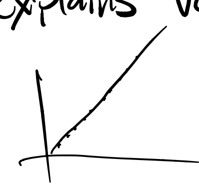
Coefficient of Determination

R^2 ; used to judge the adequacy of a regression model

$$R^2 = 1 - \frac{SS_E}{SS_T}$$

$R^2 \to 0$, model does <u>not</u> explain variability of data

$R^2 \to 1$, model <u>perfectly</u> explains variability of data



∴ for O_2 purity example:

$$R^2 = 1 - \frac{21.25}{173.4} = 0.877$$

or 87.7%

∴ very high value; model is <u>great</u> at explaining variability of data (87.7% of it explained by model)

Correlation

∴ what's the difference between <u>correlation</u> and <u>regression</u>?

∴ in correlation, X and Y are both random variables, rather than X being control and Y being response

.. to determine the degree to which random variables X and Y are correlated, define <u>correlation coefficient</u>, ρ

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

, where σ_{XY} is the <u>covariance</u> between X and Y

ρ and σ_{XY} are population parameters we don't know!

.. the point estimator for ρ is R, where $R = +\sqrt{R^2}$

$\uparrow$ coefficient of determination!

∴ useful computational relationships :

$$R^2 = \hat{\beta}_1^2 \frac{S_{XX}}{SS_T} = \frac{\hat{\beta}_1 S_{XY}}{SS_T}$$

denominator term of $\hat{\beta}_1$ (S_{XX})

numerator term of $\hat{\beta}_1$ (S_{XY})

$$R = \frac{S_{XY}}{\sqrt{S_{XX} SS_T}}$$

Hypothesis Tests on ρ

∴ to test $H_0 : \rho = 0$ → totally uncorrelated!

$H_1 : \rho \neq 0$

$$T_0 = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

critical values: $\pm t_{\alpha/2, n-2}$

ex: pull strength of wire bond in semiconductor package (Y) vs. wire length (X).

given: $S_{XX} = 698.6$

$S_{XY} = 2028$

$SS_E = 220.1$

$SS_T = 6106$

$n = 25$

test $H_0: \rho = 0$

$H_1: \rho \neq 0$ @ $\alpha = 0.05$

- need R!

$$R^2 = 1 - \frac{SS_E}{SS_T} = 1 - \frac{220.1}{6106} = 0.9640$$

$$\therefore R = \sqrt{.9640} = \underline{0.9818}$$

(very strongly correlated!)

$$t_0 = \frac{.9818\sqrt{25-2}}{\sqrt{1-0.9640}} = 24.82$$

(huge)

critical values: $\pm t_{\alpha/2,\, n-2}$

$\pm t_{.025,\, 23} = 2.069$

Very strongly reject H_0;

data suggests pull strength and wire bond length are not uncorrelated!

.. to test hypotheses

some reference value of correlation coefficient

$$H_0 : \rho = \rho_0$$
$$H_1 : \rho \neq \rho_0$$

.. an approximate test statistic for $n \geq 30$ is :

$$Z_0 = \left(\tanh^{-1} R - \tanh^{-1} \rho_0\right)\sqrt{n-3}$$

inverse hyperbolic tangent function !

critical values : $\pm Z_{\alpha/2}$

C.I. on ρ

.. again, approximate for $n \geq 30$:

$$\tanh\left(\tanh^{-1} R - \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right) < \rho < \tanh\left(\tanh^{-1} R + \frac{Z_{\alpha/2}}{\sqrt{n-3}}\right)$$

· · Wire bond example : Write 95% C.I. on ρ

$$\tanh^{-1} R = \tanh^{-1}(.9818) = 2.345$$

$$Z_{\alpha/2} = Z_{.025} = 1.960$$

(if in doubt : bottom row of t-table!)

$$\sqrt{n-3} = \sqrt{22}$$

$$\tanh\left(2.345 - \frac{1.96}{\sqrt{22}}\right) < \rho < \tanh\left(2.345 + \frac{1.96}{\sqrt{22}}\right)$$

$$\tanh(1.927) < \rho < \tanh 2.763$$

ALWAYS USE RADIAN MODE IN CALCULATOR!

$$0.9585 < \rho < 0.9921$$