1.

Explain why you would or wouldn't agree with each of the following statements:

(a) $\mathrm{Cov}(Y, X)$ and $\mathrm{Cor}(Y, X)$ can take values between $-\infty$ and $+\infty$.

(b) If $\mathrm{Cov}(Y, X) = 0$ or $\mathrm{Cor}(Y, X) = 0$, one can conclude that there is no relationship between Y and X.

(c) The least squares line fitted to the points in the scatter plot of Y versus $\hat{Y}$ has a zero intercept and a unit slope.

Disagree

(a) $Cov(x, Y)$ between $-\infty$ and ∞. $Corr(x, Y)$ between -1 and $+1$

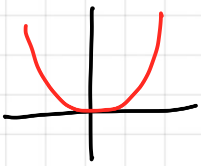
공분산은 서로 다른 변수들 사이에 얼마나 의존하는지를 수치적으로 표현하며, 두 변수들 사이 선형 관계의 관련성을 측정 표준화된 값을 가지지 X.

$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

$$-1 \le Cor(X, y) = \frac{Cov(x, y)}{\sigma_x \sigma_y} \le 1$$

Disagree

(b) NO! 예를 들어



$y = x^2$ 이라는 함수에서 x와 y의 서로 선형 관계가 없으므로 $cov(x, y) = 0 = cov(y, x)$ 갖음 but x, y 아무관계 없다고 보기 어려움 why? 2차 관계 보임

$Cor(X, Y) = 0 \Rightarrow Cov(X, Y)$ 의미

X와 Y와 linear 하지 않음을 의미, 전혀 관계가 없다는 것이 아님. (No linear relation 이면 ok.)

(c)

$$Y = \hat{Y} + e$$

$$(\hat{Y} = \beta_0 + \beta_1 X)$$

$\therefore$ $\hat{Y}$ 과 Y 간의 관계를 나타낸 직선의 절편은 0 이 아닐 수 있다.

2.

When fitting a simple linear regression model $Y = \beta_0 + \beta_1 X + \varepsilon$ to a set of data using the least squares method, suppose that $H_0 : \beta_1 = 0$ was not rejected. This implies that the model can be written simply as: $Y = \beta_0 + \varepsilon$. The least squares estimate of β_0 is $\hat{\beta}_0 = \bar{y}$. (Can you prove that?)

(a) What are the ordinary least squares residuals in this case?

(b) Show that the ordinary least squares residuals sum up to zero.

(a) residual (e_i) = Actual − Predicted $= y_i - \hat{y}_i$

$= y_i - \bar{y}$

(b) $\sum_{i=1}^{n}(y_i - \bar{y}) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \bar{y} = n\bar{y} - n\bar{y}$ $\left(\because \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}\right)$

$= 0$

3.

Table 2.9 Regression Output for Computer Repair Data

Variable	Coefficient	s.e.	t-Test	p-value
Constant	4.162	3.355	1.24	0.2385
Units	15.509	0.505	30.71	< 0.0001

1)

Using the regression output in Table 2.9, test the following hypotheses using $\alpha = 0.1$:

(a) $H_0 : \beta_1 = 15$ versus $H_1 : \beta_1 \neq 15$
(b) $H_0 : \beta_1 = 15$ versus $H_1 : \beta_1 > 15$

→ t검정 통계량 필요

(c) $H_0 : \beta_0 = 0$ versus $H_1 : \beta_0 \neq 0$
(d) $H_0 : \beta_0 = 5$ versus $H_1 : \beta_0 \neq 5$

2)

Using the regression output in Table 2.9, construct the 99% confidence interval for β_0.

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$\hat{\beta}_0 = 4.162$

$\hat{\beta}_1 = 15.509$

⊕ $n = 14$

$p = 1$

$\therefore df = n-p-1$
$= 14-1-1$
$= 12$

$t_{0.1/2, 12} = 1.78$

$t_{0.1, 12} = 1.36$

1)

(a) Test statistics : $\dfrac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} = \dfrac{15.509 - 15}{0.505} = 1.0079$ p-value : $P(t_{12} > 1.0079) * 2 = 0.333$

$0.333 > 0.1$ not reject null hypothesis

(b) Test statistics : (a)와 같음 p-value : $P(t_{12} > 1.0079) = 0.1667 > 0.1$

not reject null hypothesis

(c) Test statistics : $\dfrac{\hat{\beta}_0 - \beta_0}{s.e.(\hat{\beta}_0)} = \dfrac{4.162 - 0}{3.355} = 1.2402$ p-value : $P(t_{12} > 1.2402) \cdot 2 = 0.2385 > 0.1$

not reject null hypothesis

(d) Test statistics : $\dfrac{4.162 - 5}{3.355} = -0.2498$ p-value : $P(t_{12} > |-0.2498|) \cdot 2 = 0.807 > 0.1$

not reject null hypothesis

2)

$(\hat{\beta}_0 - t_{\frac{\alpha}{2}}(n-2)\, s.e.(\hat{\beta}_0),\ \hat{\beta}_0 + t_{\frac{\alpha}{2}}(n-2)\, s.e.(\hat{\beta}_0))$

$(4.162 - 3.0545 \cdot 3.355,\ 4.162 + 3.0545 * 3.355)$ critical value ≒ 3.0545