

PROBLEM 1A

$$P - 1 = 5 - 1 = 4$$

↓ P represents the number of predictors

Dummy Variables need to create to analyze the effect of variable region.

PROBLEM 1b

0 Dummy Variables needed to create to analyze the effect of variable region.

PROBLEM 1c

$$E(Y) = \beta_0 + \beta_1(\text{Age}) + \beta_2 I(\text{Region} = \text{East Europe}) + \beta_3 I(\text{Region} = \text{Asia}) + \beta_4 I(\text{Region} = \text{Sub Saharan Africa}) + \beta_5 I(\text{Region} = \text{South America}) + \beta_6 I(\text{Sex} = \text{Female}) + \beta_7 I(\text{Sex} = \text{Declined to Answer}) + \epsilon_i$$

I Removed Region = North America and Sex = Male, one less for each categorical variable

# PROBLEM 1d

Regression Degrees of Freedom:  $p = 7$   
Error Degrees of Freedom:  $n - (p + 1)$   
 $1000 - (7 + 1)$   
 $= 1000 - 8$   
 $992$

$$F_{(1-\alpha, \underline{p}, \underline{n-p-1})}$$
$$\underline{F(1-\alpha, 7, 992)}$$

## PROBLEM 1e

Model 1:

$$\begin{aligned}
 Y_i = & \beta_0 + \beta_1 (\text{Age}) \\
 & \beta_2 I(\text{Region} = \text{East Europe}) + \\
 & \beta_3 I(\text{Region} = \text{Asia}) + \\
 & \beta_4 I(\text{Region} = \text{Sub Saharan Africa}) + \\
 & \beta_5 I(\text{Region} = \text{South America}) + \\
 & \beta_6 I(\text{Sex} = \text{Female}) + \\
 & \beta_7 I(\text{Sex} = \text{Declined to Answer}) + \\
 & \varepsilon_i
 \end{aligned}$$

$I$  Removed Region = North America and Sex = Male,  
one less for each categorical variable

Model 2:

$$\begin{aligned}
 Y_i = & \beta_0 + \beta_1 (\text{Age}) + \\
 & \beta_2 I(\text{Sex} = \text{Female}) + \\
 & \beta_3 I(\text{Sex} = \text{Declined to Answer})
 \end{aligned}$$

$$\begin{aligned}
 df_S &= n - p_S - 1 \\
 &= 1000 - 3 - 1 \\
 &= 996
 \end{aligned}$$

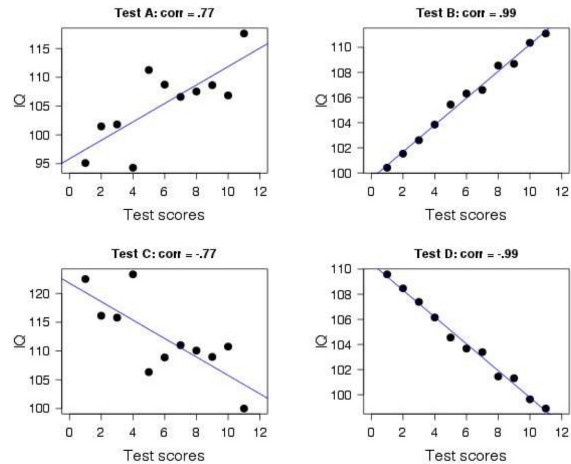
$$\begin{aligned}
 df_L &= n - p_L - 1 \\
 &= 1000 - 7 - 1 \\
 &= 992
 \end{aligned}$$

$$\begin{aligned}
 & \left. \begin{array}{l} df_S \\ df_L \end{array} \right\} F_{df_L - df_S, df_L} \\
 & \quad \quad \quad F_{4, 992}
 \end{aligned}$$

## PROBLEM 2a

### Problem 2

Researchers investigated the relationships between the results of four different visual tests (Test A, B, C, and D) and IQ. They randomly selected 44 subjects and split them into 4 groups of equal size. Each subject had his/her IQ evaluated and was then given one of the four tests. For each test, the researchers made a scatter plot of the subjects' test scores and their IQs and computed the corresponding sample correlation.



a) [Select one correct answer.] Which of the following statements correctly describes the above figure? (2 points)

- I. Higher scores on Tests C and D correspond to higher IQ.
- II. Higher scores on Tests A and C correspond to higher IQ.
- III. The relationship between scores on Test A and IQ is stronger than the relationship between scores on Test D and IQ.
- IV. Subjects with similar scores on Test A have a larger spread of IQs than subjects with similar scores on Test B.

## Problem 2b

Researchers fit a simple linear regression relating IQ to test score using data from one of the four groups of subjects. Here is part of the regression output:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	121.7847	2.9582	41.169	1.47e <sup>-11</sup>
Test score	-1.6031	0.4362		

Model Information:

Residual standard error: 4.574 with 9 degrees of freedom
R-squared: 0.5929
Adjusted R-squared: 0.5538
F-statistic: 13.51 with 1 and 9 DF and a p-value of 0.00511

b) [Select one correct answer.] Which test's scores did researchers use as the predictor in the regression? (4 points)

- I. Test A
- II. Test B
- III. Test C
- IV. Test D

## Problem 2c

**2c:**

We know researchers fit a simple linear regression.

Compute & Interpret 95% CI

For the slope from the regression output above:

A  $(1-\alpha)100\%$  Confidence Interval for the true slope is given by:

$$\hat{\beta}_1 \pm t_{n-2, 1-\alpha/2} \cdot se(\hat{\beta}_1),$$

$$\text{where } se(\hat{\beta}_1) = \sqrt{MSE / \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_1 \pm t_{n-2, 1-\alpha/2} \cdot se(\hat{\beta}_1)$$

$$-1.6031 \pm t_{11-2, 1-0.05/2} \cdot 0.4362$$

$$t_{9, 0.975}$$

$$qt(.975, 9)$$

$$-1.6031 - 2.26 \cdot 0.4362 \} -2.5889$$

$$-1.6031 + 2.26 \cdot 0.4362 \} -0.62$$

**Interpretation:** With 95% Confidence Interval, we estimate that the IQ score decreases somewhere between 0.62 and 2.5889 for each additional 1 point increase in test score.

## Problem 2d

**2d:** Two-Sided Hypothesis test

$$H_0: \beta_{\text{Test Score}} = 0$$

$$H_1: \beta_{\text{Test Score}} \neq 0$$

Test statistics has following distribution:

$$t = \frac{\hat{\beta}_1 - \beta_{10}}{\text{se}(\hat{\beta}_1)} \sim t_{n-2} \text{ under } H_0$$

$$t = \frac{-1.6031 - 0}{0.4362} = -3.675$$

$$t = -3.675$$

$$|-3.675| > 2.262 \leftarrow$$

$$\begin{aligned} & t_{11-2, 1-\alpha/2} \\ & t_{9, 1-0.05/2} \\ & t_{9, 0.975, \text{ under } H_0} \\ & qt(.975, 9) \\ & 2.262 \end{aligned}$$

**Interpretation:** At 5% Significance level,  $|-3.675| > t_{9, .975} = 2.262$ , we reject the null and conclude that there is a significant linear association between test score and IQ score.