

Calorie Consumption During Bicycle Work: A Statistical Analysis of an Incomplete Dataset

Nuno Chicória, Boris Shilov, Murat Cëm Kose, Yibing Liu, Robin Vermote

Dataset

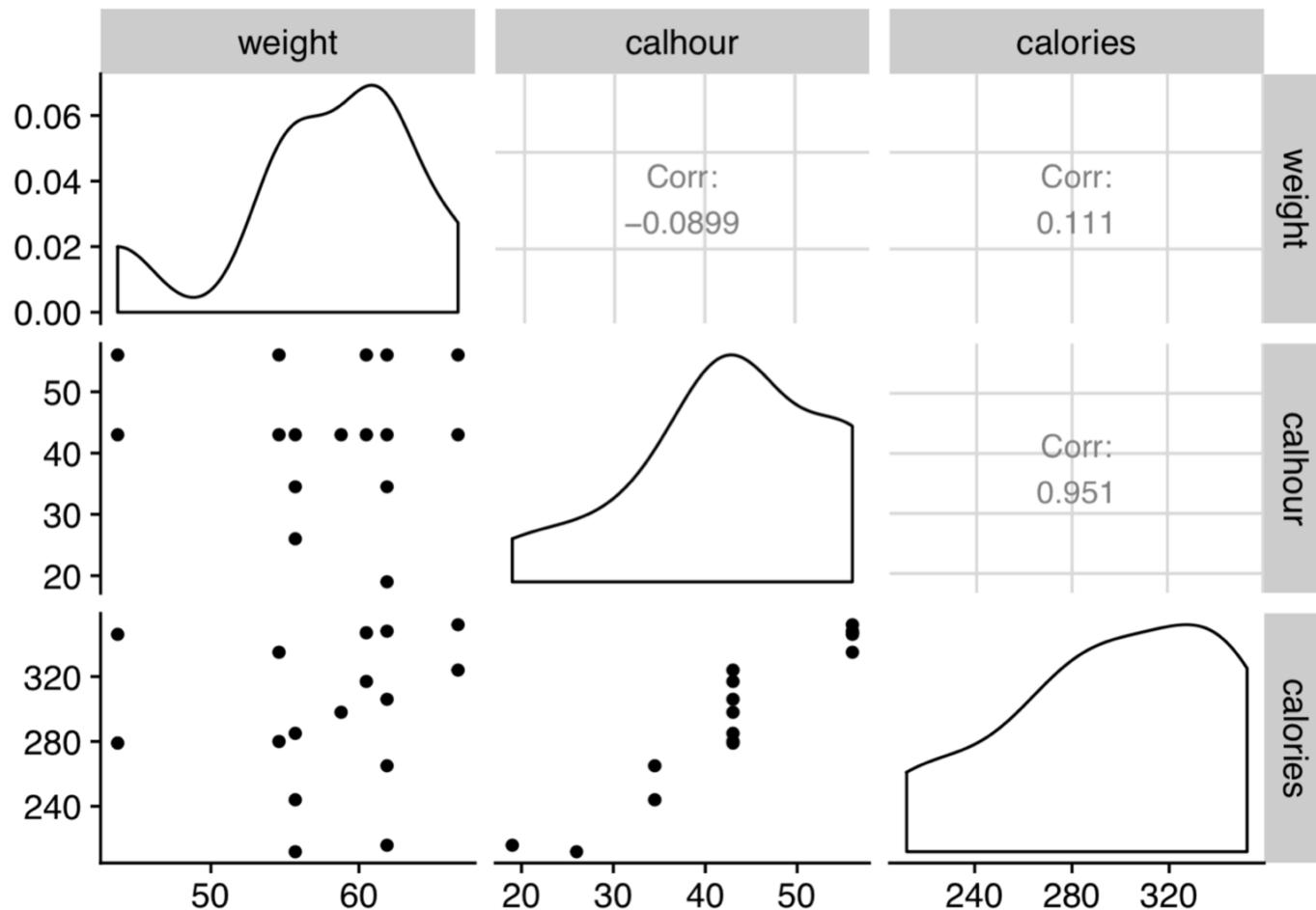
- 24 individuals
- Performing work using a bicycle ergometer
- Response Variable: heat production (calories)
- Explanatory Variables: workout intensity (calhour), weight

```
##      weight calhour calories
## 1     43.7    19.0      NA
## 2     43.7    43.0     279
## 3     43.7    56.0     346
## 4     54.6    13.0      NA
## 5     54.6    19.0      NA
## 6     54.6    43.0     280
## 7     54.6    56.0     335
## 8     55.7    13.0      NA
## 9     55.7    26.0     212
## 10    55.7    34.5     244
## 11    55.7    43.0     285
## 12    58.8    13.0      NA
## 13    58.8    43.0     298
## 14    60.5    19.0      NA
## 15    60.5    43.0     317
```

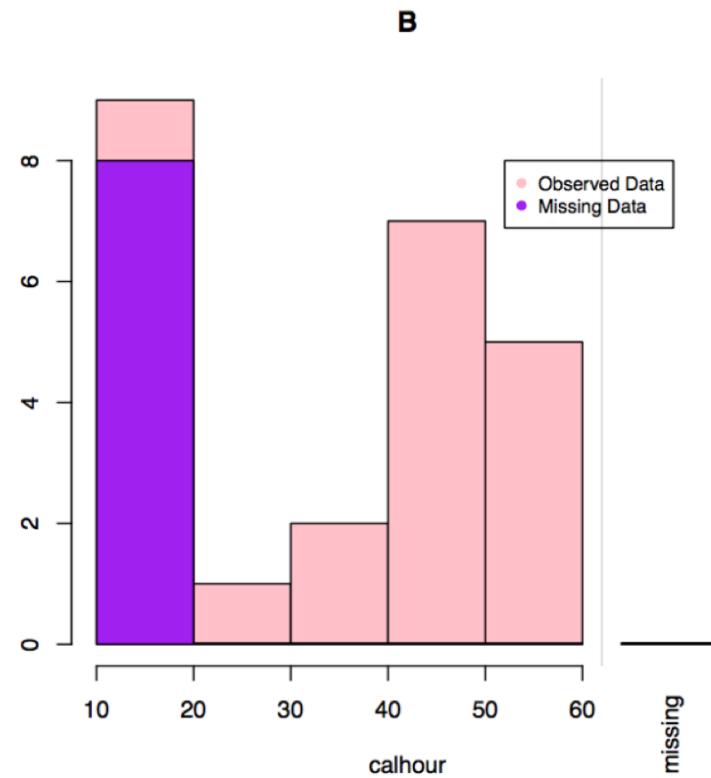
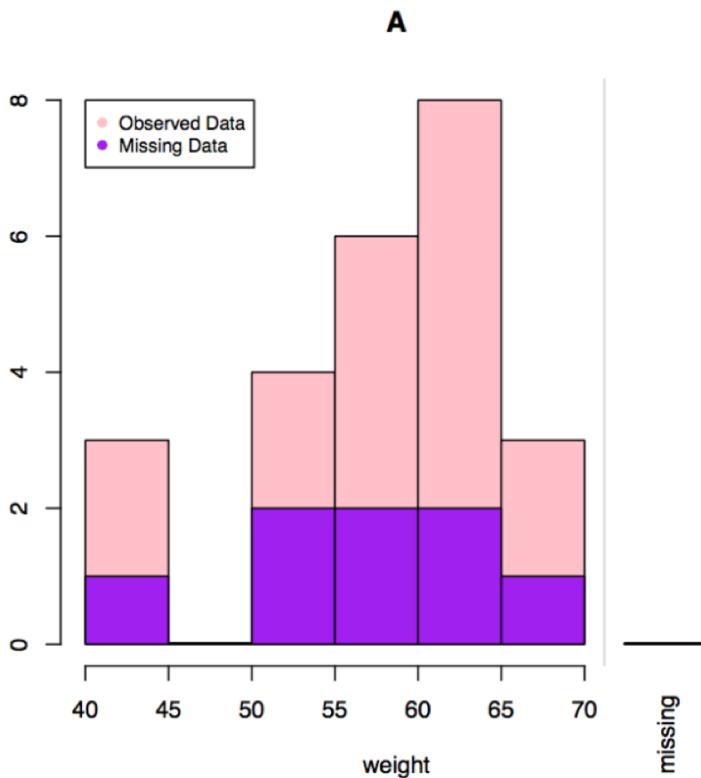
Data Exploration

```
##                                     weight  calhour  calories
## nbr.val                  24.0000  24.0000   16.0000
## nbr.null                 0.0000   0.0000   0.0000
## nbr.na                   0.0000   0.0000   8.0000
## min                      43.7000  13.0000  212.0000
## max                      66.7000  56.0000  352.0000
## range                     23.0000  43.0000  140.0000
## sum                      1381.0000 817.0000 4754.0000
## median                    58.8000  38.7500  302.0000
## mean                      57.5417  34.0417  297.1250
## SE.mean                   1.3453   3.3396  11.4669
## CI.mean.0.95              2.7829   6.9085  24.4412
## var                       43.4338  267.6721 2103.8500
## std.dev                   6.5904   16.3607  45.8677
## coef.var                  0.1145   0.4806   0.1544
```

Data Exploration



Missing Data Exploration



Model Construction

```
## Start: AIC=123.4
## calories ~ 1
##
##           Df Sum of Sq   RSS   AIC
## + calhour  1    28544  3014 87.8
## <none>            31558 123.4
## + weight   1      392 31166 125.2
##
## Step: AIC=87.81
## calories ~ calhour
##
##           Df Sum of Sq   RSS   AIC
## + weight   1     1234  1780 81.4
## <none>            3014 87.8
## - calhour  1    28544 31558 123.4
##
## Step: AIC=81.39
## calories ~ calhour + weight
##
##           Df Sum of Sq   RSS   AIC
## + weight:calhour  1      782  998 74.1
## <none>            1780 81.4
## - weight         1     1234  3014 87.8
## - calhour        1    29386 31166 125.2
##
## Step: AIC=74.13
## calories ~ calhour + weight + calhour:weight
##
##           Df Sum of Sq   RSS   AIC
## <none>            998 74.1
## - calhour:weight  1      782 1780 81.4
```

$$calories_i = \beta_0 + \beta_1 * weight_i + \beta_2 * calhour_i + \beta_3 * (weight_i * calhour_i) + \epsilon_i$$

Model Construction (Complete Case)

```
##  
## Call:  
## lm(formula = calories ~ weight + calhour + weight * calhour,  
##      data = muscledata_edit)  
##  
## Residuals:  
##    Min     1Q Median     3Q    Max  
## -12.48  -5.70  -1.04   2.39  16.95  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -330.884    124.674  -2.65  0.02102 *  
## weight        7.728     2.106   3.67  0.00321 **  
## calhour       11.787    2.548   4.63  0.00058 ***  
## weight:calhour -0.132     0.043  -3.07  0.00977 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 9.12 on 12 degrees of freedom  
## Multiple R-squared:  0.968, Adjusted R-squared:  0.96  
## F-statistic: 123 on 3 and 12 DF,  p-value: 2.89e-09
```

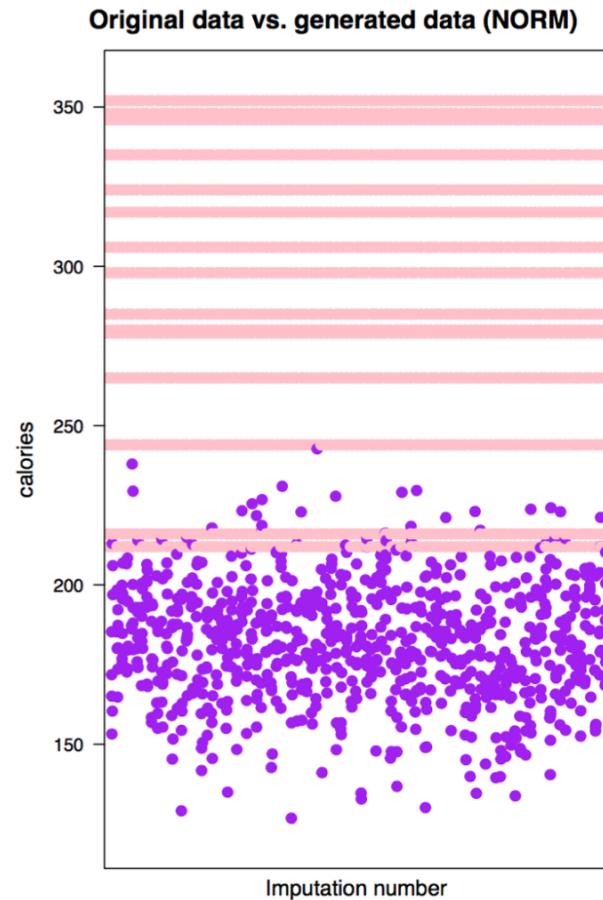
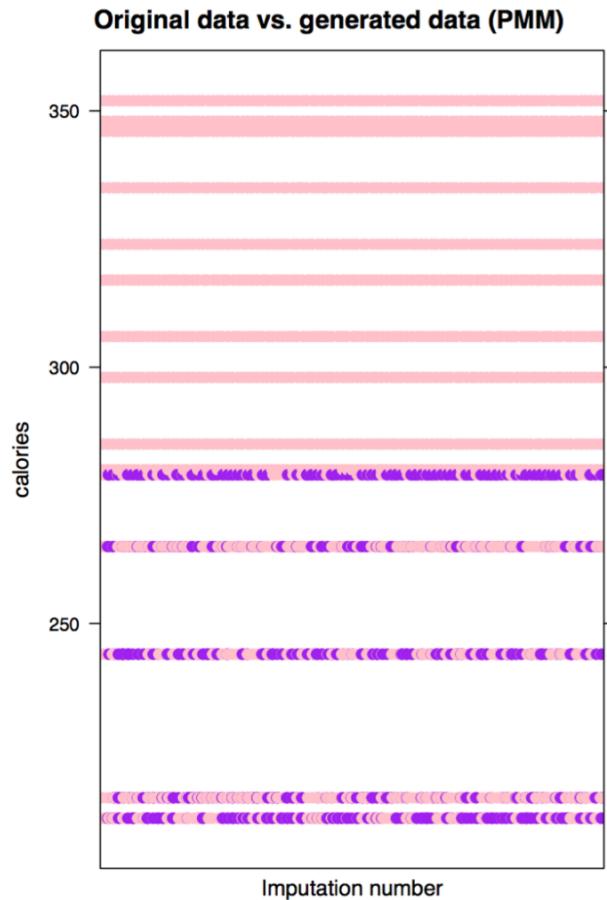
Model Construction (Multiple Imputation)

- Predictive Mean Matching

```
##          est        se      t    df Pr(>|t|)    lo 95
## (Intercept) 1.563e+02 151.38514 1.0322 10.40  0.3254 -179.3146
## weight       6.354e-01  2.59370 0.2450 10.44  0.8112 -5.1107
## calhour      2.101e+00  3.41041 0.6161 12.15  0.5492 -5.3194
## weight:calhour 7.408e-03  0.05846 0.1267 12.18  0.9012 -0.1198
##                  hi 95 nmis      fmi lambda
## (Intercept)   491.8273   NA 0.5065 0.4199
## weight         6.3814    0 0.5041 0.4174
## calhour       9.5220    0 0.4150 0.3260
```
- Bayesian Linear Regression

```
##          est        se      t    df Pr(>|t|)    lo 95
## (Intercept) -6.56646 86.90787 -0.07556 7.114  0.94185 -211.40727
## weight        2.21926  1.44880  1.53179 7.551  0.16633 -1.15657
## calhour       5.33508  1.88510  2.83013 8.946  0.01983  1.06680
## weight:calhour -0.02248  0.03154 -0.71284 9.425  0.49322 -0.09335
##                  hi 95 nmis      fmi lambda
## (Intercept)  198.27434   NA 0.6792 0.6001
## weight        5.59509    0 0.6562 0.5758
## calhour       9.60336    0 0.5827 0.4988
## weight:calhour  0.04838   NA 0.5576 0.4727
```

Models Comparisons (MI PMM vs MI NORM)



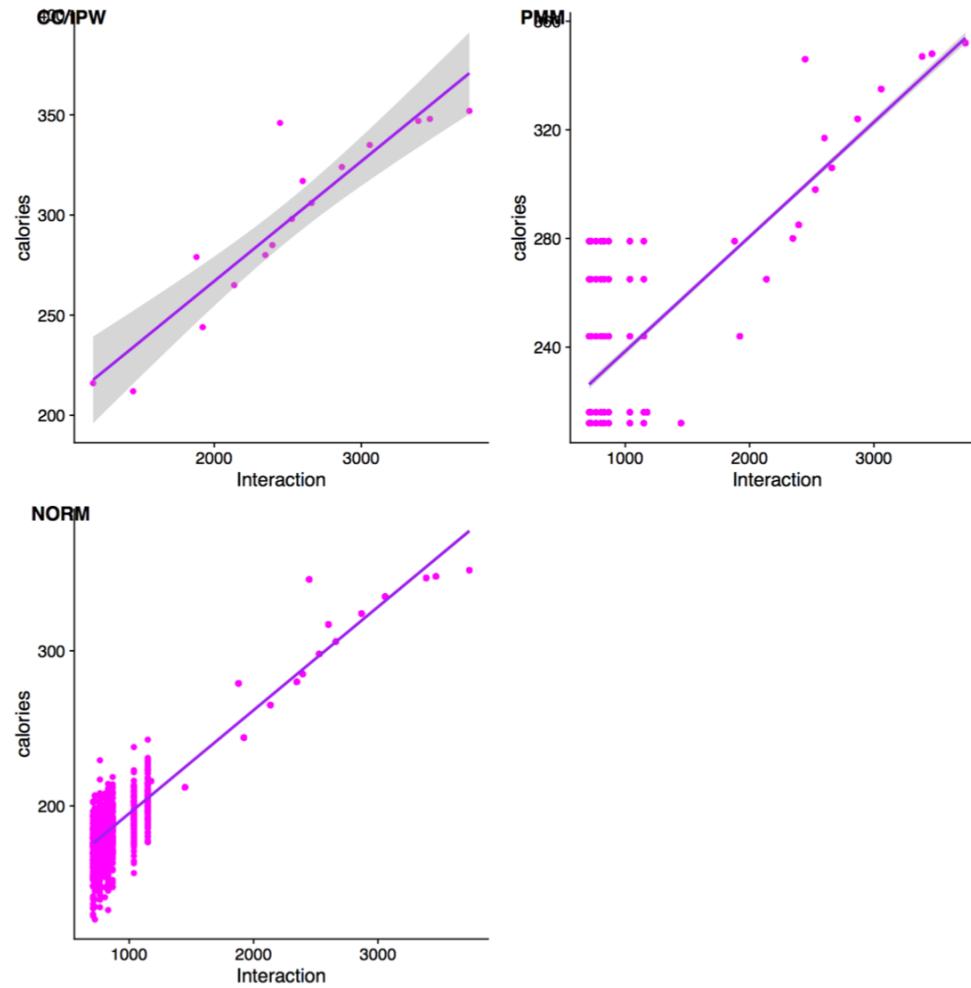
Model Construction (IPW)

```
##  
## Call:  
## lm(formula = calories ~ weight + calhour + weight * calhour,  
##      data = IPWal_muscledata, weights = muscledata$w)  
##  
## Weighted Residuals:  
##    Min     1Q Median     3Q    Max  
## -91.0  -40.5 -11.0   20.1 129.8  
##  
## Coefficients:  
##                 Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -353.7928   129.1577  -2.74  0.01796 *  
## weight        8.1131    2.1698   3.74  0.00283 **  
## calhour       12.1321    2.6513   4.58  0.00064 ***  
## weight:calhour -0.1378    0.0445  -3.10  0.00926 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 68.2 on 12 degrees of freedom  
##   (8 observations deleted due to missingness)  
## Multiple R-squared:  0.97,  Adjusted R-squared:  0.962  
## F-statistic: 128 on 3 and 12 DF,  p-value: 2.25e-09
```

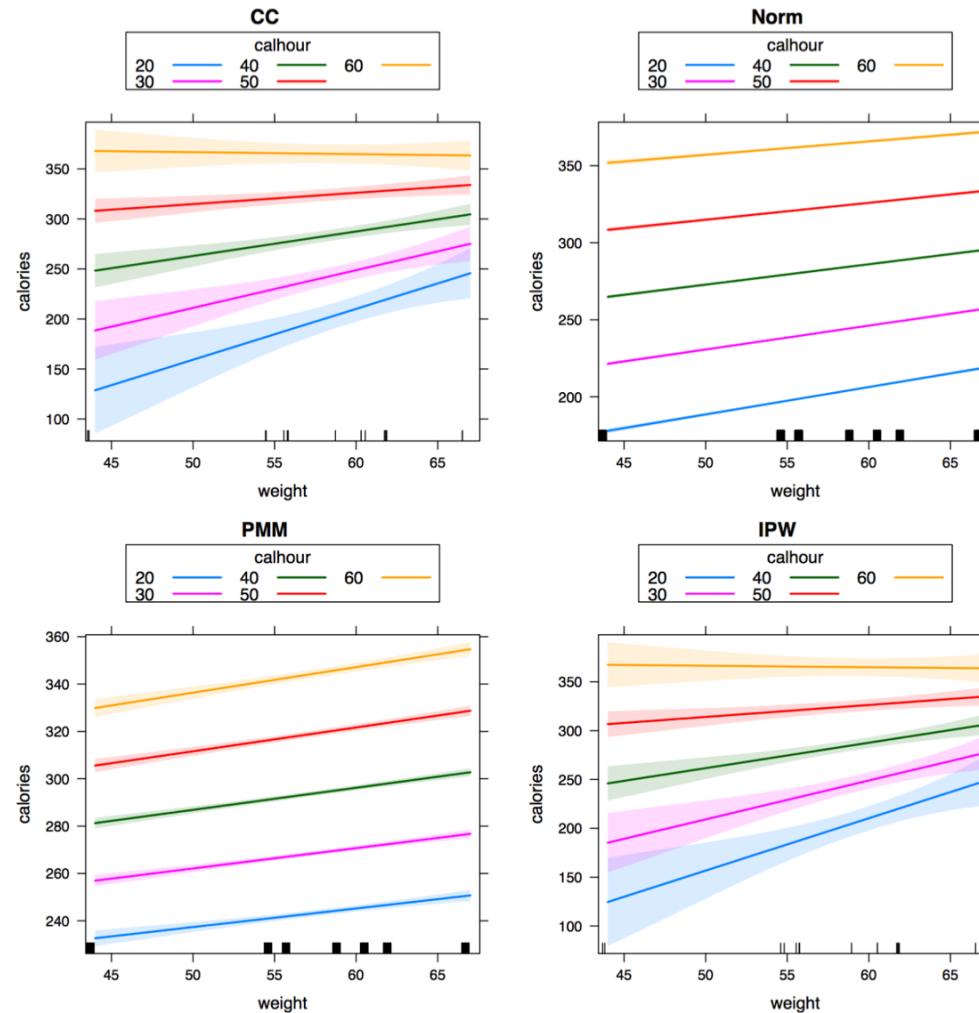
Models Summary

Coefficient	Explanation	CC			IPW		
		Estimate	Std. Error	p-value	Estimate	Std. Error	p-value
β_0	Intercept	-330.884	124.674	0.02102	-351.223	110.4449	0.00792
β_1	Weight	7.728	2.106	0.00321	8.0343	1.8406	0.00092
β_2	Calhour	11.787	2.548	0.00058	12.205	2.344	0.00022
β_3	Weight*Calhour	-0.132	0.043	0.00977	-0.1382	0.0391	0.0041
		MI-PMM			MI-NORM		
		171.837	158.033	0.387	-2.572	84.992	0.96315
		0.384	2.731	0.7772	2.175	1.436	0.19008
		1.775	3.525	0.5408	5.258	1.854	0.2313
		0.013	0.061	0.9508	-0.022	0.031	0.54274

Models Comparisons (Interaction vs Calories)



Model Comparisons (Interaction Effect)



Final Model

$$calories_i = -351.223 + 8.0343 * weight_i + 12.205 * calhour_i - 0.1382 * (weight_i * calhour_i) + \epsilon_i$$

Confidence Intervals

```
##                      2.5 %    97.5 %
## (Intercept)      -635.2033 -72.38233
## weight           3.3855   12.84065
## calhour          6.3555   17.90875
## weight:calhour -0.2347  -0.04081
```