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

error

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

where β_0 is the intercept.

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

β_1 is the slope.

It is the expected increase in y for every 1 unit change in x.

ϵ is some error. In fact,

$\varepsilon \sim N(0, \sigma^2)$ iid.

(recall checking if residuals are normally distributed is one of our model assessment techniques)

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.

inference

9.1.1 Variance estimation

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

β_0

least squares principal

β_1

σ^2 — ?

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

This is also called the 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 to calculate it (β_0 and β_1)

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. $(1-\alpha)100\%$ Confidence interval
2. Formal hypothesis (significance) test

Both of these require

① an estimate for β_1 (b_1) and ② a "standard error" for β_1

standard deviation of an estimate (statistic)

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

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

we never know this, we must estimate it using $\sqrt{MSE} = S_{LF}$

So, a $(1-\alpha)100\%$ CI for β_1 is

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

standard error for β_1

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

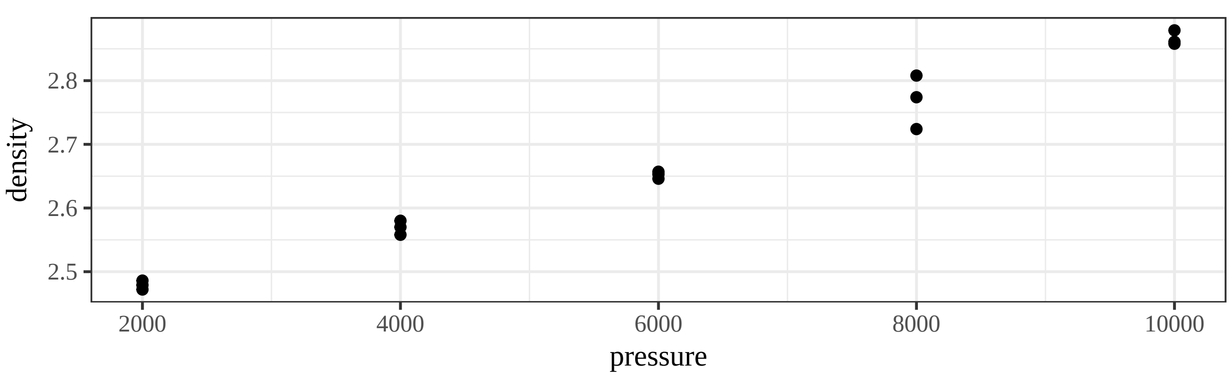
$$K = \frac{b_1 - \#}{\left(\frac{S_{LF}}{\sqrt{\sum_i (x_i - \bar{x})^2}}\right)} \sim t_{n-2} \text{ if } H_0 \text{ is true.}$$

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.

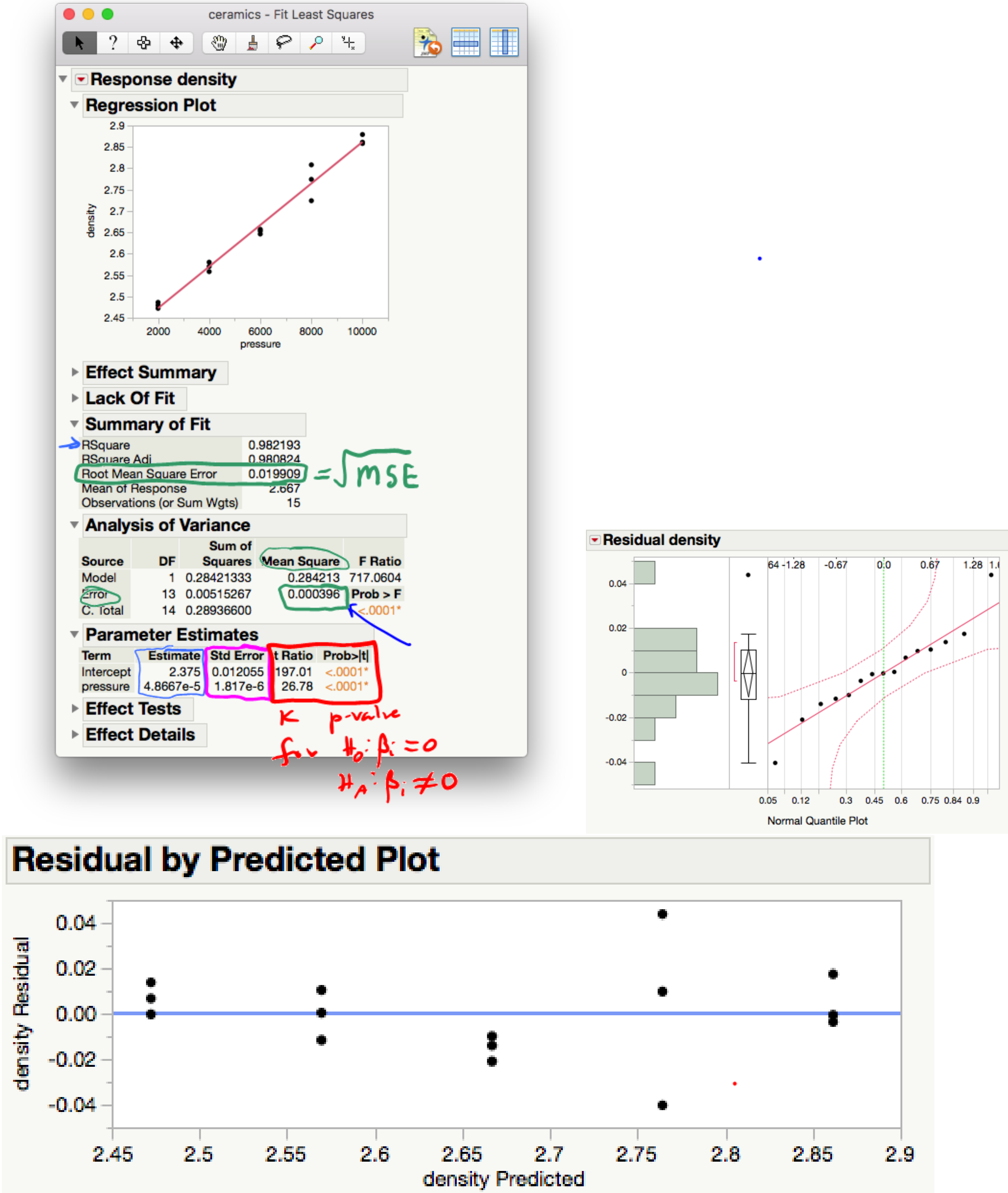


Figure 1: Least squares regression of density on pressure of ceramic cylinders.

1. Write out the model with the appropriate estimates.

$$\hat{y} = 2.375 + 4.8667 \times 10^{-5} x$$

2. Are the assumptions for the model met?

Yes. The residual plot shows random scatter around 0 and the Normal QQ plot looks relatively linear, indicating the residuals are Normally distributed.

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

$$R^2 = 0.9821$$

4. What is the correlation between x and y?

For SLR, $r = \sqrt{R^2} = \sqrt{.9821} = .9911$

5. Estimate σ^2.

$$\hat{\sigma}^2 = s_{LF}^2 = MSE = .000396$$

6. Estimate Var(b_1)

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

2-sided

7. Calculate and interpret the 95% CI for β_1

$$b_1 \pm t_{n-2, 1-\alpha/2} \frac{S_{LF}}{\sqrt{\Sigma(x_i - \bar{x})^2}} = 4.8667\times10^{-5} \pm t_{15-2, .975}(1.817\times10^{-6})$$

$$= 4.8667\times10^{-5} \pm 2.160(1.817\times10^{-6})$$

$$= (.00004474, .00005259)$$

We are 95% confident that for every 1 psi increase in pressure, density will increase between .00004474 g/cc and .00005259 g/cc on average.

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

① $H_0: \beta_1 = 0 \quad H_A: \beta_1 \neq 0$

② $\alpha = .05$

I could use my CI from 7.

③ I will use the test statistic $K = \frac{b_1 - 0}{\frac{S_{LF}}{\sqrt{\Sigma(x_i - \bar{x})^2}}}$ which has a t_{n-2} dsn assuming H_0 is true and the regression model is valid.

Done for us in JMP

④ $K = \frac{4.8667\times10^{-5}}{1.817\times10^{-6}} = 26.7843 > t_{13,.975} = 2.160$

So, p-value $= P(|T| > K) < .05 = \alpha$

⑤ Since $K = 26.7843 > 2.160 = t_{13,.975} \Rightarrow$ p-value $< .05$
$\Rightarrow$ we reject H_0.

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

9.1.3 Inference for mean response

Recall our model

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

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

$$E(\beta_0 + \beta_1 x_i + \varepsilon_i) = \beta_0 + \beta_1 x_i + \underbrace{E\varepsilon_i}_{=0}$$

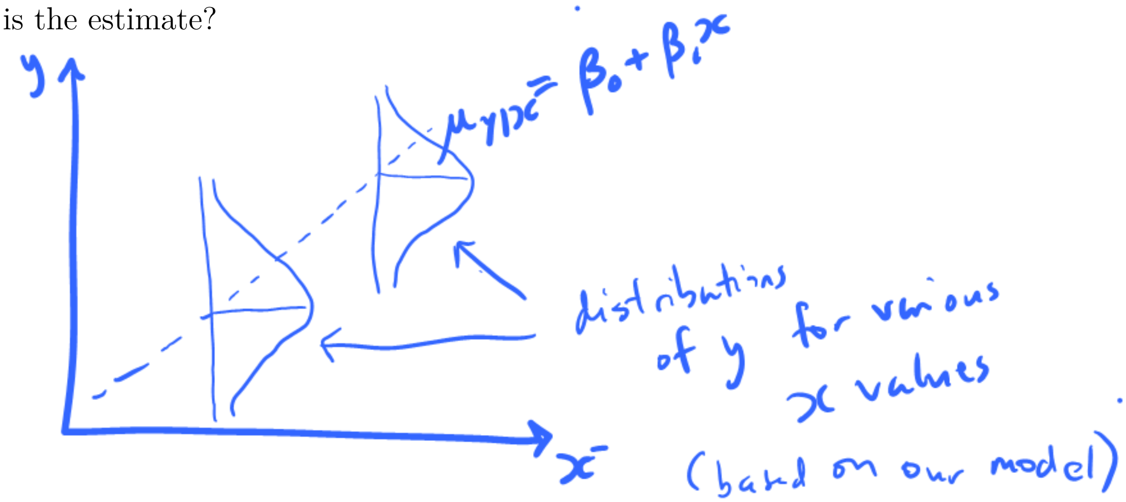
$$\mu_{y|x_i} = \beta_0 + \beta_1 x_i$$

we don't extrapolate

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

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

But how good is the estimate?



Under the model,

$\hat{\mu}_{Y|x}$ is Normally distributed with

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

individual value of x that we care about estimating $\mu_{Y|x}$ at

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

all x_i's in our data

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

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

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

replace σ with $s_{LF} = \sqrt{\frac{1}{n-2}\sum_i (y_i-\hat{y})^2}$ (can get from jmp as root MSE)

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

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

$$K = \frac{\hat{\mu}_{y|x} - \#}{S_{LF}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}}}$$

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

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

$$\hat{\mu}_{y|x} \pm t_{n-2,1-\alpha/2}\, S_{LF}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

(one-sided intervals are analogous).

This is not given by default in JMP.

(JMP shortcuts)

Notice

$$S_{LF}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}} = \sqrt{\frac{\overbrace{S_{LF}^2}^{\text{is MSE in JMP}}}{n} + \underbrace{(x-\bar{x})^2}_{\text{easy to calculate}}\boxed{\frac{S_{LF}^2}{\sum_i (x_i - \bar{x})^2}}}$$

$\boxed{\frac{S_{LF}^2}{\sum_i (x_i - \bar{x})^2}} = \hat{Var}(b_1)$ can get from JMP as $(SE(b_1))^2$

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. Note: $\bar{x} = 6000$

$$\hat{\mu}_{y|x=4000} = 2.375 + 4.8667\times 10^{-5}(4000) = 2.569668 \text{ g/cc}$$

$$S_{LF}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i-\bar{x})^2}} = \sqrt{\frac{S_{LF}^2}{n} + (x-\bar{x})^2 \frac{S_{LF}^2}{\sum_i (x_i-\bar{x})^2}} \qquad \frac{S_{LF}^2}{\sum_i (x_i-\bar{x})^2} = (SE(b_1))^2$$

$$= \sqrt{\frac{.000396}{15} + (4000-6000)^2(1.817\times 10^{-6})^2}$$

$$= \sqrt{.00039606}$$

$$= .0062933$$

Then $\hat{\mu}_{y|x=4000} \pm t_{n-2,1-\alpha/2} S_{LF}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i-\bar{x})^2}}$

$$= 2.569668 \pm t_{13,.975}(.0062933)$$

$$= 2.569668 \pm 2.160(.0062933)$$

$$= 2.569668 \pm .01359 = (2.5561, 2.5833)$$

We are 95% confident that the true mean density of the ceramics at 4000psi is between 2.5561 g/cc and 2.5833 g/cc.

on page 4, the range of X's is 2000 to 10000

So both 4000 and 5000 are reasonable values to estimate the mean response for

We are not extrapolating.

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

$$\hat{\mu}_{Y|x=5000} = 2.375 + 4.8667 \times 10^{-5}(5000) = 2.618335 \text{ g/cc}$$

$$S_{LF}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i-\bar{x})^2}} = \sqrt{\frac{S_{LF}^2}{n} + (x-\bar{x})^2 \frac{S_{LF}^2}{\sum_i (x_i-\bar{x})^2}} = (SE(b_1))^2$$

$$= \sqrt{\frac{\underset{MSE}{.000395}}{15} + (5000-6000)^2 (1.817 \times 10^{-6})^2}$$

$$= \sqrt{.00002970}$$

$$= .005449$$

Then $\hat{\mu}_{Y|x=5000} \pm t_{n-2,1-\alpha/2}\, S_{LF}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum_i (x_i-\bar{x})^2}}$

$$= 2.618335 \pm t_{13,.975}(.005449)$$

$$= 2.618335 \pm 2.160(.005449)$$

$$= 2.618335 \pm 0.01177 = (2.60656, 2.63011)$$

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

9.2 Multiple regression

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

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

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

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

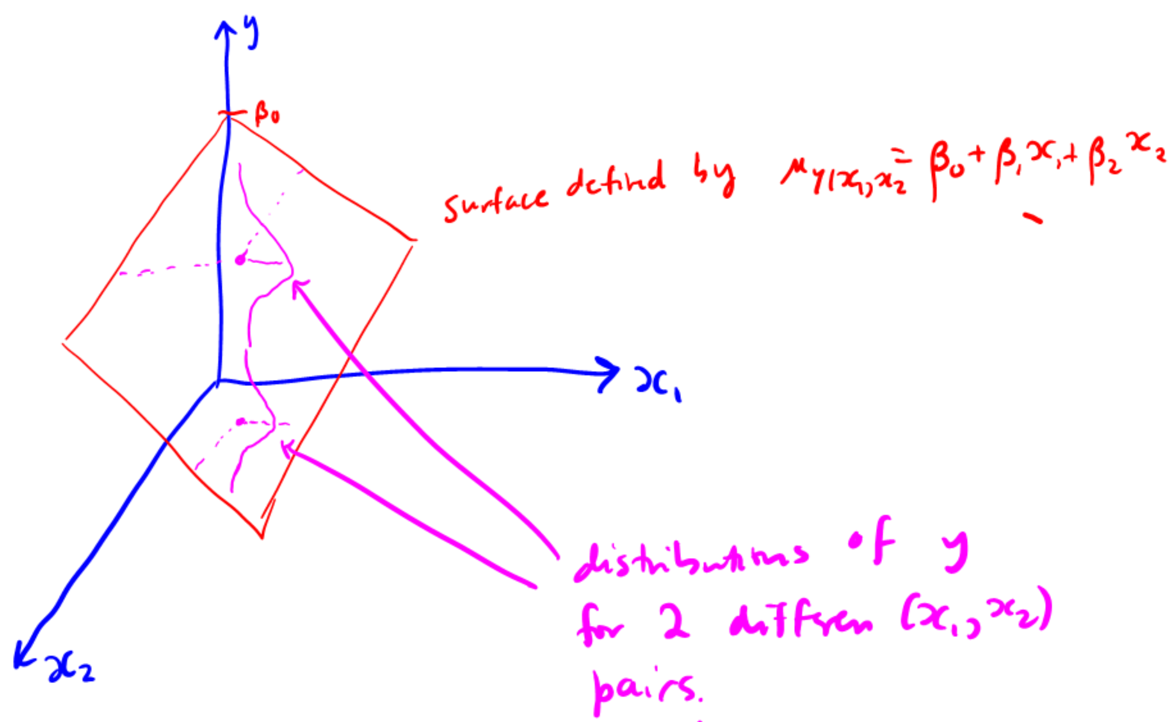
to find the estimates $b_0, \ldots, b_{p-1}$.

We can formalize this now as

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

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

For $p = 3$



9.2.1 Variance estimation

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

$$e_i = y_i - \hat{y}_i$$

$$= y_i - (b_0 + b_1 x_{1i} + \ldots + b_{p-1} x_{p-1\,i})$$

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

Definition 9.2. For a set of n data vectors $(x_{11}, x_{21}, \ldots, x_{p-11}, y), \ldots, (x_{1n}, x_{2n}, \ldots, x_{p-1n}, y)$ where least squares fitting is used to fit a surface,

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

also called mean square error (MSE)

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

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

Example 9.3 (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

Summary of Fit

RSquare	0.975006
RSquare Adj	0.969238
Root Mean Square Error	1.252714
Mean of Response	14.47059
Observations (or Sum Wgts)	17

$S_{SF} = \sqrt{S^2_{SF}}$

n

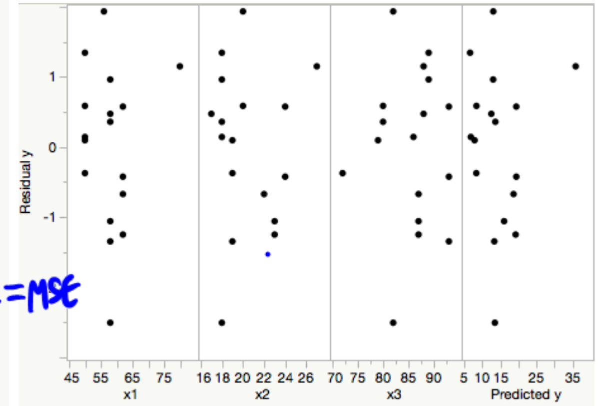
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	795.83449	265.278	169.0432
Error	13	20.40080	1.569	Prob > F
C. Total	16	816.23529		<.0001*

$S^2_{SF} = MSE$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-37.65246	4.732051	-7.96	<.0001*
x1	0.7976856	0.067439	11.83	<.0001*
x2	0.5773405	0.165969	3.48	0.0041*
x3	-0.06706	0.061603	-1.09	0.2961



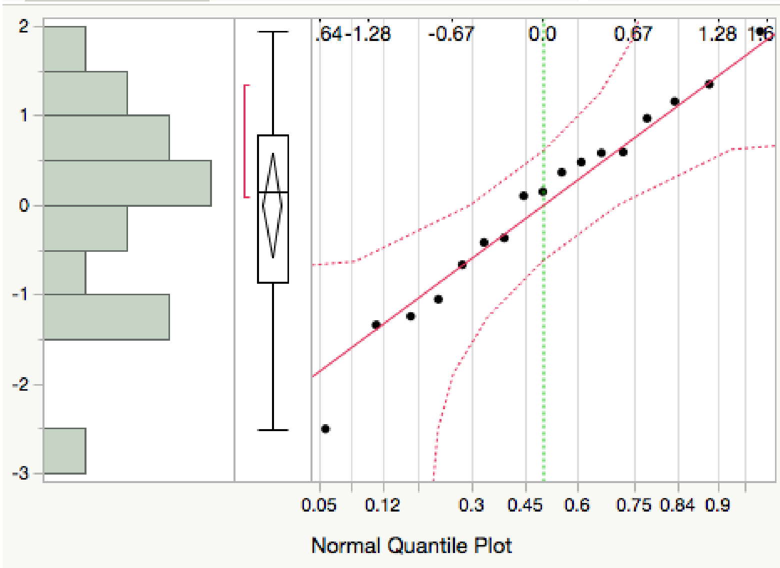


Figure 2: Least squares regression of stack loss on air flow, inlet temperature, and modified acid concentration.

$$\hat{y} = -37.65246 + 0.7977x_1 + 0.5773x_2 - 0.0671x_3$$

The residual plots vs x_1, x_2, x_3, and $\hat{y}$ look like random scatter around 0 and the QQ-plot of the residuals looks linear, indicating the residuals are Normally distributed.

This model is valid.

9.2.2 Inference for parameters

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

Under our model assumptions,

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

for some positive constant $d_i, i = 0, 1, \ldots, p-1$. (that are hard to compute/describe analytically. But JMP can help).

That means

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

So, a test statistic for $\mathrm{H}_0 : \beta_i = \#$ is

$$K = \frac{b_i - \#}{s_{SF}\sqrt{d_i}} = \frac{b_i - \#}{SE(b_i)} \sim t_{n-p}$$ if ① H_0 is true and ② the model is valid.

and a 2-sided $(1-\alpha)100\%$ CI for β_i is

$$b_i \pm t_{n-p, 1-\alpha/2} \cdot s_{SF}\sqrt{d_i}$$

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

Example 9.4 (Stack loss, cont'd)**.** Using the model fit on page 15, 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 $\alpha = .1$.
2. Is the there a significant relationship between stack loss (y) and modified acid concentation (x_3) (holding all else constant)? Use a significance testing framework with $\alpha = .05$.
3. Construct and interpret a 99% confidence interval for β_3.
4. Construct and interpret a 90% confidence interval for β_2.

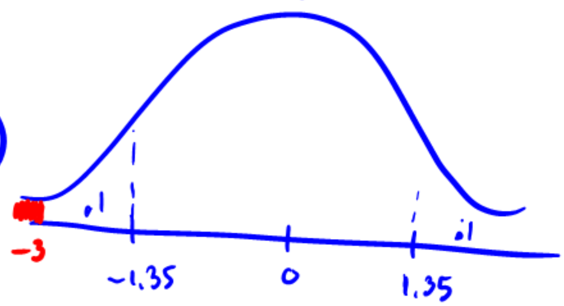
1. ① $H_0: \beta_1 = 1 \quad H_A: \beta_1 < 1$

② $\alpha = .1$

③ I will use the test statistic $K = \dfrac{b_1 - 1}{SE(b_1)}$ which, under the assumptions that ① H_0 is true and ② the model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$, $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ is correct, is distributed $t_{n-p} = t_{17-4} = t_{13}$

④ $K = \dfrac{0.7977 - 1}{0.06744} = -3.00$ and $t_{13,.9} = 1.35$

p-value: $P(T < K) = P(T < -3) < .1 = \alpha$



⑤ With $K = -3 < -1.35 = -t_{13,.9} \Rightarrow$ p-value $< \alpha \Rightarrow$ We reject Ho and conclude in favor of H_A

⑥ There is enough evidence that the true slope on airflow is less than 1 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 1 unit.

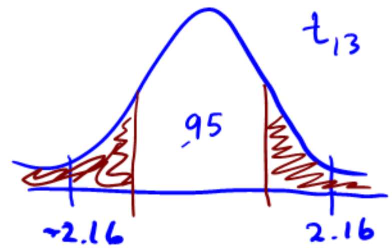
2. ① $H_0: \beta_3 = 0 \quad H_A: \beta_3 \neq 0$

② $\alpha = 0.05$

③ I will use $K = \frac{b_3 - 0}{SE(b_3)}$. If H_0 is true and the model $Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$

where $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ $i = 1, \ldots, 17$ holds, then $K \sim t_{n-p} = t_{17-4} = t_{13}$.

④ $K = \frac{-0.06706 - 0}{0.0616} = -1.09$



p-value $= P(|T| > |K|) = P(|T| > 1.09)$

$> P(|T| > t_{13,.975}) = .05 = \alpha$ $\quad t_{13,.975} = 2.16$

⑤ Since our p-value $> \alpha \Rightarrow$ 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.

3. For a two-sided 99% CI, $\alpha = .01$ and $t_{n-p, 1-\alpha/2} = t_{13,.995} = 3.012$

Then $b_3 \pm t_{n-p,1-\alpha/2} SE(b_3) = -0.06706 \pm 3.012(0.0616)$

$= (-0.2525, 0.1185)$

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 0.1165 units.

4 For a 90% two-sided CI for β_2, $\alpha = .1$, $t_{n-p, 1-\alpha/2} = t_{13,.95} = 1.77$

Then $b_2 \pm t_{n-p,1-\alpha/2} SE(b_2) = 0.5773 \pm 1.77(0.166)$

$= (.28348, 0.87127)$

We are 90% confident that for every 1 degree increase in temperature w/ all other covariates held constant, stack loss is expected to increase by anywhere from 0.2834 units to 0.8713 units.

9.2.3 Inference for mean response

We can also estimate the mean response at the set of covariate values, $(x_1, x_2, \ldots, x_{p-1})$. Under the model assumptions, the estimated mean response, $\widehat{\mu_{y|\boldsymbol{x}}}$, at $\boldsymbol{x} = (x_1, x_2, \ldots, x_{p-1})$ is

Normally distributed

with:

$$E(\hat{\mu}_{y|x}) = \mu_{y|x} = \beta_0 + \beta_1 x_1 + \ldots + \beta_{p-1} x_{p-1}$$

$$Var(\hat{\mu}_{y|x}) = \sigma^2 A^2$$ for some constant A (that is hard to compute by hand)

Then, under the model assumptions

$$Z = \frac{\hat{\mu}_{y|x} - \mu_{y|x}}{\sigma A} \sim N(0,1) \quad \text{and} \quad T = \frac{\hat{\mu}_{y|x} - \mu_{y|x}}{S_{SF} A} \sim t_{n-p}$$

And a test statistic for testing $\mathrm{H}_0 : \mu_{y|\boldsymbol{x}} = \#$ is

$$K = \frac{\hat{\mu}_{y|x} - \#}{S_{SF} A}$$ which has a t_{n-p} distribution under H_0 if the model holds.

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

$$\hat{\mu}_{y|x} \pm t_{n-p,1-\alpha/2} S_{SF} A$$ (1-sided CI analogous)

Note: $S_{SF} A = SE(\hat{\mu}_{y|x})$ we can use JMP to get this.

Example 9.5 (Stack loss, cont'd)**.** We can use JMP to compute a 2-sided 95% CI around the mean response at point 3:

$$x_1 = 62, x_2 = 23, x_3 = 87, y = 18$$

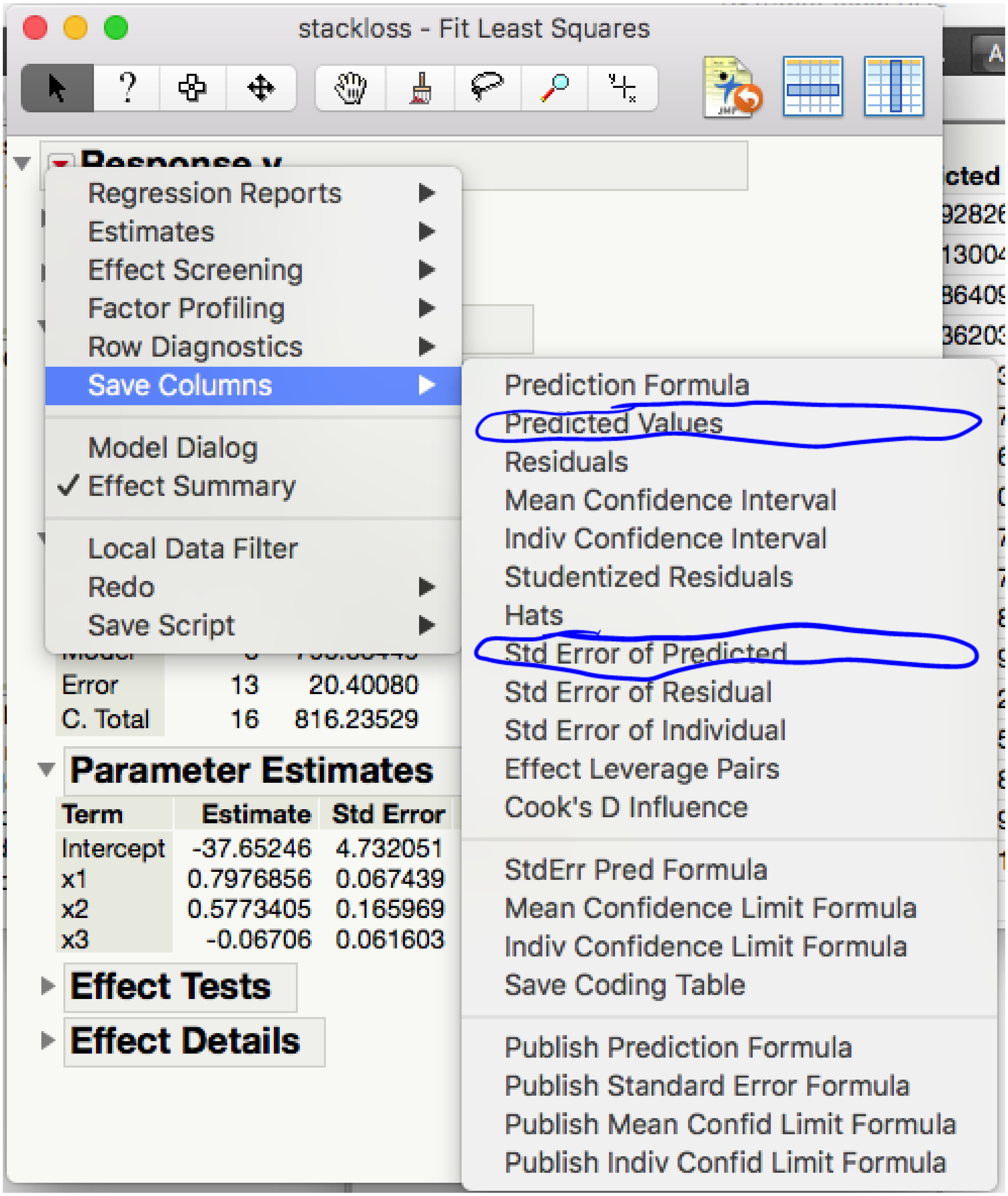


Figure 3: How to get predicted values and standard errors.

stackloss

stackloss
Source

Columns (6/0)
x1
x2
x3
y
Predicted y
StdErr Pred y

Rows

All rows	17
Selected	1
Excluded	0
Hidden	0
Labelled	0

	x1	x2	x3	y	Predicted y	StdErr Pred y
1	80	27	88	37	35.849282687	1.0461642094
2	62	22	87	18	18.671300496	0.35771273
3	62	23	87	18	19.248640953	0.417845385
4	62	24	93	19	19.423620349	0.6295687471
5	62	24	93	20	19.423620349	0.6295687471
6	58	23	87	15	16.057898713	0.5204068064
7	58	18	80	14	13.640617664	0.6090546656
8	58	18	89	14	13.037076072	0.5582571612
9	58	17	88	13	12.526795792	0.6739851764
10	58	18	82	11	13.50649731	0.5519432283
11	58	19	93	12	13.346175822	0.6055705716
12	50	18	89	8	6.6555915917	0.5876767248
13	50	18	86	7	6.8567721223	0.4891659484
14	50	19	72	8	8.3729550563	0.8232400377
15	50	19	79	8	7.903533818	0.5302896274
16	50	20	80	9	8.4138140985	0.5769617708
17	56	20	82	15	13.065807105	0.3632418427

Figure 4: Predicted values and standard errors.

With $t_{n-p,1-\alpha/2} = t_{13,.975} = 2.16$

The 95% CI is

$$\hat{\mu}_{y|x} \pm t_{n-p,1-\alpha/2}\, SE(\hat{\mu}_{y|x})$$

$$= 19.2486 \pm 2.16(0.41785)$$

$$= (18.343, 20.151)$$

We are 95% confident that when air flow is 62 units, temperature is 23 degrees, and the adjusted percentage of circulating acid is 87 units, the true mean stack loss is between 18.343 and 20.151 units.