(2 pts) 1. **No.** (1 pt)  
Subjects were not randomly assigned to groups. (1 pt)

(1 pt) 2. **.72** \[p(\text{not minority}) = 52/72 = .72\]

(3 pts) 3. **.0311** \[z = (50 - 40)/8 = 1.25; p(X \geq 50) = \text{area in smaller portion for } z \text{ of } 1.25 = .1056;\
p(\text{all 3 scores } \geq 50) = .1056(1 -.1056)(1 -.1056) + (1 -.1056)(.1056) = .0299; so p(\text{at least 2 scores } \geq 50) = .0012 + .0299 = .0311\]

(2 pts) 4. **c** [.42 probably overestimates the true correlation because it was computed using extreme groups, omitting scores in the middle of the range of the distribution; the omitted scores generally make smaller contributions to covariance so omitting them leads to overestimation of covariance]

(3 pts) 5. (a) **-3.33** (2 pts) \[\text{sum of cross products of deviations from the means: } (11\!-\!15)(10\!-\!9) + (13\!-\!15)(8\!-\!9) + (16\!-\!15)(11\!-\!9) + (20\!-\!15)(7\!-\!9) = -4+2+2+(-10) = -10;\
cov_{XY} = -10/3 = -3.33\]

(b) **-.46** (1 pt) \[r = \frac{\text{cov}_{XY}}{s_X s_Y} = -3.33/(3.92)(1.83)\]

(1 pt) 6. (b) [not (a) because the intercept is the value of \(\hat{Y}\) when \(X = 0\); (b) is true because after converting to \(z\) scores, each unit on \(X\) (and \(Y\)) is one standard deviation]  
-part mark scheme: 0.5 for not selecting (a) and 0.5 for selecting (b)

(2 pts) 7. (a) **2** (1 pt) [this data point will contribute a large positive cross product to \(\text{cov}_{XY}\)]

(b) **3** (1 pt) [this data point will contribute a large negative cross product to the current positive value of \(\text{cov}_{XY}\) because the \(X\) score is above the \(X\) mean, but the \(Y\) score is below the \(Y\) mean; this will substantially reduce \(\text{cov}_{XY}\)]
(8 pts) 8. (a) \( b = 1.44 \) (1 pt) \([b = .64(9/4) = 1.44]\)

\[ a = 21.20 \] (1 pt) \([a = 50-(1.44)20 = 21.20]\)

(b) \(6.97\) (1 pt) \([s_{Y-Y} = 9\sqrt{(1-.64^2)} (59/58)]\)

standard deviation of \(Y\) scores for any subgroup of subjects in the population sharing a specific \(X\) score (1 pt)

(c) \(47.12\) (1 pt) \(\hat{Y} = 21.20 + (1.44)18 = 47.12\)

(d) \(0.0329\) (3 pts) \(\hat{Y} = 21.20+ (1.44)25 = 57.20; z = (70–57.20)/6.97 = 1.84; area in smaller portion = .0329; note that for people with an \(X\) score of 25, their mean \(Y\) score is estimated to be 57.20, so when working with the distribution of \(Y\) scores for these people, the mean is 57.20 and the value 70 serves as the defined cutoff\)

(2 pts) 9. equal if \(r = 0\) (1 pt) [the regression line would be flat and would pass through \(M_Y\) so the distance between any data point and the regression line would be the same as the distance between that point and \(M_Y\)]

smaller (or 0) if \(r = 1\) (1 pt) [each data point would fall on the regression line, so there would be zero error of prediction for each data point]

(3 pts) 10. (a) move it to the right (closer to the regression line) without any vertical change (2 pts) [to keep \(SS_Y\) the same, the data point cannot change its \(Y\) value, so it must not move up or down; shifting it to the right will reduce its vertical deviation from the regression line, reducing the prediction error for this point and thereby reducing \(SS_{error}\)]

(b) increase (1 pt) [\(r\) increases because this data point is now closer to the regression line; a reduction in \(SS_{error}\) without changing \(SS_Y\) means that \(SS_{Y-Y}\) is increased, so \(SS_{Y-Y}/SS_Y\) (which equals \(r^2\)) is increased]