Solutions L8

```
knitr::opts_chunk$set(echo = T, message = F,
  fig.width = 7, fig.height=3.5)
options(scipen=999)
```

```
library(dplyr)
library(rgl)
```

1 Concepts

1.1 Question 1

In equation 1, the parameter β_0 measures the value of $f(X)$ when each of the indicators is zero (i.e. for the baseline region). The parameter β_k measures the difference in $f(X)$ between the baseline region and the region corresponding to the kth indicator.

2 Applications

```
################################
### Import and process data ###
################################

### Import and clean the air quality data
data("airquality")
AQ.raw = na.omit(airquality[,1:4])

### Construct new variables
AQ = AQ.raw
AQ$TWcp = with(AQ.raw, Temp * Wind)
AQ$TWrat = with(AQ.raw, Temp / Wind)

########################
### Helper Functions ###
########################

### Create function to compute MSPEs
get.MSPE = function(Y, Y.hat){
  return(mean((Y - Y.hat)^2))
}

### Create function which constructs folds for CV
### n is the number of observations, K is the number of folds
get.folds = function(n, K) {
```

```
  ### Get the appropriate number of fold labels
  n.fold = ceiling(n / K) # Number of observations per fold (rounded up)
  fold.ids.raw = rep(1:K, times = n.fold)
  fold.ids = fold.ids.raw[1:n]

  ### Shuffle the fold labels
  folds.rand = fold.ids[sample.int(n)]

  return(folds.rand)
}
```

2.1 Question 1

```
### Compute medians
med.wind = median(AQ$Wind)
med.temp = median(AQ$Temp)
```

The medians of wind speed and temperature are 9.7 and 79 respectively.

2.2 Question 2

The following is a list of all wind speed and temperature values, along with indicators for both variables which are true iff the corresponding variable exceeds its median.

```
### Split wind and temp at their medians, and construct indicators
data = AQ %>%
  mutate(wind.hilo = Wind > med.wind, temp.hilo = Temp > med.temp)

### Print values of wind, temp and their indicators
print(select(data, Wind, Temp, wind.hilo, temp.hilo))
```

```
##     Wind Temp wind.hilo temp.hilo
## 1    7.4   67     FALSE     FALSE
## 2    8.0   72     FALSE     FALSE
## 3   12.6   74      TRUE     FALSE
## 4   11.5   62      TRUE     FALSE
## 7    8.6   65     FALSE     FALSE
## 8   13.8   59      TRUE     FALSE
## 9   20.1   61      TRUE     FALSE
## 12   9.7   69     FALSE     FALSE
## 13   9.2   66     FALSE     FALSE
## 14  10.9   68      TRUE     FALSE
## 15  13.2   58      TRUE     FALSE
## 16  11.5   64      TRUE     FALSE
## 17  12.0   66      TRUE     FALSE
## 18  18.4   57      TRUE     FALSE
## 19  11.5   68      TRUE     FALSE
## 20   9.7   62     FALSE     FALSE
## 21   9.7   59     FALSE     FALSE
## 22  16.6   73      TRUE     FALSE
## 23   9.7   61     FALSE     FALSE
## 24  12.0   61      TRUE     FALSE
```

```
## 28  12.0   67      TRUE     FALSE
## 29  14.9   81      TRUE      TRUE
## 30   5.7   79     FALSE     FALSE
## 31   7.4   76     FALSE     FALSE
## 38   9.7   82     FALSE      TRUE
## 40  13.8   90      TRUE      TRUE
## 41  11.5   87      TRUE      TRUE
## 44   8.0   82     FALSE      TRUE
## 47  14.9   77      TRUE     FALSE
## 48  20.7   72      TRUE     FALSE
## 49   9.2   65     FALSE     FALSE
## 50  11.5   73      TRUE     FALSE
## 51  10.3   76      TRUE     FALSE
## 62   4.1   84     FALSE      TRUE
## 63   9.2   85     FALSE      TRUE
## 64   9.2   81     FALSE      TRUE
## 66   4.6   83     FALSE      TRUE
## 67  10.9   83      TRUE      TRUE
## 68   5.1   88     FALSE      TRUE
## 69   6.3   92     FALSE      TRUE
## 70   5.7   92     FALSE      TRUE
## 71   7.4   89     FALSE      TRUE
## 73  14.3   73      TRUE     FALSE
## 74  14.9   81      TRUE      TRUE
## 76  14.3   80      TRUE      TRUE
## 77   6.9   81     FALSE      TRUE
## 78  10.3   82      TRUE      TRUE
## 79   6.3   84     FALSE      TRUE
## 80   5.1   87     FALSE      TRUE
## 81  11.5   85      TRUE      TRUE
## 82   6.9   74     FALSE     FALSE
## 85   8.6   86     FALSE      TRUE
## 86   8.0   85     FALSE      TRUE
## 87   8.6   82     FALSE      TRUE
## 88  12.0   86      TRUE      TRUE
## 89   7.4   88     FALSE      TRUE
## 90   7.4   86     FALSE      TRUE
## 91   7.4   83     FALSE      TRUE
## 92   9.2   81     FALSE      TRUE
## 93   6.9   81     FALSE      TRUE
## 94  13.8   81      TRUE      TRUE
## 95   7.4   82     FALSE      TRUE
## 99   4.0   89     FALSE      TRUE
## 100 10.3   90      TRUE      TRUE
## 101  8.0   90     FALSE      TRUE
## 104 11.5   86      TRUE      TRUE
## 105 11.5   82      TRUE      TRUE
## 106  9.7   80     FALSE      TRUE
## 108 10.3   77      TRUE     FALSE
## 109  6.3   79     FALSE     FALSE
## 110  7.4   76     FALSE     FALSE
## 111 10.9   78      TRUE     FALSE
## 112 10.3   78      TRUE     FALSE
## 113 15.5   77      TRUE     FALSE
```

```
## 114 14.3   72      TRUE     FALSE
## 116  9.7   79     FALSE     FALSE
## 117  3.4   81     FALSE      TRUE
## 118  8.0   86     FALSE      TRUE
## 120  9.7   97     FALSE      TRUE
## 121  2.3   94     FALSE      TRUE
## 122  6.3   96     FALSE      TRUE
## 123  6.3   94     FALSE      TRUE
## 124  6.9   91     FALSE      TRUE
## 125  5.1   92     FALSE      TRUE
## 126  2.8   93     FALSE      TRUE
## 127  4.6   93     FALSE      TRUE
## 128  7.4   87     FALSE      TRUE
## 129 15.5   84      TRUE      TRUE
## 130 10.9   80      TRUE      TRUE
## 131 10.3   78      TRUE     FALSE
## 132 10.9   75      TRUE     FALSE
## 133  9.7   73     FALSE     FALSE
## 134 14.9   81      TRUE      TRUE
## 135 15.5   76      TRUE     FALSE
## 136  6.3   77     FALSE     FALSE
## 137 10.9   71      TRUE     FALSE
## 138 11.5   71      TRUE     FALSE
## 139  6.9   78     FALSE     FALSE
## 140 13.8   67      TRUE     FALSE
## 141 10.3   76      TRUE     FALSE
## 142 10.3   68      TRUE     FALSE
## 143  8.0   82     FALSE      TRUE
## 144 12.6   64      TRUE     FALSE
## 145  9.2   71     FALSE     FALSE
## 146 10.3   81      TRUE      TRUE
## 147 10.3   69      TRUE     FALSE
## 148 16.6   63      TRUE     FALSE
## 149  6.9   70     FALSE     FALSE
## 151 14.3   75      TRUE     FALSE
## 152  8.0   76     FALSE     FALSE
## 153 11.5   68      TRUE     FALSE
```

2.3 Question 3

The following is a summary of the linear model of ozone on the indicators for wind speed and temperature.

```
fit.main = lm(Ozone ~ wind.hilo + temp.hilo, data = data)

summary(fit.main)
```

```
##
## Call:
## lm(formula = Ozone ~ wind.hilo + temp.hilo, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.256 -13.065  -1.874   9.435  98.744
##
```

```
## Coefficients:
##                Estimate Std. Error t value        Pr(>|t|)
## (Intercept)      36.216      4.547   7.966 0.0000000000018 ***
## wind.hiloTRUE  -21.342      4.973  -4.292 0.0000387938694 ***
## temp.hiloTRUE   33.040      4.970   6.648 0.0000000012523 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.82 on 108 degrees of freedom
## Multiple R-squared:  0.4538, Adjusted R-squared:  0.4437
## F-statistic: 44.86 on 2 and 108 DF,  p-value: 0.00000000000000657
```

```
### Extract p-values
anova.table = summary(fit.main)$coef
p.vals = anova.table[,4]
p.val.wind = p.vals["wind.hiloTRUE"]
p.val.temp = p.vals["temp.hiloTRUE"]
```

Both predictors have statistically significant effects on ozone (wind speed: $p =$ 3.88e-05, temperature: $p =$ 1.25e-09).

```
### Construct a grid in Temp and Wind
vals.wind = seq(from = 2, to = 21, by = 1)
vals.temp = seq(from = 56, to = 98, by = 2)
pred.grid = expand.grid(Wind = vals.wind, Temp = vals.temp)

### Get indicators on grid
pred.grid.ind = mutate(pred.grid, wind.hilo = Wind > med.wind,
  temp.hilo = Temp > med.temp)

### Get fitted values on our grid
pred.ozone = predict(fit.main, newdata = pred.grid.ind)

### Plot fitted surface with scatterplot
open3d()
persp3d(x = vals.wind, y = vals.temp, z = pred.ozone, col = "orange",
  xlab = "Wind Speed", ylab = "Temperature", zlab = "Ozone")
with(data, points3d(Ozone ~ Wind + Temp))
```

See Figure 1 for a plot of the predicted values from our model. We see that ozone increases for large values of temperature, and decreases for large values of wind speed.

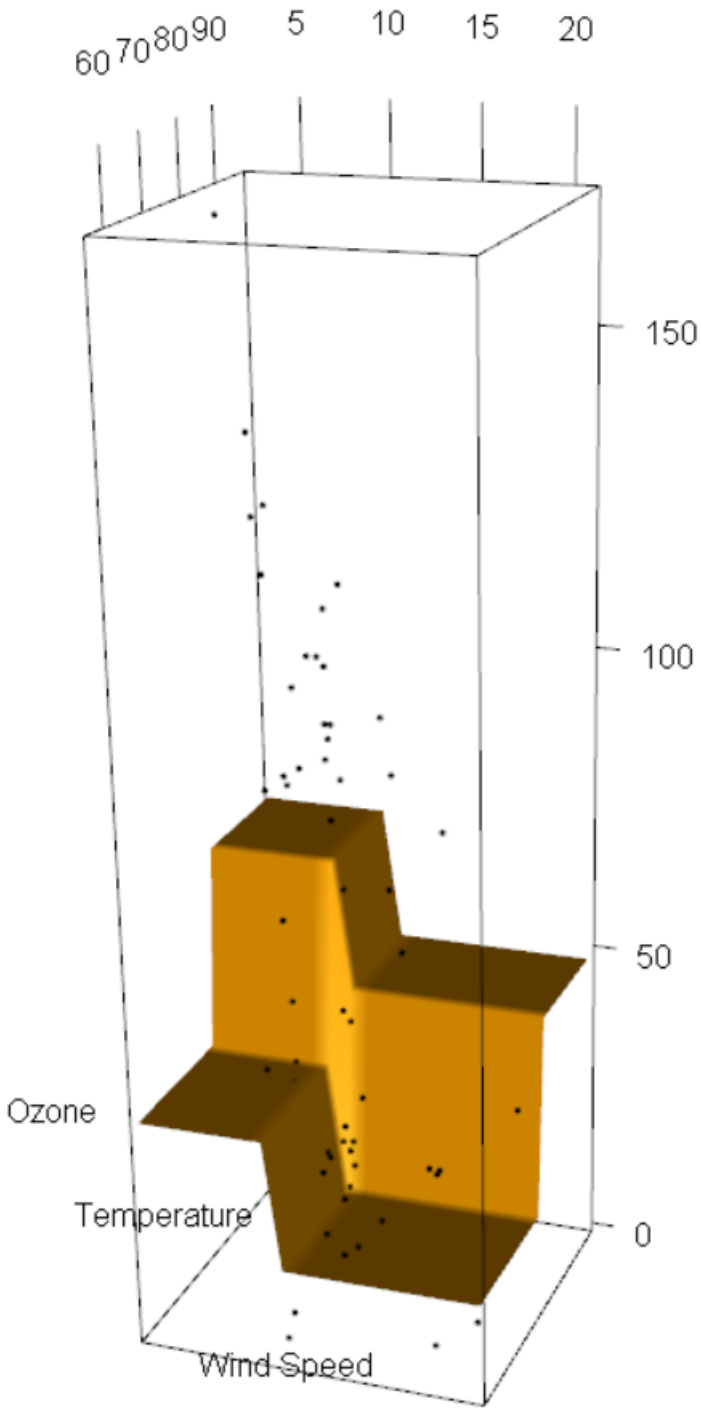
2.4 Question 4

The following is a summary of the linear model of ozone on the indicators for wind speed and temperature with interactions.

```
fit.int = lm(Ozone ~ wind.hilo * temp.hilo, data = data)

summary(fit.int)
```

```
##
## Call:
## lm(formula = Ozone ~ wind.hilo * temp.hilo, data = data)
##
```



Effects Plot.png

Figure 1: Step function with only main effects.

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -56.865 -10.220  -0.583  10.135  95.135
##
## Coefficients:
##                             Estimate Std. Error t value      Pr(>|t|)
## (Intercept)                   29.857      5.322   5.610 0.00000016038 ***
## wind.hiloTRUE                -11.274      6.697  -1.683        0.0952 .
## temp.hiloTRUE                 43.008      6.664   6.454 0.00000000326 ***
## wind.hiloTRUE:temp.hiloTRUE  -21.532      9.794  -2.198        0.0301 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.39 on 107 degrees of freedom
## Multiple R-squared:  0.4774, Adjusted R-squared:  0.4627
## F-statistic: 32.58 on 3 and 107 DF,  p-value: 0.000000000000004851
```

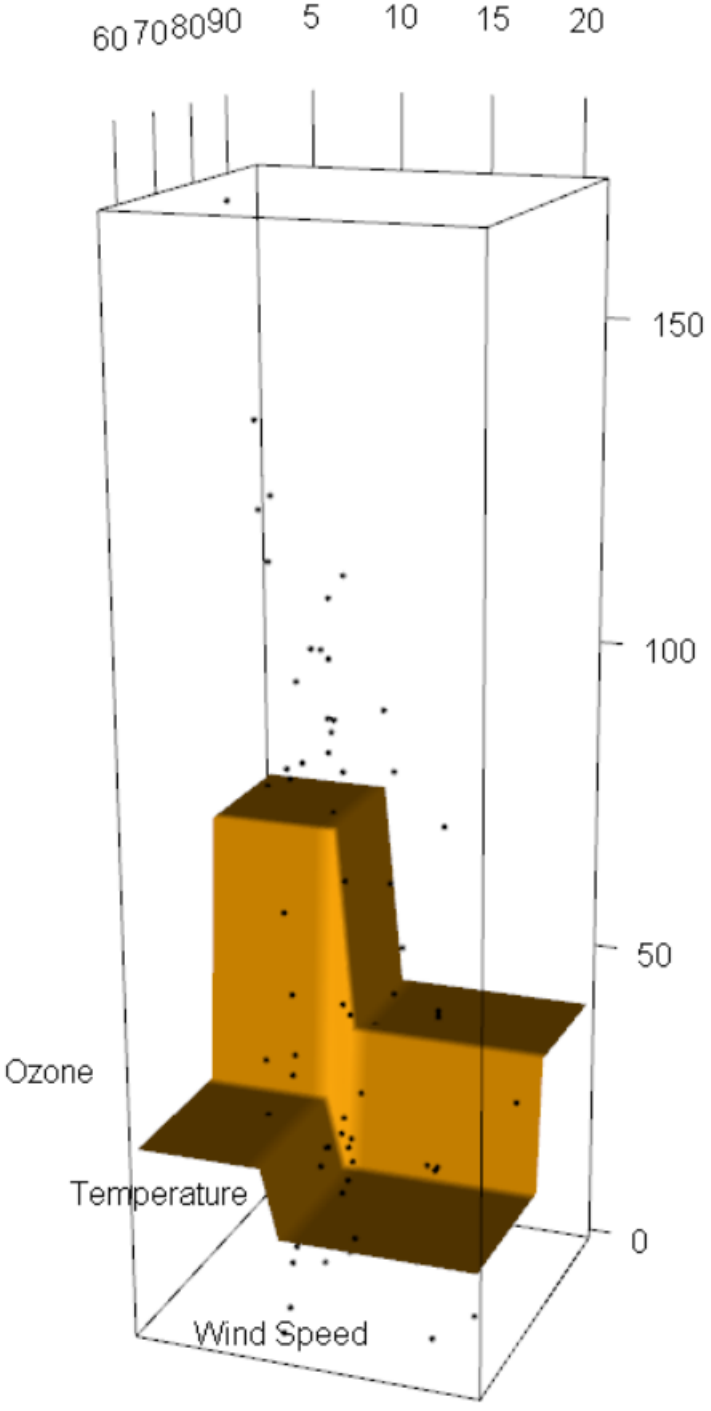
```
### Extract p-values
anova.table = summary(fit.int)$coef
p.vals = anova.table[,4]
p.val.int = p.vals["wind.hiloTRUE:temp.hiloTRUE"]
```

The interaction term has a statistically significant effect on ozone ($p = 3.01\text{e-}02$).

```
### Get fitted values on our grid
pred.ozone.int = predict(fit.int, newdata = pred.grid.ind)

### Plot fitted surface with scatterplot
open3d()
persp3d(x = vals.wind, y = vals.temp, z = pred.ozone.int, col = "orange",
  xlab = "Wind Speed", ylab = "Temperature", zlab = "Ozone")
with(data, points3d(Ozone ~ Wind + Temp))
```

See Figure 2 for a plot of the predicted values from our model. The interaction appears to lessen the effect of temperature on ozone at large values of wind speed.



Plot.png

Figure 2: Step function with interaction.