TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**



**BÁO CÁO CUỐI KỲ**

**MÔN HỌC: THỐNG KÊ TRONG KHOA HỌC SỰ SỐNG**

**Mã môn học: C01145**

**TP. HỒ CHÍ MINH, THÁNG … NĂM 2022**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**



**BÁO CÁO CUỐI KỲ**

**MÔN HỌC: THỐNG KÊ TRONG KHOA HỌC SỰ SỐNG**

**Mã môn học: C01145**

Họ và tên sinh viên:

Mã số sinh viên:

Ngành học:

Email:

**TP. HỒ CHÍ MINH, THÁNG … NĂM 2022**

**Lời cảm ơn**

**Nhận Xét**

**Mục lục**

[**Lời mở đầu** 9](#_Toc105607202)

[**Câu 1. Cho bảng dữ liệu sau: (3,5 điểm)** 10](#_Toc105607203)

[a) Nhập dữ liệu vào R (Họ và tên, phái có thể nhập không dấu để tránh lỗi hiển thị) 10](#_Toc105607251)

[b) Tạo cột Phân loại theo điều kiện sau: 11](#_Toc105607252)

[c) Tạo cột lương bằng: hệ số nhân với (1 000 000 + d\* 100 000+e\*10 000) 12](#_Toc105607257)

[d) Tạo cột Thưởng với điều kiện: 12](#_Toc105607258)

[e) Tạo cột thực lĩnh bằng: Lương + Thưởng 13](#_Toc105607262)

[f) Tính tổng lương thực lĩnh theo phái 14](#_Toc105607263)

[g) Vẽ biểu đồ Histogram cho tiền lương thực lĩnh. 14](#_Toc105607264)

[**Câu 2: (1,5 điểm)** Một khách hàng mua xe tại một đại lý. 15](#_Toc105607265)

[a) Tính xác suất để có **(1+c)** xe bị trả lại. 15](#_Toc105607266)

[b) Tính xác suất để có nhiều nhất **(2+d)** xe bị trả lại. 15](#_Toc105607267)

[c) Biết xác suất trả lại của đại lý là không cao hơn **(1+e)%** thì số xe bị trả lại của đại lý là bao nhiêu? 15](#_Toc105607268)

[**Câu 3: (1,5 điểm). Thời gian để sản xuất một sản phẩm loại A là một biến ngẫu nhiên tuân theo luật phân phối chuẩn với các tham số và (đơn vị là phút)** 15](#_Toc105607269)

[a) Tính xác suất để một sản phẩm loại A nào đó được sản xuất trong khoảng thời gian ít hơn **(11+c\*0.1)** phút 15](#_Toc105607270)

[b) Tính xác suất để một sản phẩm loại A nào đó được sản xuất trong khoảng thời gian từ **(9+b\*0.1)** phút đén **(12+d\*0.1)** phút 15](#_Toc105607271)

[c) Thời gian ít hơn bao nhiêu để xác suất được sản phẩm loại A là **(90+e)%** 15](#_Toc105607272)

[**Câu 4 (2,5 điểm):** Người ta đo chiều cao của cây dầu sau 6 tháng tuổi được bảng số liệu sau: 15](#_Toc105607273)

[a) Hãy ước lượng chiều cao trung bình của toàn bộ cây dầu sau 6 tháng tuổi, với độ tin cậy là **(90+d)%**. 15](#_Toc105607274)

[b) Với độ tin cậy 99%, nếu muốn sai số của chiều cao là **(2,5+d\*0.1)** =(2,5+5\*0.1)=3cm thì cần phải khảo sát bao nhiêu cây. 16](#_Toc105607283)

[c) Nếu chiều cao trung bình của cây dầu là **(40+e+a\*0.1)=(40+8+0\*0.1)=48** thì dựa vào kết quả trên có thể nghi ngờ rằng kỹ thuật trồng chưa đạt yêu cầu không? với mức ý nghĩa là 0.01. 16](#_Toc105607292)

[d) Chiều cao của cây dưới 36 cm được phân vào nhóm cây thấp. Hãy ước lượng tỉ lệ cây thấp của cây dầu sau 6 tháng tuổi, với độ tin cậy **(90+c)%.** =(90+1)=91% 17](#_Toc105607297)

[e) Có thể kết luận được rằng tỉ lệ cây thấp là (13+ d\*2+b\*0.1)/100 =(13+5\*2+1\*0.1)/100=0.231 hay không với mức ý nghĩa 5%. 17](#_Toc105607303)

[**Tài liệu tham khảo** 19](#_Toc105607309)

# **Lời mở đầu**

**Nội dung đề tài viết báo cáo**

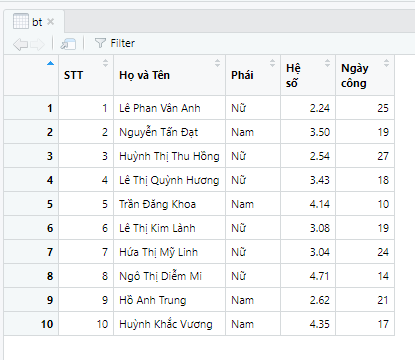
# **Câu 1. Cho bảng dữ liệu sau: (3,5 điểm)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **STT** | **Họ và Tên** | **Phái** | **Hệ số** | **Ngày công** |
| 1 | Lê Phan Vân Anh | Nữ | 2.24 | 25 |
| 2 | Nguyễn Tấn Đạt | Nam | 3.5 | 19 |
| 3 | Huỳnh Thị Thu Hồng | Nữ | 2.54 | 27 |
| 4 | Lê Thị Quỳnh Hương | Nữ | 3.43 | 18 |
| 5 | Trần Đăng Khoa | Nam | 4.14 | 10 |
| 6 | Lê Thị Kim Lành | Nữ | 3.08 | 19 |
| 7 | Hứa Thị Mỹ Linh | Nữ | 3.04 | 24 |
| 8 | Ngô Thị Diễm Mi | Nữ | 4.71 | 14 |
| 9 | Hồ Anh Trung | Nam | 2.62 | 21 |
| 10 | Huỳnh Khắc Vương | Nam | 4.35 | 17 |

## Nhập dữ liệu vào R (Họ và tên, phái có thể nhập không dấu để tránh lỗi hiển thị)

File -> Import Dataset -> From Excel -> Select Direct -> Import

Kết quả:



Kết luận:

Dữ liệu nhập vào có 5 cột và 10 dòng

## b) Tạo cột Phân loại theo điều kiện sau:

### + Nếu ngày công từ 24 ngày trở lên xếp loại A.

Lệnh:

### bt$Xep\_loai[bt["Ngày công"]>=24]<-c("A")

### + Nếu ngày công từ 18 ngày đến 23 ngày xếp loại B.

Lệnh:

bt$Xep\_loai[bt["Ngày công"]<24 & bt["Ngày công"]>17]<-c("B")

### + Nếu ngày công từ 17 ngày trở xuống xếp loại C.

Lệnh:

bt$Xep\_loai[bt["Ngày công"]<18]<-c("C")

Kết quả:

A picture containing text, white

Description automatically generated

## c) Tạo cột lương bằng: hệ số nhân với (1 000 000 + d\* 100 000+e\*10 000)

Lệnh:

a<-c(0)

b<-c(1)

c<-c(1)

d<-c(5)

e<-c(8)

bt["Lương"]<-bt["Hệ số"]\*(1000000 + d\*100000+e\*10000)

Kết quả:

A picture containing text, white

Description automatically generated

## d) Tạo cột Thưởng với điều kiện:

### + Nếu xếp loại A được thưởng 650 000 đồng.

Lệnh:

bt$"Thưởng"[bt["Xep\_loai"]=="A"]<-c(650000)

### + Nếu xếp loại B được thưởng 450 000 đồng.

Lệnh:

bt$"Thưởng"[bt["Xep\_loai"]=="B"]<-c(450000)

### + Nếu xếp loại C bị phạt 250 000 đồng.

bt$"Thưởng"[bt["Xep\_loai"]=="C"]<-c(-250000)

Kết quả:

Table

Description automatically generated

## e) Tạo cột thực lĩnh bằng: Lương + Thưởng

Lệnh:

bt["Thực lĩnh"]<-bt["Lương"]+bt["Thưởng"]

Kết quả:

Table

Description automatically generated with medium confidence

## f) Tính tổng lương thực lĩnh theo phái

Lệnh:

Sex<-c("Male","Female")

sum\_1=sum(bt[bt["Phái"]=="Nam","Thực lĩnh"])

sum\_2=sum(bt[bt["Phái"]=="Nữ","Thực lĩnh"])

sum\_salary<-c(sum\_1,sum\_2)

thongkeluong<-data.frame(Sex,sum\_salary)

Kết quả:



## g) Vẽ biểu đồ Histogram cho tiền lương thực lĩnh.

Lệnh:

install.packages("Hmisc")

library(Hmisc)

hist.data.frame(bt["Thực lĩnh"])

Chart, histogram

Description automatically generated

# **Câu 2: (1,5 điểm)** Một khách hàng mua xe tại một đại lý, nếu xe có sự cố kỹ thuật thì được quyền trả xe trong vòng 3 ngày sau khi mua và được lấy lại nguyên số tiền mua xe. Có 50 xe được bán ra. Xác suất để một xe bị trả lại là 0,1.

## a) Tính xác suất để có **(1+c)** xe bị trả lại.

## b) Tính xác suất để có nhiều nhất **(2+d)** xe bị trả lại.

## c) Biết xác suất trả lại của đại lý là không cao hơn **(1+e)%** thì số xe bị trả lại của đại lý là bao nhiêu?

# **Câu 3: (1,5 điểm). Thời gian để sản xuất một sản phẩm loại A là một biến ngẫu nhiên tuân theo luật phân phối chuẩn với các tham số và (đơn vị là phút)**

## a) Tính xác suất để một sản phẩm loại A nào đó được sản xuất trong khoảng thời gian ít hơn **(11+c\*0.1)** phút

## b) Tính xác suất để một sản phẩm loại A nào đó được sản xuất trong khoảng thời gian từ **(9+b\*0.1)** phút đén **(12+d\*0.1)** phút

## c) Thời gian ít hơn bao nhiêu để xác suất được sản phẩm loại A là **(90+e)%**

# **Câu 4 (2,5 điểm):** Người ta đo chiều cao của cây dầu sau 6 tháng tuổi được bảng số liệu sau:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Độ caoXi(cm) | 24 – 30 | 30 – 36 | 36 – 42 | 42 – 48 | 48 – 54 | 54 – 60 | 60 – 66 |
| Số lượng ni | 12 | 24 | 35 | 47 | 43 | 32 | 7 |

1. Hãy ước lượng chiều cao trung bình của toàn bộ cây dầu sau 6 tháng tuổi, với độ tin cậy là **(90+d)%**.

d=5

Height=c((24+30)/2,(30+36)/2,(36+42)/2,(42+48)/2, (48+54)/2,(54+60)/2,(60+66)/2)

Number=c(12,24,35,47,43,32,7)

Frame<-rep(Do.cao, So.luong)

t.test(Frame, conf.level = 0.95)

Kết quả:

Text, letter

Description automatically generated

**Nhận xét:**khoảng ước lượng chiều cao trung bình của toàn bộ cây dầu sau 6 tháng tuổi, với độ tin cậy 97**%** là từ 43.98 cm đến 46.56 cm

1. Với độ tin cậy 99%, nếu muốn sai số của chiều cao là **(2,5+d\*0.1)** =(2,5+5\*0.1)=3cm thì cần phải khảo sát bao nhiêu cây.

m=3

alpha=1-0.99

z = qnorm(alpha/2, lower.tail = FALSE)

s = sd(Frame)

n = (z\*s/m)^2

Kết quả: 

**Nhận xét:**

Vậy với độ tin cậy 99%, nếu muốn sai số của chiều cao 3.2 cm thì cần phải khảo sát 64 cây.

1. Nếu chiều cao trung bình của cây dầu là **(40+e+a\*0.1)=(40+8+0\*0.1)=48** thì dựa vào kết quả trên có thể nghi ngờ rằng kỹ thuật trồng chưa đạt yêu cầu không? với mức ý nghĩa là 0.01.

t.test(Frame, mu = 48, alternative = "two.sided", conf.level = 0.99)

Kết quả:

Text

Description automatically generated

**Nhận xét:** : Ta thấy chỉ số p — value= 4.445e-10 < = 0.01, nên bác bỏ giả thuyết Ho. Vậy, với mức ý nghĩa 1%, ta chưa thể nghi ngờ rằng kỹ thuật trồng chưa đạt yêu cầu

1. Chiều cao của cây dưới 36 cm được phân vào nhóm cây thấp. Hãy ước lượng tỉ lệ cây thấp của cây dầu sau 6 tháng tuổi, với độ tin cậy **(90+c)%.** =(90+1)=91%

prop.test(x = 36, n = 200, conf.level = 0.91)

Kết quả:

Text, letter

Description automatically generated

**Nhận xét:**

Vậy với độ tin cậy là 91%, khoảng ước lượng tỉ lệ cây thấp của cây dầu sau 6 tháng tuổi dao động trong khoảng từ 13.63% đến 23.31%

1. Có thể kết luận được rằng tỉ lệ cây thấp là (13+ d\*2+b\*0.1)/100 =(13+5\*2+1\*0.1)/100=0.231 hay không với mức ý nghĩa 5%.

Mức ý nghĩa 𝛼=5 🡪 độ tin cậy là 95%. Ta có thể đặt giả thuyết kiểm định thống kê 𝐻0:𝑝=0.231 và 𝐻𝛼:𝑝≠0.231.

Lệnh:

prop.test(x= 36, n= 200, p= 0.231, alternative= "two.sided", conf.level = 0.95)

Kết quả:

Text

Description automatically generated

**Nhận xét:**

Ta thấy chỉ số p — value =0.1037>𝛼=0.05, nên chấp nhận giả thuyết Ho. Vậy với mức ý nghĩa 5%, ta biết rằng tỷ lệ cây thấp 23.1 đạt yêu cầu.

# **Tài liệu tham khảo**