Задачи по мат.статистике

Nina Alexeyeva

16.03.2020

Про подходящий размер выборки

A survey of low-incomes families aws designed to determine the average heating costs for a family of 4 during January and February. Heating costs are known to have a standard deviation of $25,75. The economists conducting the study wish to cinstruct a 95 percent confidence interval with a margin for sampling error of no more than $3.95. Find the appropriate sample size.

Опрос семей с низкими доходами, предназначенный для определения средних затрат на отопление семьи из 4 человек в январе и феврале. Как известно, стандартное отклонение затрат на отопление составляет 25,75 долл. США. Экономисты, проводящие исследование, хотят построить 95-процентный доверительный интервал с пределом погрешности выборки не более 3,95 долларов США. Найдите подходящий размер выборки.

$$\frac{S}{\sqrt{n}} Z_{1-\alpha/2} = 3.95, \implies n = \left(\frac{SZ_{1-\alpha/2}}{3.95}\right)^2$$

```
S<-25.75
Z_a<-qnorm(0.975);Z_a
```

```
## [1] 1.959964
```

```
n<-(S*Z_a)^2/3.95^2;n
```

```
## [1] 163.2512
```

Построение доверительного интервала

```
CI<-function(x_,S,n,P,sigma.known)
{
  if(sigma.known) Z_a<-qnorm(1-(1-P)/2) else Z_a<-qt(1-(1-P)/2,n-1)

  M<-Z_a*S/sqrt(n)

  c(x_-M,x_+M)
}
```

Про выборы

A poll reported that 48 percent of probable votes seem determined to vote against the President. Assume that this sample was based on a random selection of 789 probable voters. Construct a 99 percent confidence interval for the probable voters who seem determined to vote aganst the President.

Опрос показал, что 48 процентов вероятных голосов, похоже, настроены голосовать против президента. Предположим, что эта выборка была основана на случайной выборке из 789 вероятных избирателей.

Построить доверительный интервал в 99 процентов для вероятных избирателей, которые, похоже, настроены голосовать против президента.

$$\bar{x} - Z_{1-\alpha/2}\frac{\sqrt{pq}}{\sqrt{n}} < \mu < \bar{x} + Z_{1-\alpha/2}\frac{\sqrt{pq}}{\sqrt{n}}$$

```
n<-789
x_ <-0.48
P<-0.99
S<-sqrt(x_*(1-x_))

CI(x_,S,n,P,sigma.known=TRUE)
```

```
## [1] 0.4341857 0.5258143
```

Построение доверительного интервала для увеличения расходов (при известном среднем)

A survey indicated that companies with fewer than 1,000 employers are expect to increase their spending by 20,4%. Form a 99 percent confidence interval for the unknown mean increase, assuming that the sample standart deviation is 6,8% and the sample size is 346.

Опрос показал, что компании с числом работодателей менее 1000, как ожидается, увеличат свои расходы на 20,4%. Сформируйте 99-процентный доверительный интервал для неизвестного среднего увеличения, предполагая, что стандартное отклонение выборки составляет 6,8%, а размер выборки составляет 346.

$$\bar{x} - Z_{1-\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + Z_{1-\alpha/2}\frac{S}{\sqrt{n}}$$

```
x_<-20.4
P<-0.99
S<-6.8
n<-346
CI(x_,S,n,P,sigma.known=TRUE)
```

```
## [1] 19.45835 21.34165
```

Построение доверительного интервала для фрезерования этой детали (при известном или неизвестном среднем)

A company has just installed a new automatic milling machine. The time it takes the machine to mill a particular part is recorded for a sample of 9 observations. the mean time if found to be $\bar{x} = 8,50$ seconds, $S^2 = 0,0064$. Find a 90 percent confidence interval for the unknown mean time milling this part.

Компания только что установила новый автоматический фрезерный станок. Время, необходимое машине для фрезерования конкретной детали, записывается для выборки из 9 наблюдений. среднее время, если установлено, что $\bar{x} = 8,50$ секунд, $S^2 = 0,0064$. Найдите 90-процентный доверительный интервал для неизвестного среднего времени фрезерования этой детали.

$$\bar{x} - Z_{1-\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + Z_{1-\alpha/2}\frac{S}{\sqrt{n}},$$

$$\bar{x} - t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{1-\alpha/2}(n-1)\frac{S}{\sqrt{n}}$$

```
x_<-8.50
S2<-0.0064
n<-9
P<-0.95


CI(x_,S,n,P,TRUE)
```

```
## [1]  4.057415 12.942585
```

```
CI(x_,S,n,P,FALSE)
```

```
## [1]  3.273057 13.726943
```

Про баскетболистов

A sample of 100 former basketball players form Slam Dunk University shows that 55 of the players graduated in 4 years. Construct a 90 percent confidence interval for the proportion of basketball players graduating in 4 years from Slam Dunc U.

Выборка из 100 бывших баскетболистов из университета Slam Dunk показывает, что 55 игроков получили высшее образование за 4 года. Постройте 90-процентный доверительный интервал для доли баскетболистов, которые через 4 года заканчивают Slam Dunc U.

```
n<-100
x_ <-55/100
P<-0.9
S<-sqrt(x_*(1-x_))

CI(x_,S,n,P,sigma.known=TRUE)
```

```
## [1] 0.4681696 0.6318304
```

Про кофе

A sample of 20 cups of coffee from a coffee machine has a mean amount of coffee of 6 ounces. The standard deviation is known to be 0.5. Construct a 99 percent confidence interval for the mean amoumt of coffee per cup. The second variant, 0.5 is the sample standard deviation.

Образец 20 чашек кофе из кофемашины имеет среднее количество кофе 6 унций. Стандартное отклонение известно как 0,5. Составьте 99-процентный доверительный интервал для среднего количества кофе на чашку. Второй вариант, 0,5 - стандартное отклонение выборки.

```
CI(x_=6,S=0.5,n=20,P=0.99,TRUE)
```

```
## [1] 5.712014 6.287986
```

```
CI(x_=6,S=0.5,n=20,P=0.99,FALSE)
```

```
## [1] 5.680138 6.319862
```

```
CI(x_=6,S=0.5,n=20,P=0.9,TRUE)
```

```
## [1] 5.8161 6.1839
```

```
CI(x_=6,S=0.5,n=20,P=0.9,FALSE)
```

```
## [1] 5.806677 6.193323
```

Про автомобили и расход топлива

A concumer rights organization tests a new cars to estimate the car's average gasoline mileage. Because its budget is limited, the organization can test only 25 cars. The standard deviation of the cars tested is 2. What is the range of the 90 percent confidence interval?

{Организация по защите прав потребителей тестирует новые автомобили, чтобы оценить средний пробег бензина. Поскольку его бюджет ограничен, организация может протестировать только 25 автомобилей. Стандартное отклонение протестированных автомобилей составляет 2. Каков диапазон 90-процентного доверительного интервала?}

```
ci<-CI(x_=11,S=2,n=20,P=0.90,TRUE)
ci[2]-ci[1]
```

```
## [1] 1.471202
```

Про рогалики

The owner of a local bakery feels that too many bagels are thrown out every night, so he decides o estimate the demand for bagels. After a month's observation, he collected 30 days'sales and ascertainted that the average sales were 120 and the standard deviation of daily sales was 10. Assume that the daily bagel sale follow a normal distribution. Construct a 90 percent confidence interval for the demand for bagels.

{Владелец местной пекарни чувствует, что каждую ночь выбрасывают слишком много бубликов, поэтому он решает оценить потребность в бубликах. После месячного наблюдения он собрал 30-дневные продажи и установил, что средние продажи составляли 120, а стандартное отклонение ежедневных продаж составляло 10. Предположим, что ежедневные продажи бублика следуют нормальному распределению. Построить 90-процентный доверительный интервал для спроса на рогалики.}

```
ci<-CI(x_=120,S=10,n=30,P=0.90,FALSE)
ci
```

```
## [1] 116.8978 123.1022
```

Проверка равенства средних

```
T.test<-function(means,SD,nn)
{
  df<-data.frame(Means=means,SD=SD,nn=nn);df
  df<-df[order(df$SD,decreasing=TRUE),]
  F.<-df$SD[1]^2/df$SD[2]^2;F.

alpha.f<-1-pf(F.,df$nn[1]-1,df$nn[2]-1);alpha.f

n1<-df$nn[1];n2<-df$nn[2];S1<-df$SD[1];S2<-df$SD[2];
x_<-df$Means[1];y_<-df$Means[2]
```

```
S<-sqrt( ((n1-1)*S1^2+(n2-1)*S2^2 )/(n1+n2-2) );S
T1<-(x_-y_)/S/sqrt(1/n1+1/n2);
k1<-n1+n2-2;
p.value.1<-2*(1-pt(abs(T1),k1));

T2<-(x_-y_)/sqrt(S1^2/n1+S2^2/n2);
k2<-(S1^2/n1+S2^2/n2)^2/( (S1^2/n1)^2/(n1-1) +(S2^2/n2)^2/(n2-1));
p.value.2<-2*(1-pt(abs(T2),k2));

list(
  alpha.f=alpha.f,

  Var.equal=c(T=T1,k=k1,p=p.value.1),

  Var.not.equal=c(T=T2,k=k2,p=p.value.2))
}
```

Про фермера и удобрения

Suppose a farmer is interested in testing two fertilizers to see which is more effective. He uses two fertilizers and gets the following results.

$Fertilizer$	$MeanGrowth$	$StandardDeviation$	$SizeofSample$
A	7	0.5	100
B	6	0.2	125

Test, atthe 10 percent level of significance, the hypotesis that the mean difference in growth between the two fertilizers is not significant.

Предположим, фермер заинтересован в тестировании двух удобрений, чтобы увидеть, какие из них более эффективны. Он использует два удобрения и получает следующие результаты. Проверьте на уровне значимости 10% гипотезу о том, что средняя разница в росте между двумя удобрениями несущественна.

В случае неизвестных неодинаковых дисперсий $\sigma_1^2 \neq \sigma_2^2$ ста-тис-ти-ка T имеет приближенно распределение Стьюдента с не-це-лым числом степеней свободы.

$$T_1 = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T(n_1 + n_2 - 2), \quad S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$T_2 = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathbf{T}(k), \quad \text{где} \quad k = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{\frac{(S_1^2/n_1)^2}{n_1-1} + \frac{(S_2^2/n_2)^2}{n_2-1}}.$$

```
T.test(means=c(7,6),SD=c(0.5,0.2),nn=c(100,125))
```

```
## $alpha.f
## [1] 0
##
## $Var.equal
##         T         k         p
##  20.42043 223.00000   0.00000
##
## $Var.not.equal
```

```
##          T          k          p
## 18.83109 124.33956   0.00000
```

Про менеджера производственного отдела

A production manager interested in the number of defects in batches derived from different production processes. He examines a random sample drawn from each process and records the following data.

$Process$	$MeanDefects$	$StandardDeviation$	$SizeofSample$
A	221	25	90
B	300	80	110

Test, at the 1 recent level of significance, the hypothesis that the mean difference in number of defects between the two production processes is not significant.

Менеджер по производству заинтересован в количестве дефектов в партиях, полученных из разных производственных процессов. Он проверяет случайную выборку, взятую из каждого процесса, и записывает следующие данные.

Проверьте на первом уровне значимости гипотезу о том, что средняя разница в количестве дефектов между двумя производственными процессами несущественна.

```
T.test(means=c(221,300),SD=c(25,80),nn=c(90,110))
```

```
## $alpha.f
## [1] 0
##
## $Var.equal
##            T            k            p
## 9.011559e+00 1.980000e+02 2.220446e-16
##
## $Var.not.equal
##          T          k          p
##   9.789241 134.230794   0.000000
```

Построение критериев

For each of the following, test the indicated hypothesis.

Для каждого из следующего, проверьте указанную гипотезу.

a) $n = 16$, $\bar{x} = 1.55$, $S^2 = 12$, $H_0 : \mu = 1.5$, $H_1 : \mu > 1.5$, $\alpha = 0.01$.

b) $n = 9$, $\bar{x} = 10.1$, $S^2 = 0.81$, $H_0 : \mu = 12$, $H_1 : \mu \neq 12$, $\alpha = 0.05$.

c) $n = 49$, $\bar{x} = 17$, $S^2 = 1$, $H_0 : \mu \geq 18$, $H_1 : \mu < 18$, $\alpha = 0.05$.

$$H_0 : \mu = 1.5, \quad H_1 : \mu > 1.5$$

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \;\sim\; T(n-1)$$

Критическая область

$$T = \frac{\bar{x} - \mu}{S/\sqrt{n}} > T_{1-\alpha}^{(n-1)}$$

```
x_<-1.55
S2<-12
n<-16
alpha<-0.01
mu<-1.5
t_<-(x_-mu)/sqrt(S2)*n;t_
```

[1] 0.2309401

Критическое значение

$$T_{1-\alpha}^{(n-1)}$$

```
qt(1-alpha,n-1)
```

[1] 2.60248

$$t_* = 0.2309401 < 2.60248 = T_{1-\alpha}^{(n-1)} = 2.60248,$$

следовательно, нет оснований отвергнуть нулевую гипотезу.

Или через оверительный уровень вероятности

$$\alpha_* = P\{T > t_*\}$$

```
p.value<-1-pt(t_,n-1)
```

$p.value = 0.4102409 > 0.05$, также нет оснований отвергнуть нулевую гипотезу. В случае $H_1 : \mu \neq 12$ критическая область имеет вид

$$\left|\frac{\bar{x} - \mu}{S/\sqrt{n}}\right| > T_{1-\alpha/2}^{(n-1)},$$

критическое значение равно

```
alpha<-0.05
Tkr<-qt(1-alpha/2,n-1); Tkr
```

[1] 2.13145

```
x_<-10.1
S2<-0.81
n<-9
mu<-12

t_<-(x_-mu)/sqrt(S2)*sqrt(n);t_
```

[1] -6.333333

По абсолютной величине значение статистики превышает критическое значение, следовательно, нулевая гипотеза отвергается.

Значимость считается как

$$\alpha_* = P\{|T| > |t_*|\}$$

```
t_<-(x_-mu)/sqrt(S2)*n;t_
```

[1] -19

```
p.value<-2* ( 1-pt(abs(t_),n-1));p.value
```

```
## [1] 6.094161e-08
```

В третьей задаче критическая область

$$\frac{\bar{x}-\mu}{S/\sqrt{n}} < T_{\alpha}^{(n-1)}, \iff \bar{x} < \mu + T_{\alpha}^{(n-1)} S/\sqrt{n} \iff$$

$$\iff \bar{x} < \mu - T_{1-\alpha}^{(n-1)} S/\sqrt{n}$$

```
alpha<-0.05
Tkr<-qt(1-alpha,n-1); Tkr
```

```
## [1] 1.859548
```

```
x_<-17
S2<-1
n<-49
mu<-18
```

Выборочное среднее $\bar{x} = 17$ попадает в критическую область, следовательно, гипотеза отвергается в пользу альтернативной.

```
mu-Tkr*sqrt(S2)/sqrt(n)
```

```
## [1] 17.73435
```

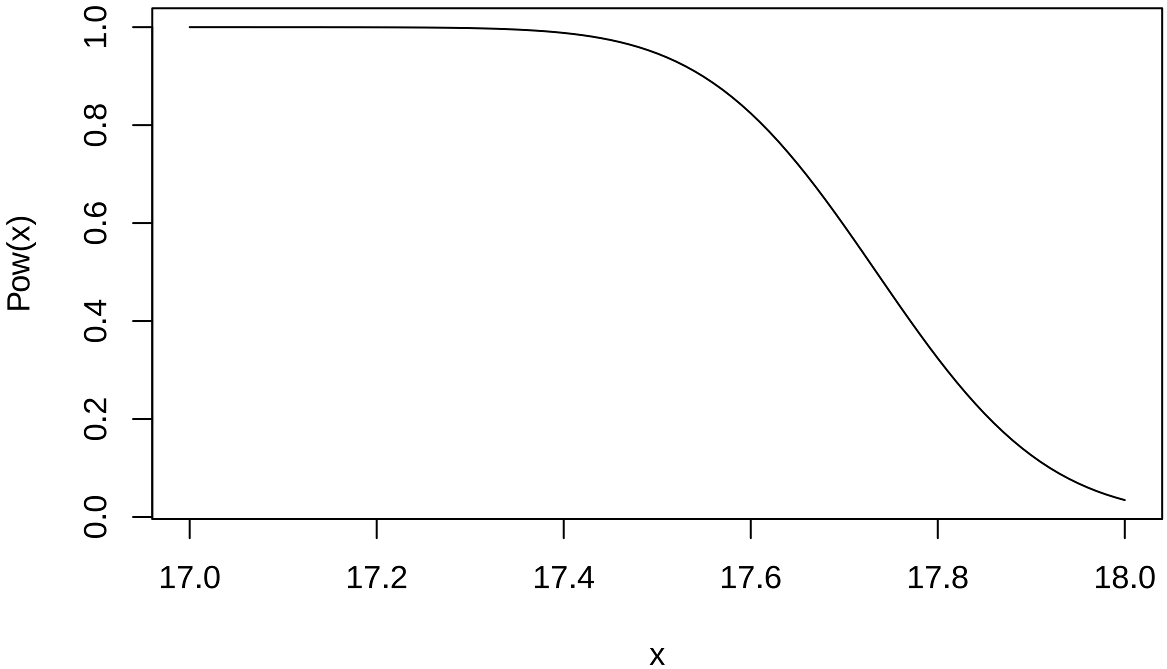
Мощность критерия определяется вероятностью попасть в критическую область при справедливости альтернативной гипотезы. При $\mu < 17$ считаем вероятность

$$P\{\bar{x} < \mu - T_{1-\alpha}^{(n-1)} S/\sqrt{n}\}$$

```
Pow<-function(x){
  pt((mu-Tkr*sqrt(S2)/sqrt(n)-x)/(sqrt(S2)/sqrt(n)),n-1)
}
Pow(17)
```

```
## [1] 0.9999975
```

```
curve(Pow,from=17,to=18)
```


Pow(x)
1.0
0.8
0.6
0.4
0.2
0.0
17.0
17.2
17.4
17.6
17.8
18.0
x