Daniel Prusinski

412

Appendix 1 for Assignment 4 – Hierarchical Models

Output 1: Metadata

'data.frame': 2421 obs. of 14 variables:

$ Year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...

$ Team : Factor w/ 30 levels "Arizona Diamondbacks",..: 3 3 3 3 3 3 3 3 3 3 ...

$ Month : Factor w/ 7 levels "April","August",..: 1 1 1 1 1 1 1 1 1 1 ...

$ Day : int 6 7 8 9 10 11 24 25 26 27 ...

$ DayofWeek: Factor w/ 7 levels "Friday","Monday",..: 1 3 4 2 6 7 6 7 5 1 ...

$ Opponent : Factor w/ 30 levels "Arizona Diamondbacks",..: 17 17 17 19 19 19 29 29 29 20 ...

$ Temp : int 59 63 68 65 62 53 60 70 64 60 ...

$ TypeOfDay: Factor w/ 4 levels "Clear Skies",..: 1 1 1 2 1 2 2 1 2 1 ...

$ Night : int 0 1 0 1 1 1 1 1 1 1 ...

$ Attend : int 46773 31532 14738 25478 24659 22919 11058 10415 13725 18297 ...

$ BobbleHd : int 0 0 0 0 0 0 0 0 0 0 ...

$ Headgear : int 0 0 0 0 0 0 0 0 0 0 ...

$ Shirts : int 0 0 0 0 0 0 0 0 1 0 ...

$ Firewks : int 0 0 0 0 0 0 0 0 0 0 ...

Output 2: Splitting Data into Training and Testing

> dodgers.train <- subset(dodgers, training\_test == "TRAIN")

> print(str(dodgers.train)) # check training data frame

'data.frame': 1614 obs. of 15 variables:

$ Year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...

$ Team : Factor w/ 30 levels "Arizona Diamondbacks",..: 3 3 3 3 3 3 3 3 3 3 ...

$ Month : Factor w/ 7 levels "April","August",..: 1 1 1 1 1 1 1 1 1 1 ...

$ Day : int 6 7 8 9 11 24 25 26 27 29 ...

$ DayofWeek : Factor w/ 7 levels "Friday","Monday",..: 1 3 4 2 7 6 7 5 1 4 ...

$ Opponent : Factor w/ 30 levels "Arizona Diamondbacks",..: 17 17 17 19 19 29 29 29 20 20 ...

$ Temp : int 59 63 68 65 53 60 70 64 60 64 ...

$ TypeOfDay : Factor w/ 4 levels "Clear Skies",..: 1 1 1 2 2 2 1 2 1 1 ...

$ Night : int 0 1 0 1 1 1 1 1 1 0 ...

$ Attend : int 46773 31532 14738 25478 22919 11058 10415 13725 18297 31793 ...

$ BobbleHd : int 0 0 0 0 0 0 0 0 0 0 ...

$ Headgear : int 0 0 0 0 0 0 0 0 0 0 ...

$ Shirts : int 0 0 0 0 0 0 0 1 0 0 ...

$ Firewks : int 0 0 0 0 0 0 0 0 0 0 ...

$ training\_test: Factor w/ 2 levels "TRAIN","TEST": 1 1 1 1 1 1 1 1 1 1 ...

NULL

> dodgers.test <- subset(dodgers, training\_test == "TEST")

> print(str(dodgers.test)) # check test data frame

'data.frame': 807 obs. of 15 variables:

$ Year : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...

$ Team : Factor w/ 30 levels "Arizona Diamondbacks",..: 3 3 3 3 3 3 3 3 3 3 ...

$ Month : Factor w/ 7 levels "April","August",..: 1 1 5 5 5 4 4 4 4 4 ...

$ Day : int 10 28 8 11 27 9 10 26 29 30 ...

$ DayofWeek : Factor w/ 7 levels "Friday","Monday",..: 6 3 6 1 4 3 4 6 1 3 ...

$ Opponent : Factor w/ 30 levels "Arizona Diamondbacks",..: 19 20 28 27 12 21 21 13 8 8 ...

$ Temp : int 62 56 68 71 86 86 90 78 100 91 ...

$ TypeOfDay : Factor w/ 4 levels "Clear Skies",..: 1 4 2 1 2 2 1 1 1 2 ...

$ Night : int 1 1 1 1 0 0 0 1 1 0 ...

$ Attend : int 24659 26926 11263 26669 33919 46611 45267 24296 24779 35335 ...

$ BobbleHd : int 0 0 0 0 0 0 0 0 0 0 ...

$ Headgear : int 0 1 0 0 0 0 0 1 0 0 ...

$ Shirts : int 0 0 0 0 1 0 0 0 0 0 ...

$ Firewks : int 0 0 0 0 0 0 0 0 1 0 ...

$ training\_test: Factor w/ 2 levels "TRAIN","TEST": 2 2 2 2 2 2 2 2 2 2 ...

NULL

Output 3: Attendance by Day



Output 4: Attendance by Month



Output 5: Team and Night/Day Attendance



Output 6: Promotion Attendance



Output 7: Stepwise Regression Model

Start: AIC=27602.39

Attend ~ Year + Team + Month + Day + DayofWeek + Opponent + Temp +

TypeOfDay + Night + BobbleHd + Headgear + Shirts + Firewks

Step: AIC=27602.39

Attend ~ Team + Month + Day + DayofWeek + Opponent + Temp + TypeOfDay +

Night + BobbleHd + Headgear + Shirts + Firewks

Df Sum of Sq RSS AIC

- TypeOfDay 3 4.4051e+07 3.9088e+10 27598

- Temp 1 1.1105e+07 3.9055e+10 27601

- Day 1 1.3294e+07 3.9057e+10 27601

<none> 3.9044e+10 27602

- Headgear 1 1.7928e+08 3.9223e+10 27608

- Shirts 1 5.4490e+08 3.9589e+10 27623

- Night 1 6.9885e+08 3.9743e+10 27629

- Firewks 1 9.5820e+08 4.0002e+10 27640

- BobbleHd 1 1.0506e+09 4.0095e+10 27643

- Month 6 1.8941e+09 4.0938e+10 27667

- Opponent 29 6.3496e+09 4.5394e+10 27788

- DayofWeek 6 7.4966e+09 4.6541e+10 27874

- Team 29 8.5221e+10 1.2427e+11 29413

Step: AIC=27598.21

Attend ~ Team + Month + Day + DayofWeek + Opponent + Temp + Night +

BobbleHd + Headgear + Shirts + Firewks

Df Sum of Sq RSS AIC

- Temp 1 3.2946e+06 3.9091e+10 27596

- Day 1 1.5556e+07 3.9104e+10 27597

<none> 3.9088e+10 27598

- Headgear 1 1.8411e+08 3.9272e+10 27604

- Shirts 1 5.4744e+08 3.9635e+10 27619

- Night 1 7.0028e+08 3.9788e+10 27625

- Firewks 1 9.6636e+08 4.0054e+10 27636

- BobbleHd 1 1.0464e+09 4.0134e+10 27639

- Month 6 1.8910e+09 4.0979e+10 27663

- Opponent 29 6.3241e+09 4.5412e+10 27782

- DayofWeek 6 7.4958e+09 4.6584e+10 27869

- Team 29 9.9401e+10 1.3849e+11 29582

Step: AIC=27596.35

Attend ~ Team + Month + Day + DayofWeek + Opponent + Night +

BobbleHd + Headgear + Shirts + Firewks

Df Sum of Sq RSS AIC

- Day 1 1.4866e+07 3.9106e+10 27595

<none> 3.9091e+10 27596

- Headgear 1 1.8591e+08 3.9277e+10 27602

- Shirts 1 5.4517e+08 3.9636e+10 27617

- Night 1 6.9707e+08 3.9788e+10 27623

- Firewks 1 9.6412e+08 4.0055e+10 27634

- BobbleHd 1 1.0489e+09 4.0140e+10 27637

- Month 6 2.1451e+09 4.1236e+10 27671

- Opponent 29 6.3389e+09 4.5430e+10 27781

- DayofWeek 6 7.4970e+09 4.6588e+10 27868

- Team 29 1.0033e+11 1.3942e+11 29591

Step: AIC=27594.96

Attend ~ Team + Month + DayofWeek + Opponent + Night + BobbleHd +

Headgear + Shirts + Firewks

Df Sum of Sq RSS AIC

<none> 3.9106e+10 27595

- Headgear 1 1.8324e+08 3.9289e+10 27601

- Shirts 1 5.5516e+08 3.9661e+10 27616

- Night 1 6.9467e+08 3.9801e+10 27621

- Firewks 1 9.6516e+08 4.0071e+10 27632

- BobbleHd 1 1.0569e+09 4.0163e+10 27636

- Month 6 2.1364e+09 4.1243e+10 27669

- Opponent 29 6.3349e+09 4.5441e+10 27779

- DayofWeek 6 7.4855e+09 4.6592e+10 27866

- Team 29 1.0116e+11 1.4026e+11 29598

> confint(lower.lm.model)

2.5 % 97.5 %

(Intercept) 30485.1 31284.29

> Anova(lower.lm.model)

Anova Table (Type III tests)

Response: Attend

Sum Sq Df F value Pr(>F)

(Intercept) 2.3093e+12 1 22971 < 2.2e-16 \*\*\*

Residuals 2.4329e+11 2420

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Warning message:

In Anova.lm(lower.lm.model) :

the model contains only an intercept: Type III test substituted

> vif(lower.lm.model)

Error in vif.lm(lower.lm.model) : model contains fewer than 2 terms

> r.rmse1 <- sqrt(mean(lower.lm.model$residuals^2))

> print (r.rmse1)

[1] 10024.48 – Training

> dodgers.test$predAttend <- predict(lower.lm.model, newdata = dodgers.test)

> dodgers.test$residuals <- dodgers.test$Attend - dodgers.test$predAttend

> test.r.rmse1 <- sqrt(mean(dodgers.test$residuals^2))

> print (test.r.rmse1)

[1] 9936.314 – Testing

> test.r.rmse2 <- sqrt(mean(dodgers.test$residuals^2))

> print (test.r.rmse2)

[1] 9936.314

> # exploratory data analysis with standard graphics: attendance by Promotion

> with(data=dodgers,plot(Promotion, Attend/1000,

+ xlab = "Promotion", ylab = "Attendance (thousands)",

+ col = "violet", las = 1))

>

Call:

lm(formula = Attend ~ Team + Month + DayofWeek + Opponent + Night +

BobbleHd + Headgear + Shirts + Firewks, data = dodgers.train)

Residuals:

Min 1Q Median 3Q Max

-13724.1 -3218.9 -275.4 2766.0 26393.0

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 27082.26 1189.38 22.770 < 2e-16 \*\*\*

TeamAtlanta Braves 2048.00 1032.94 1.983 0.047579 \*

TeamBaltimore Orioles -833.39 1066.12 -0.782 0.434505

TeamBoston Red Sox 11963.43 1085.33 11.023 < 2e-16 \*\*\*

TeamChicago Cubs 8865.42 1002.96 8.839 < 2e-16 \*\*\*

TeamChicago White Sox -2715.07 1108.68 -2.449 0.014439 \*

TeamCincinnati Reds 2487.65 1017.42 2.445 0.014595 \*

TeamCleveland Indians -7014.42 1097.92 -6.389 2.21e-10 \*\*\*

TeamColorado Rockies 5947.02 1005.78 5.913 4.13e-09 \*\*\*

TeamDetroit Tigers 10278.31 1063.56 9.664 < 2e-16 \*\*\*

TeamHouston Astros -7681.34 1044.28 -7.356 3.08e-13 \*\*\*

TeamKansas City Royals -5708.66 1107.45 -5.155 2.87e-07 \*\*\*

TeamLos Angeles Angels 11373.59 1067.27 10.657 < 2e-16 \*\*\*

TeamLos Angeles Dodgers 13966.32 1036.06 13.480 < 2e-16 \*\*\*

TeamMiami Marlins 1110.27 1052.29 1.055 0.291548

TeamMilwaukee Brewers 8754.72 1046.17 8.368 < 2e-16 \*\*\*

TeamMinnesota Twins 8334.12 1071.58 7.777 1.35e-14 \*\*\*

TeamNew York Mets 1420.80 1006.87 1.411 0.158416

TeamNew York Yankees 17825.74 1095.36 16.274 < 2e-16 \*\*\*

TeamOakland Athletics -7804.82 1094.27 -7.132 1.51e-12 \*\*\*

TeamPhiladelphia Phillies 17991.31 1019.37 17.650 < 2e-16 \*\*\*

TeamPittsburgh Pirates -399.48 1051.92 -0.380 0.704174

TeamSan Diego Padres 800.78 1020.31 0.785 0.432671

TeamSan Francisco Giants 15851.81 1061.74 14.930 < 2e-16 \*\*\*

TeamSeattle Mariners -4871.90 1106.79 -4.402 1.15e-05 \*\*\*

TeamSt. Louis Cardinals 14042.13 1014.39 13.843 < 2e-16 \*\*\*

TeamTampa Bay Rays -8939.38 1094.09 -8.171 6.34e-16 \*\*\*

TeamTexas Rangers 16509.84 1054.83 15.652 < 2e-16 \*\*\*

TeamToronto Blue Jays -714.38 1078.67 -0.662 0.507894

TeamWashington Nationals 4480.16 1039.30 4.311 1.73e-05 \*\*\*

MonthAugust 580.57 453.97 1.279 0.201138

MonthJuly 2616.92 461.59 5.669 1.71e-08 \*\*\*

MonthJune 2920.17 473.42 6.168 8.81e-10 \*\*\*

MonthMay 473.77 453.04 1.046 0.295831

MonthOctober 2004.72 1097.76 1.826 0.068015 .

MonthSeptember 33.30 460.18 0.072 0.942314

DayofWeekMonday -4508.27 541.02 -8.333 < 2e-16 \*\*\*

DayofWeekSaturday 1866.25 494.35 3.775 0.000166 \*\*\*

DayofWeekSunday -1453.47 573.62 -2.534 0.011380 \*

DayofWeekThursday -3686.22 566.57 -6.506 1.04e-10 \*\*\*

DayofWeekTuesday -4298.47 506.47 -8.487 < 2e-16 \*\*\*

DayofWeekWednesday -4203.99 503.00 -8.358 < 2e-16 \*\*\*

OpponentAtlanta Braves 284.13 1000.25 0.284 0.776399

OpponentBaltimore Orioles 792.89 1048.85 0.756 0.449787

OpponentBoston Red Sox 4314.82 1038.10 4.156 3.41e-05 \*\*\*

OpponentChicago Cubs 1745.40 1049.76 1.663 0.096581 .

OpponentChicago White Sox -514.38 1062.57 -0.484 0.628389

OpponentCincinnati Reds -178.29 1011.97 -0.176 0.860171

OpponentCleveland Indians 179.28 1055.91 0.170 0.865201

OpponentColorado Rockies -1167.73 1001.03 -1.167 0.243581

OpponentDetroit Tigers 3303.65 1018.47 3.244 0.001205 \*\*

OpponentHouston Astros -789.49 989.27 -0.798 0.424962

OpponentKansas City Royals -1678.93 1053.87 -1.593 0.111341

OpponentLos Angeles Angels 2379.67 1046.47 2.274 0.023104 \*

OpponentLos Angeles Dodgers 1932.30 988.86 1.954 0.050874 .

OpponentMiami Marlins 260.00 999.01 0.260 0.794699

OpponentMilwaukee Brewers 106.88 1046.89 0.102 0.918700

OpponentMinnesota Twins -18.39 1042.59 -0.018 0.985927

OpponentNew York Mets 2382.58 982.03 2.426 0.015373 \*

OpponentNew York Yankees 8463.10 1068.12 7.923 4.40e-15 \*\*\*

OpponentOakland Athletics 782.05 1059.91 0.738 0.460723

OpponentPhiladelphia Phillies 2254.64 1013.20 2.225 0.026208 \*

OpponentPittsburgh Pirates -69.06 979.63 -0.070 0.943806

OpponentSan Diego Padres -6.52 991.73 -0.007 0.994755

OpponentSan Francisco Giants 2799.94 996.61 2.809 0.005025 \*\*

OpponentSeattle Mariners -1296.88 1035.23 -1.253 0.210492

OpponentSt. Louis Cardinals 2582.45 1043.97 2.474 0.013480 \*

OpponentTampa Bay Rays -1147.27 1022.11 -1.122 0.261845

OpponentTexas Rangers 1387.66 1081.23 1.283 0.199540

OpponentToronto Blue Jays 415.39 1066.54 0.389 0.696982

OpponentWashington Nationals 2226.39 977.15 2.278 0.022836 \*

Night -1877.87 359.27 -5.227 1.96e-07 \*\*\*

BobbleHdYES 4351.33 674.92 6.447 1.52e-10 \*\*\*

HeadgearYES 1935.24 720.88 2.685 0.007341 \*\*

ShirtsYES 2788.05 596.67 4.673 3.23e-06 \*\*\*

Firewks 3604.61 585.06 6.161 9.21e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5042 on 1538 degrees of freedom

Multiple R-squared: 0.761, Adjusted R-squared: 0.7493

F-statistic: 65.28 on 75 and 1538 DF, p-value: < 2.2e-16