Simple Linear Regression Model – Part II

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 1 – Lecture 1E

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

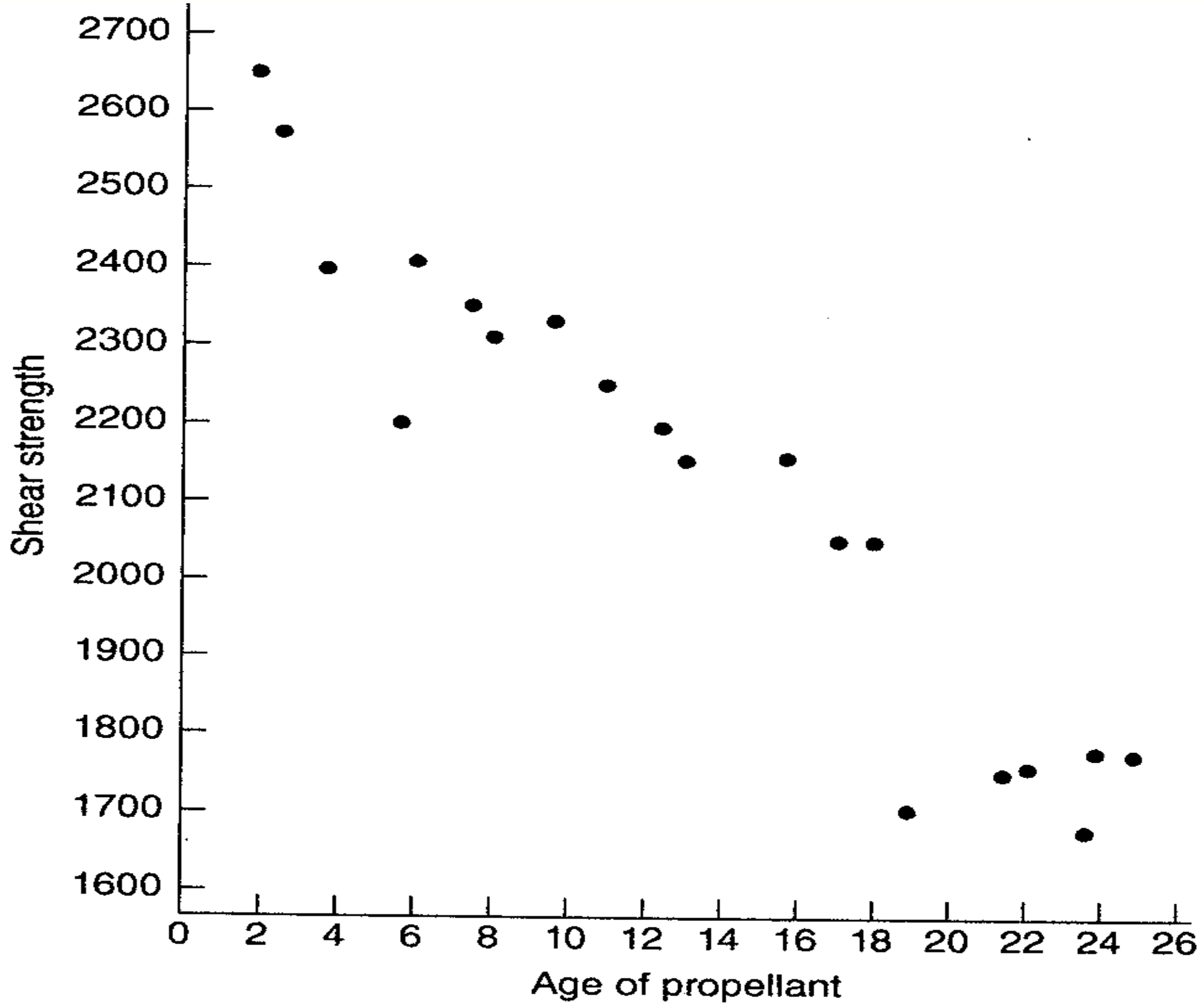
Example 2.1 The Rocket Propellant Data

Need to bond an igniter propellant and a sustainer propellant together in manufacturing a rocket motor.

Suspected that shear strength (in psi) of the bond is related to the age (in weeks) of the batch of sustainer propellant 20 observations are in Table 2.1.

Reference: page 15 of Textbook

Example 2.1 The Rocket Propellant Data – Scatter Diagram



Example 2.1 The Rocket Propellant Data - Estimation

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 4677.69 - \frac{71,422.56}{20} = 1106.56$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} = 528,492.64 - \frac{(267.25)(42,627.15)}{20} = -41,112.65$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-41,112.65}{1106.56} = -37.15$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2131.3575 - (-37.15)13.3625 = 2627.82$$

Example 2.1 The Rocket Propellant Data - Estimation

$$SS_{\mathrm{T}} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$= 92,547,433.45 - \frac{(42,627.15)^2}{20} = 1,693,737.60$$

$$SS_{\mathrm{Res}} = SS_{\mathrm{T}} - \hat{\beta}_1 S_{xy}$$

$$= 1,693,737.60 - (-37.15)(-41,112.65) = 166,402.65$$

$$\hat{\sigma}^2 = \frac{SS_{\mathrm{Res}}}{n-2} = \frac{166,402.65}{18} = 9244.59$$

Example 2.1 The Rocket Propellant Data - Check Model Fit

TABLE 2.2 Data, Fitted Values, and Residuals for Example 2.1

Observed Value, y_i	Fitted Value, $\hat{y}_i$	Residual, e_i
2158.70	2051.94	106.76
1678.15	1745.42	−67.27
2316.00	2330.59	−14.59
2061.30	1996.21	65.09
2207.50	2423.48	−215.98
1708.30	1921.90	−213.60
1784.70	1736.14	48.56
2575.00	2534.94	40.06
2357.90	2349.17	8.73
2256.70	2219.13	37.57
2165.20	2144.83	20.37
2399.55	2488.50	−88.95
1799.80	1698.98	80.82
2336.75	2265.58	71.17
1765.30	1810.44	−45.14
2053.50	1959.06	94.44
2414.40	2404.90	9.50
2200.50	2163.40	37.10
2654.20	2553.52	100.68
1753.70	1829.02	−75.32
$\sum y_i = 42{,}627.15$	$\sum \hat{y}_i = 42{,}627.15$	$\sum e_i = 0.00$

Variances of OLS Estimators

$$\operatorname{Var}\left(\hat{\beta}_1\right) = \operatorname{Var}\left(\sum_{i=1}^{n} c_i y_i\right) = \sum_{i=1}^{n} c_i^2 \operatorname{Var}(y_i)$$

$$= \frac{\sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2}{S_{xx}^2} = \frac{\sigma^2}{S_{xx}}$$

Variances of OLS Estimators

$$\mathrm{Var}\left(\hat{\beta}_0\right) = \mathrm{Var}\left(\bar{y} - \hat{\beta}_1\bar{x}\right)$$

$$= \mathrm{Var}(\bar{y}) + \bar{x}^2\mathrm{Var}\left(\hat{\beta}_1\right) - 2\bar{x}\mathrm{Cov}\left(\bar{y}, \hat{\beta}_1\right)$$

$$\mathrm{Var}(\bar{y}) = \sigma^2/n$$

$$\mathrm{Cov}\left(\bar{y}, \hat{\beta}_1\right) = 0$$

$$\mathrm{Var}\left(\hat{\beta}_0\right) = \mathrm{Var}(\bar{y}) + \bar{x}^2\mathrm{Var}\left(\hat{\beta}_1\right) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$$

Estimation of Variances of OLS Estimators

To estimate the variances of the least-squares estimators, just replace σ^2 with the unbiased estimator $\hat{\sigma}^2$ in the variance formulas.

Some Important Facts:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i) = \sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \qquad \sum_{i=1}^{n} x_i e_i = 0 \qquad \sum_{i=1}^{n} \hat{y}_i e_i = 0$$

The least-squares regression line always passes through the centroid – the point $(\bar{y}, \bar{x})$

Gauss-Markov Theorem

For the linear regression with $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$, and uncorrelated errors, the least-squares estimators are unbiased and has the minimum variance when compared to any estimator that is a linear combination of y_i .

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING