

Your Name (Please **PRINT CLEARLY**):

Your *Signature*:

For each question, *choose the best answer*.

1. Consider *categorical data* in the 4×3 contingency table (with row totals, column totals, and the grand total shown in **bold**)

19	14	9	42
5	10	6	21
91	41	9	141
15	20	9	44
120	85	33	238

For the (1, 1) cell with *observed count* $O_{11} = 19$, the *expected count* is $E_{11} =$

- (A) 238 (B) $\frac{120+42}{238}$ (C) $\frac{120-42}{238}$ (D) 19×238 (E) $\frac{120 \times 42}{238}$

2. (Continuation) For testing homogeneity (identical distributions) or independence using the above data, the test statistic

$$\sum_{i=1}^4 \sum_{j=1}^3 \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

has approximate distribution *chi square* with degrees of freedom

- (A) $(4 - 1) + (3 - 1) = 5$ (B) $(4 - 1) \times (3 - 1) = 6$ (C) $4 + 3 = 7$

3. (Continuation) If the three column totals **120**, **85**, and **33** are sample sizes *fixed in advance*, then the hypothesis of interest is

- (A) homogeneity (identical distributions) of 3 populations
 (B) independence of 2 variables
 (C) independence of 12 variables

4. (Continuation) If the sample size is the grand total **238**, and the row totals and column totals are random, then the hypothesis of interest is

- (A) homogeneity (identical distributions) of 3 populations
 (B) independence of 2 variables
 (C) homogeneity (identical distributions) of 12 populations

5. In the linear regression model for the relation between independent variable X and dependent variable Y ,

- (A) The conditional means $E(Y | X = x)$ for different values of x fall on a straight line.
 (B) All data points (X_i, Y_i) , $1 \leq i \leq n$, always fall on a straight line.
 (C) X and Y are assumed to be independent.