The Simple Linear Regression Model (Reference: Chapter 10 of Montgomery) My STAT3610 Notes Maghsoodloo

The objective of simple linear regression (SLR or SLREG) is to determine if an output y is linearly related to a single input x. The SLREG model is given by $y_i = \beta_0 + \beta_1 x_i + \in_i$, where $\in_i$'s are assumed NID(0, $\sigma_{\in}^2$).

Regression	Versus	Correlation
Levels of x are fixed		(x, y) is a bivariate random vector

As an example, consider the Problem 1 on p. 414 of Montgomery (p. 401 0f the 6th edition), where the tensile strength (TS) of paper is thought to be linearly related to the amount of hardwood concentration in the pulp. In a pilot plant, 10 sample pairs (x_i, y_i), i = 1, 2, …, 10, were produced leading to the following data:

x_i :	10	15	15	20	20	20	25	25	28	30%
y_i:	160	171	175	182	184	181	188	193	195	200 psi

where x = Hardwood Concentration is fixed, and y = Paper TS is a rv. Thus,

$$V(y_i) = V(\beta_0 + \beta_1 x_i + \in_i) = V(\mu_i + \in_i) = \sigma_{\in}^2 = \sigma^2,$$

and as a result y_i's are assumed NID($\mu_i = E(y_i|x_i) = \beta_0 + \beta_1 x_i$, $\sigma_{\in}^2$). The objective is to estimate the parameters β_0 and β_1 such that the least squares function (LSF):

$$L(\beta_0, \beta_1) = \sum_{i=1}^{n} \in_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \qquad (52)$$

is minimized wrt (with respect to) β_0 and β_1. For example, at (x_1 = 10, y = y_1), the expression for the error before experimentation is given by $\in_1 = y_1 - \beta_0 - 10\beta_1$, while at ($x_2$, y_2), $\in_2 = y_2 - \beta_0 - 15\beta_1$, ..., and at ($x_{10}$, y_{10}) the prior error is given by $\in_{10} = y_{10} - \beta_0 - 30\beta_1$. However, after data have been gathered, as shown above, the LSF in Eq. (52) reduces to:

$$L(\beta_0, \beta_1) = (160 - \beta_0 - 10\beta_1)^2 + (171 - \beta_0 - 15\beta_1)^2 + \ldots + (200 - \beta_0 - 30\beta_1)^2$$

In order to minimize the LSF in Eq. (52) wrt the unknown parameters β_0 and β_1, we partially differentiate the LSF in Eq. (52) and equate its partial derivatives to zero in order to obtain the optimum point.

$\partial L / \partial \beta_0$ = $2(160 - \beta_0 - 10\beta_1)(-1) + 2(171 - \beta_0 - 15\beta_1)(-1) + \ldots +$

$2(200 - \beta_0 - 30\beta_1)(-1) \xrightarrow{\text{Set to}} = 0$

$\partial L / \partial \beta_1$ = $2(160 - \beta_0 - 10\beta_1)(-10) + 2(171 - \beta_0 - 15\beta_1)(-15) + \ldots + 2(200 -$

$\beta_0 - 30\beta_1)(-30) \xrightarrow{\text{Set to}} = 0$

The above 2 equations simplify to

$(160 - \hat{\beta}_0 - 10\hat{\beta}_1) + (171 - \hat{\beta}_0 - 15\hat{\beta}_1) + \ldots + (200 - \hat{\beta}_0 - 30\hat{\beta}_1) = 0$, and

$(160 - \hat{\beta}_0 - 10\hat{\beta}_1)(10) + (171 - \hat{\beta}_0 - 15\hat{\beta}_1)(15) + \ldots + (200 - \hat{\beta}_0 - 30\hat{\beta}_1)(30) = 0$.

After some algebraic simplification, the above two equations reduce to

$$\sum_{i=1}^{n} y_i - \hat{\beta}_0 \sum_{i=1}^{10} 1 - \hat{\beta}_1 \sum_{i=1}^{10} x_i = 0$$

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{10} x_i - \hat{\beta}_1 \sum_{i=1}^{10} x_i^2 = 0.$$

The above heterogeneous system of 2 equations with 2 unknowns when written as

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{10} x_i = \sum_{i=1}^{n} y_i \quad \rightarrow \quad 10\hat{\beta}_0 + 208\hat{\beta}_1 = 1829, \text{ and}$$

$$\hat{\beta}_0 \sum_{i=1}^{10} x_i + \hat{\beta}_1 \sum_{i=1}^{10} x_i^2 = \sum_{i=1}^{n} x_i y_i \quad \rightarrow \quad 208\hat{\beta}_0 + 4684\hat{\beta}_1 = 38715,$$

is called the LS (Least-Squares) normal equations. Note that in matrix form, the above system of equations can be written as

$$\begin{bmatrix} 10 & 208 \\ 208 & 4684 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 1829 \\ 38715 \end{bmatrix}, \quad \text{or} \rightarrow \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Note that the 2×2 symmetric matrix on the LHS of this last equation is $\mathbf{X}^T\mathbf{X}$, where $\mathbf{X}$ is the 10×2 design matrix given atop the next page:

X = $\begin{bmatrix} 1 & 1 & 1 & \ldots & 1 \\ 10 & 15 & 15 & \ldots & 30 \end{bmatrix}^T$; the 2×1 vector estimator $\hat{\beta}$ = $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$, the 2×1 vector

$\begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$ on the RHS is equal to $\mathbf{X^T Y}$, and **Y** is a 10×1 = n×1 response vector. Clearly,

the solution vector is given by $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \mathbf{(X^T X)^{-1}(X^T Y)} = \mathbf{(X^T X)^{-1}} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$.

The 1st normal equation gives rise to $\hat{\beta}_0 + \hat{\beta}_1 \overline{x} = \overline{y} \rightarrow \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \rightarrow$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \overline{y} - \hat{\beta}_1 \overline{x} + \hat{\beta}_1 x_i \quad \rightarrow \hat{y}_i = \overline{y} + \hat{\beta}_1 (x_i - \overline{x}). \qquad (53a)$$

Note that $\hat{y}_i$ is called the fitted value of the LS model, and if the slope $\hat{\beta}_1$ is zero, then the best fit for each y_i is $\overline{y}$. Substituting $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$ into

the 2nd normal equation $\left(\hat{\beta}_0 \sum_{i=1}^{10} x_i + \hat{\beta}_1 \sum_{i=1}^{10} x_i^2 = \sum_{i=1}^{n} x_i y_i\right)$ yields

$$(\overline{y} - \hat{\beta}_1 \overline{x}) \sum_{i=1}^{10} x_i + \hat{\beta}_1 \sum_{i=1}^{10} x_i^2 = \sum_{i=1}^{n} x_i y_i \longrightarrow$$

$$(-\hat{\beta}_1 \overline{x}) \sum_{i=1}^{10} x_i + \hat{\beta}_1 \sum_{i=1}^{10} x_i^2 = \sum_{i=1}^{n} x_i y_i - \overline{y} \sum_{i=1}^{10} x_i \longrightarrow$$

$$\hat{\beta}_1 \left(\sum_{i=1}^{10} x_i^2 - \overline{x} \sum_{i=1}^{10} x_i\right) = \sum_{i=1}^{n} x_i (y_i - \overline{y}) \longrightarrow$$

$$\hat{\beta}_1 \sum_{i=1}^{10} x_i (x_i - \overline{x}) = \sum_{i=1}^{n} x_i (y_i - \overline{y}) - \overline{x} \sum_{i=1}^{n} (y_i - \overline{y}) \longrightarrow$$

$$\hat{\beta}_1 \sum_{i=1}^{10} (x_i - \overline{x})^2 = \sum_{i=1}^{n} [(x_i - \overline{x})(y_i - \overline{y})] = S_{xy} \rightarrow$$

$\hat{\beta}_1 = S_{xy} / S_{xx}$, where $S_{xx} = \sum_{i=1}^{10} (x_i - \overline{x})^2 = \sum_{i=1}^{n=10} x_i^2 - (\sum_{i=1}^{10} x_i)^2 / n$,

S_{xx} = 4684 – $(208)^2$ /10 = 357.6, $\overline{x}$ = 20.80, and $\overline{y}$ = 182.90,

S_{xy} = $\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{10} x_i)(\sum_{i=1}^{n} y_i)$/10 = 38715 – (208)(1829)/10 $\rightarrow$

S_{xy} = 671.8 $\rightarrow$ $\hat{\beta}_1$ = 671.8 /357.6 = 1.87864 $\rightarrow \hat{\beta}_0$ = $\overline{y} - \hat{\beta}_1 \overline{x}$ = 182.9 – 1.87864×20.8 = 143.824385. Hence, the fitted regression model is:

$$\hat{y} = 143.8244 + 1.87864x. \qquad (53b)$$

Note that $\hat{y}$ is an estimate of the (curve) of regression of y on x, which is given by E(y | x) = $\mu_{y|x} = \beta_0 + \beta_1 x$. For example, when x = 20, then $\hat{E}(y|20)$ = $\hat{\mu}_{y|x=20} = \hat{y}_4$ = 143.8244 + 1.87864×20 = 181.3971 $\rightarrow$ $e_4 = y_4 - \hat{y}_4$ = 182 – 181.3971 = 0.60291 (the 4th residual). Similarly, e_1 = – 2.6107, ..., $e_{10} = y_{10} - \hat{y}_{10}$ = – 0.1834 (the 10th residual error). Further, it is necessary that $\sum_{i=1}^{n} e_i \equiv 0$ for all statistical models. For example, in the case of SLREG $\sum_{i=1}^{n} e_i$ = The sum of all n residuals = $\sum_{i=1}^{n}(y_i - \hat{y}_i)$ = $\sum_{i=1}^{n}[(y_i - \overline{y}) - \hat{\beta}_1(x_i - \overline{x})] \equiv 0$ because $\hat{y}_i = \overline{y} + \hat{\beta}_1 (x_i - \overline{x})$, $\sum_{i=1}^{n}(y_i - \overline{y}) \equiv 0$ and $\sum_{i=1}^{n}(x_i - \overline{x}) \equiv 0$.

A test of Significance in Regression is 1st conducted through an ANOVA Table. To this end, we compute the SS's for the ANOVA.

USS = $\sum_{i=1}^{n} y_i^2$ = 335825 (with n = 10 *df*), CF = $(\sum_{i=1}^{n} y_i)^2/n$ = 1829^2/10=334524.10 (with 1 *df*) $\rightarrow$ $SS_T = SS_{Total} = \sum_{i=1}^{n}(y_i - \overline{y})^2$ =USS – CF (with 9 *df*) = 1300.90. SS(Residuals) =

$\sum_{i=1}^{n} e_i^2$ = 38.83277405 (with n – 2 = 8 *df* because there are 2 constraints; try to determine what the 2 constraints are for bonus points).

Now, SS(Residuals) = $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} [(y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})]^2 =$

$\sum_{i=1}^{n} (y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum_{i=1}^{n} [(y_i - \bar{y})(x_i - \bar{x})] + \hat{\beta}_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 = S_{yy} - 2\hat{\beta}_1 S_{xy} + \hat{\beta}_1^2 S_{xx} =$

$S_{yy} - 2\hat{\beta}_1 S_{xy} + \hat{\beta}_1 (S_{xy}/S_{xx}) S_{xx} \rightarrow$

$$SS_{RES} = SS(\text{Total}) - \hat{\beta}_1 S_{xy} = SS_T - SS_{Model}.$$

Exercise 93. Show that $SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 =$

$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = SS_{RES} + SS_{REG}$ (where REG = Regression),

and as a result $SS_{REG} = SS(\text{Model}) = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \hat{\beta}_1 S_{xy}$.

For Problem 10.1 of Montgomery, $SS_{REG} = SS_{Model} = \hat{\beta}_1 S_{xy} = 1.87864 \times 671.8 =$ 1262.0672. Note that $SS_{REG} + SS_{RES}$ = 1262.0672 + 38.8328 = 1300.90 = $SS_{Total} = SS_T$. The ANOVA Table is given below.

Source	*df*	SS	MS	F_0
Total	9	1300.90		
Model or Regression	1	1262.06723	1262.06723	260.0004
Residuals	8	38.83277	4.8541	$F_{0.01,1,8}$ = 11.25862

Since the *P-value* of the F-test is almost zero, then strongly reject H_0: $\beta_1 = 0$. Thus, the regressor variable x accounts for 97.015% (= SS_{REG}/SS_T) of variability in the response y, i.e., $R^2_{Model} = R^2_{REG}$ = 1262.06723/1300.90 = 97.015%.

To determine how well the model $\hat{y}$ = 143.8244 + 1.87864x fits the 10 ordered pairs (x_i, y_i), we proceed as follows:
SS(Pure Error) = $(171^2 + 175^2 - 346^2/2) + (182^2 + 184^2 + 181^2 - 547^2/3) + (188^2 + 193^2 - 381^2/2)$ = 25.16667 (with 1 + 2 + 1 = 4 *df*). → SS(LOF = Lack of Fit) = SS_{RES} –SS(PE = Pure Error) = 38.832774 – 25.16667 = 13.66611 (with 8 – 4 = 4 *df*). The augmented ANOVA Table is given atop the next page. Since the *P-value* for LOF is $\hat{\alpha}$ = $P(F_{4,4} \geq 0.54302)$ = 0.715623, the regression model $\hat{y}$ = 143.8244 + 1.87864x fits the 10 points extremely well. The Complete ANOVA Table for Problem 10.1 on page 414 of Montgomery is given atop the next page.

The Complete ANOVA for Problem 10.1 on p. 414 of Montgomery 7(e)

Source	*df*	SS	MS	F_0
Total	9	1300.90		
Model	1	1262.06723	1262.06723	260.0004
Residuals	8	38.83277	4.8541	$F_{0.01,1,8}$ = 11.25862
Pure Error (PE)	4	25.16667	6.29167	
Lack of Fit (LOF = x^2, x^3, x^4, x^5)	4	13.66611	3.41653	0.54302

Exercise 94. Let $Y_1 = \sum_{i=1}^{n} a_i U_i$ and $Y_2 = \sum_{j=1}^{m} b_j W_j$ be two linear combinations with $E(U_i) = \mu_{1i}$, $E(W_j) = \mu_{2j}$, and $COV(U_i, W_j) = \sigma_{ij}$. Show that the $COV(Y_1, Y_2)$ = $\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \sigma_{ij}$.

As an application of the above Exercise, suppose U_1, U_2, W_1, W_2, and W_3 are variates (or rvs) with $COV(U_i, W_j) = (-1/2)^{i-j}$. Our objective is to compute the $COV(-U_1 + 3U_2, 2W_1 - W_2 - 2W_3) = COV(Y_1, Y_2) = (-1)(2)(-1/2)^0 + (-1)(-1)(-2) + (-1)(-2)(-1/2)^{-2} + (3)(2)(-1/2) + (3)(-1)(1) + (3)(-2)(-1/2)^{-1}$ →

$COV(Y_1, Y_2) = -2 - 2 + 8 - 3 - 3 + 12 = 10$.

Exercise 95. Show the following properties of SLREG estimators: (you must assume that $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$). (1) Both $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased. (2) $V(\hat{\beta}_1) = \sigma^2_\epsilon / S_{xx}$, $COV(\bar{y}, \hat{\beta}_1) = 0$, $V(\hat{\beta}_0) = [\frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}}]\sigma^2_\epsilon$, $COV(\hat{\beta}_0, \hat{\beta}_1) = -\bar{x}\sigma^2_\epsilon / S_{xx}$, $E(e_i) = 0$, $V(e_i) = [\frac{n-1}{n} - \frac{(x_i - \bar{x})^2}{S_{xx}}]\sigma^2_\epsilon$, and $V(\hat{y}_0) = V(\hat{\beta}_0 + \hat{\beta}_1 x_0) = [\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}]\sigma^2_\epsilon$. (3) $E(SS_{RES}) = (n-2)\sigma^2_\epsilon$, which shows that MS_{RES} is an unbiased estimator of σ^2_ϵ because $MS_{RES} = SS_{RES}/(n-2)$.

Exercise 96. The strength of paper, y, used in the manufacture of cardboard boxes is related to % of hardwood in the original pulp (x). Under controlled conditions, a pilot plant manufactures 14 samples, each from a different batch of pulp and measures the TS. The resulting data is provided atop the next page. (a) Fit a SLR model to the data and check your answer via Minitab. (b) Provide the complete ANOVA table (with LOF test) and conduct all tests of significance. (c) After studying through pp. 163-166 of these notes, obtain the 95% CIs for β_0, β_1 and $E(Y \mid x = 2.5)$. (d) Obtain a 95% prediction interval (PI) for a single future observation made at $x_0 \equiv 2.5$. Then, compare the length of your PI with that of the corresponding CI and draw statistical conclusions.

x_i	1.0	1.5	1.5	1.5	2.0	2.0	2.2
y_i	101.4	117.4	117.1	106.2	131.9	146.9	146.8
x_i	2.4	2.8	2.8	3.0	3.0	3.2	3.3
y_i	133.9	135.1	145.2	134.3	144.5	143.7	146.9

Exercise 97. Show that in SLREG with n_i observations at m distinct levels SS_{RES} =

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}[(y_{ij}-\overline{y}_{i.}) + (\overline{y}_{i.}-\hat{y}_i)]^2$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\overline{y}_{i.})^2 + \sum_{i=1}^{m}n_i(\overline{y}_i-\hat{y}_i)^2 = \text{SS(PE)} + \text{SS(LOF)}.$$

Note that the terminology regression is due to Francis Galton who first observed (in the late 19th century) that the height of sons tended to be closer to the mean height of species than their fathers, i.e., the height of off-springs tended to regress back toward the mean height of the entire species, no matter what the height of fathers were. This is called regressing toward the mean.

Statistical Inference in SLREG

Recall that if a rv is N(μ, unknown σ^2), then the statistic [rv – E(rv)]/*se*(rv) ~ T_ν, where ν is the *df* of the *se*(rv). Therefore, to test H_0: $\beta_1 = 0$ VS $H_1 : \beta_1 \neq 0$,
we may use the fact that

$$\frac{\hat{\beta}_1 - E(\hat{\beta}_1)}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{MS_{RES}/S_{xx}}}$$

has a Student's t-distribution with ν = n –2 *df*. We now apply the above t-test to the regression model of Problem 10.1 of Montgomery. We 1st compute the $se(\hat{\beta}_1)$ = $\sqrt{MS_{RES}/S_{xx}}$ = $\sqrt{4.8541/357.60}$ = 0.11651 (see Exercise 95). Since under H_0 the slope, β_1, is hypothesized to be zero, then our null test statistic becomes

$$t_0 = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{MS_{RES}/S_{xx}}} = \frac{1.87864-0}{0.11651} = 16.12455,$$

which far exceeds $t_{0.025,8}$ = 2.3060, and hence the null hypothesis of zero slope must be strongly rejected. Note that $(t_0)^2 = F_0$ = 260.0004 of the ANOVA Table on page 161 of these notes. The discrepancy in the 4th decimal is strictly due to rounding.

The next SI (Statistical Inference) is to obtain either a 95% or 99% CI for β_1.

Figure 27 shows the sampling distribution (SMD) of $\frac{\hat{\beta}_1 - \beta_1}{\sqrt{MS_{RES} / S_{xx}}}$ for a confidence level of $1 - \alpha = 0.95$. Figure 27 clearly shows that the Pr($-2.3060 \le T_8 \le 2.3060$) $= \Pr(-2.3060 \le (\hat{\beta}_1 - \beta_1)/ se(\hat{\beta}_1) \le 2.3060) = 0.95$, (note that 2.3060 = $t_{0.025,8}$), or

$$\Pr[\,-2.3060\, se(\hat{\beta}_1) - \hat{\beta}_1 \le -\beta_1 \le -\hat{\beta}_1 + 2.3060\, se(\hat{\beta}_1)\,] = 0.95.$$

In order to solve for β_1 we need to multiply the above 2 inequalities by -1, which results in

$$\Pr[\,\hat{\beta}_1 - 2.3060\, se(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + 2.3060\, se(\hat{\beta}_1)\,] = 0.95.$$

The above equation clearly shows that the 95% confidence limits are $L(\beta_1) = \hat{\beta}_1 - 2.3060\, se(\hat{\beta}_1)$ and $U(\beta_1) = \hat{\beta}_1 + 2.3060\, se(\hat{\beta}_1)$; further, the 95% HCIL (Half-CI-Length) is given by $t_{0.025,8} \times se(\hat{\beta}_1) = 2.3060 \times 0.11651 = 0.2687$ and hence the requisite CI is $\hat{\beta}_1 \pm 0.2687 = (1.6100, 2.1473)$. Since this 95% CI: $1.6100 \le \beta_1 \le 2.1473$, clearly

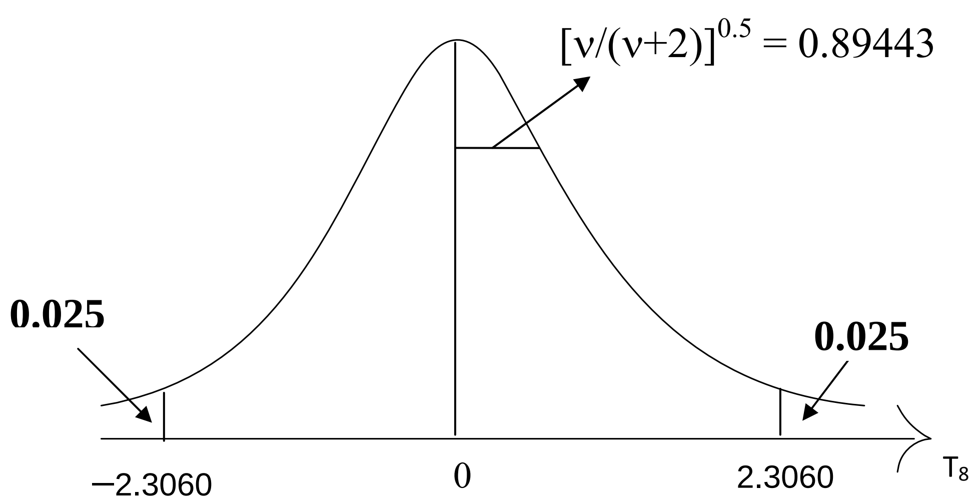

Figure 27. The SMD of $(\hat{\beta}_1 - \beta_1)/ se(\hat{\beta}_1)$

excludes the hypothesized value of zero, it is consistent with the rejection of the above t-test on $\beta_1 = 0$ at the 5% level.

Applying the same procedure as the above to $\hat{\beta}_0$, and using V($\hat{\beta}_0$) = $[\frac{1}{n} + \frac{(\overline{x})^2}{S_{xx}}]\sigma_\epsilon^2$, we obtain the following 95% CI for the parameter β_0.

$$138.00973 \le \beta_0 \le 149.63904$$

Since $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ is a point unbiased estimate of $E(Y \mid x_0)$, then using one of the results of Exercise 95, the 95% CI for $\mu_0 = \beta_0 + \beta_1 x_0$ is given by

$$\hat{y}_0 \pm 2.3060 \times se(\hat{y}_0) = \hat{\beta}_0 + \hat{\beta}_1 x_0 \pm 2.3060 \times \sqrt{[\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}] \times MS_{RES}} \quad .$$

As an example, if we wish to obtain a 95% CI for E(Y|x = 22) = $\beta_0 + 22\beta_1$, the above formula yields 185.1544 $\pm$ 2.3060$\times$0.7106 = (183.5158, 186.7930), or $183.5158 \le \beta_0 + 22\beta_1 \le 186.7930$. Note that this last 95% CI does not have a 0.95 Pr of containing the mean of Y at x = 22; that Pr is either 0 or 1. The length of this CI is 2$\times$2.3060$\times$ $se(\hat{\beta}_0 + \hat{\beta}_1 x_0)$ = 3.2773.

PREDICTION INTERVAL FOR a FUTURE OBSERVATION at a SPECIFIED X_0

As an example, reconsider the regression model of Eq. (53b) on p. 158 of these notes. Suppose we are to make N (the most common value of N is one observation in the future) observations at x_0 = 22% hardwood concentration in the future. Let y_0 be an actual future observation (not yet observed and hence is a rv). Clearly, the best single point forecast from our model is $\hat{y}_0$ = 143.8244 + 1.87864$\times$22 = 185.1545. How do we use our rvs y_0 and $\hat{y}_0$ to obtain a prediction interval for y_0?

To accomplish this task, we must 1st define the forecast error rv as $\psi = y_0 - \hat{y}_0$. Because E(ψ) = E($\beta_0 + \beta_1 x_0 + \varepsilon_0 - \hat{\beta}_0 - \hat{\beta}_1 x_0$) = 0, and the variance of forecast error at N = 1 is given by

$$V(\psi) = V[y_0 - \overline{y} - \hat{\beta}_1 (x_0 - \overline{\mathbf{x}})] = V(y_0) + V(\overline{y}) + (x_0 - \overline{x})^2 V(\hat{\beta}_1)$$

Thus, V(ψ) = $[1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}]\sigma_\epsilon^2$. Since $\psi = y_0 - \hat{y}_0$ is a LC (Linear Combination) of NID rvs, then the forecast error ψ is also N(0, V(ψ)), and as a result the rv

$$\frac{\Psi - E(\Psi)}{se(\Psi)} = \frac{y_0 - \hat{y}_0}{\sqrt{[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}] \times MS_{RES}}}$$

has a student's t-distribution with (n – 2) *df*. Therefore, for N = 1 future observation at x_0 = 22 we obtain the following 95% prediction Pr statement:

$$Pr[-2.3060 \le \frac{y_0 - \hat{y}_0}{\sqrt{[1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}] MS_{RES}}} \le 2.3060] = 0.95$$

Rearranging the inequality inside the above brackets yields the 95% PI (prediction interval) for y_0.

$$Pr[\hat{\beta}_0 + \hat{\beta}_1 x_0 - 2.3060 se(\psi) \le y_0 \le \hat{\beta}_0 + \hat{\beta}_1 x_0 + 2.3060 se(\psi)] = 0.95.$$

To obtain the actual 95% PI, we 1^{st} compute the *se*(ψ).

$$se(\psi) = se(y_0 - \hat{y}_0) = \sqrt{[1 + \frac{1}{10} + \frac{(22 - 20.8)^2}{357.6}] 4.8541} = 2.3150$$

Next, we insert $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 = 185.1545$ and the *se*(ψ) into the above Pr statement.

$$Pr[-2.3060 \le \frac{y_0 - 185.1545}{2.3150} \le 2.3060] = 0.95$$

Finally, rearranging the inequality inside the above brackets results in Pr[185.1545 $-2.3060 \times 2.3150 \le y_0 \le 185.1545 + 5.3383$] = 0.95 $\rightarrow$

$$Pr(179.8160 \le y_0 \le 190.4927] = 0.95.$$

The length of the above prediction band is 10.6766. Note that, unlike a CI, the above PI actually has a Pr of 0.95 to contain a future observation because y_0 is still a random variable, as it has not yet been observed. Further, the length of the 95% PI is always larger than the corresponding 95% CI because it contains 2 sources of error unlike a CI.