

Trường Chính sách công và Quản lý Fulbright

Học kỳ Xuân năm 2022

THỐNG KÊ VÀ PHÂN TÍCH DỮ LIỆU

BÀI TẬP 2

Ngày Phát: Thứ hai 14/02/2022

Ngày Nộp: Thứ tư 16/02/2022

Học viên nộp bản điện tử bao gồm Word và Excel trước 8h20 trên ứng dụng Microsoft Teams

Học viên cần tính toán trên Excel và trình bày rõ ràng trên Word

Câu 1. (Hiểu về suy luận theo phân phối) (20 điểm)

Dưới đây là số liệu mô phỏng chiều cao của 1000 người dân Việt Nam, học viên trả lời một số câu hỏi sau (**File BT2_c1**):

Bảng 1. Chiều cao mô phỏng 1000 người Việt Nam

Chiều cao (cm)	Tần số	Xác suất
<150	25	2.5%
151-155	25	2.5%
156-160	150	15%
161-165	180	18%
166-170	240	24%
171-175	180	18%
176-180	150	15%
181-185	25	2.5%
>185	25	2.5%

Nguồn: ban giảng viên mô phỏng

- a) Hãy vẽ đồ thị tần suất (histogram) với dữ liệu bảng trên. Nhận xét về phân phối của chiều cao Việt Nam. (10 điểm)

- b) Có bao nhiêu người (%) có chiều cao trong khoảng từ dưới 175cm; trên 155cm và trong khoảng từ [151, 185]? (5 điểm)
- c) Có ý kiến cho rằng: “Chiều cao trung bình của người Việt Nam là 153 cm”. Với dữ liệu tại bảng trên, phát biểu này có đáng tin không? (5 điểm)

Câu 2. (Suy luận phân phối với mẫu lớn) (25 điểm)

Có nhiều nghiên cứu chỉ ra mối quan hệ giữa hiệu quả công việc và làm việc quá giờ. Theo đó, năng suất sẽ bắt đầu giảm nếu số giờ làm việc trung bình đến ngưỡng 60 giờ/tuần, tương đương khoảng 18 giờ làm thêm/tăng ca mỗi tuần. Công ty A thực hiện khảo sát 200 nhân viên về số giờ làm tăng ca mỗi tuần và dữ liệu file đính kèm (**BT2_c2**) (EW), được lựa chọn ngẫu nhiên. Hãy ước tính 95% độ tin cậy số giờ làm tăng ca của công ty A mỗi tuần qua một số gợi ý sau:

- a) Kiểm tra phân phối của dữ liệu được cung cấp (trung bình, độ xiên và kurtosis) và nhận xét. (10 điểm)
- b) Dựa trên các thông tin trên, xác định chiến lược (phân phối nào) xác định 95% độ tin cậy của EW tại công ty A. Với các số liệu về giờ làm tăng ca này, liệu có đáng quan ngại về văn hóa làm việc tại công ty A không? (15 điểm)

Câu 3 (Xác suất) (25 điểm)

Xổ số miền Bắc mỗi ngày có 27 kết quả từ giải đặc biệt đến giải 7. Một trò chơi yêu cầu bạn dự đoán đúng 1 số trong khoảng từ 00 – 99 mà số đó sẽ xuất hiện ở 2 chữ số cuối cùng trong 27 kết quả.

- a) Năm nay là năm 2022, anh Tiền dự đoán xổ số sẽ về 22 trong hôm nay. Theo bạn xác suất để kết quả hôm đó có 22 là bao nhiêu? (5 điểm)
- b) Từ kết quả có được, anh Tiền cảm thấy xác suất để thắng khá thấp nên bạn quyết định chọn thêm số 20. Xác suất để kết quả có ít nhất một trong hai số 20 và 22 là bao nhiêu? (5 điểm)
- c) Xác suất để kết quả có cả 20 và 22 là bao nhiêu? (7 điểm)
- d) Khi biết chắc chắn kết quả có 22 thì xác suất để kết quả có 20 là bao nhiêu? (8 điểm)

Câu 4. (30 điểm)

Một công ty về sản xuất ô tô (X), ứng dụng công nghệ mới - sử dụng bức xạ mặt trời để sạc pin cho loại ô tô mới đang nghiên cứu. Một trong những vấn đề quan trọng mà họ phải đối mặt là tính ổn định về khả năng hấp thụ bức xạ mặt trời trong các loại thời tiết khác nhau. CEO sẽ quyết định

tiếp tục dự án nghiên cứu này và thương mại hóa trong năm nay nếu như tiêu chí ổn định được đảm bảo. Thiết kế nghiên cứu 1: Một nhóm nghiên cứu đo lường số km chạy được trong 1 phút hấp thụ bức xạ mặt trời của 15 ô tô mới loại này, tại hai loại thời tiết ***có nắng*** và ***không nắng***, được ghi chép tại file **BT2_c4**.

- a) Thiết kế nghiên cứu 2: Một nhóm nghiên cứu khác được thuê tiến hành nghiên cứu như sau: (i) lựa chọn ngẫu nhiên 30 chiếc ô tô loại này và chia làm 2 nhóm bất kỳ; (ii) nhóm 1 đo số km chạy trong 1 phút sặc trong thời tiết *không có nắng* rồi ghi nhận kết quả, nhóm 2 đo lường trong thời tiết *có nắng* và ghi nhận kết quả. Theo anh/chị cách thiết kế nghiên cứu nào khả hiệu quả hơn? Vì sao? (5 điểm)
- b) (Thiết kế mẫu ghép nối - paired-sample design) Từ dữ kiện theo cách thiết kế thứ nhất, liệu có sự khác biệt trong số km chạy được mỗi 1 phút sặc giữa thời tiết có nắng và không nắng? Trả lời theo một số gợi ý sau: (10 điểm)
- ✓ Xác định sự khác biệt giữa số km chạy được mỗi ô tô chạy trong hai loại thời tiết (gọi tắt là D)
 - ✓ Xác định giả thuyết H_0 , chiến lược kiểm định giả thuyết và các giả định đi kèm (nếu có)
 - ✓ Xác định hệ số kiểm định và kết luận
- c) Xác định 95% độ tin cậy về sự khác biệt về số km chạy được trong 1 phút sặc bởi bức xạ mặt trời của loại ô tô mới này giữa hai loại thời tiết (D) trong tổng thể. (5 điểm)
- d) (Thiết kế đo lường độc lập - Independent Measures Design) Giả sử CEO công ty cho rằng việc thiết kế nghiên cứu thứ 2 (theo câu a) là không nghiêm trọng; nên thuê nhóm nghiên cứu này tiếp tục tiến hành đo lường và kiểm định. Giả sử kết quả nghiên cứu được đo lường tương tự như dữ liệu tại các câu trên (**BT2_c4**), trong trường hợp này hãy xác định kiểm định phù hợp với thiết kế nghiên cứu mới và kiểm định lại giả thuyết câu b. So sánh kết luận với câu b và đưa ra nhận xét (*lưu ý: giả định phương sai tương đồng*). Nếu là CEO của công ty anh/chị sẽ thực hiện quyết định như thế nào? (10 điểm)