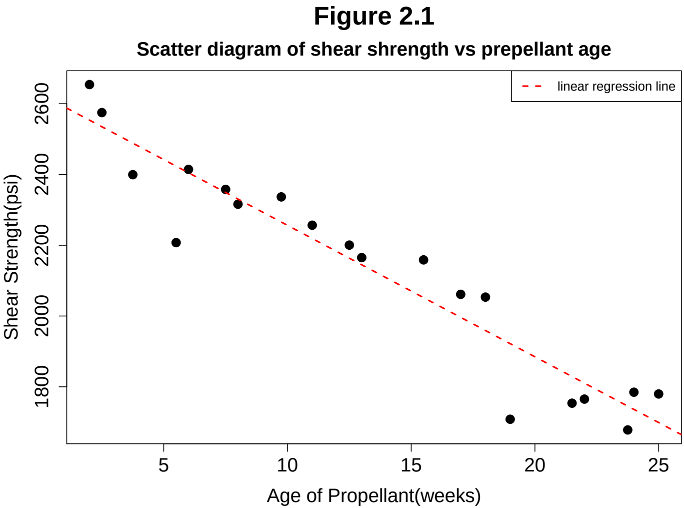
Chapter 2
Simple Linear Regression

Example 2.1: The Rocket Propellant Data

1. data & plot

Obs	Shear Strength(psi)	Age of Propellant(weeks)
1	2158.70	15.50
2	1678.15	23.75
3	2316.00	8.00
4	2061.30	17.00
5	2207.50	5.50
6	1708.30	19.00
7	1784.70	24.00
8	2575.00	2.50
⋮	⋮	⋮
17	2414.40	6.00
18	2200.50	12.50
19	2654.20	2.00
20	1753.70	21.50

Example 2.1: The Rocket Propellant Data (cont.)



Example 2.1: The Rocket Propellant Data (cont.)

2. model fit & summary

```
> # load data file
> library(RCurl)
> library(gdata)
> tf <- paste(tempfile(), "xls", sep = ".")
> http <- "https://raw.github.com/dongikjang/regression/master/"
> download.file(paste(http, "data-ex-2-1.xls", sep=""), tf,
+              method="curl")
  % Total    % Received % Xferd  Average Speed   Time    Time     Time
                                 Dload  Upload   Total   Spent    Left
  0     0    0     0    0     0      0      0 --:--:-- --:--:-- --:--:

> source(paste(http, "reg_r_source.txt", sep=""))
> # If OS is Windows then install "xlsReadWrite" package
> # If OS is Mac or Linux then install "gdata" package
> data_2.1 <- read.xls2(tf,  header=TRUE)
> View(data_2.1)
```

Example 2.1: The Rocket Propellant Data (cont.)

```
> # scatter plot
> par(mar=c(4.5,5,5,2),cex.main=2, cex.lab=1.5, cex.axis=1.5)
> plot(data_2.1[,3:2], xlab="Age of Propellant(weeks)",
+ ylab="Shear Strength(psi)", pch=19, cex=1.5)
> title(main="Figure 2.1", line=3)
> par(cex.main=1.5)
> title(main="Scatter diagram of shear shrength vs prepellant age",
+ line=1)
> abline(lm(data_2.1[,2:3]), col=2, lty=2, lwd=2)
> legend("topright", legend="linear regression line",
+ lty=2, col=2, lwd=2)
```

Example 2.1: The Rocket Propellant Data (cont.)

```
> # linear model fit
> colnames(data_2.1) <- c("obs", "yi", "xi")
> attach(data_2.1)
> lmfit <- lm(yi~xi)
> lmfit

Call:
lm(formula = yi ~ xi)

Coefficients:
(Intercept)           xi
    2627.82       -37.15

> names(lmfit)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "xlevels"       "call"          "terms"         "model"
```

Example 2.1: The Rocket Propellant Data (cont.)

```
> # values of linear  model object
> lmfit$coefficients
(Intercept)          xi
 2627.82236   -37.15359
> lmfit$residuals
          1           2           3           4           5  ...
 106.758301  -67.274574  -14.593631   65.088687 -215.977609  ...
...            19          20
...   100.684823  -75.320154
> lmfit$fitted.values
       1        2        3        4        5        6 ...
2051.942 1745.425 2330.594 1996.211 2423.478 1921.904 ...
...        19       20
... 2553.515 1829.020
```

Example 2.1: The Rocket Propellant Data (cont.)

```
> # summary values of linear  model object
> sfit <-summary(lmfit)
> sfit

Residuals:
    Min      1Q  Median      3Q     Max
-215.98  -50.68   28.74   66.61  106.76

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2627.822     44.184   59.48  < 2e-16 ***
xi           -37.154      2.889  -12.86 1.64e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 96.11 on 18 degrees of freedom
Multiple R-squared: 0.9018,       Adjusted R-squared: 0.8964
F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10
```

Example 2.1: The Rocket Propellant Data (cont.)

```
> names(sfit)
 [1] "call"          "terms"         "residuals"     "coefficients"
 [5] "aliased"       "sigma"         "df"            "r.squared"
 [9] "adj.r.squared" "fstatistic"    "cov.unscaled"
> sfit$coeff
              Estimate Std. Error   t value     Pr(>|t|)
(Intercept) 2627.82236  44.183912  59.47464 4.063559e-22
xi           -37.15359   2.889107 -12.85989 1.643344e-10
> sfit$sigma
[1] 96.1061
> sfit$df
[1]  2 18  2
> sfit$r.squared
[1] 0.9018414
> sfit$fstatistic
   value    numdf    dendf
165.3768   1.0000  18.0000
```

Example 2.1: The Rocket Propellant Data (cont.)

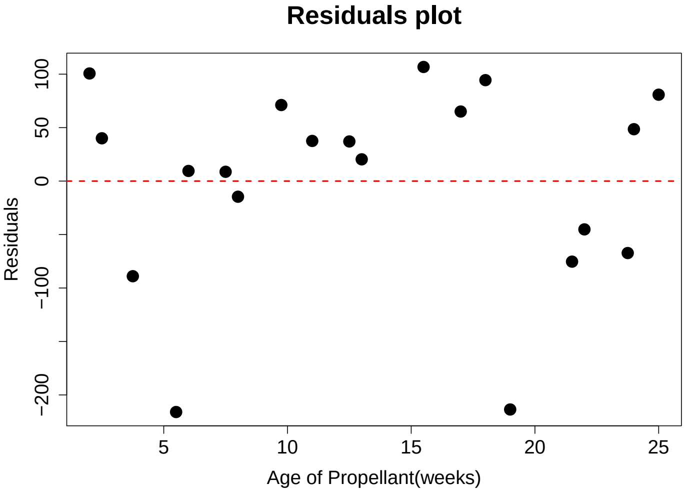
3. fitted values & residuals

```
> fit<-fitted(lmfit)
> res <-residuals(lmfit)
> ta2.2 <- cbind(xi,fit,res)
> ta2.2    #  table 2.2
      xi      fit         res
1  15.50 2051.942  106.758301
2  23.75 1745.425  -67.274574
3   8.00 2330.594  -14.593631
4  17.00 1996.211   65.088687
...
17  6.00 2404.901    9.499187
18 12.50 2163.402   37.097528
19  2.00 2553.515  100.684823
20 21.50 1829.020  -75.320154
```

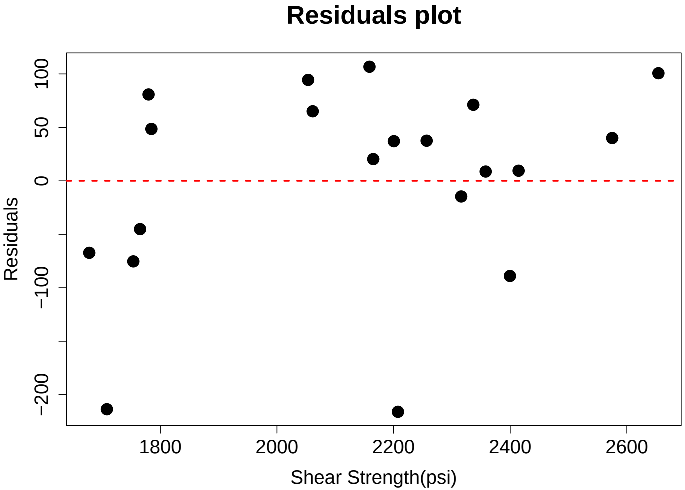
Example 2.1: The Rocket Propellant Data (cont.)

```
> par(mar=c(4.5,5,5,2),cex.main=2, cex.lab=1.5, cex.axis=1.5)
> plot(xi, res,  xlab="Age of Propellant(weeks)",
+ ylab="Residuals", main="Residuals plot", pch=19, cex=2)
> abline(h=0, col=2, lty=2, lwd=2)
> plot(yi, res,  xlab="Shear Strength(psi)",
+ ylab="Residuals", main="Residuals plot", pch=19, cex=2)
> abline(h=0, col=2, lty=2, lwd=2)
```

Example 2.1: The Rocket Propellant Data (cont.)



Example 2.1: The Rocket Propellant Data (cont.)



Example 2.1: The Rocket Propellant Data (cont.)

4. ANOVA table

```
> anova(lmfit)
Analysis of Variance Table

Response: yi
          Df  Sum Sq Mean Sq F value    Pr(>F)
xi         1 1527483 1527483  165.38 1.643e-10 ***
Residuals 18  166255    9236
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> names(anova(lmfit))
[1] "Df"      "Sum Sq"  "Mean Sq" "F value" "Pr(>F)"
```

Example 2.1: The Rocket Propellant Data (cont.)

5. Estimation of coefficient

```
> x <- model.matrix( ~ xi) #design matrix
> y <- yi
> x
   (Intercept)    xi
1            1 15.50
2            1 23.75
3            1  8.00
4            1 17.00
...
18           1 12.50
19           1  2.00
20           1 21.50
attr(,"assign")
[1] 0 1
```

Example 2.1: The Rocket Propellant Data (cont.)

```
> xtxi <- solve(t(x) %*% x)
> xtxi %*% t(x) %*% y
                  [,1]
(Intercept) 2627.82236
xi           -37.15359
> solve(crossprod(x,x),crossprod(x,y))
                  [,1]
(Intercept) 2627.82236
xi           -37.15359
> lmfit$coefficients
(Intercept)          xi
 2627.82236   -37.15359
> coefficients(lmfit)
(Intercept)          xi
 2627.82236   -37.15359
```

Example 2.1: The Rocket Propellant Data (cont.)

6. Estimation of σ

```
> sqrt(sum(lmfit$res^2)/df.residual(lmfit))
[1] 96.1061

> sqrt(deviance(lmfit)/df.residual(lmfit))
[1] 96.1061

> sfit$sigma
[1] 96.1061
```

Example 2.1: The Rocket Propellant Data (cont.)

7. Estimation of s.e.($\hat{\beta}$): $cov(\hat{\beta}) = (X^T X)^{-1}\sigma^2$

```
> xtxi <- sfit$cov.unscaled  #solve(t(x) %*% x)

> sqrt(diag(xtxi))*sfit$sigma
(Intercept)          xi
  44.183912    2.889107

> sfit$coef[,2]
(Intercept)          xi
  44.183912    2.889107
```

Example 2.1: The Rocket Propellant Data (cont.)

8. R-square & Adjusted R-square: $R^2 = 1 - \frac{SSE}{SST}, R^2_{adj} = 1 - \frac{SSE/(n-2)}{SST/(n-1)}$

```
># r-square
> 1-deviance(lmfit)/sum((y-mean(y))^2)  # deviance is
[1] 0.9018414                           # Sum of Square Residual

> sfit$r.squared
[1] 0.9018414

> #adjust r-square
> num <- (deviance(lmfit)/df.residual(lmfit))
> den <-(sum((y-mean(y))^2)/(length(y)-1))
> 1- num/den
[1] 0.8963882

> sfit$adj.r.squared
[1] 0.8963882
```

Example 2.1: The Rocket Propellant Data (cont.)

9. 95% confidence interval of β: $\hat{\beta} \pm t_{n-k-1,0.025} s.e.(\hat{\beta})$

```
> t.025 <- qt(0.975, df.residual(lmfit)) #quantile of t-dstribution
> l.beta1 <- coefficients(lmfit)[2]-t.025*sfit$coef[2,2]
> u.beta1 <- coefficients(lmfit)[2]+t.025*sfit$coef[2,2]
> c(l.beta1, u.beta1)
       xi        xi
-43.22338 -31.08380
```

Example 2.1: The Rocket Propellant Data (cont.)

10. 95% confidence interval of σ:

$$\left[(n-2)\frac{MS_E}{\chi^2_{0.025,n-2}},(n-2)\frac{MS_E}{\chi^2_{0.975,n-2}}\right]$$

```
> sigma2 <- anova(lmfit)[2,3]
> q.025 <- qchisq(.975, df.residual(lmfit))
> q.975 <- qchisq(.025, df.residual(lmfit))
> l.sigma2 <- df.residual(lmfit)*sigma2/q.025
> u.sigma2 <- df.residual(lmfit)*sigma2/q.975
> c(l.sigma2,u.sigma2)
[1]  5273.516 20199.245
```

Example 2.1: The Rocket Propellant Data (cont.)

11. Interval estimation of mean response

$$\hat{\mu}_{y|x_0} \pm t_{0.025,n-2}\sqrt{MS_E\ x_0(X^TX)^{-1}x_0^T}$$

```
> x0 <- 13.3625
> x0 <- model.matrix( ~x0)
> t.025 <- qt(0.975,df.residual(lmfit))
> err <- t.025*sfit$sigma*sqrt(x0%*%xtxi%*%t(x0))

> c(coefficients(lmfit)%*%t(x0)-err,
+   coefficients(lmfit)%*%t(x0)+err)
[1] 2086.209 2176.506
```

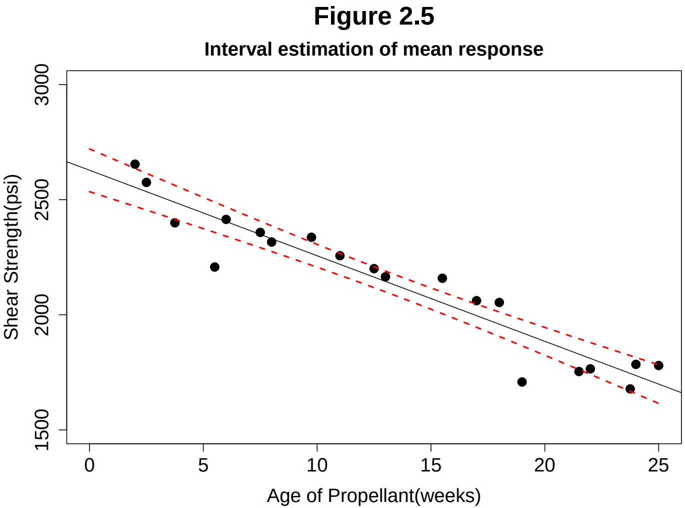
Example 2.1: The Rocket Propellant Data (cont.)

```
> predict(lmfit, data.frame(xi=13.3625), interval="confidence")
       fit      lwr      upr
1 2131.357 2086.209 2176.506
> x1 <- seq(0, 25, length=100)
> interval.est <- predict(lmfit, data.frame(xi=x1),
+ interval="confidence")
> head(interval.est)
       fit      lwr      upr
1 2627.822 2534.995 2720.649
2 2618.440 2526.949 2709.931
3 2609.058 2518.897 2699.219
4 2599.676 2510.838 2688.513
5 2590.293 2502.773 2677.814
6 2580.911 2494.700 2667.123
```

Example 2.1: The Rocket Propellant Data (cont.)

```
> # plot
> par(mar=c(4.5,5,5,2),cex.main=2, cex.lab=1.5, cex.axis=1.5)
> plot(xi, yi, pch=19, xlim=c(0,25), ylim=c(1500,3000),
+ xlab="Age of Propellant(weeks)", ylab="Shear Strength(psi)",
+ cex=1.5)
> title(main="Figure 2.5", line=3)
> par(cex.main=1.5)
> title(main="Interval estimation of mean response", line=1)
> abline(lmfit)
> lines(x1,interval.est[,2],lty=2,col=2,lwd=2)
> lines(x1,interval.est[,3],lty=2,col=2,lwd=2)
```

Example 2.1: The Rocket Propellant Data (cont.)



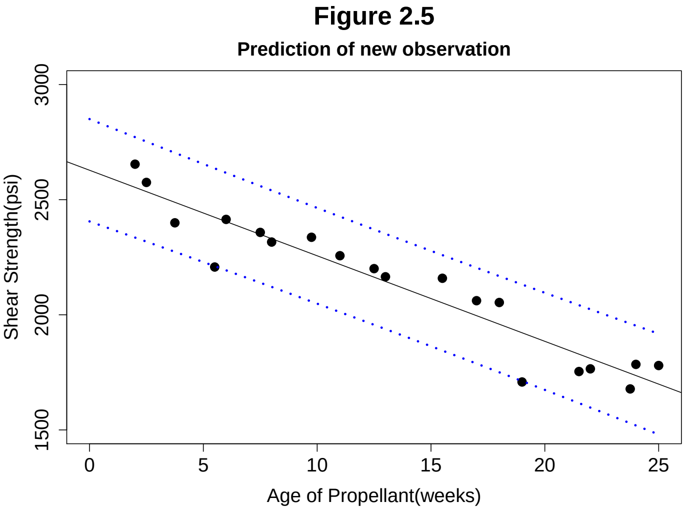
Example 2.1: The Rocket Propellant Data (cont.)

12. Prediction of new observation

$$\hat{\mu}_{y|x_0} \pm t_{0.025,n-2}\sqrt{MS_E\ \{1 + x_0(X^TX)x_0^T\}}$$

```
> err1 <- t.025*sfit$sigma*sqrt(1+x0%*%xtxi%*%t(x0))
> c(coefficients(lmfit)%*%t(x0)-err1,
+ coefficients(lmfit)%*%t(x0)+err1)
[1] 1924.460 2338.255
> predict(lmfit, data.frame(xi=13.3625), interval="prediction")
       fit     lwr      upr
1 2131.357 1924.46 2338.255
```

Example 2.1: The Rocket Propellant Data (cont.)



Example 2.1: The Rocket Propellant Data (cont.)

```
> x1<-seq(0, 25, length=100)
> pred <- predict(lmfit, data.frame(xi=x1), interval="prediction")
> par(mar=c(4.5,5,5,2), cex.main=2, cex.lab=1.5, cex.axis=1.5)
> plot(xi, yi, pch=19, xlim=c(0,25), ylim=c(1500,3000),
+ xlab="Age of Propellant(weeks)", ylab="Shear Strength(psi)",
+ cex=1.5)
> title(main="Figure 2.5", line=3)
> par(cex.main=1.5)
> title(main="Interval va Prediction", line=1)
> abline(lmfit)
> lines(x1, interval.est[,2], lty=2, col=2, lwd=3)
> lines(x1, interval.est[,3], lty=2, col=2, lwd=3)
> lines(x1, pred[,2], lty=3, col=4, lwd=3)
> lines(x1, pred[,3], lty=3, col=4, lwd=3)
> legend("bottomleft", legend=c("interval","prediction"),
+ lty=c(2,3), col=c(2,4), lwd=3, cex=1.5)
```

Example 2.1: The Rocket Propellant Data (cont.)