car prices

Chris Parrish

January 18, 2016

car prices

reference:
- Cannon, et al., Stat2, chapter 04, examples 4.7-4.8

Import the data.

```
data <- read.csv("ThreeCars.csv", header=TRUE)
head(data)
```

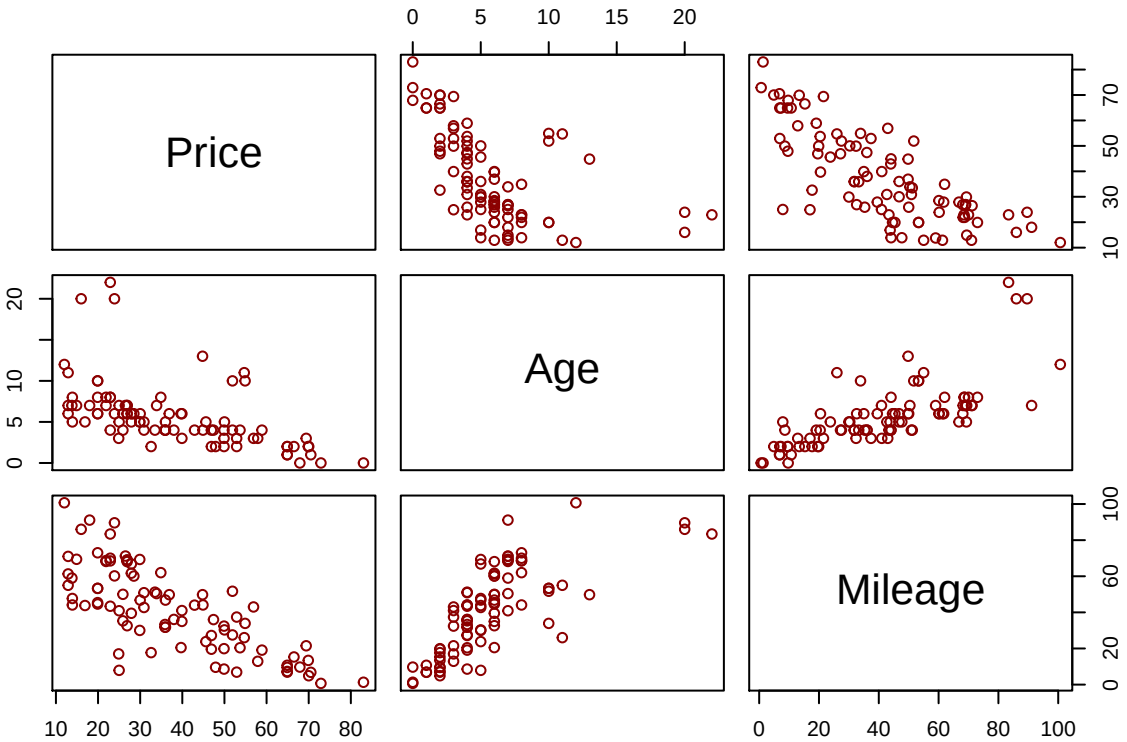
```
##   CarType Price Age Mileage Car Porsche Jaguar BMW
## 1 Porsche  69.4   3    21.5   0       1      0   0
## 2 Porsche  56.9   3    43.0   0       1      0   0
## 3 Porsche  49.9   2    19.9   0       1      0   0
## 4 Porsche  47.4   4    36.0   0       1      0   0
## 5 Porsche  42.9   4    44.0   0       1      0   0
## 6 Porsche  36.9   6    49.8   0       1      0   0
```

```
dim(data)
```

```
## [1] 90  8
```

Scatterplot matrix.

```
pairs(~ Price + Age + Mileage, data=data, col="darkred")
```



Average car prices

```
prices.lm1 <- lm(Price ~ Porsche + Jaguar, data=data)
```

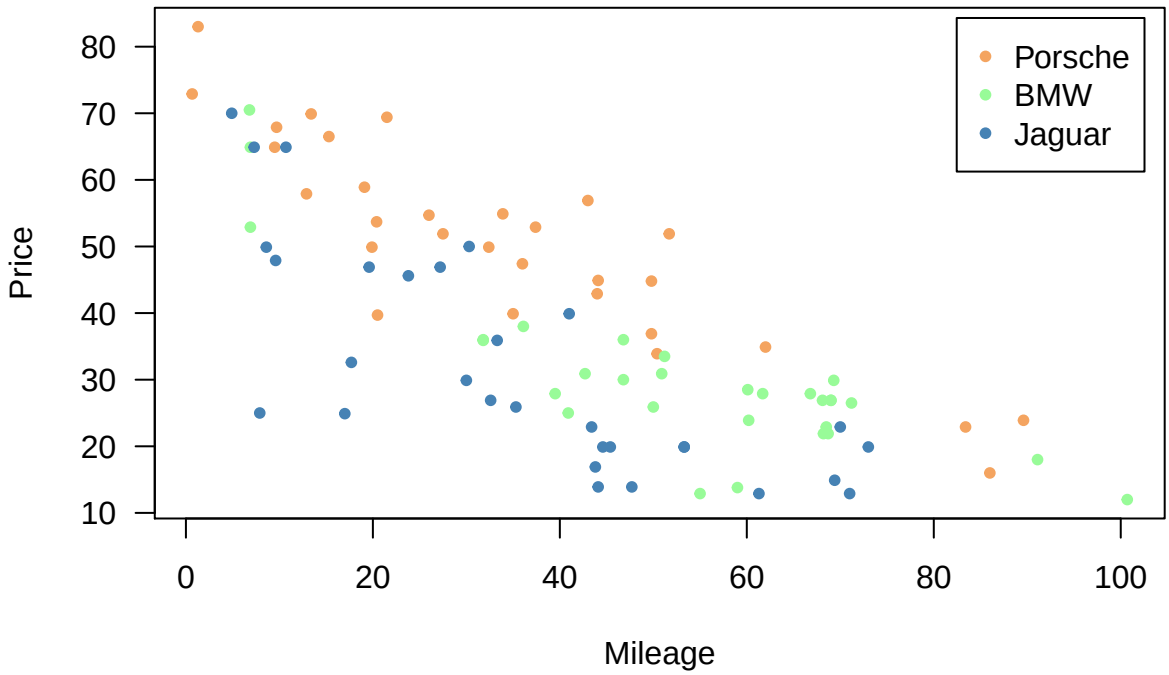
BMW: 30.233
Porsche: 50.537
Jaguar: 31.957

```
options(show.signif.stars=FALSE)
summary(prices.lm1)
```

```
##
## Call:
## lm(formula = Price ~ Porsche + Jaguar, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.537 -10.242  -2.333   7.113  40.267
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   30.233      2.775  10.896  < 2e-16
## Porsche       20.303      3.924   5.174 1.45e-06
## Jaguar         1.723      3.924   0.439    0.662
##
## Residual standard error: 15.2 on 87 degrees of freedom
## Multiple R-squared:  0.2745, Adjusted R-squared:  0.2579
## F-statistic: 16.46 on 2 and 87 DF,  p-value: 8.646e-07
```

Price vs. Mileage.

```
plot(Price ~ Mileage, data=data, las=1, type="n")
points(data[data$Porsche=="1", c(4, 2)], pch=20, col="sandybrown")
points(data[data$BMW=="1", c(4, 2)], pch=20, col="palegreen")
points(data[data$Jaguar=="1", c(4, 2)], pch=20, col="steelblue")
legend(x="topright", pch=20, inset=0.02,
       legend=c("Porsche", "BMW", "Jaguar"),
       col=c("sandybrown", "palegreen", "steelblue"))
```



Multiple regression: same slope, separate intercepts

$$Price \sim Mileage + Porsche + Jaguar$$

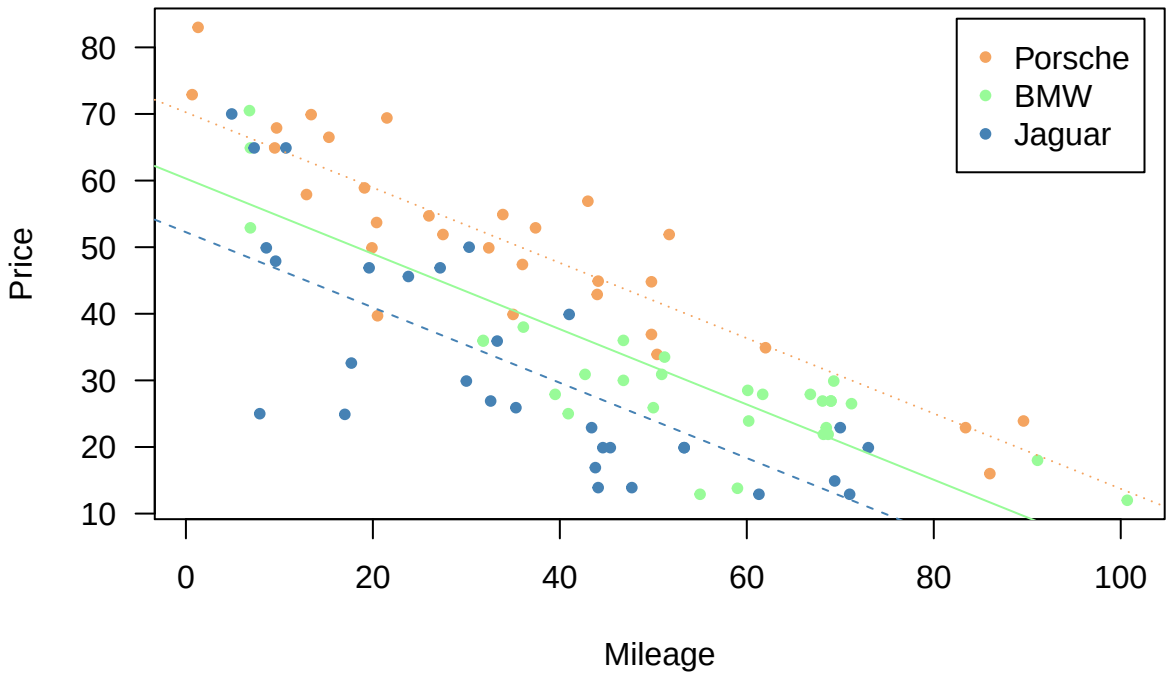
```
prices.lm2 <- lm(Price ~ Mileage + Porsche + Jaguar, data=data)
summary(prices.lm2)
```

```
##
## Call:
## lm(formula = Price ~ Mileage + Porsche + Jaguar, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.787  -5.361  -0.092   5.543  20.518
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 60.30070    2.71824  22.184  < 2e-16
## Mileage     -0.56528    0.04166 -13.570  < 2e-16
## Porsche      9.94868    2.35395   4.226 5.89e-05
## Jaguar      -8.04851    2.34038  -3.439 0.000903
##
## Residual standard error: 8.625 on 86 degrees of freedom
## Multiple R-squared:  0.7691, Adjusted R-squared:  0.761
## F-statistic: 95.46 on 3 and 86 DF,  p-value: < 2.2e-16
```

Illustration.

```
plot(Price ~ Mileage, data=data, las=1, type="n")
points(data[data$Porsche=="1", c(4, 2)], pch=20, col="sandybrown")
points(data[data$BMW=="1", c(4, 2)], pch=20, col="palegreen")
```

```
points(data[data$Jaguar=="1", c(4, 2)], pch=20, col="steelblue")
legend(x="topright", pch=20, inset=0.02,
       legend=c("Porsche", "BMW", "Jaguar"),
       col=c("sandybrown", "palegreen", "steelblue"))
abline(a=coef(prices.lm2)[1] + coef(prices.lm2)[3], b=coef(prices.lm2)[2],
       lty=3, col="sandybrown")
abline(a=coef(prices.lm2)[1], b=coef(prices.lm2)[2],
       lty=1, col="palegreen")
abline(a=coef(prices.lm2)[1] + coef(prices.lm2)[4], b=coef(prices.lm2)[2],
       lty=2, col="steelblue")
```



Multiple regression: separate slopes, separate intercepts

$$Price \sim Mileage + Porsche + Jaguar + Porsche \cdot Mileage + Jaguar \cdot Mileage$$

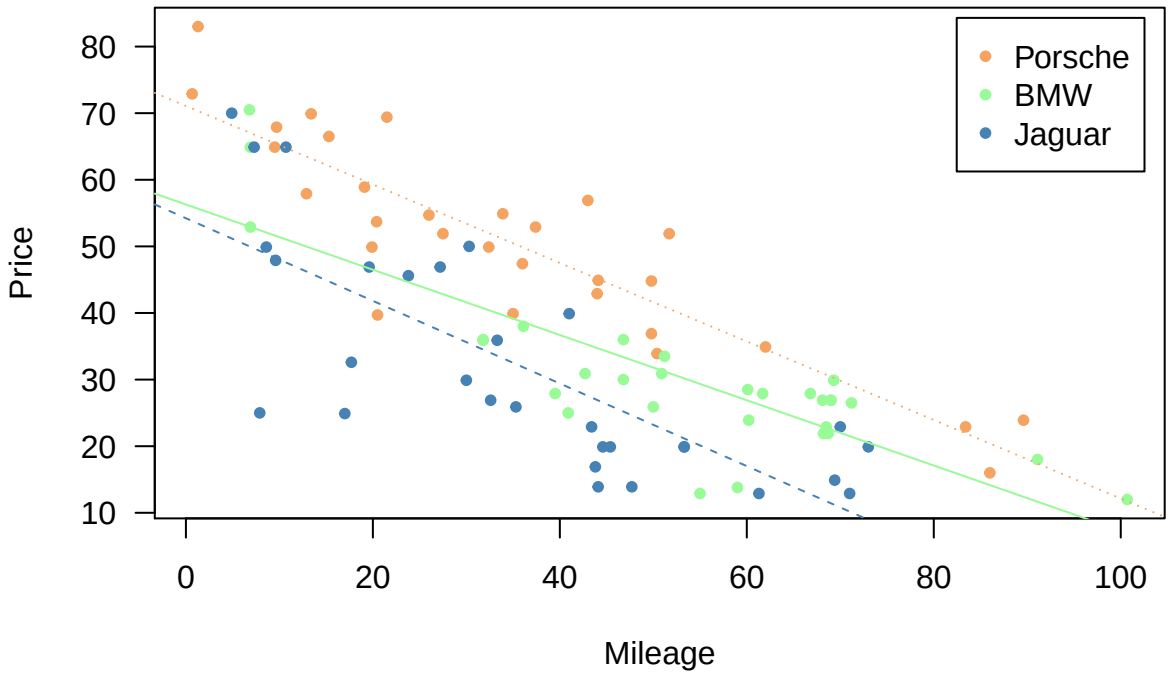
```
prices.lm3 <- lm(Price ~ Mileage + Porsche + Jaguar +
                         Porsche:Mileage + Jaguar:Mileage, data=data)
summary(prices.lm3)
```

```
##
## Call:
## lm(formula = Price ~ Mileage + Porsche + Jaguar + Porsche:Mileage +
##     Jaguar:Mileage, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -24.3271  -4.8317  -0.2852   4.4235  18.8120
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     56.29007    4.15512  13.547  < 2e-16
```

```
## Mileage          -0.48988    0.07227  -6.778 1.58e-09
## Porsche          14.80038    5.04149   2.936  0.00429
## Jaguar           -2.06261    5.23575  -0.394  0.69462
## Mileage:Porsche -0.09952    0.09940  -1.001  0.31962
## Mileage:Jaguar  -0.13042    0.10567  -1.234  0.22057
##
## Residual standard error: 8.638 on 84 degrees of freedom
## Multiple R-squared:  0.7737, Adjusted R-squared:  0.7602
## F-statistic: 57.43 on 5 and 84 DF,  p-value: < 2.2e-16
```

Illustration.

```
plot(Price ~ Mileage, data=data, las=1, type="n")
points(data[data$Porsche=="1", c(4, 2)], pch=20, col="sandybrown")
points(data[data$BMW=="1", c(4, 2)], pch=20, col="palegreen")
points(data[data$Jaguar=="1", c(4, 2)], pch=20, col="steelblue")
legend(x="topright", pch=20, inset=0.02,
       legend=c("Porsche", "BMW", "Jaguar"),
       col=c("sandybrown", "palegreen", "steelblue"))
abline(a=coef(prices.lm3)[1] + coef(prices.lm3)[3],
       b=coef(prices.lm3)[2] + coef(prices.lm3)[5],
       lty=3, col="sandybrown")
abline(a=coef(prices.lm3)[1],
       b=coef(prices.lm3)[2],
       lty=1, col="palegreen")
abline(a=coef(prices.lm3)[1] + coef(prices.lm3)[4],
       b=coef(prices.lm3)[2] + coef(prices.lm3)[6],
       lty=2, col="steelblue")
```



Nested (or incremental) F test.

$$Price \sim Mileage + Porsche + Jaguar$$

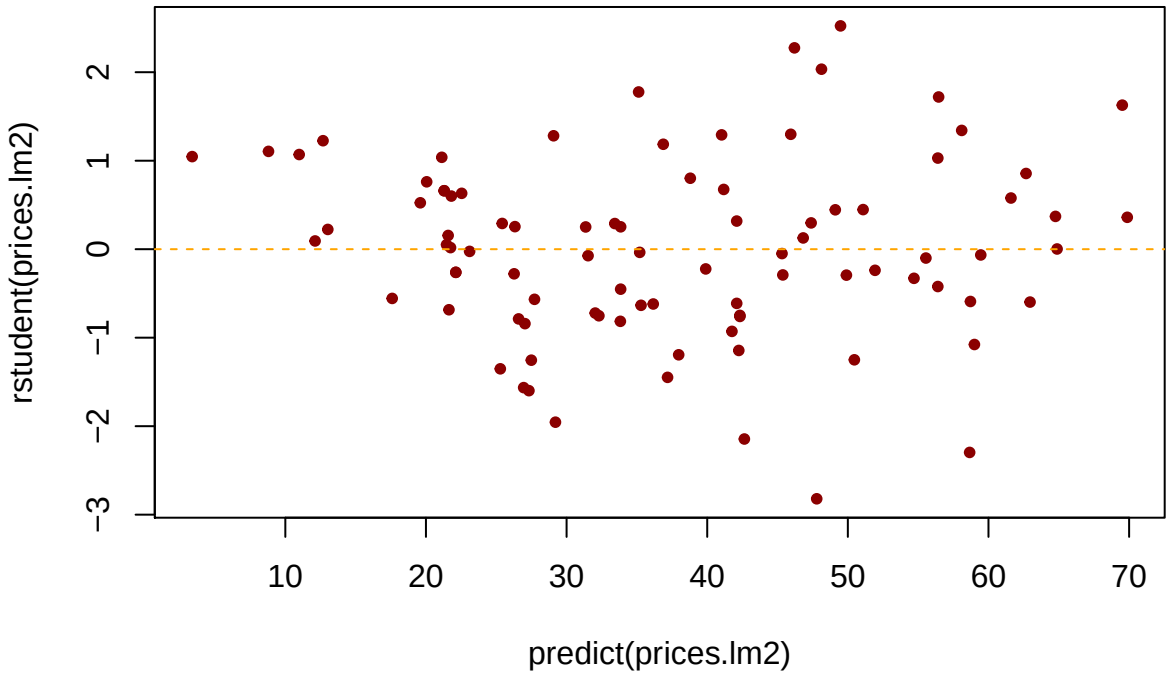
$$Price \sim Mileage + Porsche + Jaguar + Porsche \cdot Mileage + Jaguar \cdot Mileage$$

```
anova(prices.lm2, prices.lm3)
```

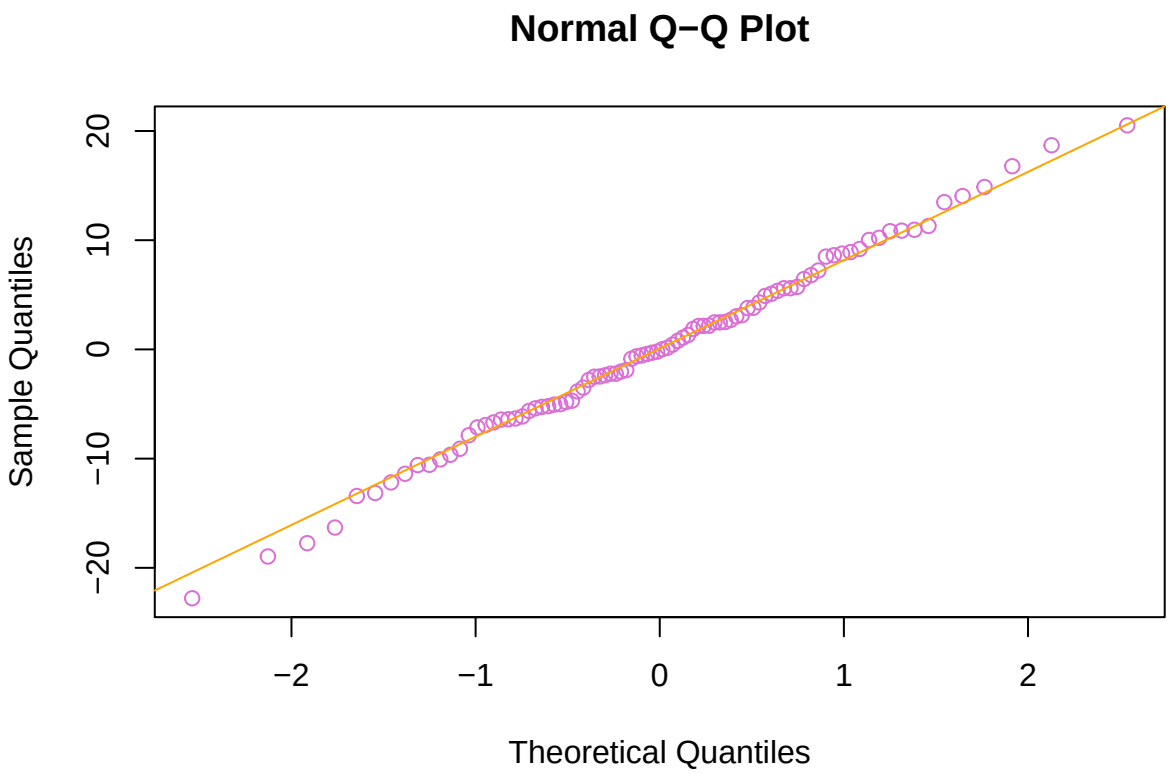
```
## Analysis of Variance Table
##
## Model 1: Price ~ Mileage + Porsche + Jaguar
## Model 2: Price ~ Mileage + Porsche + Jaguar + Porsche:Mileage + Jaguar:Mileage
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     86 6396.9
## 2     84 6268.3  2    128.55 0.8613 0.4263
```

Residuals.

```
plot(predict(prices.lm2), rstudent(prices.lm2),   #  studentized residuals
     pch=20, col="darkred")
abline(h=0, col="orange", lty="dashed")
```




```
qqnorm(resid(prices.lm2), col="orchid")           #  residuals
qqline(resid(prices.lm2), col="orange")
```



Prediction.

Exercise:
Write a function which will produce one row of the table at a time,
and then call it three times to produce the table.

```
new.data <- data.frame(Mileage=50, Porsche=1, Jaguar=0, BMW=0)
porsche.ci <- predict(prices.lm2, new.data, interval="confidence")
porsche.pi <- predict(prices.lm2, new.data, interval="prediction")
new.data <- data.frame(Mileage=50, Porsche=0, Jaguar=1, BMW=0)
jaguar.ci <- predict(prices.lm2, new.data, interval="confidence")
jaguar.pi <- predict(prices.lm2, new.data, interval="prediction")
new.data <- data.frame(Mileage=50, Porsche=0, Jaguar=0, BMW=1)
BMW.ci <- predict(prices.lm2, new.data, interval="confidence")
BMW.pi <- predict(prices.lm2, new.data, interval="prediction")
prices <- rbind(porsche=c(porsche.ci, porsche.pi[2:3]),
                jaguar=c(jaguar.ci, jaguar.pi[2:3]),
                BMW=c(BMW.ci, BMW.pi[2:3]))
rownames(prices) <- c("Porsche", "Jaguar", "BMW")
colnames(prices) <- c("fit", "CI:lwr", "CI:upr", "PI:lwr", "PI:upr")
prices
```

```
##              fit   CI:lwr   CI:upr    PI:lwr   PI:upr
## Porsche 41.98526 38.61367 45.35686 24.511942 59.45858
## Jaguar  23.98807 20.64726 27.32888  6.520662 41.45548
## BMW     32.03658 28.89523 35.17794 14.606226 49.46694
```