

# STAT 305: Chapter 9

## Inference for curve and surface fitting

Amin Shirazi

[ashirazist.github.io/stat305.github.io](http://ashirazist.github.io/stat305.github.io)

# Chapter 9:

## Inference for curve and surface fitting

## Simple Linear regression

## 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.

### 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

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

$\epsilon \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$

# Simple Linear Regression

where  $\beta_0$  is the intercept.

| It is the expected value for  $y$  when  $x = 0$ .

→  $\beta_1$  is the slope.

| It is the expected increase (decrease) in  $y$  for every one unit change in  $x$

$\epsilon$  is some error. In fact,

$$\epsilon \sim \text{iid } N(0, \sigma^2)$$

Recall:

→ Cheking if residuals are normally distributed is one of our model assessment technique.

**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.

Note, in SLR  $\epsilon \sim N(0, \sigma^2)$ , if  $\beta_1 = 0$ ,  $y = \beta_0 + \epsilon$ . i.e

there's no relationship between  $x, y$ . → So it's important to test if  $\beta_1 = 0$ .

# Variance Estimation

# Simple Linear Regression

## Variance Estimation

### Variance estimation

In the simple linear regression  $y = \beta_0 + \beta_1 x + \epsilon$ , the parameters are  $\beta_0$ ,  $\beta_1$  and  $\sigma^2$ .

$$\epsilon \sim N(0, \sigma^2)$$

unknown parameter.

We already know how to estimate  $\beta_0$  and  $\beta_1$  using least squares.

We need an estimate for  $\sigma^2$  in a *regression*, or "*line-fitting*" context.

#### Definition:

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

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

is the **line-fitting sample variance**.

because there are  $\beta_0, \beta_1$  in the model

# Simple Linear Regression

## Variance Estimation

### MSE

#### Variance estimation

Associated with  $s_{LF}^2$  are  $\nu = n - 2$  degrees of freedom and an estimated standard deviation of response

$$s_{LF} = \sqrt{s_{LF}^2}.$$

This is also called **Mean Square Error (MSE)** ✓ and can be found in JMP output.

It has  $\nu = n - 2$  degrees of freedom because we must estimate 2 quantities  $\beta_0$  and  $\beta_1$  to calculate it.

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

# Inference for Parameter $\beta_0$ and $\beta_1$

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

$\uparrow$   
 $\ominus 5x:$

## Inference for parameters

### Inference for $\beta_1$ :

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.  $100^* (1 - \alpha) \%$  confidence interval

\* 2. Formal hypothesis tests

Both of these require

$$\begin{cases} H_0: \beta_1 = 0 \\ H_a: \beta_1 \neq 0 \end{cases}$$

1. An estimate for  $\beta_1$  and  $\underline{\underline{b_1}} \rightarrow b_1 = \hat{\beta}_1$

2. a **standard error** for  $\beta_1$   $\underline{\underline{s_e(\hat{\beta}_1)}}$

$$s_e(\hat{\beta}_1) = \underline{\underline{s_e(b_1)}}$$

# Simple Linear Regression

## Variance Estimation

MSE

## Inference for Parameters

### Inference for $\beta_1$ :

It can be shown that since  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  and

$\epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ , then

$$\hat{\beta}_1 = b_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum(x - \bar{x})^2}\right)$$

variance of  $b_1$

$$E b_1 = \beta_1$$

Note that we never know  $\sigma^2$ , so we must estimate it using  
 $\sqrt{\text{MSE}} = S_{LF}$ .

So, a  $(1 - \alpha)100\%$  CI for  $\beta_1$  is

$$\rightarrow b_1 \pm t_{(n-2, 1-\alpha/2)} \frac{s_{LF}}{\sqrt{\sum(x_i - \bar{x})^2}}$$

standard error for  $\beta_1$

and the test statistic for  $H_0 : \beta_1 = \#$  is

$$\rightarrow K = \frac{b_1 - \#}{\frac{s_{LF}}{\sqrt{\sum(x_i - \bar{x})^2}}} \sim t_{(n-2)}$$

$s_{\text{EC}}(b_1)$

under  $H_0$

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

### Example:[Ceramic powder pressing]

A mixture of  $\text{Al}_2\text{O}_3$ , 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

$\left. \begin{array}{l} x = \text{the pressure setting used (psi)} \\ y = \text{the density obtained (g/cc)} \end{array} \right\}$

in the dry pressing of a ceramic compound into cylinders.

# Simple Linear Regression

## Variance Estimation

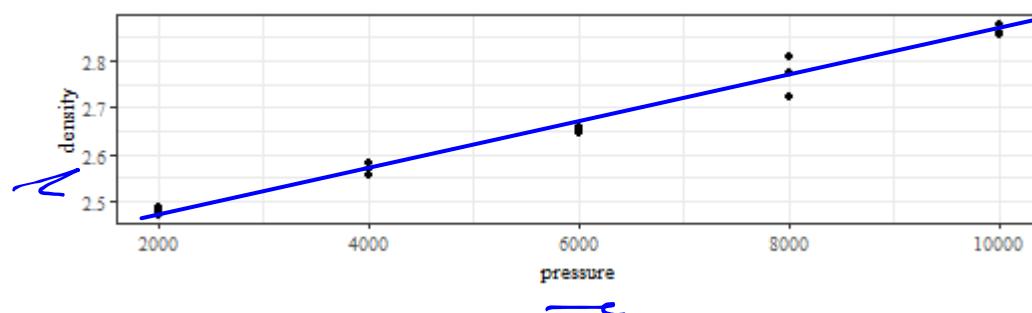
## MSE

## Inference for Parameters

Example:[Ceramic powder pressing]

pressure	density	pressure	density
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		

$n = 15$



# Simple Linear Regression

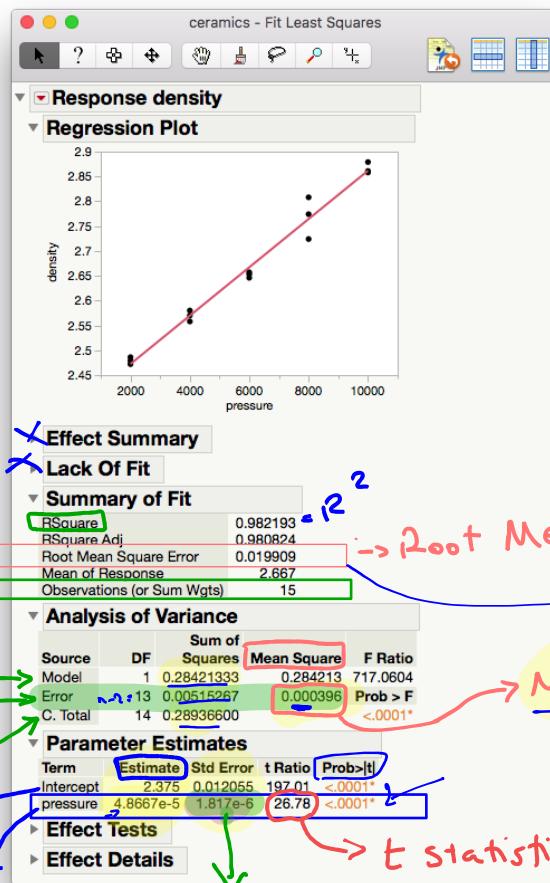
## Variance Estimation

MSE

## Inference for Parameters

Example:[Ceramic powder pressing]

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



$$\hat{\beta}_0 = b_0$$

$$\hat{\beta}_1 = b_1$$

$$SS_R$$

$$SS_E$$

$$SS_T$$

$$\rightarrow SE(b_1)$$

$$MSE = S_{LF}^2 = 0.0199^2$$

$$= \sqrt{S_{LF}^2}$$

$$t \text{ statistic : } k = \frac{b_1 - 0}{SE(b_1)} = 26.78$$

$$\left\{ \begin{array}{l} H_0: b_1 = 0 \\ H_a: b_1 \neq 0 \end{array} \right.$$

$$\frac{\sum (x_i - \bar{x})^2}{n}$$

# Simple Linear Regression

## Variance Estimation

MSE

## Inference for Parameters

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

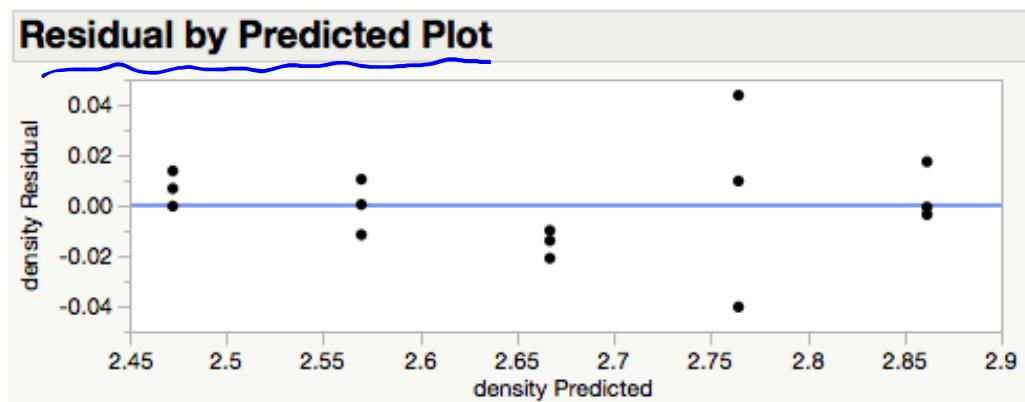
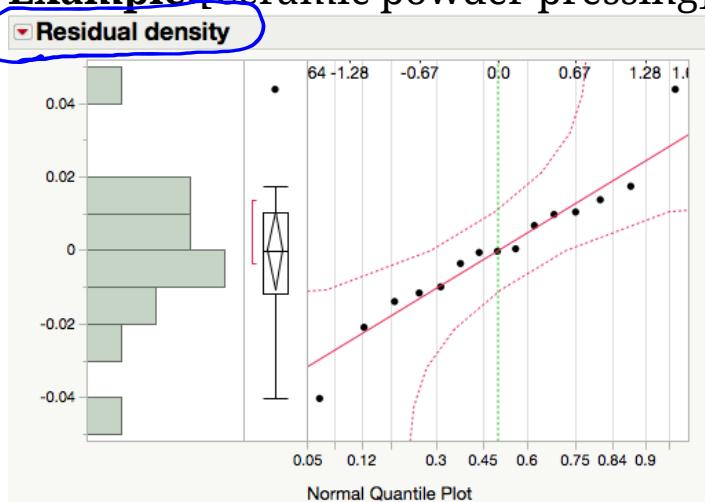
density                      pressure.

ma 511

P-value < 0.001



we'll reject  $H_0$ . i.e  
 $\beta_1 \neq 0$  and there's  
a significant  
relationship between  
pressure & density



Least squares regression of density on pressure of ceramic cylinders

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

**Example:** [Ceramic powder pressing]

1. Write out the model with the appropriate estimates.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = b_0 + b_1 x \\ = 2.375 + 4.8667 \times 10^{-5} x$$

2. Are the assumptions for the model met?

Yes, the residual plot shows random scatter around zero & Normal QQ-Plot looks relatively linear indicating that residuals are normal

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

$$R^2 = 98.21\%$$

# Simple Linear Regression

## Variance Estimation

MSE

## Inference for Parameters

**Example:** [Ceramic powder pressing]

4. What is the correlation between  $x$  and  $y$ ?

$$\text{In SLR : } r = \sqrt{R^2} \\ = \sqrt{0.9821} = 0.9911$$

5. Estimate  $\sigma^2$ .

$$\hat{\sigma}^2 = S_{\text{LF}}^2 = \text{MSE} = 0.000396$$

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

$$\text{Var}(b_1) = \frac{S_{\text{LF}}^2}{\sum(x_i - \bar{x})^2} = (\text{SE}(b_1))^2 \\ = (1.817 \times 10^{-6})^2 \\ = 3.3015 \times 10^{-12}$$

JMP doesn't give us  
this.

# Simple Linear Regression

## Variance Estimation

MSE

## Inference for Parameters

**Example:** [Ceramic powder pressing]

7. Calculate and interpret the 95% CI for  $\beta_1$

$$\hat{b}_1 \pm t_{(n-2, 1-\alpha/2)} \cdot \frac{s_{LF}}{\sqrt{\sum(x_i - \bar{x})^2}}$$
$$= 4.8667 \times 10^{-5} \pm t_{(15, 0.975)} \cdot (3.3 \text{ or } 10^{-6})$$

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

1-  $H_0 : \beta_1 = 0$  vs.  $H_1 : \beta_1 \neq 0$

2-  $\alpha = 0.05$

3- I will use the test statistics  $K = \frac{\hat{b}_1 - \#}{s_{LF}}$

which has a  $t_{n-2}$  distribution assuming that

- $H_0$  is true and
- The regression model is valid

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

Example:[Ceramic powder pressing]

4-

$$K = \frac{4.8667 \exp -5}{1.817 \exp -6}$$

So,

$SE(b_1)$

$b_1$

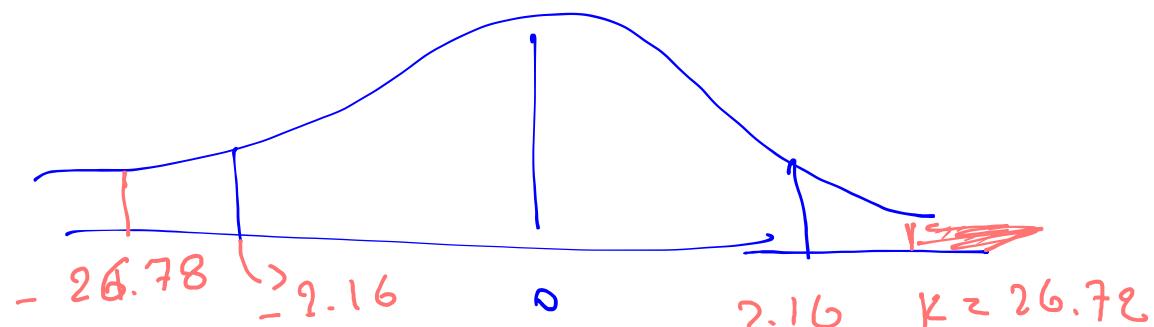
$$= 26.7843 > t_{(13,.975)} = 2.160$$

$\checkmark$

$$p\text{-value} = P(|T| > K) < 0.05 = \alpha$$

5- Since  $K = 26.7843 > 2.160 = t_{(13,.975)}$ , we reject  $H_0$ .

6- There is **enough evidence** to conclude that there is a linear relationship between density and pressure



# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

### Inference for mean response

Recall our model

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

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

$$\begin{aligned} E(\beta_0 + \beta_1 x_i + \epsilon_i) &= \beta_0 + \beta_1 x_i + E(\epsilon_i) \\ \Rightarrow \mu_{Y|x} &= \beta_0 + \beta_1 x_i \end{aligned}$$

Now, if some new covariate value  $x$  is within the range of the  $x_i$ 's (we don't extrapolate), we can estimate the true mean response at this new  $x$ . i.e

$$\hat{\mu}_{Y|x} = \hat{y} = b_0 + b_1 x$$

But how good is the estimate?

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

### Inference for mean response

Under the model,  $\hat{\mu}_{Y|x}$  is Normally distributed with

$$E(\hat{\mu}_{Y|x}) = \mu_{Y|x} = \beta_0 + \beta_1 x$$

and

$$\text{Var}\hat{\mu}_{Y|x} = \sigma^2 \left( \frac{1}{n} + \frac{(\mathbf{x} - \bar{x})^2}{\sum(\mathbf{x}_i - \bar{x})^2} \right)$$

Where  $\mathbf{x}$  is the individual value of  $x$  that we care about estimating  $\mu_{Y|x}$  at, and  $\mathbf{x}_i$  are all  $x_i$ 's in our data.

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

$$Z = \frac{\hat{\mu}_{Y|x} - \mu_{Y|x}}{\sigma \sqrt{\left( \frac{1}{n} + \frac{(\mathbf{x} - \bar{x})^2}{\sum(\mathbf{x}_i - \bar{x})^2} \right)}} \sim N(0, 1)$$

# Simple Linear Regression

## Variance Estimation

### MSE

## Inference for Parameters

## Inference for mean response

## Inference for mean response

And when  $\sigma$  is unknown (i.e. basically always), we replace  $\sigma$  with  $S_{LF} = \sqrt{\frac{1}{n-2} \sum (y_i - \hat{y}_i)^2}$  where we can get from JMP as **root mean square error (MSE)**. Then

$$T = \frac{\hat{\mu}_{Y|x} - \mu_{Y|x}}{s_{LF} \sqrt{\left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)}} \sim t_{(n-2)}$$

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

$$K = \frac{\hat{\mu}_{Y|x} - \#}{s_{LF} \sqrt{\left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)}}$$

which has a  $t_{n-2}$  distribution if 1)  $H_0$  is true and 2) the model is correct.

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

### Inference for mean response

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

$$\hat{\mu}_{Y|x} \pm t_{(n-2, 1-\alpha/2)} * s_{LF} \sqrt{\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}$$

and the one-sided the CI are analogous.

Note:

in the above formula,  $\sum(x_i - \bar{x})^2$  is not given by default in JMP.

## JMP Shortcut Notice

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

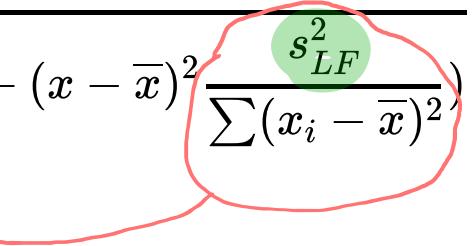
## Inference for mean response

### Inference for mean response

Using JMP we can get

$$s_{LF} \sqrt{\left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right)} = \sqrt{\left( \frac{s_{LF}^2}{n} + (x - \bar{x})^2 \frac{s_{LF}^2}{\sum(x_i - \bar{x})^2} \right)}$$

$\hat{Var}(b_1)$



Note that:

We can get  $\hat{Var}(b_0)$  from JMP as  $(SE(b_1))^2$

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

### Example:[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. (Note:  $\bar{x} = 6000$ )

**solution:**

$$\begin{aligned}\hat{\mu}_{Y|x=4000} &= \hat{y} = b_0 + b_1 x \\ &= 2.375 + 4.8667 \times 10^{-5} \times (4000) = 2.569668\end{aligned}$$

and

$$s_{LF} \sqrt{\left( \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2} \right)}$$

$$= \sqrt{\left( \frac{s_{LF}^2}{n} + (x - \bar{x})^2 \frac{s_{LF}^2}{\sum(x_i - \bar{x})^2} \right)}$$

$\text{Var}(\hat{y}) = (\text{SE}(\hat{y}))^2$

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

**Example:** [Ceramic powder pressing]

$$= \sqrt{\frac{0.000396}{15} + (4000 - 6000)^2 (1.817 \times 10^{-6})^2}$$

$$= \sqrt{0.000039606}$$

jmp output

$$= 0.0062933$$

Therefore, a two-sided 95% confidence interval for the true mean density at 4000 psi is

$$\hat{\mu}_{Y|x=4000} \pm t_{(n-2, 1-\alpha/2)} \times s_{LF} \sqrt{\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}$$

$$= 2.569648 \pm t_{(15-2, 0.975)} \times (0.0062933)$$

$$= 2.569648 \pm 2.160 \times (0.0062933)$$

$$= (2.5561, 2.5833)$$

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

**Example:** [Ceramic powder pressing]

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

$$\begin{aligned}\hat{\mu}_{Y|x=5000} &= \hat{y} = b_0 + b_1 x \\ &= 2.375 + 4.8667 \times 10^{-5} \times (5000) = 2.618335\end{aligned}$$

and

$$S_{LF} \sqrt{\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)} =$$

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

**Example:**[Ceramic powder pressing]

Therefore, a two-sided 95% confidence interval for the true mean density at 4000 psi is

$$\hat{\mu}_{Y|x=5000} \pm t_{(n-2,1-\alpha/2)} \times s_{LF} \sqrt{\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum(x_i - \bar{x})^2}\right)}$$

$$= 2.618335 \pm t_{(15-2,0.975)} \times (0.005449)$$

$$= 2.618335 \pm 2.160 \times (0.005449)$$

$$= (2.60656, 2.63011)$$

We are 95% confident that the true mean density of the ceramics at 5000 psi is between 2.60656 and 2.63011

# Multiple Linear Regression

## Simple Linear Regression

### Variance Estimation

### MSE

### Inference for Parameters

### Inference for mean response

### MLR

## Multiple linear regression

Recall the summarization the effects of several different quantitative variables  $x_1, \dots, x_{p-1}$  on a response  $y$ .

$$y_i \approx \beta_0 + \beta_1 x_{1,i} + \dots + \beta_{p-1} x_{p-1,i}$$

Where we estimate  $\beta_0, \dots, \beta_{p-1}$  using the *least squares principle* by minimizing the function

$$S(b_0, \dots, b_{p-1}) = \sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1,i} - \dots - \beta_{p-1} x_{p-1,i})^2$$

to find the estimates  $b_0, \dots, b_{p-1}$ .

We can formalize this now as

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_{p-1} x_{p-1,i} + \epsilon_i$$

where we assume  $\epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ .

# Variance Estimation in MLR

# Simple Linear Regression

## Variance Estimation

### MSE

## Inference for Parameters

## Inference for mean response

### MLR

## Variance estimation

Based on our multiple regression model, the residuals are of the form

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_{1i} + \cdots + b_{p-1} x_{pi})$$

And we can estimate the variance similarly to the SLR case.

### Definition:

For a set of  $n$  data vectors

$(x_{11}, x_{21}, \dots, x_{p-11}, y), \dots, (x_{1n}, x_{2n}, \dots, x_{p-1n}, y)$   
where least squares fitting is used to fit a surface,

$$s_{SF}^2 = \frac{1}{n-p} \sum (y - \hat{y})^2 = \frac{1}{n-p} \sum e_i^2$$

is the **surface-fitting sample variance** (also called **mean square error**, MSE). Associated with it are  $\nu = n - p$  degrees of freedom and an estimated standard deviation of response  $s_{SF} = \sqrt{s_{SF}^2}$ .

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

## Variance estimation

**Note:** the SLR fitting sample variance  $s_{LF}^2$  is the special case of  $s_{SF}^2$  for  $p = 2$ .

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

### Example:[Stack loss]

Consider a chemical plant that makes nitric acid from ammonia. We want to predict stack loss ( $y$ , 10 times the % of ammonia lost) using

$x_1$ : air flow into the plant

$x_2$ : inlet temperature of the cooling water

$x_3$ : modified acid concentration (% circulating acid -50%)  $\times 10$

# Simple Linear Regression

## Variance Estimation

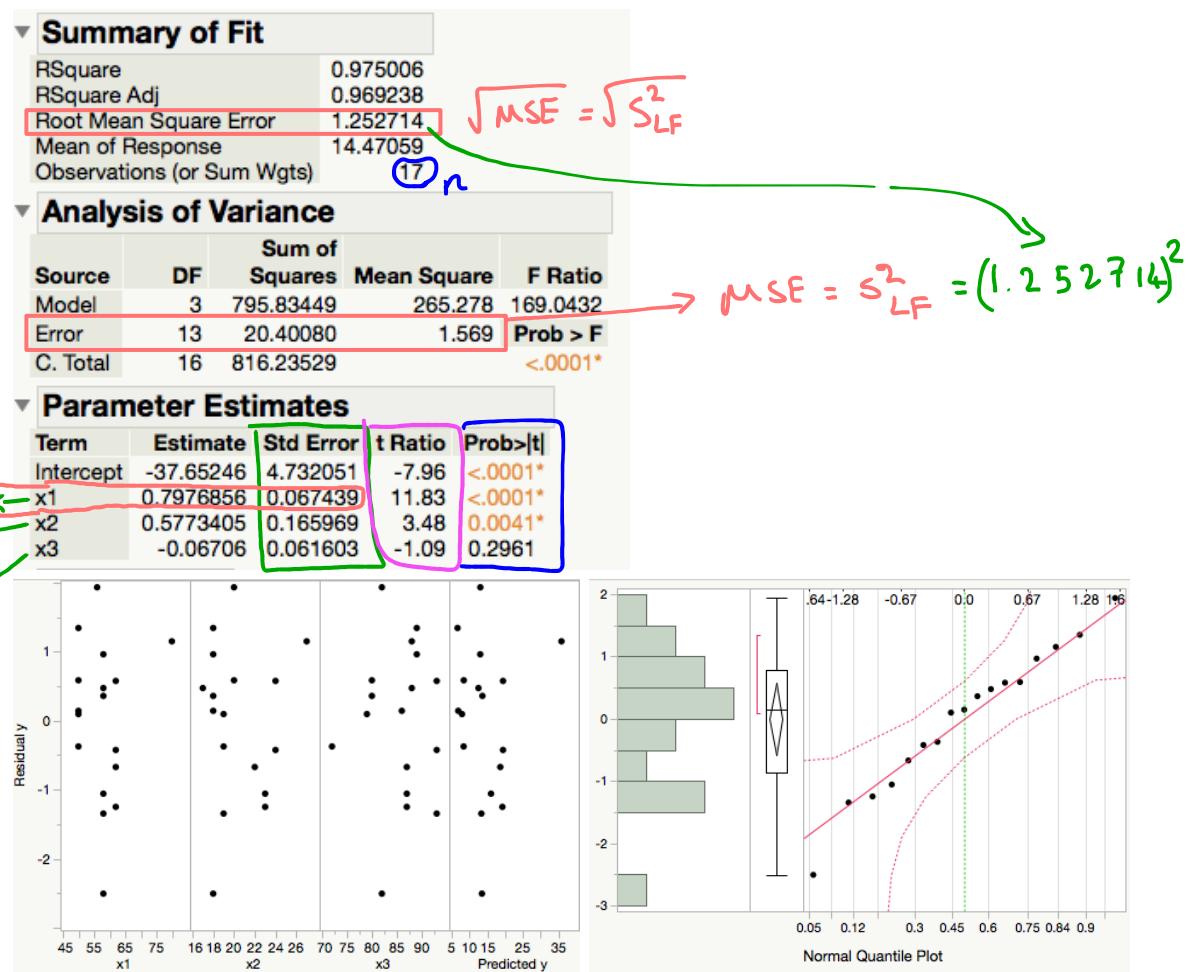
MSE

## Inference for Parameters

## Inference for mean response

MLR

### Example:[Stack loss]



## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

**Example:**[Stack loss]

Then we have the fitted model as

$$\hat{y} = -37.65246 + 0.7977x_1 + 0.5773x_2 - 0.0971x_3$$

The residual plots VS.  $x_1$ ,  $x_2$   $x_3$  and  $\hat{y}$  look like random scatter around zero.

The QQ-plot of the residuals looks linear, indicating that the residuals are Normally distributed.

This model is valid.

# Inference for Parameters in MLR

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

## Inference for parameters

We are often interested in answering questions (doing formal inference) for  $\beta_0, \dots, \beta_{p-1}$  individually. For example, we may want to know if there is a significant relationship between  $y$  and  $x_2$  (holding all else constant).

\vspace{.2in}

Under our model assumptions,

$$b_i \sim N(\beta_i, d_i \sigma^2)$$

for some positive constant  $d_i, i = 0, 1, \dots, p - 1$ . That are hard to compute analytically, but JMP can help)

That means

$$\frac{b_i - \beta_i}{s_{LF} \sqrt{d_i}} = \frac{b_i - \beta_i}{SE(b_i)} \sim t_{(n-p)}$$

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

## Inference for parameters

So, a test statistic for  $H_0 : \beta_i = \#$  is

$$K = \frac{b_i - \#}{s_{LF} \sqrt{d_i}} = \frac{b_i - \#}{SE(b_i)} \sim t_{(n-p)}$$

if 1)  $H_0$  is true and 2) the model is valid, and a 2-sided  $(1 - \alpha)$  100% CI for  $\beta_i$  is

$$b_i \pm t_{(n-p, 1-\alpha/2)} \times s_{LF} \sqrt{d_i}$$

or

$$b_i \pm t_{(n-p, 1-\alpha/2)} \times SE(b_i)$$

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

### Example:[Stack loss, cont'd]

Using the model fit on slide 35, answer the following questions:

1. Is the average change in stack loss ( $y$ ) for a one unit change in air flow into the plant ( $x_1$ ) less than 1 (holding all else constant)? Use a significance testing framework with  $\alpha = .1$ .

$\beta_1$  is associated with  $x_1$

solution:

$$1- H_0 : \beta_1 = 1 \text{ vs. } H_1 : \beta_1 < 1$$

$$2- \alpha = 0.1$$

3- I will use the test statistics  $K = \frac{b_1 - 1}{SE(b_1)}$  which has a  $t_{n-p} = t_{17-4}$  distribution assuming that

- $H_0$  is true and
- The regression model  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$  is valid

# Simple Linear Regression

Variance Estimation

SEE(b)  
from JMP

MSE

Inference for Parameters

Inference for mean response

MLR

## Example:[Stack loss, cont'd]

4-  $K = \frac{0.7977 - 1}{0.06744} = -3$  and  $t_{(13,9)} = 1.35$ . So,

p-value

$$= P(T < K) < P(T < -3) < 0.1 = \alpha$$



5- Since  $K = -3 < -1.35 = -t_{(13,9)}$ , we reject  $H_0$ .

6- There is enough evidence to conclude that the slope on airflow is less than one unit stackloss/unit airflow. With each unit increase in airflow and all other covariates held constant, we expect stack loss to increase by less than one unit.

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

### Example:[Stack loss, cont'd]

2. Is there a significant relationship between stack loss ( $y$ ) and modified acid concentration ( $x_3$ ) (holding all else constant)? Use a significance testing framework with  $\alpha = .05$ .

#### solution:

$$1- H_0 : \beta_3 = 0 \text{ vs. } H_1 : \beta_3 \neq 0$$

$$2- \alpha = 0.05$$

3- I will use the test statistics  $K = \frac{b_3 - 1}{SE(b_3)}$  which has a  $t_{n-p} = t_{17-4}$  distribution assuming that

- $H_0$  is true and
- The regression model  $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \epsilon_i$  is valid

# Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

### Example:[Stack loss, cont'd]

4-  $K = \frac{-0.06706 - 0}{0.0616} = -1.09$  and  
 $t_{(13,.975)} = 2.16$ . So,

$$\text{p-value} = P(|T| > |K|) =$$

$$P(|T| > 1.09) > P(|T| > t_{(13,.975)}) = 0.05\alpha$$

5- Since p-value >  $\alpha$ , we **fail to reject  $H_0$** .

6- There is **not enough evidence** to conclude that, with all other covariates held constant, there is a significant linear relationship between stack loss and acid concentration.

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

**Example:**[Stack loss, cont'd]

3. Construct and interpret a 99% two-sided confidence interval for  $\beta_3$ .

**solution:**

$$t_{(n-p,1-\alpha/2)} = t_{(13,.995)} = 3.012$$

then

$$\begin{aligned} b_3 \pm t_{(n-p,1-\alpha/2)} SE(b_3) &= -.06706 \pm 3.62(0.0616) \\ &= (-0.2525 \ 0.1185) \end{aligned}$$

We are 99% confident that for every unit increase in acid concentration, **with all other covariates held constant**, we expect stack loss to increase anywhere from -0.2525 units to 0.1185 units.

## Simple Linear Regression

## Variance Estimation

## MSE

## Inference for Parameters

## Inference for mean response

## MLR

### Example:[Stack loss, cont'd]

4. Construct and interpret a two-sided 90% confidence interval for  $\beta_2$

**solution:**

For a 90% two-sided CI for  $\beta_2$ ,

$$\alpha = 0.1 , t_{(n-p,1-\alpha/2)} = t_{(13,0.95)} = 1.77$$

Then

$$\begin{aligned} b_2 \pm t_{(n-p,1-\alpha/2)} \times SE(b_2) &= 0.5773 \pm 1.77(0.166) \\ &= (0.2834 \ 0.87127) \end{aligned}$$

We are 90% confident that for every one degree increase in temperature **with all other covariates held constant**, stack loss is expected to increase by anywhere from 0.2834 units to 0.8713 units.