

**KHÓA TẬP HUẤN GIẢNG VIÊN
KHU VỰC MIỀN BẮC**
Thống kê hiện đại với phần mềm thống kê R 2023

Bài tập nhóm
Gợi ý chủ đề



Hà Nội, 12/2023

Chủ đề 1: Dữ liệu Professors Salaries (cô Bình Đào)

Dữ liệu Professors Salaries đã tích hợp sẵn trong package: carData

```
library(carData)
```

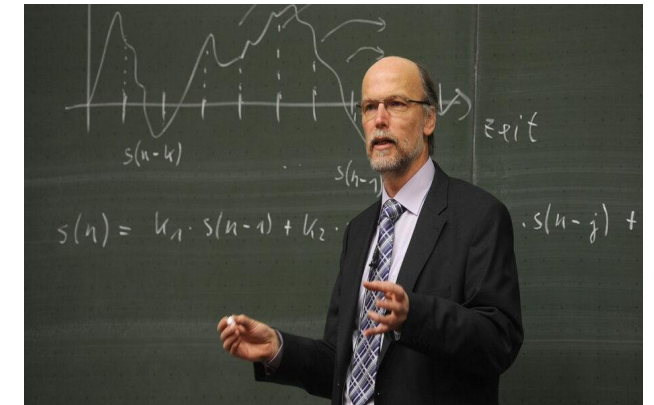
```
Salaries
```

```
dim(Salaries)
```

Key words in google search:

“linear regression Professors Salaries dataset in r”; “Professors Salaries dataset in r”

Link tham khảo: <https://rpubs.com/paulasanta/769382>



Chủ đề 2: Giá thuê nhà tại New York trên AirBnB (Cô Tuyệt Mai + Thầy Đỗ Văn Cường)

Dữ liệu download miễn phí tại:

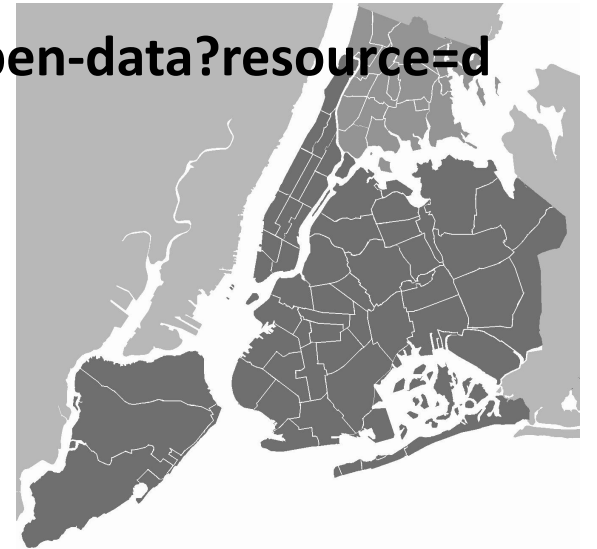
<https://www.kaggle.com/datasets/dgomonov/new-york-city-airbnb-open-data?resource=download>

Hoặc download trong folder tư liệu của khóa tập huấn

```
airbnb_data <- read.csv("AB_NYC_2019.csv")
```

Key words in google search:

“new york city airbnb linear regression in r”



Link tham khảo: https://rpubs.com/SayakChakraborty/NewYorkAirBnB_Modelling



Chủ đề 3: Personal Medical Cost Dataset (Cô Ngô Nga + Cô Trịnh Hường)

File: insurance.csv

This dataset was inspired by the book Machine Learning with R by Brett Lantz. The data contains medical information and costs billed by health insurance companies. It contains 1338 rows of data and the following columns: age, gender, BMI, children, smoker, region, insurance charges.

Key words in google search: “Personal Medical Cost Dataset in r”;



Link tham khảo: https://rstudio-pubs-static.s3.amazonaws.com/369885_f143f456b243444a8e687028fcfab3be.html
<https://rpubs.com/koki25ando/medicalcost>
<https://www.kaggle.com/mirichoi0218/insurance>

Reference:

Lantz, Brett. *Machine learning with R*. Packt publishing ltd, 2013.

https://edu.kpfu.ru/pluginfile.php/278552/mod_resource/content/1/MachineLearningR__Brett_Lantz.pdf

Chủ đề 4: Dữ liệu Boston (cô Nhung Nguyễn)

Dữ liệu đã tích hợp sẵn trong package MASS

library(MASS)

Boston

dim(Boston)

Key words in google search:

“linear regression boston dataset in r”; “boston dataset in r”

Link tham khảo:

https://rpubs.com/Rashmi_Subrahmanya/371719

Reference:

Harrison, D. and Rubinfeld, D.L. *'Hedonic prices and the demand for clean air'*, J. Environ. Economics & Management, vol.5, 81-102, 1978.



Chủ đề 5: Thiết kế thí nghiệm (thầy Tùng Phạm)

Thiết kế thí nghiệm để đánh giá các nhân tố ảnh hưởng tới thành tích của người bơi chuyên nghiệp.

Mục tiêu: Tìm các nhân tố ảnh hưởng đến thành tích (thời gian) của người bơi.

Đối tượng: Người bơi chuyên nghiệp trong bể bơi dài 20 m.

Nhân tố nghiên cứu: Thời gian (sáng/chiều), Thức ăn (đã ăn/ chưa ăn trước khi bơi), Vị trí xuất phát (chỗ nông/ chỗ sâu), trang phục (quần bơi/quần sóc thường) được coi là biến block (controllable factor).

Hướng dẫn:

Sử dụng thiết kế 2^4 nhân tố đầy đủ và phần mềm R để phân tích. Có thể sử dụng dữ liệu đã có trên mạng để phân tích.

Học viên cần làm báo cáo đầy đủ và chi tiết.

Liên hệ giảng viên khi cần tư vấn thêm: phamdinhtung@hus.edu.vn.

Chủ đề 6: Thiết kế thí nghiệm (thầy Hoàng Hà)

Thiết kế thí nghiệm để đánh giá các nhân tố ảnh hưởng tới chỉ số khúc xạ (refractive index) trong các tấm pin mặt trời

Mục tiêu: Tìm các nhân tố ảnh hưởng đến chỉ số khúc xạ của các tấm pin mặt trời được phủ bởi chất silicon nitride (SiN_x).

Đối tượng: Các tấm pin mặt trời được phủ bởi SiN_x .

Nhân tố nghiên cứu: A. Silan to Ammonia Flow Rate Ratio; B. Total Gas Flow Rate (sccm); C. Press (mtorr); D. Temp (độ C); E. Power (W).

Biến đáp ứng: Refract. Index

Hướng dẫn:

Sử dụng thiết kế 2^5 nhân tố đầy đủ và phần mềm R để phân tích. File dữ liệu: project_DOE_SiNx.csv. Học viên có thể xem mô tả chi tiết về thí nghiệm ở trong project_DOE_SiNx.pdf.

Học viên cần làm báo cáo đầy đủ và chi tiết.

Liên hệ giảng viên khi cần tư vấn thêm: hvha@hcmus.edu.vn.