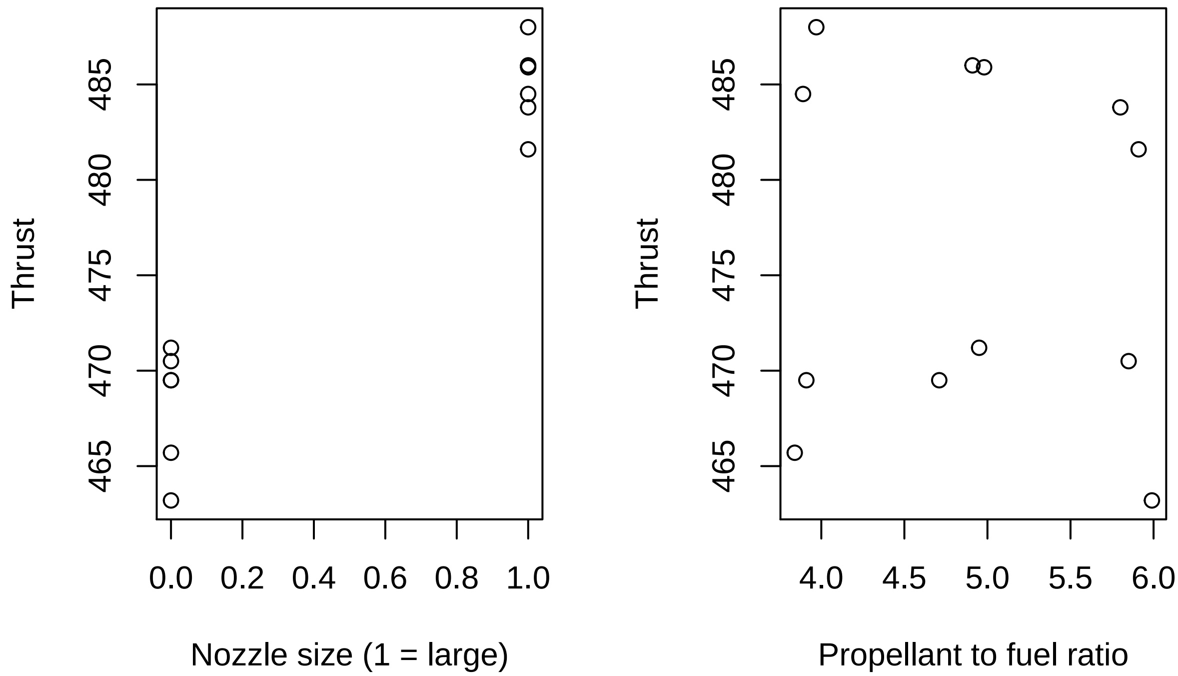
```r
## NASA rocket data example

## From: R.S. Jankovsky, T.D. Smith, A.J. Pavli (1999). "High-Area-Ratio Rocket
## Nozzle at High Combustion Chamber Pressure-Experimental and Analytical
## Validation".

# setwd(...) first if your CSV file is somewhere else
rocket <- read.csv("csv/rocket.csv")
# output all data in rocket vector
rocket
```

```
##    thrust nozzle propratio
## 1   488.0      1      3.97
## 2   481.6      1      5.91
## 3   485.9      1      4.98
## 4   486.0      1      4.91
## 5   484.5      1      3.89
## 6   483.8      1      5.80
## 7   463.2      0      5.99
## 8   471.2      0      4.95
## 9   469.5      0      3.91
## 10  470.5      0      5.85
## 11  469.5      0      4.71
## 12  465.7      0      3.84
```

Y (thrust) is the response variable, and there are two explanatory variables x_1, x_2 (nozzle, propratio) where nozzle is coded as 1 if it's large.

```r
# Scatter plots where mfrow is used to put multiple plots on one image
par(mfrow = c(1, 2))
plot(rocket$nozzle,
     rocket$thrust,
     ylab = "Thrust",
     xlab = "Nozzle size (1 = large)")
plot(rocket$propratio,
     rocket$thrust,
     ylab = "Thrust",
     xlab = "Propellant to fuel ratio")
```



Left is nozzle size vs thrust. Right is propellant relationship vs thrust.

```r
# Fit MLR using lm
m1 <- lm(thrust ~ nozzle + propratio, data = rocket)
summary(m1)
```

```
##
## Call:
## lm(formula = thrust ~ nozzle + propratio, data = rocket)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8459 -1.7555  0.5934  1.2906  3.3008
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 473.6039     4.7158 100.430 4.88e-15 ***
## nozzle       16.7383     1.5329  10.919 1.71e-06 ***
## propratio    -1.0948     0.9414  -1.163    0.275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.655 on 9 degrees of freedom
## Multiple R-squared:  0.9303, Adjusted R-squared:  0.9148
## F-statistic: 60.05 on 2 and 9 DF,  p-value: 6.238e-06
```

```r
m2 <- lm(thrust ~ 0 + nozzle, data = rocket)
summary(m2)
```

```
##
## Call:
## lm(formula = thrust ~ 0 + nozzle, data = rocket)
##
## Residuals:
```

```
##     Min      1Q Median     3Q    Max
##   -3.37   0.58 233.12 469.50 471.20
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## nozzle    485.0      141.2   3.435  0.00558 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 345.8 on 11 degrees of freedom
## Multiple R-squared:  0.5175, Adjusted R-squared:  0.4736
## F-statistic:  11.8 on 1 and 11 DF,  p-value: 0.005575
```

```r
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: thrust
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## nozzle     1 836.67  836.67 118.7377 1.743e-06 ***
## propratio  1   9.53    9.53   1.3524    0.2748
## Residuals  9  63.42    7.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On the left it's Y (response variable) and on the right it's x_1, x_2 (explanatory variables). From summary, we get the estimate vector $\hat{\beta} = (473.6039, 16.7383, -1.0948)^\top$.

```r
# Manual beta estimates where rep is used to make the columns of 1s
X <- cbind(rep(1, 12), rocket$nozzle, rocket$propratio) # X matrix
y <- matrix(rocket$thrust, ncol = 1) # response vector
beta_hat <- solve(t(X) %*% X) %*% t(X) %*% y
beta_hat
```

```
##             [,1]
## [1,] 473.603924
## [2,]  16.738319
## [3,]  -1.094822
```

`solve` is used for the inverse. `%*%` is used for matrix-matrix multiplication, and `t(X)` is used for transposing X.

```r
# Manual sigma estimate
mu_hat <- X %*% beta_hat # fitted values
e <- y - mu_hat # residuals
sigma_hat <- sqrt((t(e) %*% e) / 9) # Note n-p-1 = 12-2-1 = 9
sigma_hat
```

```
##        [,1]
## [1,] 2.6545
```

```r
sigma_hat <- sqrt(sum(e ^ 2) / 9) # equivalent
sigma_hat
```

```
## [1] 2.6545
```

- $\hat{\boldsymbol{\mu}} = X\hat{\boldsymbol{\beta}}$
- $\boldsymbol{e} = \boldsymbol{y} - \hat{\boldsymbol{\mu}}$

- $\hat{\sigma} = \sqrt{\left(\sum_{i=1}^{n} e_i^2\right)/9} = 2.6545$, or
- $\hat{\sigma} = \sqrt{(\boldsymbol{e}^\top \boldsymbol{e})/9} = 2.6545$

```
# Covariance matrix of beta_hat
vcov(m1)
```

```
##             (Intercept)      nozzle   propratio
## (Intercept)   22.238325 -1.02316688 -4.32080608
## nozzle        -1.023167  2.34987593 -0.03102117
## propratio     -4.320806 -0.03102117  0.88631920
```

```
sqrt(diag(vcov(m1))) # SEs of individual betas
```

```
## (Intercept)      nozzle   propratio
##   4.7157528   1.5329305   0.9414453
```

```
# Manual
se_beta <- sigma_hat * sqrt(diag(solve(t(X) %*% X)))
se_beta
```

```
## [1] 4.7157528 1.5329305 0.9414453
```

- $Se(\hat{\boldsymbol{\beta}}) = \hat{\sigma}\sqrt{(X^\top X)^{-1}} = (4.71, 1.53, 0.94)^\top$

```
# Estimate the mean response for units with small nozzle and propellant ratio 5.5
# include a 95% CI
predict(
  object = m1,
  newdata = data.frame(nozzle = 0, propratio = 5.5),
  interval = "confidence",
  level = 0.95
)
```

```
##        fit      lwr      upr
## 1 467.5824 464.7929 470.3719
```

Therefore, $\hat{y}_0 = 467.58$. The 95% confidence interval for the mean response given $\boldsymbol{x}_0$ is $[464.7929, 470.3719]$.

```
# Manual calculation
x0 <- matrix(c(1, 0, 5.5), nrow = 1)
y0_hat <- x0 %*% beta_hat
y0_hat
```

```
##          [,1]
## [1,] 467.5824
```

```
# mu0 is also known as \hat{Y}_0
se_mu0 <- sigma_hat * sqrt(x0 %*% solve(t(X) %*% X) %*% t(x0))
se_mu0
```

```
##          [,1]
## [1,] 1.233132
```

```
crit_val <- qt(0.975, 9)
ci_lo <- y0_hat - crit_val * se_mu0
ci_hi <- y0_hat + crit_val * se_mu0
c(y0_hat, ci_lo, ci_hi)
```

```
## [1] 467.5824 464.7929 470.3719
```

- $\boldsymbol{x}_0 = \begin{bmatrix} 1 & 0 & 5.5 \end{bmatrix}$
- $\hat{y}_0 = \boldsymbol{x}_0 \hat{\boldsymbol{\beta}} = 467.5824$
- $Se(\hat{Y}_0) = \hat{\sigma}\sqrt{\boldsymbol{x}_0 (X^\top X)^{-1} \boldsymbol{x}_0^\top} = 1.233132$

Therefore, $\hat{y}_0 = 467.58$. The 95% confidence interval for the mean response given $\boldsymbol{x}_0$ is $[464.7929, 470.3719]$.

```
# Predict the value of the response for a unit with small nozzle and propellant ratio 5.5
# include a 95% PI
predict(
  object = m1,
  newdata = data.frame(nozzle = 0, propratio = 5.5),
  interval = "prediction",
  level = 0.95
)
```

```
##        fit      lwr      upr
## 1 467.5824 460.9612 474.2036
```

Therefore, $y_0 = 467.5824$. The 95% prediction interval for the response (y_0) given $\boldsymbol{x}_0$ is $[460.9612474.2036]$.

```
# Manual calculation for an individual
x0 <- matrix(c(1, 0, 5.5), nrow = 1)
y0_hat <- x0 %*% beta_hat
se_y0 <- sigma_hat * sqrt(1 + x0 %*% solve(t(X) %*% X) %*% t(x0))
se_y0
```

```
##          [,1]
## [1,] 2.926941
```

```
crit_val <- qt(0.975, 9)
pi_lo <- y0_hat - crit_val * se_y0
pi_hi <- y0_hat + crit_val * se_y0
c(y0_hat, pi_lo, pi_hi)
```

```
## [1] 467.5824 460.9612 474.2036
```

- $Se(Y_0 - \hat{Y}_0) = \hat{\sigma}\sqrt{1 + \boldsymbol{x}_0 (X^\top X)^{-1} \boldsymbol{x}_0^\top} = 2.926941$