Домашна работа 3

Изготвена от:

Веселина Венциславова Кръстева,

ИС, II курс, фн. 71283

Зад. 7

n = 16; x = 2; y = 8; z = 3

prob = numeric(0)

for (i in 1:100000){ # Провеждаме 100 000 симулации за хвърляне на 16 зара

a = sample(1:6, n, replace = T)

# Намираме броят на падналите се 6, 5, 4

six = length(which(a == 6))

five = length(which(a == 5))

four = length(which(a == 4))

if (six == x & five == y & four == z){ prob[i] = 1}

else {prob[i] = 0}

}

mean(prob)

[1] 7e-05

Зад. 6

Използваме chi square тест за хомогенност

Нулева хипотеза: H0 – трите множества от данни идват от равномерно разпределение

Алтернативна хипотеза: НА – от различни разпределения.

В контекста на задачата тестваме дали НА е в сила – разликата в степента на успешност на А, В и С говори, че те са част от различни разпределения

> res\_a = c(63, 37)

> res\_b = c(71, 29)

> res\_c = c(55, 45)

> table = rbind(res\_a, res\_b, res\_c)

# Chi square тестът работи върху групирани данни

> chisq.test(table)

 Pearson's Chi-squared test

data: table

X-squared = 5.4912, df = 2, p-value = 0.06421

Стойността на на p-value е по-голяма от α=0.05 => на база случайната извадка за покупките нямаме основания да отхвърлим нулевата хипотеза. Това означава, че тези данни не са достатъчно доказателство за различие в нивото на успеваемост на управителите.

Зад. 2

a)

>data = c(7.4, 7.1, 6.5, 7.5, 7.6, 6.3, 6.9, 7.7, 6.5, 7.0)

>m = mean(data) # Намиране на средното

[1] 7.05

>sd = sd(data) # Намиране на стандартното отклонение

[1] 0.4994441

б) Функция, намираща 99% доверителен интервал за m

simple.z.test = function(x, sigma, conf.level=0.99)

{

 n = length(x) # В нашия случай n = 10

 xbar = mean(x)

 a = 1 - conf.level

 zstar = qnorm(1 - a/2)

 SE = sigma/sqrt(n)

 xbar + c(-zstar\*SE, zstar\*SE)

}

>a = simple.z.test(data, sd)

>a

[1] 6.643178 7.456822

в) Тестване на H0: µ = 7.5 срещу НА: µ<7.5

>t = (m - 7.5)/(sd/sqrt(10))

>pt(t, df = 9) # Намиране на p-value

[1] 0.009556784

p-value < α = 0.01 => имаме основания да отхвърлим H0 в полза на НА

Зад. 5

Функция, която прави симулация на едно разиграване между двамата играчи. При равен резултат функцията се извиква наново. Тя връща 1 или 0 взависимост от това кой е спечелил.

play = function(){

res = 0

a = sample(1:2, 3, replace = T) # Играч А хвърля 3 монети

b = sample(1:2, 2, replace = T) # Играч В хвърля 2 монети

heads\_a = length(which(a == 1))

heads\_b = length(which(b == 1))

if(heads\_a > heads\_b) res = 1

if(heads\_a < heads\_b) res = 0

if(heads\_a == heads\_b) res = play()

res

}

Функция, която прави симулация на n игри (в нашия случай n = 1000) и връща вектор с дължина n състоящ се от 0 и 1.

sim = function(n){

res = numeric(0)

for (i in 1:n) res[i] = play()

res

}

Функция, която повтаря 100 пъти горната процедура. За всяка симулация на 1000 игри намираме вероятността играч А да спечели и връщаме вектор с усреднените вероятности от всяко от 100-те повторения на опита.

prob = function(){

x = numeric(0)

for(i in 1:100){

a = sim(1000)

p = length(which(a==1))/length(a)

x[i] = p

}

x

}

>phat = prob()

 [1]0.704 0.753 0.722 0.734 0.759 0.695 0.725 0.725 0.722 0.717 0.734 0.732

 [13] 0.700 0.733 0.733 0.730 0.730 0.739 0.716 0.720 0.739 0.720 0.754 0.732

 [25] 0.749 0.699 0.724 0.743 0.712 0.737 0.724 0.705 0.716 0.760 0.709 0.734

 [37] 0.750 0.692 0.711 0.748 0.727 0.733 0.748 0.711 0.714 0.709 0.717 0.717

 [49] 0.727 0.724 0.748 0.717 0.742 0.723 0.702 0.738 0.719 0.764 0.744 0.731

 [61] 0.732 0.726 0.719 0.736 0.717 0.732 0.713 0.702 0.722 0.739 0.743 0.748

 [73] 0.729 0.733 0.688 0.701 0.749 0.722 0.707 0.737 0.745 0.736 0.716 0.718

 [85] 0.728 0.742 0.727 0.737 0.736 0.727 0.750 0.729 0.704 0.737 0.708 0.710

 [97] 0.727 0.713 0.747 0.732

Теоретичната вероятност първият играч да спечели е 0.(72)

Построяваме 95% доверителните интервали:

>SE = sqrt(phat\*(1-phat)/1000)

>alpha = 0.05;zstar = qnorm(1-alpha/2)

>matplot(rbind(phat - zstar\*SE, phat + zstar\*SE),rbind(1:100,1:100),type="l",lty=1)

>abline(v=0.728)

Теоретично 95% от построените интервали съдържат истинската вероятност. В нашата симулация виждаме, че 94% от тях съдържат р (6 хоризонтални линии не пресичат вертикалната линия х = 0.728)



Зад. 1

а)

H0 – нулева хипотеза: pH = µ < 7.5

НА – алтернативна хипотеза: pH = µ >= 7.5

=> Това е едностранен тест.

б)

Пресмятаме t – тест статистика

>xbar = 7.3; s = 0.2; n = 30; mu = 7.5

>t = (xbar - mu) / (s/sqrt(n))

[1] -5.477226

>t.alpha = qnorm(1 -0.05)

[1] 1.644854

Тази стойност > t => нямаме основание да отхвърлим нулевата хипотеза

в)

pvalue = pnorm(z, lower.tail=FALSE)

Зад. 3:

1 – контролна група, 2 – третирана група

а) H0 : µ1 = µ2

 HA : µ1 > µ2

б)

m1 = 1.26; m2 = 0.78; var = 0.32\*0.32

alpha = 0.05

zalpha = qnorm(1-alpha) # Критичната стойност, за която стойностите на T\*(tvalue) по-големи от нея навлизат в rejection region.

[1] 1.644854

Стойностите на стандартните отклонения от двете извадки са равни затова за намиране tvalue използваме следната формула:

s\_square = sqrt(((7 - 1) \* var + (7 - 1) \* var)/(7 + 7 - 2))

tvalue = (m1 – m2)/(s\_square \* sqrt(1/7 + 1/7))

[1] 2.806243

Tvalue > zalpha => можем да отхвърлим нулевата хипотеза в полза на алтернативната, което доказва, че третирането с лекарството наистина дава резултат.

в) pvalue = 2 \* pt(tvalue, df = 12, lower.tail = F)

[1] 0.01585955

Pvalue < α = 0.05 => отново потвърждаваме получения резултат, че можем да отхвърлим нулевата хипотеза