9 Inference for curve and surface fitting

Previously, we have discussed how to describe relationships between variables (Ch. 4). We now move into formal inference for these relationships starting with relationships between two variables and moving on to more.

9.1 Simple linear regression

Recall, in Ch. 4, we wanted an equation to describe how a dependent (response) variable, y, changes in response to a change in one or more independent (experimental) variable(s), x.

We used the notation

$$y = \beta_0 + \beta_1 x + \epsilon$$

where β_0 is the intercept.

β_1 is the slope.

ϵ is some error. In fact,

Goal: We want to use inference to get interval estimates for our slope and predicted values and significance tests that the slope is not equal to zero.

9.1.1 Variance estimation

What are the parameters in our model, and how do we estimate them?

We need an estimate for σ^2 in a regression, or "line-fitting" context.

Definition 9.1. For a set of data pairs $(x_1, y_1), \ldots, (x_n, y_n)$ where least squares fitting of a line produces fitted values $\hat{y}_i = b_0 + b_1 x_i$ and residuals $e_i = y_i - \hat{y}_i$,

$$s_{LF}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2$$

is the *line-fitting sample variance*. Associated with it are $\nu = n - 2$ degrees of freedom and an estimated standard deviation of response $s_{LF} = \sqrt{s_{LF}^2}$.

s_{LF}^2 estimates the level of basic background variation σ^2, whenever the model is an adequate description of the data.

9.1.2 Inference for parameters

We are often interested in testing if $\beta_1 = 0$. This tests whether or not there is a *significant linear relationship* between x and y. We can do this using

1.

2.

Both of these require

It can be shown that since $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ and $\epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, then

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum(x - \bar{x})^2}\right)$$

So, a $(1 - \alpha)100\%$ CI for β_1 is

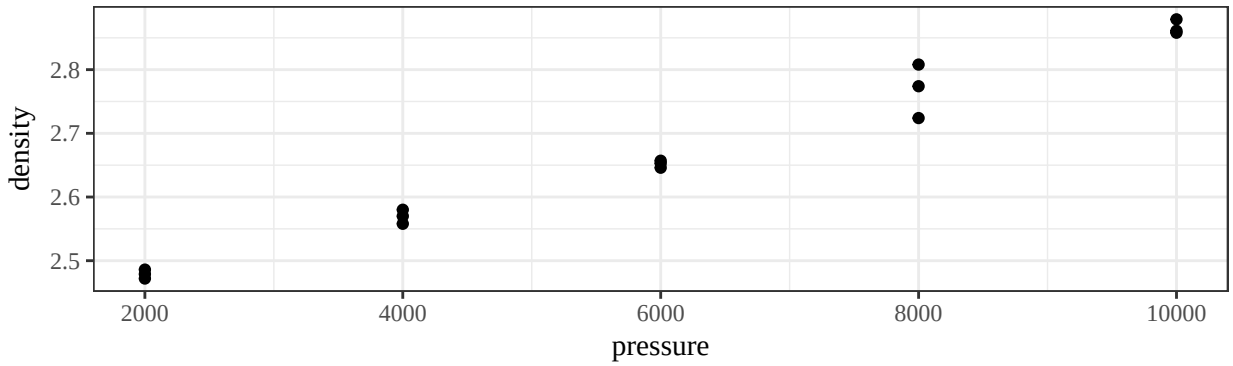
and the test statistic for $\text{H}_0 : \beta_1 = \#$ is

Example 9.1 (Ceramic powder pressing)**.** A mixture of Al_2O_3, polyvinyl alcohol, and water was prepared, dried overnight, crushed, and sieved to obtain 100 mesh size grains. These were pressed into cylinders at pressures from 2,000 psi to 10,000 psi, and cylinder densities were calculated. Consider a pressure/density study of $n = 15$ data pairs representing

$$x = \text{ the pressure setting used (psi)}$$
$$y = \text{ the density obtained (g/cc)}$$

in the dry pressing of a ceramic compound into cylinders.

pressure	density p	ressure d	ensity
2000	2.486	6000	2.653
2000	2.479	8000	2.724
2000	2.472	8000	2.774
4000	2.558	8000	2.808
4000	2.570	10000	2.861
4000	2.580	10000	2.879
6000	2.646	10000	2.858
6000	2.657		



A line has been fit in JMP using the method of least squares.

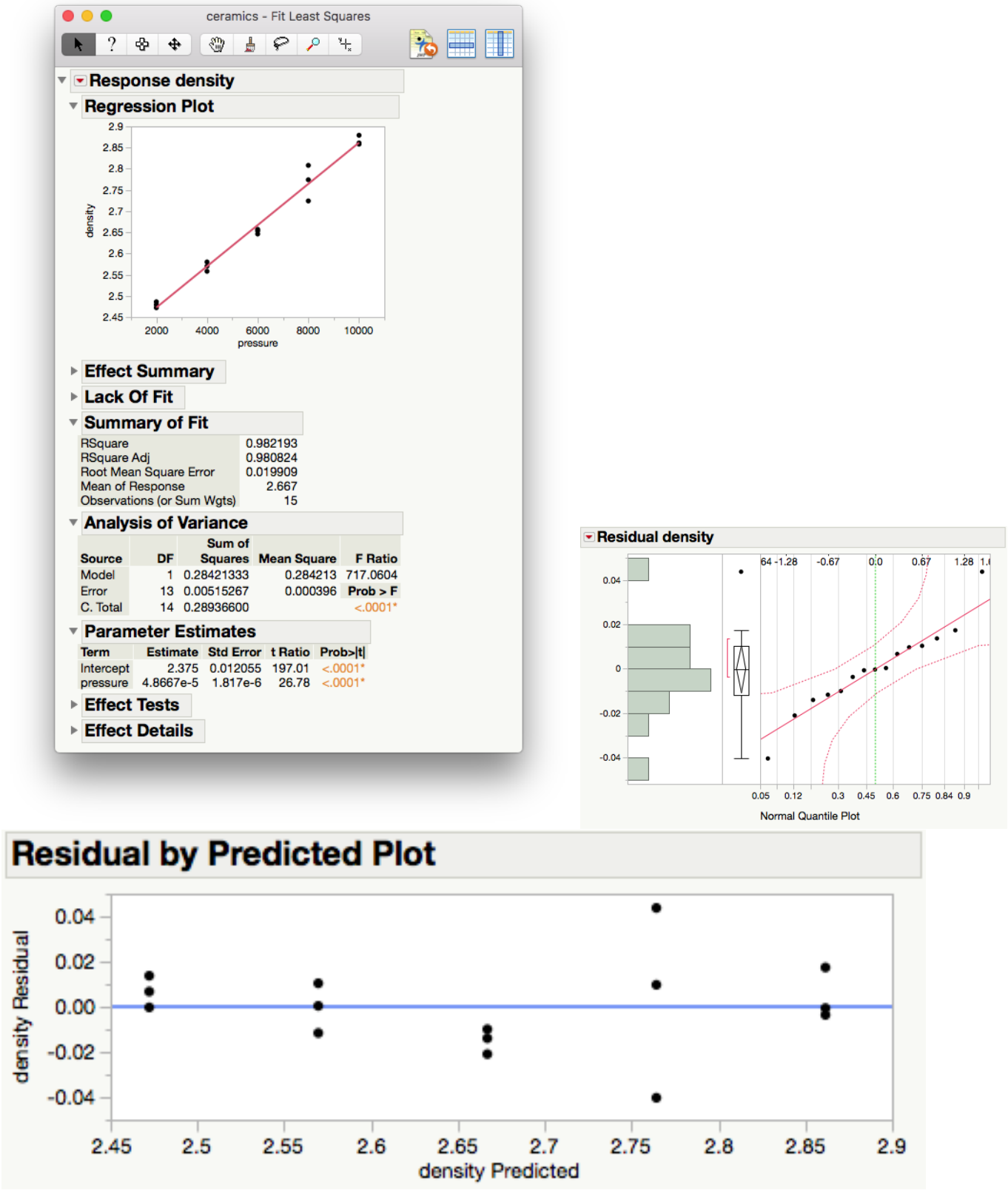


Figure 1: Least squares regression of density on pressure of ceramic cylinders.

1. Write out the model with the appropriate estimates.

2. Are the assumptions for the model met?

3. What is the fraction of raw variation in y accounted for by the fitted equation?

4. What is the correlation between x and y?

5. Estimate σ^2.

6. Estimate $\text{Var}(b_1)$.

7. Calculate and interpret the 95% CI for β_1

8. Conduct a formal hypothesis test at the $\alpha = .05$ significance level to determine if the relationship between density and pressure is significant.

9.1.3 Inference for mean response

Recall our model

$$y_1 = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

Under the model, the true mean response at some observed covariate value x_i is

Now, if some new covariate value x is within the range of the x_i's, we can estimate the true mean response at this new x

But how good is the estimate?

Under the model,

So we can construct a $N(0,1)$ random variable by standardizing.

And when σ is unknown (i.e. basically always),

To test $H_0 : \mu_{y|x} = \#$, we can use the test statistics

$$K =$$

which has a t_{n-2} distribution if H_0 is true and the model is correct.

A 2-sided $(1-\alpha)100\%$ CI for $\mu_{y|x}$ is

Example 9.2 (Ceramic powder pressing)**.** Return to the ceramic density problem. We will make a 2-sided 95% confidence interval for the true mean density of ceramics at 4000 psi and interpret it.

Now calculate and interpret a 2-sided 95% confidence interval for the true mean density at 5000 psi.

9.2 Multiple regression

9.2.1 Variance estimation

9.2.2 Standardized residuals

9.2.3 Inference for parameters

9.2.4 Inference for mean response