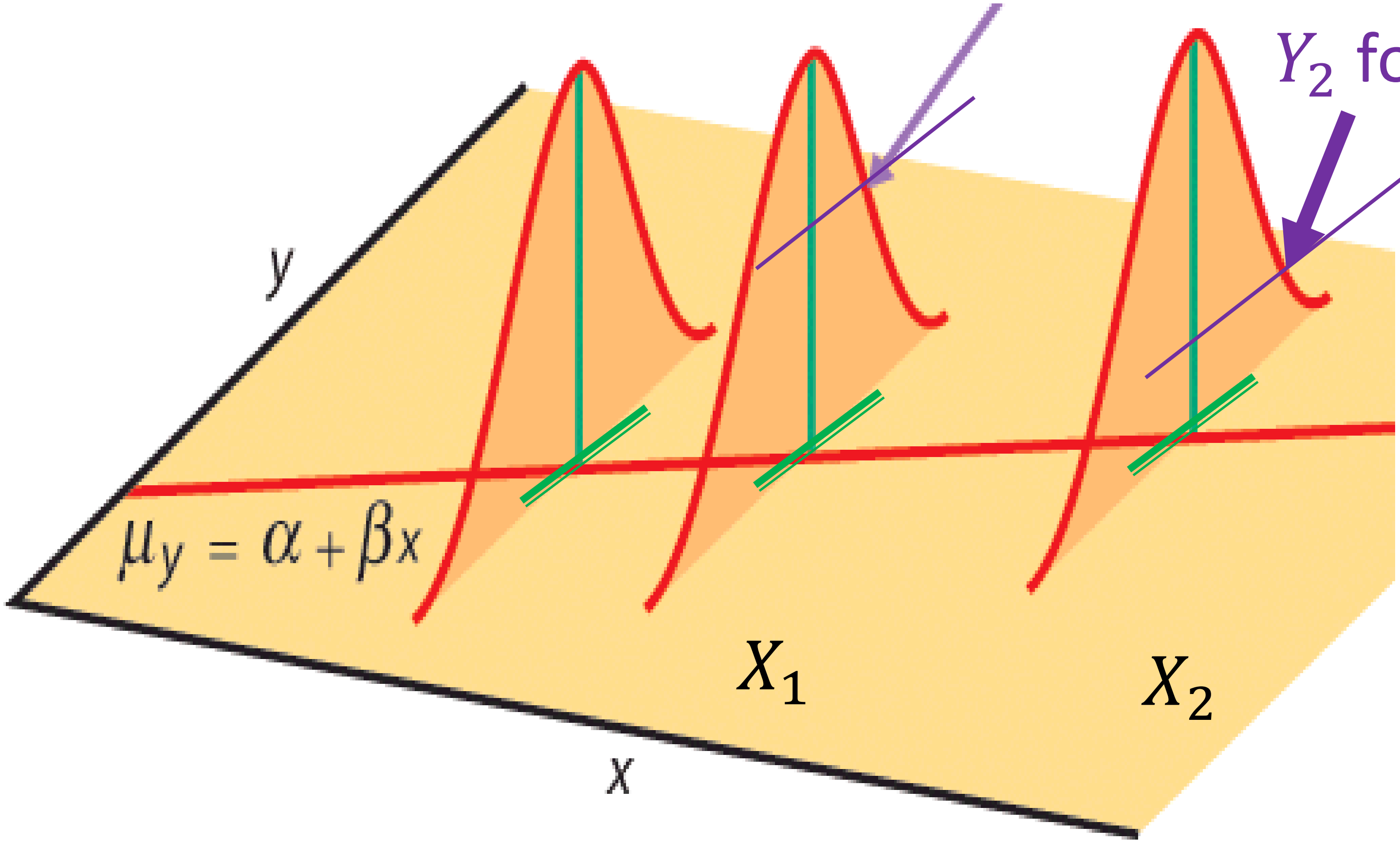
Interval Estimation of mean response $E\{Y_h\}, or\ \hat{Y}_h$ and single response $\hat{Y}_h$(new) when $X = X_h$

Mean response vs. single response

Y_1 for a given X_1 = mean + random error

Y_2 for a given X_2= mean + random error



The mean interval

single Y $\pm$ *margin error*

is larger than

The mean interval

$\boldsymbol{\mu \pm margin\ error}$

- Predict the mean response of Y on X

$$E\{Y_h\}\ or\ \hat{\mu}_h$$

- Predict the single response of Y on X

$$\hat{Y}_h$$

Predict in the same manner,
Same value;
But different precision.

Recall that

The best point estimates of β_1 and β_0 given the data (X, Y) are:

$$b_1 = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-\bar{X})^2} = \frac{\mathrm{SS_{XY}}}{\mathrm{SS_X}} = \Sigma c_i Y_i \qquad E(b_1) = \beta_1 \; and \quad Var(b_1) = \sigma^2 \frac{1}{\Sigma\,(X_i - \bar{X})^2}$$

$$b_0 = \bar{Y} - b_1\bar{X} = \frac{\Sigma \mathrm{Y_i}}{\mathrm{n}} - \Sigma c_i \bar{X} Y_i = \Sigma d_i Y_i \qquad E(b_0) = \beta_0 \; and \quad Var(b_0) = \sigma^2 [\frac{1}{n} + \frac{\bar{X}^2}{\Sigma\,(X_i - \bar{X})^2}]$$

Hence, $\hat{Y}_h = b_0 + b_1 X_h$ is a linear combination of the observations Y_i

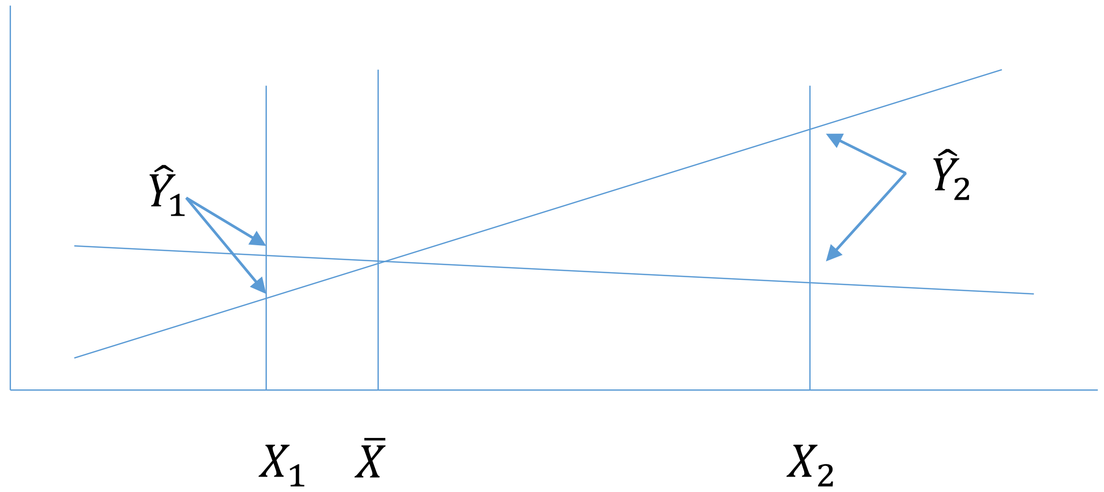
Question1: Does $\hat{Y}_h$ follow normal distribution?

Question2: are b_0 and b_1 independent?

Prediction of the mean response

$$\hat{Y}_h = b_0 + b_1 X_h$$

- For normal error (ε) regression model, $\hat{Y}_h \sim Normal$, with mean and variance:
- $E\{\hat{Y}_h\} = E\{Y_h\} = \mu_h$
- $\sigma^2\{\hat{Y}_h\} = \sigma^2[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}]$
- $\hat{Y}_h$ is normal because $b_0 + b_1 X_h$ is a linear combination of independent, normal Y_i's.
- Its variance is affected by how far X_h is from $\bar{X}$, through the term $(X_h - \bar{X})^2$.
- Estimation is more precise near $\bar{X}$.



Prediction of the mean response

$$\hat{Y}_h = b_0 + b_1 X_h$$

- For normal error (ε) regression model, $\hat{Y}_h \sim Normal$, with mean and variance:

$$E\{\hat{Y}_h\} = E\{Y_h\} = \mu_h$$

- $\sigma^2\{\hat{Y}_h\} = \sigma^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$

- When replace σ^2 with MSE $s^2\{\hat{Y}_h\} = MSE\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$

Therefore, it follows that

$$\frac{\hat{Y}_h - E\{Y_h\}}{s\{\hat{Y}\}} \sim t(n-2)$$

Prediction confidence interval of mean response, $\boldsymbol{E\{Y_h\}}$

$$\frac{\hat{Y}_h - E\{Y_h\}}{s\{\hat{Y}_h\}} \sim t(n-2)$$

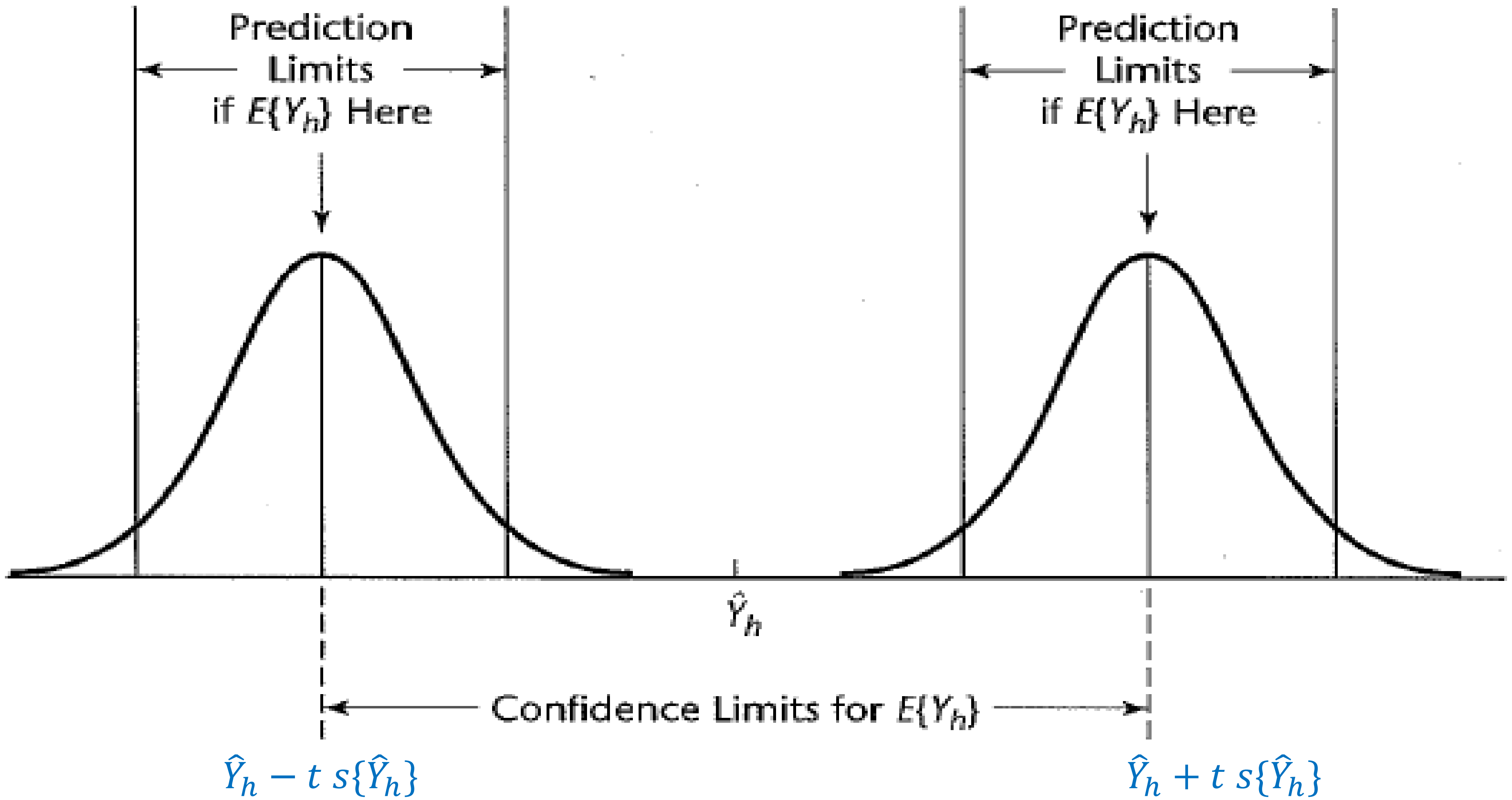
The confidence interval of $E\{Y_h\}$

$$\hat{Y}_h \pm t\left(1 - \frac{\alpha}{2}; n-2\right) s\{\hat{Y}_h\}$$

$$where\ s^2_{\{\hat{Y}_h\}} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$$

$s\{\hat{Y}_h\}$ is the “standard error of the mean response value at $X = X_h$”
s is the “standard error of the residuals”

Prediction of single response $\hat{Y}_{h(new)}$



$$\sigma^2\{Y_{h(new)}\} = \sigma^2\{\hat{Y}_h\} + \sigma^2$$

The variance of prediction = variance in possible location of the distribution + variance within the distribution

We estimate the variance of the single prediction as

$$s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2} + 1\right]$$

For normal error regression model

$$\frac{Y_{h(new)} - \hat{Y}_h}{s_{\{pred\}}} \sim t\ (n\text{-}2)$$

$s\{pred\}$ is the "standard error for predicting one new response value at X_h."

Prediction interval of single response $\hat{Y}_{h(new)}$

$$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2}; n-2\right) s\{pred\}$$

$$s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 = s^2\left[\frac{1}{n} + \frac{(X_h-\bar{X})^2}{\Sigma(X_i-\bar{X})^2} + 1\right]$$

- More sensitive to departure of normal in error terms distribution.
- Predictions are more precise near $\bar{X}$ because $\sigma^2_{\{\hat{Y}_h\}}$ decreases with $|X_h - \bar{X}|$.

Prediction interval of mean of m new response $\bar{Y}_{h\{new\}}$ not $\hat{Y}_h$, or $\hat{Y}_h\{new\}$

$$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2}; n-2\right) s\{predmean\}$$

Where: $s^2\{predmean\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\}$

$$= MSE\left[\frac{1}{m} + \frac{1}{n} + \frac{(X_h-\bar{X})^2}{\Sigma(X_i-\bar{X})^2}\right]$$

- Predict the mean of m new observations on Y for a given level of the predictor variables.
- The variance $s^2\{predmean\}$ has two components: variance between the distribution and variance within a distribution.

$s\{predmean\}$ is the "standard error for predicting the mean of m new response value."

The Diamond example, if $X_h = 0.43$, compute

1. The confidence interval for the mean predicted value $E\left(\hat{Y}_h\right)$

$\hat{Y}_h \pm t_c s_{\{\hat{Y}_h\}}$ Where $s^2_{\{\hat{Y}_h\}} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$

2. The confidence interval for the single predicted value $\left(\hat{Y}_h\right)$

$\hat{Y}_h \pm t_c s_{\{pred\}}$ Where $s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2} + 1\right]$

3. The confidence interval for the mean price $\bar{Y}_{h(new)}$ of three diamonds with the same weight (0.43)

$\hat{Y}_h \pm t_c s\{predmean\}$ Where: $s^2\{predmean\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = MSE[\frac{1}{m} + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}]$

Recall that in the Diamond example

The lm output

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -259.63      17.32  -14.99   <2e-16 ***
weight       3721.02      81.79   45.50   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31.84 on 46 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9783,    Adjusted R-squared:  0.9778
F-statistic:  2070 on 1 and 46 DF,  p-value: < 2.2e-16
```

$\boldsymbol{MSE = s^2 = 31.84^2 = 1013.8}$

$\bar{X} = 0.204, s_X = 0.0568,$ n=48

Recall that in the diamond example

Where $\alpha = 0.05, n = 48, df = 46\ round\ down\ to\ 40$

TABLE C *t* distribution critical values

Degrees of freedom	Confidence level C											
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
40	0.681	0.851	1.050	1.303	1.684	$\boxed{2.021}$	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
One-sided *P*	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
Two-sided *P*	.50	.40	.30	.20	.10	.05	.04	.02	.01	.005	.002	.001

$$t\left(1-\frac{\alpha}{2}, n-2\right) = t(0.975, 46)$$

$= 2.021$ (estimation using the t table)

$= 2.013$ (exact value using R)

Or use R

```
qt(1-0.5*alpha,n-2)
```

```
[1] 2.012896
```

The Diamond example, if $X_h = 0.43$, compute

1. The confidence interval for the mean predicted value $E\left(\hat{Y}_h\right)$

$$\hat{Y}_h \pm t_c s_{\{\hat{Y}_h\}} \qquad \text{Where} \quad s^2_{\{\hat{Y}_h\}} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$$

$$SS_X = s_X^2(n-1) =$$

$$t\left(1 - \frac{\alpha}{2}, n-2\right) = t(0.975, 46) = 2.021 \ \text{(estimation from T} - \text{table) or} \ \ 2.013 \ \text{(exact value from R)}$$

$$s^2_{\{\hat{Y}_h\}} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] =$$

$$\boldsymbol{s\{\hat{Y}_h\}} =$$

$$\hat{Y}_h \pm t_c s_{\{\hat{Y}_h\}} = 1340.415 \pm 2.013(19.03) = 1302.1, \ \ 1378.73.$$

```
ci.reg(diamond.mod, new, type='m',alpha=0.05)
```

The Diamond example, if $X_h = 0.43$, compute

1. The confidence interval for the mean predicted value $E(\hat{Y}_h)$

$$\hat{Y}_h \pm t_c s_{\{\hat{Y}_h\}} \qquad \text{Where} \quad s^2_{\{\hat{Y}_h\}} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right]$$

$$SS_X = s_X^2(n-1) = 0.0568^2(48-1) = 0.152$$

$$t\left(1 - \frac{\alpha}{2}, n-2\right) = t(0.975, 46) = 2.021 \ (\text{estimation from T} - \text{table}) \text{ or } \ 2.013 \ (\text{exact value from R})$$

$$s^2_{\{\hat{Y}_h\}} = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right] = \quad 31.84^2\left[\frac{1}{48} + \frac{(0.43 - 0.204)^2}{0.152}\right] = 362.14 \qquad s\{\hat{Y}_h\} = \sqrt{362.14} = 19.03$$

$$\hat{Y}_h \pm t_c s_{\{\hat{Y}_h\}} = 1340.415 \pm 2.013(19.03) = 1302.1, \ 1378.73.$$

```
ci.reg(diamond.mod, new, type='m',alpha=0.05)
```

The Diamond example, if $X_h = 0.43$, compute

2. The confidence interval for the single predicted value $(\hat{Y}_h)$

$\hat{Y}_h \pm t_c s_{\{pred\}}$ Where $s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 = s^2 \left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2} + 1\right]$

$s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 =$

$\hat{Y}_h \pm t_c s_{\{pred\}} = 1340.415 \pm 2.013(37.093) = 1265.75, 1415.08$

```
ci.reg(diamond.mod, new, type='n',alpha=0.05)
```

The Diamond example, if $X_h = 0.43$, compute

2. The confidence interval for the single predicted value $(\hat{Y}_h)$

$$\hat{Y}_h \pm t_c s_{\{pred\}} \qquad \text{Where } s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 = s^2\left[\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2} + 1\right]$$

$$s^2_{\{pred\}} = s^2_{\{\hat{Y}_h\}} + s^2 = \quad 362.14 + 1013.78 = 1375.92$$

$$\hat{Y}_h \pm t_c s_{\{pred\}} = 1340.415 \pm 2.013(37.093) \quad = 1265.75, 1415.08$$

```
ci.reg(diamond.mod, new, type='n',alpha=0.05)
```

The Diamond example, if $X_h = 0.43$, compute

3. The confidence interval for the mean price $\bar{Y}_{h(new)}$ of three diamonds with the same weight (0.43)

$\hat{Y}_h \pm t_c s\{predmean\}$ Where: $s^2\{predmean\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = MSE[\frac{1}{m} + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}]$

$$s^2\{predmean\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} =$$

$\hat{Y}_h \pm t_c s\{predmean\}$=1340.415 $\pm 2.013(26.46) = (1287.151, 1393.679)$

```
ci.reg(diamond.mod, new, type='nm', m=3, alpha=0.05)
```

The Diamond example, if $X_h = 0.43$, compute

3. The confidence interval for the mean price $\bar{Y}_{h(new)}$ of three diamonds with the same weight (0.43)

$\hat{Y}_h \pm t_c s\{predmean\}$ Where: $s^2\{predmean\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = MSE[\frac{1}{m} + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}]$

$$s^2\{predmean\} = \frac{MSE}{m} + s^2\{\hat{Y}_h\} = \frac{31.84^2}{3} + 362.14 = 700.07$$

$$\hat{Y}_h \pm t_c s\{predmean\} = 1340.415 \pm 2.013(26.46) = (1287.151, 1393.679)$$

```
ci.reg(diamond.mod, new, type='nm', m=3, alpha=0.05)
```