model with a cross-product.

## 2.4 Question 4



The new models are competitive, but the best is still the model with all interactions and second-order terms.

## 3 Applications B

Here is the result of a regression of claims on the other variables.

```
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) -79.65011619 2.043655e+01 -3.8974339 1.008235e-04
## per         13.20869331 1.534397e+00  8.6083950 1.602788e-17
## km          -2.13394845 1.464133e+00 -1.4574827 1.451598e-01
## zone2       -10.39327248 6.415512e+00 -1.6200223 1.054048e-01
## zone3       -20.08869808 6.422986e+00 -3.1276260 1.790798e-03
## zone4       -26.92313581 6.403750e+00 -4.2042763 2.749330e-05
## zone5       -25.86991611 7.052991e+00 -3.6679355 2.517130e-04
```

## zone6	-30.22482960	6.659624e+00	-4.5385190	6.044679e-06
## zone7	-34.24375333	1.011206e+01	-3.3864256	7.235171e-04
## bonus	-6.04643784	9.846118e-01	-6.1409358	1.009599e-09
## make2	2.41163207	7.839965e+00	0.3076075	7.584170e-01
## make3	1.44241734	8.293514e+00	0.1739211	8.619472e-01
## make4	-0.91304348	8.678102e+00	-0.1052124	9.162192e-01
## make5	1.95294945	7.795020e+00	0.2505381	8.022002e-01
## make6	-6.69572971	7.562698e+00	-0.8853625	3.760809e-01
## make7	3.55328709	8.076772e+00	0.4399390	6.600347e-01
## make8	5.75420245	8.555254e+00	0.6725928	5.012938e-01
## make9	95.20745812	7.990441e+00	11.9151687	1.512788e-31
## insured	0.02895804	3.371639e-04	85.8871460	0.000000e+00

- (i) The total number of parameters we fit is 19.
- (ii) The regression intercept when make and zone are both at their first levels is -79.7.
- (iii) The intercept when these variables are at their last level is -18.7.