**3.43:**

Dịch: Để kiểm soát rủi ro hư hỏng lõi nghiêm trọng trong một nhà máy điện hạt nhân thương mại

tai nạn mất điện, độ tin cậy của máy phát điện diesel khẩn cấp khi khởi động theo yêu cầu phải

được duy trì ở mức cao. Bài báo “Ước tính Bayes theo kinh nghiệm về độ tin cậy của máy phát điện diesel khẩn cấp năng lượng hạt nhân” [Technometrics (1996) 38: 11–23] chứa dữ liệu về

lịch sử thất bại của bảy nhà máy điện hạt nhân. Dữ liệu sau đây là số lượng thành công

nhu cầu giữa những lần hỏng hóc đối với máy phát điện diesel tại một trong những nhà máy này từ năm 1982 đến năm 1988.

28 50 193 55 4 7 147 76 10 0 10 84 0 9 1 0 62 26 15 226 54 46 128 4 105 40

4 273 164 7 55 41 26 6

(Lưu ý: Sự cố của máy phát điện diesel không nhất thiết dẫn đến hư hỏng hạt nhân

cốt lõi vì tất cả các nhà máy điện hạt nhân đều có một số máy phát điện diesel khẩn cấp.)

1. Tính giá trị trung bình và trung vị của các nhu cầu thành công giữa các lần thất bại

> dl<-c(28,50,193,55,4,7,147,76,10,0,10,84,0,9,1,0,62,26,15,226,54,46,128,4,105,40,4,273,164,7,55,41,26,6)

> mean(dl)

[1] 57.52941

> median(dl)

[1] 34

1. Thước đo nào thể hiện tốt nhất trung tâm của dữ liệu?

* Trung bình thể hiện tốt nhất trung tâm của dữ liệu vì giá trị trung bình lớn hơn 24 giá trị trong dữ liệu, còn trung vị chỉ lớn hơn 17 giá trị trong dữ liệu, vì thế trung bình cho thấy chính xác nhất về trung tâm của dữ liệu.

1. Tính phạm vi và độ lệch chuẩn, s.

> range(dl)

[1] 0 273

> var(dl)

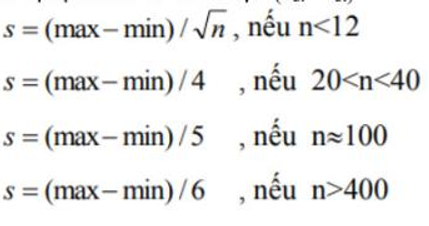
[1] 4927.408

> sd(dl)

[1] 70.1955

1. Sử dụng xấp xỉ phạm vi để ước tính s. Khoảng cách gần đúng với giá trị đích thực?

Sử sụng xấp xỉ phạm vi để ước tính s: sut≈(max-min)/4=(273-0)/4=68.25 Độ lệch chuẩn ban đầu sdt =70.1955 =>Ta thấy được độ lệch chuẩn khi đo bằng sắp xỉ phạm vi có giá trị gần đúng với độ lệch chuẩn đích thực (sut≈sdt).



1. Xây dựng các khoảng

; ;

Đếm số lần yêu cầu giữa các lần không đạt trong mỗi khoảng thời gian trong ba khoảng thời gian. Chuyển đổi những con số này thành tỷ lệ phần trăm và so sánh kết quả của bạn với

Quy tăc thực nghiệm.

* **= 28 (82%) ≠ 68% :**

> ybar<-mean(dl)

> S<-sd(dl)

> ybar-S;ybar+S

[1] -12.66609

[1] 127.7249

> t1<-function(x){

+ print('so lan thanh cong trong khoang 1: ')

+ dem<-0

+ gt<-ybar-S

+ gd<-ybar+S

+ for(i in 1:length(x)){

+ if(x[i]>=gt && x[i]<=gd){dem<-dem+1}

+ }

+ print(dem)

+ }

> t1(dl)

[1] "so lan thanh cong trong khoang 1: "

[1] 28

* **= 32 (94%) ≈ 95%**

> t1<-function(x){

+ print('so lan thanh cong trong khoang 1: ')

+ dem<-0

+ gt<-ybar-2\*S

+ gd<-ybar+2\*S

+ for(i in 1:length(x)){

+ if(x[i]>=gt && x[i]<=gd){dem<-dem+1}

+ }

+ print(dem)

+ print(gt)

+ print(gd)

+ }

> t1(dl)

[1] "so lan thanh cong trong khoang 1: "

[1] 32

[1] -82.86159

[1] 197.9204

* **= 33 (97%) ≈ 99,7%**

> t1<-function(x){

+ print('so lan thanh cong trong khoang 1: ')

+ dem<-0

+ gt<-ybar-3\*S

+ gd<-ybar+3\*S

+ for(i in 1:length(x)){

+ if(x[i]>=gt && x[i]<=gd){dem<-dem+1}

+ }

+ print(dem)

+ print(gt)

+ print(gd)

+ }

> t1(dl)

[1] "so lan thanh cong trong khoang 1: "

[1] 33

[1] -153.0571

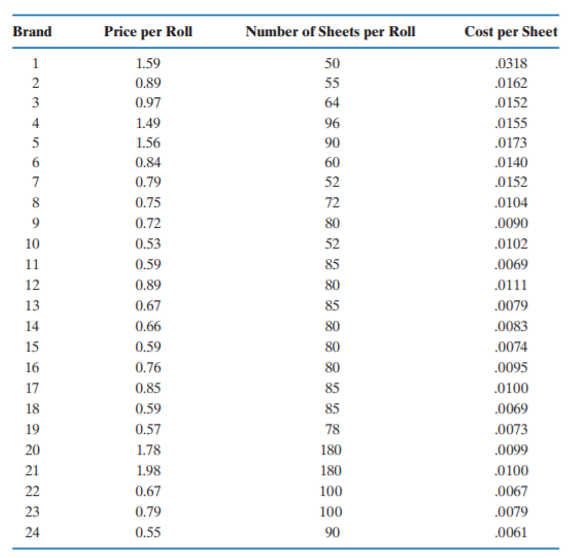
[1] 268.1159

1. Tại sao bạn cho rằng Quy tắc thực nghiệm và tỷ lệ phần trăm của bạn không khớp nhau?

Y+-s => 68% -> 82% Ta thấy ở mức 68% thì kết quả không trùng khớp với Qui tắc thực nghiệm.

**3.45**

Dịch: Số tháng 2 năm 1998 của Báo cáo Người tiêu dùng cung cấp dữ liệu về giá của 24 nhãn hiệu khăn giấy. Giá được đưa ra theo cả chi phí trên mỗi cuộn và giá mỗi tờ vì các thương hiệu đã số lượng tờ mỗi cuộn khác nhau.



1. Tính toán độ lệch chuẩn cho cả giá mỗi cuộn và giá mỗi tờ giấy.

> gc<-c(1.59,0.89,0.97,1.49,1.56,0.84,0.79,0.75,0.72,0.53,0.59,0.89,0.67,0.66,0.59,0.76,0.85,0.59,0.57,1.78,1.98,0.67,0.79,0.55)

> length(gc)

[1] 24

> tg<-c(.0318,.0162,.0152,.0155,.0173,.0140,.0152,.0104,.0090,.0102,.0069,.0111,.0079,.0083,.0074,.0095,.0100,.0069,.0073,.0099,.0100,.0067,.0079,.0061)

> length(tg)

[1] 24

> var(gc)

[1] 0.1791868

> var(tg)

[1] 3.051216e-05

b. Cái nào thay đổi nhiều hơn, giá mỗi cuộn hay giá mỗi tờ?

- Giá mỗi cuộn thay đổi nhiều hơn.

**3.46**

Refer to Exercise 3.45. Use a scatterplot to plot the price per roll and number of sheets per roll. a. Do the 24 points appear to fall on a straight line? b. If not, is there any other relation between the two prices? c. What factors may explain why the ratio of price per roll to number of sheets is not a constant?

Dịch: Dựa vào bài 3.45, sử dụng biểu đồ tán xạ (scatter plot) để vẽ biểu đồ giá mỗi cuộn và số tờ mỗi cuộn.

> gc<-c(1.59,0.89,0.97,1.49,1.56,0.84,0.79,0.75,0.72,0.53,0.59,0.89,0.67,0.66,0.59,0.76,0.85,0.59,0.57,1.78,1.98,0.67,0.79,0.55)

> n<-c(50,55,64,96,90,60,52,72,80,52,85,80,85,80,80,80,85,85,78,180,180,100,100,90)

> plot(gc,n,ylab='number of sheets per roll',xlab='Price per roll',pch=16)

1. 24 điểm có dường như nằm trên một đường thẳng không?

Sau khi sử dụng ngôn ngữ R để vẽ biểu đồ tán xạ (scatter plot) của giá mỗi cuộn và số tờ mỗi cuộn của 24 hãng khăn giấy, thì ta có thể nhìn thấy 24 điểm không nằm trên một đường thẳng.

1. Nếu không, có mối quan hệ nào khác giữa hai mức giá không?

Nhìn vào biểu đồ, ta có thể thấy đa số các mức giá nằm trong khoảng từ 0.5 đến 1.0, tuy nhiên lại có 5 mức giá cá biệt nằm trong khoảng từ 1.5 đến 2.0.

1. Những yếu tố nào có thể giải thích tại sao giá mỗi cuộn trên số tờ mỗi cuộn không phải là một hằng số?

Những yếu tố như sự uy tính, mức độ nổi tiếng, và chất liệu làm giấy, …của mỗi hãng khăn giấy có thể giải thích được tại sao giá mỗi cuộn trên số tờ không phải là một hằng số.

**3.49**

Suppose the random variable y has a Poisson distribution. Compute the following probabilities:

a. P(y = 4) given µ = 2

b. P(y = 4) given µ = 3.5

c. P(y ≥ 4) given µ = 2

d. P(1 ≤ y < 4) given µ = 2

Dịch: Giả sử biến ngẫu nhiên y có phân phối Poisson. Tính toán những điều sau xác

suất:

1. P(y = 4) cho 𝜇 = 2

> dpois(4,2)

[1] 0.09022352

1. P(y = 4) cho 𝜇 = 3.5

> dpois(4,3.5)

[1] 0.1888123

1. P(y > 4) cho 𝜇 = 2

> 1-ppois(4,2)

[1] 0.05265302

1. P(1 ≤ y < 4) cho 𝜇 = 2

> dpois(1,2)+dpois(2,2)+dpois(3,2)

[1] 0.7217882

**4.17**

A high accumulation of ozone gas in the lower atmosphere at ground level is air pollution and can be harmful to people, animals, crops, and various materials. Elevated levels above the national standard may cause lung and respiratory disorders. Nitrogen oxides and hydrocarbons are known as the chief “precursors” of ozone. These compounds react in the presence of sunlight to produce ozone. The sources of these precursor pollutants include cars, trucks, power plants, and factories. Large industrial areas and cities with heavy summer traffic are the main contributors to ozone formation. The United States Environmental Protection Agency (EPA) has developed procedures for measuring vehicle emission levels of nitrogen oxide. Let P denote the amount of this pollutant in a randomly selected automobile in Houston, Texas. Suppose the distribution of P can be adequately modeled by a normal distribution with a mean level of µ = 70 ppb (parts per billion) and a standard deviation of σ = 13 ppb.

a. What is the probability that a randomly selected vehicle will have an emission level less than 60 ppb?

b. What is the probability that a randomly selected vehicle will have an emission level greater than 90 ppb?

c. What is the probability that a randomly selected vehicle will have an emission level between 60 and 90 ppb?

Dịch: Sự tích tụ khí ozone cao trong bầu khí quyển thấp hơn ở mặt đất là ô nhiễm không khí và có thể gây hại cho con người, động vật, cây trồng và các vật liệu khác nhau. Mức độ cao hơn tiêu chuẩn quốc gia có thể gây rối loạn phổi và hô hấp. Nitơ oxit và hydrocarbon được gọi là "tiền chất" chính của ozone. Các hợp chất này phản ứng với sự hiện diện của ánh sáng mặt trời để tạo ra ozone. Các nguồn của các chất ô nhiễm tiền chất này bao gồm xe hơi, xe tải, nhà máy điện và nhà máy. Các khu vực công nghiệp lớn và các thành phố có lưu lượng giao thông mùa hè nặng nề là những đóng góp chính cho sự hình thành ozone. Cơ quan Bảo vệ Môi trường Hoa Kỳ (EPA) đã phát triển các quy trình để đo mức phát thải oxit nitơ của xe. Hãy để P biểu thị lượng chất gây ô nhiễm này trong một chiếc ô tô được chọn ngẫu nhiên ở Houston, Texas. Giả sử sự phân bố của P có thể được mô hình hóa đầy đủ bởi một phân bố bình thường với mức trung bình là μ = 70 ppb (phần tỷ) và độ lệch chuẩn của σ = 13 ppb.

1. Xác suất một chiếc xe được chọn ngẫu nhiên sẽ có mức khí thải dưới 60 ppb là bao cao?

P(X<60):

> x<-seq(70-13\*4, 70+13\*4, .1)

> y<-dnorm(x, mean = 70, sd = 13)

> plot(x, y, type = "l",axes=T,ylim=c(0,0.04),col="blue",lwd=2,

+ ylab="Normal density",xlab="",yaxt="n",xaxt="n")

> x<-seq(70-13\*4, 60, .1)

> y<-dnorm(x, mean = 70, sd = 13)

> polygon(c(70-13\*4, x, 60), c(-.0016, y, -.0016), col = "lightblue")

> axis(1,at=seq(70-4\*10,70+4\*10,10),labels=c("","","","60",expression(paste(mu,"=70")),"","","",""))

> axis(2,at=seq(0,0.04,0.01),labels=c("0",".1",".2",".3",".4"),las=1)

> text(x=52,y=0.005,cex=1,".2209")

b. Xác suất một chiếc xe được chọn ngẫu nhiên sẽ có mức phát thải lớn hơn 90 ppb là bao giờ?

P(X > 90) = 1 – P(X<90)

> 1-pnorm(90,70,13)

[1] 0.0619679

- Vẽ đồ thị:

> #ve khoang X > 90

> z2<-90

> t2<-subset(gh,z>=z2)

> polygon(c(rev(t2$z),t2$z),

+ c(rep(0,nrow(t2)),t2$gh),col="lightblue")

c. Xác suất một chiếc xe được chọn ngẫu nhiên sẽ có mức phát thải từ 60 đến 90 ppb là bao giờ?

P(60<X<90)= P(X<90) – P(X<60)

> pnorm(90,70,13)

[1] 0.9380321

> pnorm(90,70,13) - dnorm(60,70,13)

[1] 0.9152036

**4.23 :** In this example, the population values are known, and, hence, we can compute the exact values of the population mean, µ, and population standard deviation, σ. We will then examine the behavior of based on samples of size n = 5, 10, and 25 selected from the population. The population consists of 500 pennies from which we compute the age of each penny: Age = 2015 - Date on penny. The histogram of the 500 ages is displayed in Figure 4.20(a). The shape is skewed to the right with a very long right tail. The mean and standard deviation are computed to be µ = 13.468 years and σ = 11.164 years. In order to generate the sampling distribution of for n = 5, we would need to generate all possible samples of size n = 5 and then compute the from each of these samples. This would be an enormous task, since there are 255,244,687,600 possible samples of size 5 that could be selected from a population of 500 elements. The number of possible samples of size 10 or 25 is so large it makes even the national debt look small. Thus, we will use a computer program to select 25,000 samples of size 5 from the population of 500 pennies. For example, the first sample consists of pennies with ages 4, 12, 26, 16, and 9. The sample mean = (4 1 12 1 26 1 16 1 9)/5 = 13.4. We repeat 25,000 times the process of selecting 5 pennies; recording their ages, y1, y2, y3, y4, y5; and then computing = (y1 + y2 + y3 + y4 + y5)/5. The 25,000 values for are then plotted in a frequency histogram, called the sampling distribution of for n = 5. A similar procedure is followed for samples of size n =10 and n = 25. The sampling distributions obtained are displayed in Figures 4.20(b)–(d). Note that all three sampling distributions have nearly the same central value, approximately 13.5. (See Table 4.11.) The mean values of for the three samples are nearly the same as the population mean, µ = 13.468. In fact, if we had generated all possible samples for all three values of n, the mean of the possible values of would agree exactly with µ. The next characteristic to notice about the three histograms is their shape. All three are somewhat symmetric in shape, achieving a nearly normal distribution

Dịch: Trong ví dụ này, các giá trị dân số được biết đến, và do đó, chúng ta có thể tính toán các giá trị chính xác của trung bình dân số, μ và độ lệch chuẩn dân số, σ. Sau đó, chúng tôi sẽ kiểm tra hành vi của y ̅ dựa trên các mẫu có kích thước n = 5, 10 và 25 được chọn từ dân số. Dân số bao gồm 500 đồng xu mà từ đó chúng ta tính tuổi của mỗi xu: Tuổi = 2015 - Ngày trên penny. Biểu đồ của 500 tuổi được hiển thị trong Hình 4.20 (a). Hình dạng được nghiêng sang phải với đuôi phải rất dài. Độ lệch trung bình và chuẩn được tính là μ = 13.468 năm và σ = 11.164 năm. Để tạo ra sự phân bố lấy mẫu của y ̅ cho n = 5, chúng ta sẽ cần tạo ra tất cả các mẫu có thể có kích thước n = 5 và sau đó tính toán y ̅ từ mỗi mẫu này. Đây sẽ là một nhiệm vụ rất lớn, vì có 255.244.687.600 mẫu có thể có kích thước 5 có thể được chọn từ dân số 500 nguyên tố. Số lượng các mẫu có thể có kích thước 10 hoặc 25 lớn đến mức khiến Ngay cả nợ quốc gia cũng có vẻ nhỏ. Do đó, chúng tôi sẽ sử dụng một chương trình máy tính để chọn 25.000 mẫu có kích thước 5 từ dân số 500 đồng xu. Ví dụ, mẫu đầu tiên bao gồm đồng xu với độ tuổi 4, 12, 26, 16 và 9. Mẫu có nghĩa là y ̅ = (4 + 12 + 26 + 16 + 9)/5 = 13,4. Chúng tôi lặp lại 25.000 lần quá trình chọn 5 đồng xu; ghi lại tuổi của họ, y1, y2, y3, y4, y5; và sau đó tính toán y ̅ = (y1 + y2 + y3 + y4 + y5)/5. 25.000 giá trị cho y ̅ sau đó được vẽ theo biểu đồ tần số, được gọi là phân bố lấy mẫu của y ̅ cho n = 5. Một quy trình tương tự được thực hiện cho các mẫu có kích thước n = 10 và n = 25. Các phân phối lấy mẫu thu được được hiển thị trong Hình 4.20(b)–(d). Lưu ý rằng cả ba phân phối lấy mẫu đều có giá trị trung tâm gần như giống nhau, khoảng 13,5. (Xem Bảng 4.11.) Các giá trị trung bình của y ̅ cho ba mẫu gần giống với trung bình dân số, μ = 13.468. Trên thực tế, nếu chúng ta đã tạo ra tất cả các mẫu có thể cho cả ba giá trị của n, , ý nghĩa của các giá trị có thể có của y ̅ sẽ đồng ý chính xác với μ. Đặc điểm tiếp theo cần chú ý về ba biểu đồ của mô hình là hình dạng của chúng. Cả ba đều có hình dạng hơi đối xứng, đạt được sự phân bố gần như bình thường.

**Giải:**

Xét 5 tiêu chí:

Thí nghiệm gồm n lần thử giống nhau: cho 10 phần nước là như nhau cùng về khối lượng, số lượng. Thõa mãn

Mỗi thí nghiệm dẫn đến một trong hai kết quả: kết quả mỗi lần thí nghiệm là số vi khuẩn trong mỗi phần -> không thõa mãn

Xác xuất thành công của thí nghiệm duy nhất bằng p và p không đổi từ thí nghiệm này sang thí nghiệm khác: không thõa mãn

Các thử nghiệm đọc lập: thỏa mãn

Biến ngẫu nhiên y là số lần thành công được quan sát thấy trong n lần thử nghiệm: không thõa mãn vì biến ngẫu nhiên là số lượng vi khuẩn có trong mỗi phần (có nhiều kết quả khác nhau) không phải một phần nước có vi khuẩn hay không (có hai kết quả)

=> Kết luận: đây không phải là một thí nghiệm nhị thức

**4.43:**  Examine the accompanying newspaper clipping. Does this sampling appear to satisfy the characteristics of a binomial experiment?

Poll Finds Opposition to Phone Taps New York—People surveyed in a recent poll indicated they are 81% to 13% against having their phones tapped without a court order. The people in the survey, by 68% to 27%, were opposed to letting the government use a wiretap on citizens suspected of crimes, except with a court order. The survey was conducted for 1,495 households and also found the following results: —The people surveyed are 80% to 12% against the use of any kind of electronic spying device without a court order. —Citizens are 77% to 14% against allowing the government to open their mail without court orders. —They oppose, by 80% to 12%, letting the telephone company disclose records of longdistance phone calls, except by court order. For each of the questions, a few of those in the survey had no responses.

Dịch: Kiểm tra clipping báo chí đi kèm. Việc lấy mẫu này có thỏa mãn các đặc điểm của thí nghiệm nhị thức không? Cuộc thăm dò cho thấy sự phản đối đối với Phone Taps New York - Những người được khảo sát trong một cuộc thăm dò gần đây cho thấy họ là 81% đến 13% chống lại việc nghe lén điện thoại của họ mà không có lệnh của tòa án. Những người trong cuộc khảo sát, từ 68% đến 27%, phản đối việc cho phép chính phủ sử dụng nghe lén đối với các công dân bị nghi ngờ phạm tội, ngoại trừ với lệnh của tòa án. Cuộc khảo sát được thực hiện cho 1.495 hộ gia đình và cũng tìm thấy các kết quả sau: —Những người được khảo sát là 80% đến 12% chống lại việc sử dụng bất kỳ loại thiết bị gián điệp điện tử nào mà không có lệnh của tòa án. Công dân từ 77% đến 14% không cho phép chính phủ mở thư mà không có lệnh của tòa án. Họ phản đối, từ 80% đến 12%, cho phép công ty điện thoại tiết lộ hồ sơ về các cuộc gọi điện thoại kéo dài, ngoại trừ theo lệnh của tòa án. Đối với mỗi câu hỏi, một vài người trong cuộc khảo sát không có câu trả lời.

Trả lời:

- Có những thử nghiệm giống hệt nhau không? 1,495 người được khảo sát có thể giả định là giống hệt nhau nếu sống trong những điều kiện là như nhau. Đúng.

- Mỗi thử nghiệm dẫn đến 2 kết quả: Đúng. Mỗi người được khảo sát sẽ phản đối hoặc không phản đối đối với Phone taps New York.

- Xác suất thành công từ lần thử này sang lần thử khác có giống nhau không? Giống. Vì khi 1 người không phản đối thì xác suất người tiếp theo phản đối là không đổi p = 50%. => xác suất thành công không thay đổi.

-Xác suất thành công từ lần thử này sang lần thử khác độc lập không? CÓ Vì 1 người khi đã phản đối sẽ không ảnh hưởng đến quyết định của người khác.

- Biến ngẫu nhiên của thử nghiệm này là số người phản đối trong 1495 người được khảo sát.

**4.46:** The quality control department examines all the products returned to a store by customers. An examination of the returned products yields the following assessment: 5% are defective and not repairable, 45% are defective but repairable, 35% have small surface scratches but are functioning properly, and 15% have no problems. Compute the following probabilities for a random sample of 20 returned products:

a. All of the 20 returned products have some type of problem.

b. Exactly 6 of the 20 returned products are defective and not repairable.

c. Of the 20 returned products, 6 or more are defective and not functioning properly.

d. None of the 20 returned products has any sort of defect.

Dịch: Bộ phận kiểm soát chất lượng kiểm tra tất cả các sản phẩm được khách hàng trả lại cho cửa hàng. Một cuộc kiểm tra các sản phẩm trả lại mang lại đánh giá sau: 5% bị lỗi và không thể sửa chữa, 45% bị lỗi nhưng có thể sửa chữa, 35% có vết trầy xước bề mặt nhỏ nhưng hoạt động bình thường và 15% không có vấn đề gì. Tính toán các xác suất sau đây cho một mẫu ngẫu nhiên của 20 sản phẩm trả về:

1. Tất cả 20 sản phẩm được trả lại đều có một số vấn đề.

> dbinom(0,20,0.15)

[1] 0.03875953

1. Chính xác 6 trong số 20 sản phẩm được trả lại bị lỗi và không thể sửa chữa.

> #P(X=6)

> dbinom(6,20,0.05)

[1] 0.0002953482

c. Trong số 20 sản phẩm trả lại, có từ 6 sản phẩm trở lên bị lỗi và không hoạt động bình thường.

> #P(X>=6) = 1-P(X<5)

> 1-pbinom(5,20,0.45)

[1] 0.9793053

d. Không có sản phẩm nào trong số 20 sản phẩm trả lại có bất kỳ loại khiếm khuyết nào.

Vì xác suất sản phẩm gặp vấn đề là 15% nên xác suất không có sản phẩm nào trong 20 sản phẩm có vấn đề là:

> dbinom(20,20,0.15)

[1] 3.325257e-17

**4.51:** A firm is considering using the Internet to supplement its traditional sales methods. Using data from an industry association, the firm estimates that 1 of every 1,000 Internet hits results in a sale. Suppose the firm has 2,500 hits per day.

a. What is the probability that the firm will have more than five sales in a randomly selected day?

b. What conditions must be satisfied in order for you to make the calculation in part (a)?

c. Use the Poisson approximation to compute the probability that the firm will have more than five sales in a randomly selected day.

d. Is the Poisson approximation accurate?

Dịch: Một công ty đang xem xét sử dụng Internet để bổ sung các phương pháp bán hàng truyền thống của mình. Sử dụng dữ liệu từ một hiệp hội ngành công nghiệp, công ty ước tính rằng 1 trong số 1.000 Internet đạt được dẫn đến việc bán hàng. Giả sử công ty có 2.500 lượt truy cập mỗi ngày.

1. Xác suất mà công ty sẽ có hơn năm doanh số bán hàng trong một ngày được chọn ngẫu nhiên là bao động?

λ = = 2,5

> #P(X>5) = 1 - P(X<=5)

> 1-ppois(5,2.5)

[1] 0.04202104

b. Những điều kiện nào phải được đáp ứng để bạn thực hiện tính toán câu (a)?

▪ Mỗi một sự kiện xảy ra độc lập và ngẫu nhiên.  
▪ Các sự kiện xảy ra với một tỉ lệ không đổi (theo nghĩa là số lượng các sự kiện  
xảy ra trong một khoảng thời gian nhất định là tỷ lệ thuận với độ dài của  
khoảng thời gian đó).  
▪ Các sự kiện xảy ra riêng lẻ (mỗi thời điểm chỉ có 1 sự kiện xảy ra)

c. Sử dụng xấp xỉ Poisson để tính xác suất công ty sẽ có hơn năm doanh số bán hàng trong một ngày được chọn ngẫu nhiên.

λ=n.p=2500\*1/1000=2.5

> #P(X>5) = 1 - P(X<=5)

> 1-ppois(5,2.5)

[1] 0.04202104

d. Xấp xỉ Poisson có chính xác không?

**4.53:** Find the area under the standard normal curve between these values:

Dịch: : Tìm khu vực dưới đường cong bình thường tiêu chuẩn giữa các giá trị này:

1. z = 0 and z = 1.3

> x<-seq(0,1.3,0.1)

> y<-dunif(x,0,1.3)

> plot(x,y,col='red',type='l',lty=1)

1. z = 0 and z = 2.7

> x<-seq(0,2.7,0.1)

> y<-dunif(x,0,2.7)

> plot(x,y,col='red',type='l',lty=1)

**4.64:**  Let y be a random variable having a normal distribution with a mean equal to 50 and a standard deviation equal to 8. Find the following probabilities:

Dịch: Hãy để y là một biến ngẫu nhiên có sự phân phối chuẩn với trung bình bằng 50 và độ lệch chuẩn bằng 8. Tìm các xác suất sau:

σ = 8 , μ = 50

>muy<-50

>sig<-8

1. P(y > 50)

> 1 - pnorm(50,muy,sig)

[1] 0.5

1. P(y > 53)

> 1 - pnorm(53,muy,sig)

[1] 0.3538302

1. P(y < 58)

> pnorm(58,muy,sig)

[1] 0.8413447

1. P(38 < y < 62)

> #P(38<X<62) = P(X<62) - P(X<38)

> pnorm(62,muy,sig)-pnorm(38,muy,sig)

[1] 0.8663856

1. P(38 ≤ y ≤ 62)

> #P(38≤X≤62) = P(X≤62) - P(X≤38)

> pnorm(62,muy,sig)-pnorm(38,muy,sig)

[1] 0.8663856

**4.70:** The College Boards, which are administered each year to many thousands of high school students, are scored so as to yield a mean of 513 and a standard deviation of 130. These scores are close to being normally distributed. What percentage of the scores can be expected to satisfy each of the following conditions?

Dịch: Các Hội đồng đại học, được quản lý mỗi năm cho hàng ngàn học sinh trung học, được chấm điểm có trung bình là 513 và độ lệch chuẩn là 130. Những điểm số này gần như được phân phối bình thường. Tỷ lệ phần trăm của điểm số có thể được dự kiến để đáp ứng từng điều kiện sau đây?

μ = 513, σ = 130

a. Greater than 600 P(X>600)

> 1-pnorm(600,513,130)

[1] 0.2516741

b. Greater than 700 P(X>700)

> #P(X>700)

> 1-pnorm(700,513,130)

[1] 0.07515157

c. Less than 450 P(X<450)

> #P(X<450)

> pnorm(450,513,130)

[1] 0.3139746

d. Between 450 and 600 P(450<X<600)

> #P(450<X<600) = P(X<600)- P(X<450) = P(X<=600) - P(X<=450)

> pnorm(600,513,130) - pnorm(450,513,130)

[1] 0.4343513

**4.86:** A patient visits her doctor with concerns about her blood pressure. If the systolic blood pressure exceeds 150, the patient is considered to have high blood pressure, and medication may be prescribed. The problem is that there is a considerable variation in a patient’s systolic blood pressure readings during a given day. a. If a patient’s systolic readings during a given day have a normal distribution with a mean of 160 mm mercury and a standard deviation of 20 mm, what is the probability that a single measurement will fail to detect that the patient has high blood pressure? b. If five measurements are taken at various times during the day, what is the probability that the average blood pressure reading will be less than 150 and hence fail to indicate that the patient has a high blood pressure problem? c. How many measurements would be required so that the probability of failing to detect that the patient has high blood pressure is at most 1%.

Dịch: Một bệnh nhân đến gặp bác sĩ với những lo ngại về huyết áp của cô. Nếu huyết áp tâm thu vượt quá 150, bệnh nhân được coi là bị huyết áp cao và thuốc có thể được kê đơn. Vấn đề là có một sự khác biệt đáng kể trong chỉ số huyết áp tâm thu của bệnh nhân trong một ngày nhất định.

1. Nếu chỉ số tâm thu của bệnh nhân trong một ngày nhất định có sự phân bố bình thường với trung bình thủy ngân 160 mm và độ lệch chuẩn 20 mm, xác suất một phép đo duy nhất sẽ không phát hiện ra rằng bệnh nhân bị huyết áp cao là bao cao?

> #P(X<=150)

> muy<-160

> sig<-20

> pnorm(150,muy,sig)

[1] 0.3085375

b. Nếu năm phép đo được thực hiện vào các thời điểm khác nhau trong ngày, xác suất chỉ số huyết áp trung bình sẽ dưới 150 và do đó không chỉ ra rằng bệnh nhân có vấn đề về huyết áp cao là bao gồm?

> sig<-20/sqrt(5)

> pnorm(150,muy,sig)

[1] 0.1317762

c. Cần bao nhiêu phép đo để xác suất không phát hiện bệnh nhân bị huyết áp cao tối đa là 1%.

# huyết áp tối đa là 1%.

# Xác định n

# P(-X- < 150) <= .01

# Dựa theo bảng pp xx khi df = vô cùng và alpha = 0.01

# https://www.slideshare.net/ructruong/bang-tra-phn-phi-chun

# => P(-X- < 150) = P(z < (150-160)/(20/sqrt(n))) = P(z < - 2,326) = 0.01

# <=> (150-160)/(20/sqrt(n)) = -2,326

# => n = 21,64

n<-((20\*(-2.326))/(-10))^2

**Bài 5.8**: As part of the recruitment of new businesses, the city’s economic development department wants to estimate the gross profit margin of small businesses (under $1 million in sales) currently residing in the city. A random sample of the previous years annual reports of 15 small businesses shows the mean net profit margin to be 7.2% (of sales) with a standard deviation of 12.5%.

a. Construct a 99% confidence interval for the mean gross profit margin of m of all small businesses in the city.

b. The city manager reads the report and states that the confidence interval for m constructed in part (a) is not valid because the data are obviously not normally distributed and thus the sample size is too small. Based on just knowing the mean and standard deviation of the sample of 15 businesses, do you think the city manager is valid in his conclusion about the data? Explain your answer.

Dịch: Là một phần của việc tuyển dụng các doanh nghiệp mới, bộ phận phát triển kinh tế của thành phố muốn ước tính tỷ suất lợi nhuận gộp của các doanh nghiệp nhỏ (dưới 1 triệu đô la doanh thu) hiện đang cư trú trong thành phố. Một mẫu ngẫu nhiên của các báo cáo hàng năm năm trước của 15 doanh nghiệp nhỏ cho thấy tỷ suất lợi nhuận ròng trung bình là 7,2% (doanh thu) với độ lệch chuẩn là 12,5%.

1. Xây dựng một khoảng tin cậy 99% cho biên lợi nhuận gộp trung bình của μ của tất cả các doanh nghiệp nhỏ trong thành phố.

> xbar<-7.2

> sig<-12.5

> n<-15

> alpha<-0.01

> zstar<-qnorm(1-alpha/2,0,1)

> er<-zstar\*sig/sqrt(n)

> xbar-er;xbar+er

[1] -1.113453

[1] 15.51345

b. Người quản lý thành phố đọc báo cáo và nói rằng khoảng tin cậy cho μ được xây dựng ở câu (a) là không hợp lệ vì dữ liệu rõ ràng không được phân phối thông thường và do đó kích thước mẫu quá nhỏ. Dựa trên việc chỉ biết độ lệch trung bình và chuẩn của mẫu của 15 doanh nghiệp, bạn có nghĩ rằng người quản lý thành phố có hợp lệ trong kết luận của mình về dữ liệu không? Giải thích câu trả lời của bạn.

**Bài 5.9** : A program to reduce recidivism has been in effect for two years in a large northeastern state. A sociologist investigates the effectiveness of the program by taking a random sample of 200 prison records of repeat offenders. The records were selected from the files in the courthouse of the largest city in the state. The average length of time out of prison between the first and second offenses is 2.8 years with a standard deviation of 1.3 years.

a. Use this information to estimate the mean prison-free time between first and second offenses using a 95% confidence interval.

b. Identify the group for which the confidence interval would be an appropriate estimate of the population mean.

c. Would it be valid to use this confidence interval to estimate the mean prison-free time between first and second offenses for all two-time offenders in the whole state? In a large southern state?

Dịch: Một chương trình giảm tái phạm đã có hiệu lực trong hai năm ở một bang đông bắc lớn. Một nhà xã hội học điều tra hiệu quả của chương trình bằng cách lấy một mẫu ngẫu nhiên của 200 hồ sơ nhà tù của những người phạm tội lặp lại. Các hồ sơ được chọn từ các hồ sơ trong tòa án của thành phố lớn nhất trong tiểu bang. Thời gian ra tù trung bình giữa tội thứ nhất và lần thứ hai là 2,8 năm với độ lệch chuẩn là 1,3 năm.

1. Sử dụng thông tin này để ước tính thời gian tự do của tù nhân trung bình giữa lần phạm tội thứ nhất và thứ hai bằng cách sử dụng khoảng tin cậy 95%.

> xbar<-2.8

> sig<-1.3

> n<-200

> alpha<-0.05

> zstar<-qnorm(1-alpha,0,1)

> er<-zstar\*sig/sqrt(n)

> xbar-er;xbar+er

[1] 2.648799

[1] 2.951201

=> Vậy khoảng thời gian là 2.6 – 2.9 năm.

1. Xác định nhóm mà khoảng tin cậy sẽ là một ước tính thích hợp về ý nghĩa dân số.

* Khoảng tin cậy này ước tính thích hợp về ý nghĩa dân số.

1. Có hợp lệ khi sử dụng khoảng thời gian tin cậy này để ước tính thời gian không có tù trung bình giữa lần phạm tội thứ nhất và thứ hai cho tất cả người phạm tội hai lần trong toàn tiểu bang không? Ở một quốc gia miền Nam rộng lớn?

* Không hợp lệ vì với n = 200 là quá bé, mặc dù khoảng tin cậy 95% nhưng vẫn chưa đủ để đánh giá mức hiệu quả của chương trình.

**Bài 5.14** : The housing department in a large city monitors the rent for rent-controlled apartments in the city. The mayor wants an estimate of the average rent. The housing department must determine the number of apartments to include in a survey in order to be able to estimate the average rent to within $100 using a 95% confidence interval. From past surveys, the monthly charge for rent-controlled apartments ranged from $1,000 to $3,500. How many renters must be included in the survey to meet the requirements?

Dịch: Bộ phận nhà ở trong một thành phố lớn giám sát tiền thuê các căn hộ được cho thuê trong thành phố. Thị trưởng muốn ước tính giá thuê trung bình. Bộ phận nhà ở phải xác định số lượng căn hộ cần đưa vào một cuộc khảo sát để có thể ước tính giá thuê trung bình trong vòng 100 đô la bằng cách sử dụng khoảng tin cậy 95%. Từ các cuộc khảo sát trước đây, phí hàng tháng cho các căn hộ được thuê dao động từ 1.000 đến 3.500 đô la. Có bao nhiêu người thuê nhà phải được kiểm tra trong cuộc khảo sát để đáp ứng các yêu cầu?

Giải:

> e<-100

> alpha<-0.95

> sig<-(3500-1000)/4

> zstar<-qnorm(1-alpha)

> n<-(zstar\*sig/e)^2

> n

[1] 105.6853

=> Vậy 105 người thuê nhà được bao gồm trong cuộc khảo sát để đáp ứng các yêu cầu.

**Bài 5.15** : Refer to Exercise 5.14. Suppose the mayor's staff reviews the proposed survey and decides that in order for the survey to be taken seriously the requirements need to be increased.

a. If the level of confidence is increased to 99% with the average rent estimated within $50, how many apartments need to be included in the survey?

b. Suppose the budget for the survey will not support increasing the level of confidence to 99%. Provide an explanation to the mayor, who has never taken a statistics course, of the impact on the accuracy of the estimate of the average rent of not raising the level of confidence from 95% to 99%.

Dịch: Tham khảo bài tập 5.14. Giả sử nhân viên của thị trưởng xem xét cuộc khảo sát được đề xuất và quyết định rằng để cuộc khảo sát được thực hiện nghiêm túc, các yêu cầu cần phải được tăng lên.

a. Nếu mức độ tin tưởng được tăng lên 99% với giá thuê trung bình ước tính trong vòng 50 đô la, có bao nhiêu căn hộ cần được đưa vào cuộc khảo sát?

b. Giả sử ngân sách cho cuộc khảo sát sẽ không hỗ trợ tăng mức độ tin tưởng lên 99%. Cung cấp một lời giải thích cho thị trưởng, người chưa bao giờ tham gia một khóa học thống kê, về tác động đến tính chính xác của ước tính tiền thuê trung bình không nâng mức độ tin tưởng từ 95% lên 99%.

Giải:

**Bài 5.16** A study is designed to test the hypotheses H0: µ ≥ 26 versus H1: µ ≤ 26. A random sample of 50 units was selected from a specified population, and the measurements were summarized to ybar = 25.9 and s = 7.6.

a. With alpha = .05, is there substantial evidence that the population mean is less than 26?

b. Calculate the probability of making a Type II error if the actual value of the population mean is at most 24.

c. If the sample size is doubled to 100, what is the probability of making a Type II error if the actual value of the population mean is at most 24?

Dịch: Một nghiên cứu được thiết kế để kiểm tra các giả thuyết H0: μ ≥ 26 so với H1: μ ≤ 26. Một mẫu ngẫu nhiên của 50 đơn vị đã được chọn từ một quần thể được chỉ định, và các phép đo được tóm tắt thành ybar = 25,9 và s = 7,6.

a. Với alpha = 0,05, có bằng chứng đáng kể nào cho thấy dân số trung bình dưới 26 không?

> #H0: mu>=mu0; H1:mu<mu0

> mu0<-26

> n<-0

> n<-50

> xbar<-25.9

> s2<-7.6

> s<-sqrt(s2)

> alpha<-0.05

> z0<-(xbar-mu0)/(s/sqrt(n))

> zstar<-qnorm(1-alpha)

> z0< -zstar

[1] FALSE

> # vậy dân số trung bình không dưới 26

b. Tính xác suất mắc lỗi Loại II nếu giá trị thực tế của dân số trung bình tối đa là 24.

c. Nếu kích thước mẫu tăng gấp đôi lên 100, xác suất mắc lỗi loại II là gì nếu giá trị thực tế của dân số có nghĩa là tối đa là 24?

**Bài 5.18** : Use a computer to simulate 100 samples of n = 25 from a normal distribution with μ = 43 and alpha = 4. Test the hypotheses H0: μ = 43 versus Ha: μ ≠ 43 separately for each of the 100 samples of size 25 with alpha = .05.

a. How many of the 100 tests of hypotheses resulted in a rejection of H0?

b. Suppose 1,000 tests of hypotheses of H0: μ = 43 versus Ha: μ ≠ 43 were conducted. Each of the 1,000 data sets consists of n = 50 data values randomly selected from a population having μ = 43. Suppose alpha = .05 is used in each of the 1,000 tests. On the average, how many of the 1,000 tests would result in the rejection of H0?

c. Suppose the procedure in part (b) is repeated with 1,000 tests with n = 75 and alpha = .01. On the average, how many of the 1,000 tests would result in a rejection of H0?

Dịch: Sử dụng máy tính để mô phỏng 100 mẫu n = 25 từ một phân phối chuẩn với μ = 43 và alpha = 4. Kiểm tra các giả thuyết H0: μ = 43 so với H1: μ ≠ 43 riêng cho mỗi trong số 100 mẫu có kích thước 25 với alpha = 0,05.

# Sử dụng máy tính để mô phỏng 100 mẫu n = 25 từ phân phối chuẩn

# với u = 43 và alpha = 4. Kiểm tra các giả thuyết H0: u = 43 so

# với Ha: u <> 43 riêng biệt cho mỗi 100 mẫu cỡ 25 với alpha = 0,05.

# khoảng tin cậy:

a<-.05

u<-43

uo<-43

o<-4

n<-25

r<-100

twotailed<-1

y<-rep(0,n)

L<-rep(0,r)

U<-rep(0,r)

chk<-0

if (twotailed==1){a<-a/2}

for(i in 1:r)

{

y<-rnorm(n,u,o)

L[i]<- mean(y)-qnorm(a,0,1,lower.tail=FALSE)\*(o/sqrt(n))

U[i]<- mean(y)+qnorm(a,0,1,lower.tail=FALSE)\*(o/sqrt(n))

if (L[i]>uo | U[i]<uo)

{

chk<-chk+1

}

}

x.bar<-8900

s<-500

n<-35

alpha<-1-0.95

Zstar<-qnorm(1-alpha/2,0,1)

err<-Zstar\*s/sqrt(n)

x.bar-err;x.bar+err

1. Có bao nhiêu trong số 100 bài kiểm tra giả thuyết dẫn đến việc từ chối H0?
2. Giả sử 1.000 bài kiểm tra các giả thuyết của H0: μ = 43 so với H1: μ ≠ 43 đã được tiến hành. Mỗi bộ dữ liệu trong số 1.000 bao gồm n = 50 giá trị dữ liệu được chọn ngẫu nhiên từ một dân số có μ = 43. Giả sử alpha = .05 được sử dụng trong mỗi 1.000 bài kiểm tra. Trung bình, có bao nhiêu trong số 1.000 xét nghiệm sẽ dẫn đến việc từ chối H0?

# được tiến hành. Mỗi bộ trong số 1.000 bộ dữ liệu bao gồm n = 50 giá trị dữ

# liệu được chọn ngẫu nhiên từ một tập hợp có u = 43. Giả sử alpha = 0,05 được

# sử dụng trong mỗi 1.000 phép thử. Trung bình, có bao nhiêu trong số 1.000 bài

# kiểm tra có kết quả loại bỏ H0?

a<-.05

u<-43

uo<-43

o<-4

n<-50

r<-1000

twotailed<-1

y<-rep(0,n)

L<-rep(0,r)

U<-rep(0,n)

chk<-0

if (twotailed==1){a<-a/2}

for(i in 1:r)

{

y<-rnorm(n,u,o)

L[i]<- mean(y)-qnorm(a,0,1,lower.tail=FALSE)\*(o/sqrt(n))

U[i]<- mean(y)+qnorm(a,0,1,lower.tail=FALSE)\*(o/sqrt(n))

if (L[i]>uo | U[i]<uo)

{

chk<-chk+1

}

}

c. Giả sử thủ tục một phần (b) được lặp lại với 1.000 xét nghiệm với n = 75 và alpha = .01. Trung bình, có bao nhiêu trong số 1.000 xét nghiệm sẽ dẫn đến việc từ chối H0?

# n = 75 và a = 0,01. Trung bình, có bao nhiêu trong số 1.000 bài kiểm

# tra dẫn đến loại bỏ H0?

a<-.01

u<-43

uo<-43

o<-4

n<-75

r<-1000

twotailed<-1

y<-rep(0,n)

L<-rep(0,r)

U<-rep(0,r)

chk<-0

if (twotailed==1){a<-a/2}

for(i in 1:r)

{

y<-rnorm(n,u,o)

L[i]<- mean(y)-qnorm(a,0,1,lower.tail=FALSE)\*(o/sqrt(n))

U[i]<- mean(y)+qnorm(a,0,1,lower.tail=FALSE)\*(o/sqrt(n))

if (L[i]>uo | U[i]<uo)

{

chk<-chk+1}}

**Bài 5.24** : To evaluate the success of a 1-year experimental program designed to increase the mathematical achievement of underprivileged high school seniors, a random sample of participants in the program will be selected and their mathematics scores will be compared with the previous year’s statewide average of 525 for underprivileged seniors. The researchers want to determine whether the experimental program has increased the mean achievement level over the previous year’s statewide average. If alpha = .05, what sample size is needed to have a probability of Type II error of at most .025 if the actual mean is increased to 550? From previous results, σ ≈ 80.

Dịch: Để đánh giá sự thành công của một chương trình thực nghiệm 1 năm được thiết kế để tăng thành tích toán học của học sinh trung học kém may mắn, một mẫu ngẫu nhiên của những người tham gia chương trình sẽ được chọn và điểm toán của họ sẽ được so sánh với mức trung bình toàn tiểu bang năm trước là 525 cho người cao niên có hoàn cảnh khó khăn. Các nhà nghiên cứu muốn xác định xem chương trình thử nghiệm có làm tăng mức độ thành tích trung bình so với mức trung bình toàn tiểu bang năm trước hay không. Nếu alpha = .05, kích thước mẫu nào là cần thiết để có xác suất lỗi loại II nhiều nhất là 0,025 nếu trung bình thực tế được tăng lên 550? Từ kết quả trước, σ ≈ 80.

Giải:

= 0.05

= 0.025, power = 0.975

80

= 550 – 525 = 25

> power.t.test(delta=25,sd=80,sig.level=.05,power=.975,type='one.sample')

One-sample t test power calculation

n = 159.282

delta = 25

sd = 80

sig.level = 0.05

power = 0.975

alternative = two.sided

**Bài 5.28** :The R&D department of a paint company has developed an additive that it hopes will increase the ability of the company’s stain for outdoor decks to resist water absorption. The current formulation of the stain has a mean absorption rate of 35 units. Before changing the stain, a study was designed to evaluate whether the mean absorption rate of the stain with the additive was decreased from the current rate of 35 units. The stain with the additive was applied to 50 pieces of decking material. The resulting data were summarized to ybar = 33.6 and s = 9.2

a. Is there substantial evidence (alpha = .01) that the additive reduces the mean absorption from its current value?

b. What is the level of significance (p-value) of your test results?

c. What is the probability of a Type II error if the stain with the additive in fact has a mean absorption rate of 30?

d. Estimate the mean absorption using a 99% confidence interval. Is the confidence interval consistent with your conclusions from the test of hypotheses?

Dịch: Bộ phận R &D của một công ty sơn đã phát triển một chất phụ gia mà họ hy vọng sẽ làm tăng khả năng vết bẩn của công ty cho sàn ngoài trời để chống lại sự hấp thụ nước. Công thức hiện tại của vết bẩn có tỷ lệ hấp thụ trung bình là 35 đơn vị. Trước khi thay đổi vết bẩn, một nghiên cứu được thiết kế để đánh giá liệu tỷ lệ hấp thụ trung bình của vết bẩn với chất phụ gia có giảm từ tỷ lệ hiện tại là 35 đơn vị hay không. Vết bẩn với chất phụ gia đã được áp dụng cho 50 mảnh vật liệu sàn. Dữ liệu kết quả được tóm tắt vào ybar = 33,6 và s = 9,2

a. Có bằng chứng đáng kể (alpha = 0,01) rằng chất phụ gia làm giảm sự hấp thụ trung bình từ giá trị hiện tại của nó không?

> #H0: mu = mu0

> #H1: mu < mu0

> n<-50

> xbar<-33.6

> mu0<-35

> s<-9.2

> sd<-sqrt(9.2)

> z0<-(xbar-mu0)/(sd/sqrt(n))

> alpha<-0.01

> zstar<-qnorm(1-alpha)

> z0< -zstar

[1] TRUE

> #bac bo H0, chap nhan H1

b. Mức độ ý nghĩa (giá trị p) của kết quả kiểm tra của bạn là gì?

> p<- pnorm(z0)

> p

[1] 0.000549708

c. Xác suất của lỗi loại II là bao nhiêu nếu vết bẩn với phụ gia trên thực tế có tỷ lệ hấp thụ trung bình là 30?

d. Ước tính sự hấp thụ trung bình bằng cách sử dụng khoảng tin cậy 99%. Khoảng tin cậy có phù hợp với kết luận của bạn từ việc kiểm tra các giả thuyết không?

**Bài 5.30** : A concern to public health officials is whether a concentration of lead in the paint of older homes may have an effect on the muscular development of young children. In order to evaluate this phenomenon, a researcher exposed 90 newly born mice to paint containing a specified amount of lead. The number of Type 2 fibers in the skeletal muscle was determined 6 weeks after exposure. The mean number of Type 2 fibers in the skeletal muscles of normal mice of this age is 21.7. The n = 90 mice yielded ybar = 18.8, s = 15.3. Is there significant evidence in the data to support the hypothesis that the mean number of Type 2 fibers is different from 21.7 using an alpha = .05 test?

Dịch: Một mối quan tâm đối với các quan chức y tế công cộng là liệu sự tập trung chì trong sơn của những ngôi nhà cũ có thể có ảnh hưởng đến sự phát triển cơ bắp của trẻ nhỏ hay không. Để đánh giá hiện tượng này, một nhà nghiên cứu đã cho 90 con chuột mới sinh ra vẽ có chứa một lượng chì nhất định. Số lượng sợi loại 2 trong cơ xương được xác định 6 tuần sau khi tiếp xúc. Số lượng trung bình của các sợi loại 2 trong cơ xương của chuột bình thường ở độ tuổi này là 21,7. N = 90 con chuột mang lại ybar = 18,8, s = 15,3. Có bằng chứng đáng kể trong dữ liệu để hỗ trợ giả thuyết rằng số lượng trung bình của sợi loại 2 khác với 21,7 bằng cách sử dụng xét nghiệm alpha = 0,05?

alpha <- 0.05

xbar <- 21.7

muy0 <- 18.8

sdev <- 15.3

n <- 90

z0 <- (xbar-muy0)/(sdev/sqrt(n))

hs <- 1-alpha/2

zhs <- qnorm(hs,mean=0,sd=1)

z0

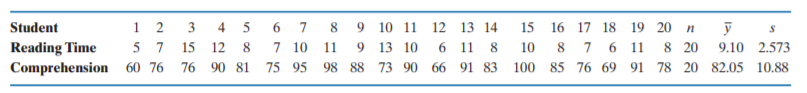
zhs

phiz0 <- pnorm(z0,mean=0,sd=1)

p <- 2\*(1-phiz0)

p

**Bài 5.36** The ability to read rapidly and simultaneously maintain a high level of comprehension is often a determining factor in the academic success of many high school students. A school district is considering a supplemental reading program for incoming freshmen. Prior to implementing the program, the school runs a pilot program on a random sample of n = 20 students. The students were thoroughly tested to determine reading speed and reading comprehension. Based on a fixed-length standardized test reading passage, the following reading times (in minutes) and comprehension scores (based on a 100-point scale) were recorded.



a. What is the population about which inferences are being made?

b. Place a 95% confidence interval on the mean reading time for all incoming freshmen in the district.

c. Plot the reading time using a normal probability plot or boxplot. Do the data appear to be a random sample from a population having a normal distribution?

d. Provide an interpretation of the interval estimate in part (b).

Dịch: Khả năng đọc nhanh và đồng thời duy trì mức độ hiểu cao thường là yếu tố quyết định thành công trong học tập của nhiều học sinh trung học. Một khu học chánh đang xem xét một chương trình đọc bổ sung cho sinh viên năm nhất sắp tới. Trước khi thực hiện chương trình, trường chạy một chương trình thí điểm trên một mẫu ngẫu nhiên của n = 20 học sinh. Các học sinh đã được kiểm tra kỹ lưỡng để xác định tốc độ đọc và đọc hiểu. Dựa trên một đoạn đọc bài kiểm tra tiêu chuẩn cố định, thời gian đọc sau đây (tính bằng phút) và điểm hiểu (dựa trên thang điểm 100) đã được ghi lại.

a. Quần thể về những suy luận nào đang được thực hiện?

b. Đặt khoảng tin cậy 95% về thời gian đọc trung bình cho tất cả sinh viên năm nhất đến trong quận.

c. Vẽ bản đồ thời gian đọc bằng một đồ thị xác suất phân phối chuẩn hoặc boxplot. Dữ liệu có phải là một mẫu ngẫu nhiên từ một dân số có phân phối chuẩn không?

d. Giải thích về ước tính khoảng thời gian một phần (b).

**Bài 5.43** : Refer to Exercise 5.36.

a. Use a computer program to obtain 10,000 bootstrap samples from the 20 comprehension scores. Use these 10,000 samples to obtain the bootstrap p-value for the t test of Ha: µ > 80.

b. Compare the p-value from part (a) to the p-value obtained in Exercise 5.37.

Dịch: Tham khảo bài tập 5.36.

a. Sử dụng một chương trình máy tính để có được 10.000 mẫu bootstrap từ 20 điểm hiểu. Sử dụng 10.000 mẫu này để có được giá trị p bootstrap cho t test của Hà: μ > 80.

b. So sánh giá trị p từ một phần (a) với giá trị p thu được trong Bài tập 5.37.

**Bài 5.51** The amount of money spent on health care is an important issue for workers because many companies provide health insurance that only partially covers many medical procedures. The director of employee benefits at a midsize company wants to determine the amount spent on health care by the typical hourly worker in the company. A random sample of 25 workers is selected, and the amounts they spent on their families’ health care needs during the past year are given here. 400 345 248 1,290 398 218 197 342 208 223 531 172 4,321 143 254 201 3,142 219 276 326 207 225 123 211 108

a. Graph the data using a boxplot or normal probability plot, and determine whether the population has a normal distribution.

b. Based on your answer to part (a), is the mean or the median cost per household a more appropriate measure of what the typical worker spends on health care needs?

c. Place a 95% confidence interval on the amount spent on health care by the typical worker. Explain what the confidence interval is telling us about the amount spent on health care needs. d. Does the typical worker spend more than $400 per year on health care needs? Use alpha = .05.

Dịch: Số tiền chi cho chăm sóc sức khỏe là một vấn đề quan trọng đối với người lao động vì nhiều công ty cung cấp bảo hiểm y tế chỉ chi trả một phần cho nhiều thủ tục y tế. Giám đốc phúc lợi của nhân viên tại một công ty cỡ trung bình muốn xác định số tiền chi cho chăm sóc sức khỏe của công nhân hàng giờ điển hình trong công ty. Một mẫu ngẫu nhiên của 25 công nhân được chọn, và số tiền họ đã chi cho nhu cầu chăm sóc sức khỏe của gia đình họ trong năm qua được đưa ra ở đây. 400 345 248 1.290 398 218 197 342 208 223 531 172 4.321 143 254 201 3.142 219 276 326 207 225 123 211 108

1. Vẽ đồ thị dữ liệu bằng cách sử dụng một lô đo hình hộp hoặc lô xác suất bình thường và xác định xem dân số có phân bố bình thường hay không.

> x<-c(400,345,248,1290,398,218,197,342,208,223,531,172,4321,143,254,201,3142,219,276,326,207,225,123,211,108)

> barplot(x)

b. Dựa trên câu trả lời của bạn cho một phần (a), là trung bình hoặc chi phí trung bình cho mỗi hộ gia đình là thước đo thích hợp hơn về những gì người lao động điển hình chi cho nhu cầu chăm sóc sức khỏe?

> mean(x)

[1] 573.12

> median(x)

[1] 225

c. Đặt một khoảng tin cậy 95% về số tiền chi cho chăm sóc sức khỏe của người lao động điển hình. Giải thích những gì khoảng tin cậy đang cho chúng ta biết về số tiền chi cho nhu cầu chăm sóc sức khỏe.

1. Người lao động điển hình có chi hơn 400 đô la mỗi năm cho nhu cầu chăm sóc sức khỏe không? Sử dụng alpha = .05.

> #H0: mu >= 400; H1: mu < 400

> xbar<-mean(x)

> mu0<-400

> s2<-var(x)

> s<-sqrt(s2)

> n<-length(x)

> z0<-(xbar-mu0)/(s/sqrt(n))

> alpha<-0.05

> zstar<-qnorm(1-alpha)

> z0< -zstar

[1] FALSE

> #bac bo h0, chap nhan H1