

Cho  $X$  và  $Y$  là hai biến ngẫu nhiên (thuộc tính), lần lượt là tuổi và cân nặng. Xét mẫu cỡ  $n = 20$  của hai biến này

$$x_i = (69, 74, 68, 70, 72, 67, 66, 70, 76, 68, 72, 79, 74, 67, 66, 71, 74, 75, 75, 76)$$

$$y_i = (153, 175, 155, 135, 172, 150, 115, 137, 200, 130, 140, 265, 185, 112, 140, 150, 165, 185, 210, 220)$$

- a) **(2đ)** Tìm trung bình, trung vị, và mode cho mẫu của  $X$ .
- b) **(2đ)** Tính phương sai mẫu của  $Y$ .
- c) **(2đ)** Tìm kỳ vọng mẫu, phương sai mẫu của  $X$ . Từ đó vẽ đồ thị phân bố chuẩn khớp với mẫu của  $X$  (gồm đồng thời biểu đồ cột của mẫu và đồ thị của hàm phân bố).
- d) **(2đ)** Tìm tỷ lệ người có tuổi  $\geq 80$ .
- e) **(2đ)** Tính kỳ vọng mẫu hai chiều  $\hat{\mu} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$ , và ma trận hiệp phương sai mẫu  $\widehat{\Sigma}$  của hai biến. Xác định hệ số tương quan mẫu giữa tuổi và cân nặng.
- f) **(1đ)** Vẽ biểu đồ phân tán để mô tả mối liên quan giữa tuổi và cân nặng.
- g) **(4đ)** Vẽ các biểu đồ
  - Biểu đồ hộp cho mẫu của  $X$ .
  - Biểu đồ cột cho mẫu của  $Y$ , gồm 5 cột chia đều khoảng giá trị từ lớn nhất tới nhỏ nhất trong mẫu của  $Y$ .
  - Biểu đồ phân vị cho mẫu của  $X$ .
  - Biểu đồ phân vị kép cho mẫu của hai biến.

**Ghi chú:** phần in nghiêng thực hiện bằng Python, các phần còn lại (1) tính tay trình bày ra giấy kiểm tra và (2) tính bằng Python trong đó dữ liệu đọc từ file “test.csv”. Sinh viên được sử dụng tài liệu, kể cả Internet. Sinh viên không được trao đổi và tuyệt đối không dùng các ứng dụng có chức năng liên lạc.