(1, 1pt) 12 (=1+2+9) eggs.

(2, 1pt) The correlation 0.26 is the correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$. It is positive because in the data the 2 "variables" LogL and LogW are *negatively* correlated (i.e., -0.257).

(3, 2pts) The test statistic is

$$T = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 3}{se(\hat{\beta}_1 + \hat{\beta}_2)} = \frac{0.728 + 1.811 - 3}{se(\hat{\beta}_1 + \hat{\beta}_2)}.$$

Because $Var(\hat{\beta}_1 + \hat{\beta}_2) = Var(\hat{\beta}_1) + Var(\hat{\beta}_2) + 2Cov(\hat{\beta}_1, \hat{\beta}_2)$, we have

$$se(\hat{\beta}_1 + \hat{\beta}_2) = \sqrt{0.267^2 + 0.546^2 + 2 \times 0.267 \times 0.546 \times 0.26} = 0.6672418.$$

and $T = -0.6909039$. Since $|T| < t_9^{(0.975)} = 2.26$, we cannot reject the null.

(4, 1pt) The C.I. of the prediction on (LogL, LogW)=(1, 0.85) would be wider than that on (0.75, 0.60). It is because the former point is far away from the "center" of (LogL, LogW), which is (0.763, 0.621), while the latter is quite close to the center. Actually, the former prediction is an extrapolation while the latter is an interpolation.

(5, 1pt) False. Model A2 has such a large R^2 (almost being 1) because it is a model without intercept. A high R^2 reported from a model without intercept does not make any sense because it compares the fit to the model $y = \epsilon$.
(Notice that model A2 is a sub-model of model A1 so that the former model has a larger RSS than the latter one.)

(6, 2pts) Let RSS_1, RSS_2, and RSS_3 be the residuals sum of squares obtained from $\hat{\epsilon}_1$, $\hat{\epsilon}_2$, and $\hat{\epsilon}_3$, respectively. Then, $RSS_1 < RSS_2$ because model A2 is a sub-model of model A1 with $\beta_0 = 0$, and $RSS_2 < RSS_3$ because $\hat{\epsilon}_2$ corresponds to the least square estimate (which minimizes RSS) of model A2 while $\hat{\epsilon}_3$ does not.

(7, 2pts) No, the analysis results do not offer enough information to identify the sum of $\hat{\epsilon}_2$. The sum of $\hat{\epsilon}_2$ equals the inner product of $\hat{\epsilon}_2$ and $\mathbf{1}$, but the space Ω spanned by LogL and LogW does not contain the $\mathbf{1}$ vector (otherwise, model A1 is unidentifiable) so that we do not know the angle between the two vectors $\hat{\epsilon}_2$ and $\mathbf{1}$.
Saying model A2 has no intercept is not enough because it is possible that a model has no intercept but the space spanned by its predictors contains $\mathbf{1}$ (check LNp.5-12), under which the sum of $\hat{\epsilon}_2$ would be zero.

(8, 2pts) The model is $LogKN = LogL + 2 \times LogW + \epsilon$, where the mean structure $LogL + 2 \times LogW$ is an offset. Because there is no parameter in the mean structure of this model, the degrees of freedom of $\hat{\epsilon}_3$ is $n - p = 12 - 0 = 12$.

(9, 1pt) Because $df_1 < df_2 < df_3$, we cannot rank the $\hat{\sigma}^2$'s based on the information available even though we know $RSS_1 < RSS_2 < RSS_3$ (from the answer to problem 6). Notice that RSS always decreases (or keeps the same) when more predictors are added, but $\hat{\sigma}^2$=RSS/$(n - p)$ could decrease or increase under the circumstance.

(10, 2pts) Negative. It can be found from Figure 1(b), in which the ellipse has a major axis with a negative slope.
The changing of correlation between $\hat{\beta}_1$ and $\hat{\beta}_2$ from positive in model A1 to negative in model A2 is because under model A2, the off-diagonal component of the 2×2 matrix $(X^T X)^{-1}$ is negative, which is a result of the off-diagonal component of $X^T X$, i.e., the inner product of the two vectors LogL and LogW, being positive. The positivity of the inner product is due to all the data of LogL and LogW being larger than zero.

(11, 2pts) Under model A1, we must test $H_0 : \beta_0 = 0, \beta_1 = 1, \beta_2 = 2$, but the analysis outputs do not offer enough information for us to perform this test.
Under model A2, we can test $H_0 : \beta_1 = 1, \beta_2 = 2$ vs. $H_A : \beta_1 \neq 1$ or $\beta_2 \neq 2$. Because in Figure 1(b), the point $(\beta_1, \beta_2) = (1, 2)$ falls in the 95% confidence region of (β_1, β_2), we cannot reject the hypothesis that hen's eggs are ellipsoids of revolution using this data.
(FYI. Actually, eggs are ovoid rather than elliptical in one cross section. Historians say that this difference between an ovoid and an ellipse held the astronomer Kepler up for several years.)

(12, 1pt) From the scatter plots between Score and each predictors, SES has the strongest linear association with Score because its points fall closest to a straight line.

(13, 1pt) The coefficients of Salary and Education are unexpected, particularly in sign. The most surprising one is Education, because we usually expect that pupils with parents of higher education level would tend to perform better in tests. For Salary, it might not be so surprising because teacher's salary might mostly or partially reflect the cost of living in an area, rather than the teaching quality and/or student achievement.

(14, 2pts) The positive association in the scatter plot of Score and Education would cause a positive slope when we fit the simple regression model Score∼Education, called model B3. However, compared to the coefficient estimate of Education under the larger model B1, the positive slope in model B3 could be regarded as a biased estimate, occurring when Education has a strong collinearity with the other predictors, of which some are important in interpreting Score.

(15, 1pt) One unit increase in mother's education level is *associated* with an average decrease of 1.8109 in pupil's score after adjusted for the other predictors (i.e., Salary, White, SES, and TScore are held constant).

(16, 1pt) The R^2 of model B1 is at least 86% because model B1 contains more predictors than model B2. When we add more predictors, the R^2 would always increase or stay the same.
(FYI. In this case, we actually can obtain the R^2=1−RSS/TSS of model B1 from the analysis outputs. The RSS can be obtained from the $\hat{\sigma}^2$ of model B1 and the TSS can be obtained from the R^2 and $\hat{\sigma}^2$ of model B2.)

(17, 1pt) $+\sqrt{0.86} = +0.9273618$. It is positive because the scatter plot shows a positive linear association.

(18, 2pts) Denote models B1 and B2 by Ω and ω, respectively. To answer this question, we can perform the test of $H_0 : \omega$ vs. $H_A : \Omega \setminus \omega$. The test statistics is

$$F = \frac{(RSS_\omega - RSS_\Omega)/(df_\omega - df_\Omega)}{RSS_\Omega/df_\Omega} = \frac{(2.24^2 \times 18 - 2.07^2 \times 14)/4}{2.07^2} = 1.769481,$$

and the null distribution is $F_{4,14}$ with a mean $14/12 = 1.166667$. Because the test statistic is not too large away from the mean of $F_{4,14}$, we would not reject the null and therefore the data supports this simplification.

(19, 2pts) Without loss of generality, suppose all the predictors have been centered (so that they are orthogonal to the intercept). Let X_1 be the model matrix containing only intercept and SES, and X_2 be the model matrix containing the other predictors except SES. Let $\hat{Y}_\Omega$ be the predictions of Y under models B1, and $\hat{\beta}_{1,\Omega}$ and $\hat{\beta}_{2,\Omega}$ be the coefficient estimates corresponding to X_1 and X_2 respectively under model B1 and $\hat{\beta}_{1,\omega}$ be the coefficient estimates under model B2. Then, we have

$$\begin{aligned}\hat{\beta}_{1,\omega} &= (X_1^T X_1)^{-1} X_1^T Y = (X_1^T X_1)^{-1} X_1^T \hat{Y}_\Omega \ \text{(by } H_\omega Y = H_\omega H_\Omega Y = H_\omega \hat{Y}_\Omega \text{ in LNp.4-7)} \\ &= (X_1^T X_1)^{-1} X_1^T (X_1 \hat{\beta}_{1,\Omega} + X_2 \hat{\beta}_{2,\Omega}) = \hat{\beta}_{1,\Omega} + (X_1^T X_1)^{-1} X_1^T X_2 \hat{\beta}_{2,\Omega}\end{aligned}$$

From this result, we can see that $\hat{\beta}_{1,\omega}$ would be about the same as $\hat{\beta}_{1,\Omega}$ if (i) $X_1^T X_2 \approx 0$ (i.e., nearly orthogonal), (ii) $X_2 \hat{\beta}_{2,\Omega} \approx 0$ (i.e., X_2 is useless in interpreting Y), or (iii) a mix of (i) and (ii).
From the scatter plots, we might expect that (i) holds for some predictors (e.g., Salary and TScore) but definitely not for the predictors White and Education. To check whether (ii) holds for White and Education, we can calculate

$$(\text{coefficient estimate})^2 \times (\text{sample variance of predictor}),$$

which is proportional to the length^2 of the vectors that form $X_2 \hat{\beta}_{2,\Omega}$ (note. it's because the predictors have been centered). The values are

Salary	White	SES	TScore	Education
0.6631576	1.2751316	28.6161129	2.1273524	1.4040311

Because SES has a much larger value than the other predictors, it explain why its coefficient estimates does not change too much in the two models.

(20, 2pts) No, this claim is inappropriate for decision-making or policy-making. The model B2 might be a model suitable for prediction. But, policy-making requires being able to interpret the "true" influence of each predictors, which is not what model B2 can achieve especially when some predictors are strongly correlated.
The following are some acceptable reasons supporting saying no to this claim:

(1) the data could be contaminated by the biases from nonresponse,

(2) SES is already an attempt to weight together several economic variables, including White and Education. It is unreasonable to claim that SES is important but White and Education (they are highly correlated with SES) have no impact on Score,

(3) it is uneasy, or even dangerous, to interpret the coefficients of predictors for an observational data, especially when collinearity exists,

(4) the sample size (i.e., 20) is not large enough for us to judge the influence of the other 4 predictors by using a test.