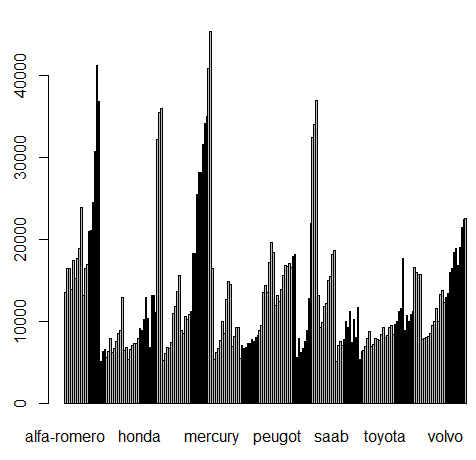
|  |  |  |
| --- | --- | --- |
| TRƯỜNG ĐẠI HỌC NHA TRANG  **KHOA CÔNG NGHỆ THÔNG TIN**  Người ra đề    Nguyễn Đức Thuần | ĐỀ THI KẾT THÚC MÔN HỌC  MÔN: **THỐNG KÊ MÁY TÍNH**  Thời gian: 75 phút  Lớp: 61 CNTT-3  *Không sử dụng tài liệu*  **ĐỀ I**  **Ngày thi: 07 /01/2022** | DUYỆT ĐỀ    Nguyễn Đức Thuần |

* **Đề này dành cho các sinh viên ký số cuối trong Mã số sinh viên là số lẻ.**

**Câu I**: (4.0đ) (**Mô tả dữ liệu**)

1. (1.0đ) Đọc File dữ liệu cars.csv từ thư mục chứa đề bài hay trong <https://pluto.huji.ac.il/~msby/StatThink/index.html> vào dataframe có tên là cars.
2. (1.0đ) Điền giá trị thiếu trong cột price bằng giá trị trung bình của các giá trị còn lại.
3. (1.0đ) Đếm số cột mang giá trị thiếu trong cars.
4. (1.0đ) Vẽ biểu đồ barplot thể hiện mối quan hệ giữa giá (price)và hãng sản xuất (make).



**Câu II**: (3.0đ) (**Ước lượng**)

* Dùng hàm table(cars$fuel.type) để biết bao nhiêu xe chạy diesel, gas.

1. (1.5đ) Ước lượng khoảng cho tỷ lê xe chạy diesel với độ tin cậy 95%.
2. (1.5đ) Cũng với dữ liệu trên, nếu muốn sai số ước lượng cho tỷ lệ xe chạy bằng diesel là 0.05 thì độ tin cậy là bao nhiêu?

**Câu III**: (3.0đ) (**So sánh 2 kỳ vọng chưa biết phương sai, mẫu nhỏ**)

Một cuộc điều tra của Hội phụ nữ để đánh giá về dư luận xã hội cho rằng lương của phụ nữ thấp hơn lương của nam giới. Một mẫu ngẫu nhiên gồm 4 đàn ông có lương trung bình là 7.8 triệu với độ lệch chuẩn của mẫu là 2.44; một mẫu ngẫu nhiên khác trên 4 phụ nữ có lương trung bình là 6.35 triệu, với độ lệch chuẩn là 2.02. Giả sử lương của cả nam và nữ đều là biến ngẫu nhiên theo luật phân phối chuẩn có cùng phương sai. Hãy kết luận về cuộc điều tra với mức ý nghĩa α=0.1.

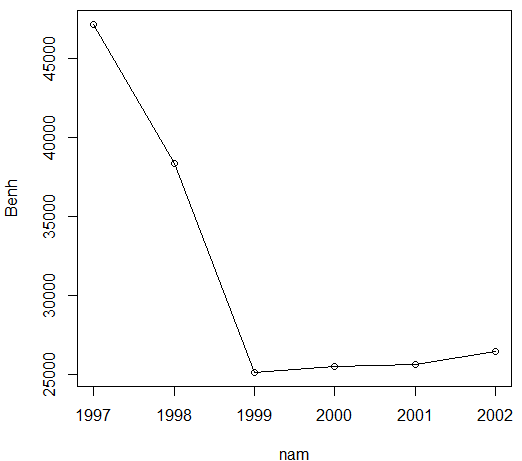
1. Phát biểu giả thuyết thống kê
2. Kiểm định giả thuyết.

|  |  |  |
| --- | --- | --- |
| TRƯỜNG ĐẠI HỌC NHA TRANG  **KHOA CÔNG NGHỆ THÔNG TIN**  Người ra đề    Nguyễn Đức Thuần | ĐỀ THI KẾT THÚC MÔN HỌC  MÔN: **THỐNG KÊ MÁY TÍNH**  Thời gian: 75 phút  Lớp: 61 CNTT-3  *Không sử dụng tài liệu*  **ĐỀ II**  **Ngày thi: 07 /01/2022** | DUYỆT ĐỀ    Nguyễn Đức Thuần |

* **Đề này dành cho các sinh viên ký số cuối trong Mã số sinh viên là số chẵn.**

**Câu I**: (4.0đ) (**Mô tả dữ liệu**)

1. (1.0đ) Đọc File dữ liệu aids.csv từ thư mục chứa đề bài hay trong <https://pluto.huji.ac.il/~msby/StatThink/index.html> vào dataframe có tên là aids.
2. (1.0đ) Cho biết số lượng trung bình người mắc bệnh từ năm 1981 đến năm 2002.
3. (1.0đ) Cho biết năm có số người mắc bệnh nhiều nhất.
4. (1.0đ) Vẽ biểu đồ thể hiện số người mắc bệnh từ năm 1997 đến 2002. (Trục hoành thể hiện năm, trục tung thể hiện số người mắc bệnh aids$ diagnosed[aids$year>=1997] tương ứng.



**Câu II**: (3.0đ) (**Ước lượng**)

1. (1.5đ) Khảo sát việc thắt dây an toàn khi lái xe, một công ty bảo hiểm thu thập mẫu là 400 tài xế. Họ thấy rằng có 320 tài xế luôn thắt dây an toàn khi lên xe. Hãy ước lượng khoảng cho tỷ lệ tài xế luôn thắt dây an toàn với độ tin cậy 80%.
2. (1.5đ) Cũng với dữ liệu ở câu a, nếu muốn sai số ước lượng er = 0.03 thì độ tin cậy là bao nhiêu?

**Câu III**: (3.0đ) (**Kiểm định giả thuyết thống kê**)

Khối lượng bao gạo (KLBG) là biến ngẫu nhiên có phân phối chuẩn N(50,0.01). Có nhiều ý kiến cho rằng khối lượng các bao gạo bị thiếu. Một nhóm thanh tra đã cân ngẫu nhiên 25 bao gạo trong kho và thu được kết quả:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| KLBG | (48;48.5) | (48.5;49) | (49;49.5) | (49.5;50) | (50;50.5) |
| Số bao gạo | 2 | 5 | 10 | 6 | 2 |

Hãy kiểm định ý kiến của khách hàng với mức ý nghĩa α=0.05.

Gợi ý: *Kiểm định trường hợp 2 mẫu không độc lập*.

* Nộp bài:
* Chụp ảnh các trang bài làm, tên các ảnh tương thích với thứ tự các trang. Ví dụ:

Page1, Page2,..

* Lưu các ảnh trang bài làm vào 1 thư mục có tên: TênSV\_Mã số SV, ví dụ:

Tuan\_62123456, nén lại và nộp lên trang Elearning.