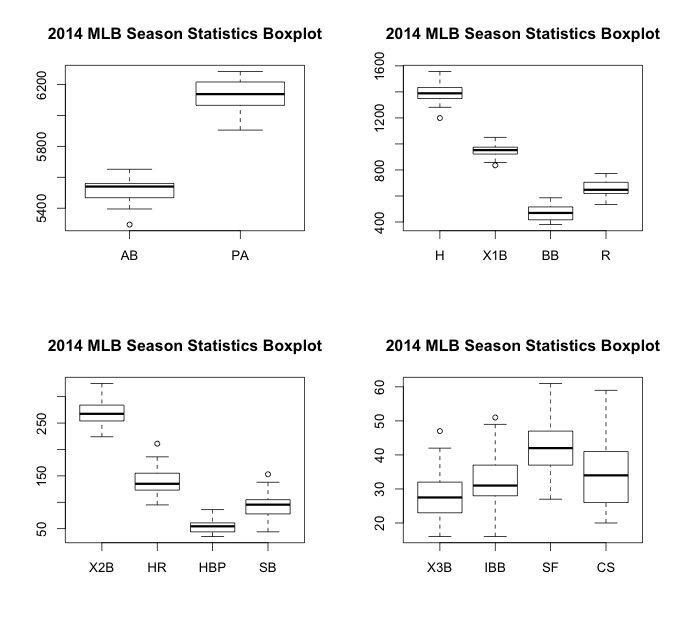
**Question 1**



**Figure 1. 2014 MLB Season Statistics Boxplot**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **AB** | **PA** | **H** | **X1B** | **X2B** | **X3B** | **HR** | **BB** | **IBB** | **HBP** | **SF** | **SB** | **CS** | **R** |
| **AB** | 1.0000 | 0.8027 | 0.8055 | 0.6426 | 0.6301 | 0.2197 | 0.4264 | 0.0869 | 0.3974 | -0.0883 | 0.3669 | -0.0981 | -0.2058 | 0.7816 |
| **PA** | 0.8027 | 1.0000 | 0.5728 | 0.4414 | 0.5646 | 0.0392 | 0.2568 | 0.6300 | 0.4110 | 0.0727 | 0.3844 | -0.1467 | -0.3052 | 0.6857 |
| **H** | 0.8055 | 0.5728 | 1.0000 | 0.8698 | 0.7713 | 0.2937 | 0.3888 | -0.1021 | 0.3786 | 0.0080 | 0.4201 | 0.0498 | 0.1334 | 0.8430 |
| **X1B** | 0.6426 | 0.4414 | 0.8698 | 1.0000 | 0.5280 | 0.2914 | -0.0246 | -0.1192 | 0.3568 | -0.0185 | 0.4023 | 0.1426 | 0.0918 | 0.5738 |
| **X2B** | 0.6301 | 0.5646 | 0.7713 | 0.5280 | 1.0000 | 0.0904 | 0.2382 | 0.0679 | 0.2807 | 0.1736 | 0.5016 | 0.1104 | 0.2693 | 0.7370 |
| **X3B** | 0.2197 | 0.0392 | 0.2937 | 0.2914 | 0.0904 | 1.0000 | -0.0936 | -0.2220 | 0.2467 | -0.3041 | 0.0616 | -0.0291 | 0.1699 | 0.2023 |
| **HR** | 0.4264 | 0.2568 | 0.3888 | -0.0246 | 0.2382 | -0.0936 | 1.0000 | -0.0635 | 0.0633 | -0.0170 | -0.0633 | -0.2364 | -0.0971 | 0.5782 |
| **BB** | 0.0869 | 0.6300 | -0.1021 | -0.1192 | 0.0679 | -0.2220 | -0.0635 | 1.0000 | 0.1264 | 0.0538 | 0.1663 | -0.0982 | -0.3549 | 0.1840 |
| **IBB** | 0.3974 | 0.4110 | 0.3786 | 0.3568 | 0.2807 | 0.2467 | 0.0633 | 0.1264 | 1.0000 | -0.0285 | 0.2002 | -0.2450 | 0.1061 | 0.3389 |
| **HBP** | -0.0883 | 0.0727 | 0.0080 | -0.0185 | 0.1736 | -0.3041 | -0.0170 | 0.0538 | -0.0285 | 1.0000 | -0.1002 | -0.0479 | 0.2573 | -0.0127 |
| **SF** | 0.3669 | 0.3844 | 0.4201 | 0.4023 | 0.5016 | 0.0616 | -0.0633 | 0.1663 | 0.2002 | -0.1002 | 1.0000 | -0.0380 | -0.0190 | 0.4140 |
| **SB** | -0.0981 | -0.1467 | 0.0498 | 0.1426 | 0.1104 | -0.0291 | -0.2364 | -0.0982 | -0.2450 | -0.0479 | -0.0380 | 1.0000 | 0.4831 | -0.0174 |
| **CS** | -0.2058 | -0.3052 | 0.1334 | 0.0918 | 0.2693 | 0.1699 | -0.0971 | -0.3549 | 0.1061 | 0.2573 | -0.0190 | 0.4831 | 1.0000 | 0.0126 |

**Table 1. Correlation between variables**

***Multiple Regression (All Factors)***

Call: lm(formula = R ~ 0 + X1B + X2B + X3B + HR + BB + IBB + HBP + SF + SB + CS)

Residuals:

Min 1Q Median 3Q Max

-37.283 -13.852 -3.274 12.384 54.429

Coefficients:

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

X1B 0.12536 0.09133 1.373 0.18504

X2B 0.93407 0.30223 3.091 0.00577 \*\*

X3B 1.26752 0.79461 1.595 0.12636

HR 1.06994 0.20952 5.107 5.38e-05 \*\*\*

BB 0.10827 0.08998 1.203 0.24290

IBB 0.63298 0.72558 0.872 0.39336

HBP -0.05058 0.51516 -0.098 0.92276

SF 0.83629 0.86767 0.964 0.34664

SB 0.25490 0.27550 0.925 0.36588

CS -0.74914 0.70436 -1.064 0.30019

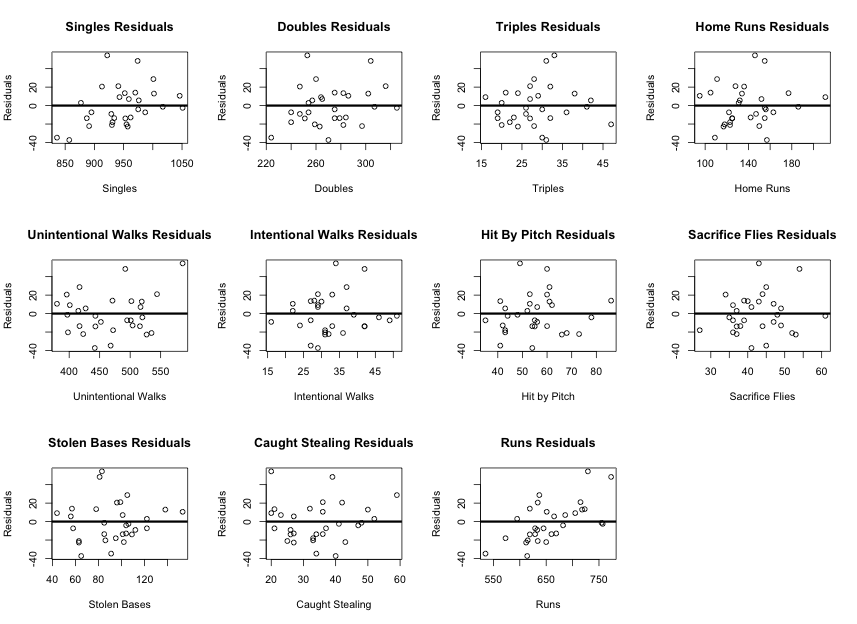
---

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

Residual standard error: 26.04 on 20 degrees of freedom

Multiple R-squared: 0.999, Adjusted R-squared: 0.9984

F-statistic: 1932 on 10 and 20 DF, p-value: < 2.2e-16



***Model Reduction***

Using the Bayesian Information Criterion method we were able to reduce the model. The initial model with all variables had BIC = 305.9548. The model was then reduced to the 9 combinations of 9 variables, and the BIC was again calculated. The combination with the lowest BIC value was determined to be best. This process repeated until reducing the model further resulted in only larger BIC values. The best-reduced model was determined to be:

*Runs = 0.16037\*Singles + 0.96242\*Doubles + 1.20950\*Triples + 1.00580\*HRs + 0.15345\*Walks*

Call:

lm(formula = R ~ 0 + X1B + X2B + X3B + HR + BB)

Residuals:

Min 1Q Median 3Q Max

-46.519 - 15.411 -2.059 13.278 60.966

Coefficients:

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

X1B 0.16037 0.07726 2.076 0.048366 \*

X2B 0.96242 0.24134 3.988 0.000511 \*\*\*

X3B 1.20950 0.68238 1.772 0.088503 .

HR 1.00580 0.18663 5.389 1.37e-05 \*\*\*

BB 0.15345 0.07834 1.959 0.061401 .

---

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

Residual standard error: 25.15 on 25 degrees of freedom

Multiple R-squared: 0.9988, Adjusted R-squared: 0.9986

F-statistic: 4140 on 5 and 25 DF, p-value: < 2.2e-16

***Model selection was confirmed with the automated AIC method.***

**PROBLEM 2**

***“Using Results from Part 1”???***

|  |  |  |
| --- | --- | --- |
| **Top RV during 2014 MLB Season** | | |
| **Rank** | **Name** | **RV** |
| 1 | Mike Trout | 226 |
| 2 | Miguel Cabrera | 210 |
| 3 | Jose Bautista | 204 |
| 4 | Adrian Gonzalez | 199 |
| 5 | Nelson Cruz | 195 |
| 6 | Anthony Rendon | 194 |
| 6 | Giancarlo Stanton | 194 |
| 6 | Albert Pujols | 194 |
| 9 | Ian Kinsler | 192 |
| 10 | Josh Donaldson | 191 |
| 10 | Michael Brantley | 191 |

|  |  |  |
| --- | --- | --- |
| **Top RR during 2014 MLB Season** | | |
| **Rank** | **Name** | **RR** |
| 1 | Mike Trout | 226 |
| 2 | Edwin Encarnacion | 210 |
| 3 | Miguel Cabrera | 204 |
| 4 | Giancarlo Stanton | 199 |
| 5 | Jose Bautista | 195 |
| 6 | Adrian Gonzalez | 194 |
| 7 | Jose Abreu | 194 |
| 8 | Victor Martinez | 194 |
| 9 | Yoenis Cespedes | 192 |
| 10 | Nelson Cruz | 191 |

**PROMBLEM 3**

*ANOVA: Effect of Position on RV*

Analysis of Variance Table

Response: RV

Df Sum Sq Mean Sq F value Pr(>F)

Position 8 19918 2489.78 2.8145 0.006406 \*\*

Residuals 137 121194 884.63

---

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

***CONTRASTS??***

**PROBLEM 4**