Задание 1

Зубарева Наталия БПИ195 Вариант 8

Выборка (n = 17 значений): [20 19 4 16 31 13 3 15 9 14 19 8 15 21 11 23 20]

А) Рассчитаем 90% доверительный интервал для средней продолжительности грудного вскармливания. Нам неизвестны матожидание x и дисперсия σ^2 распределения (предполагаем, что оно нормальное), вычислим их оценки:

$\bar{x} = (\Sigma x_i) / n = 15.35294118 \sim 15.3529$ - выборочное среднее

$\bar{\sigma}^2 = (\Sigma(x_i - \bar{x})^2) / n = 47.52249135 \sim 47.5225$ - смещённая выборочная дисперсия

$\bar{\sigma} = \sqrt{\bar{\sigma}^2} = 6.893655877 \sim 6.8937$ - выборочное среднеквадратичное отклонение

$(\bar{x} - x) \cdot \sqrt{n-1}/\bar{\sigma} \sim t(n-1)$ - центральная статистика, имеет распределение Стьюдента с 16 степенями свободы - по теореме Фишера

$P(t(n-1)(\alpha/2) < (\bar{x} - x) \cdot \sqrt{n-1}/\bar{\sigma} < t(n-1)(1-\alpha/2)) =$
$= P(t(n-1)(\alpha/2) \cdot \bar{\sigma}/\sqrt{n-1} - \bar{x} < -x < t(n-1)(1-\alpha/2) \cdot \bar{\sigma}/\sqrt{n-1} - \bar{x}) =$
$= P(\bar{x} - t(n-1)(1-\alpha/2) \cdot \bar{\sigma}/\sqrt{n-1} < x < \bar{x} - t(n-1)(\alpha/2) \cdot \bar{\sigma}/\sqrt{n-1}) =$
$= P(\bar{x} - t(n-1)(1-\alpha/2) \cdot \bar{\sigma}/\sqrt{n-1} < x < \bar{x} + t(n-1)(1-\alpha/2) \cdot \bar{\sigma}/\sqrt{n-1}) =$
$= P(15.3529 - t(16)(1-\alpha/2) \cdot 6.8937/\sqrt{16} < x < 15.3529 + t(16)(1-\alpha/2) \cdot 6.8937/\sqrt{16}) = 1-\alpha$

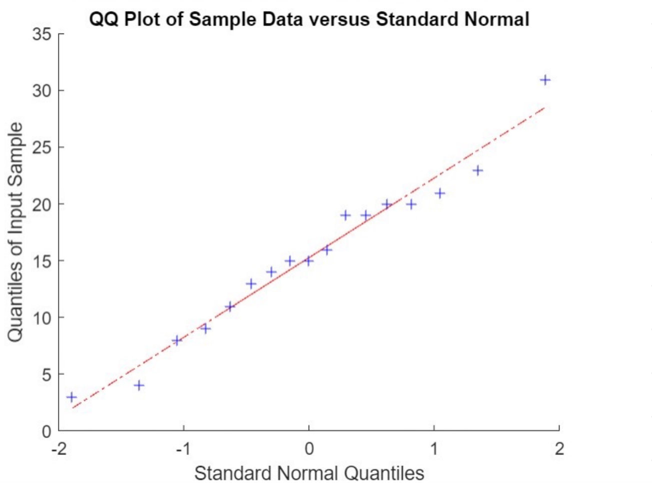
$1-\alpha = 0.9$ $\alpha = 0.1$ $1 - \alpha/2 = 0.95$ - уровень надёжности и соотв. уровень квантили

Найдём квантиль: $t(16)(0.95) = 1.746$ - по таблице

$(15.3529 - 1.746 \cdot 6.8937/4;\ 15.3529 + 1.746 \cdot 6.8937/4)$ - интервал

Доверительный интервал уровня 0.9 для среднего: (12.34379995; 18.36200005)

Б) Построим график квантиль-квантиль, чтобы сравнить с нормальным распределением $N(\bar{x}, \bar{\sigma}^2) = N(15.3529, 47.5225)$



Как можно заметить, значения примерно лежат на прямой линии.
Это значит, что наблюдаемая выборка хорошо описывается нормальным законом.

В) Сгенерируем 1000 перевыборок методом бутстрапа и посчитаем 90% доверительный интервал для средней продолжительности вскармливания:

```
Columns 921 through 940

 17.5294   15.1765   19.0588   16.0000   13.1176   16.1765   19.0000   15.0588

Columns 941 through 960

 18.0588   14.6471   14.7059   12.7059   14.9412   14.2353   13.8235   17.4706

Columns 961 through 980

 16.4706   17.6471   17.4118   15.5882   16.4706   15.5294   14.6471   16.9412

Columns 981 through 1,000

 11.9412   19.8824   17.5882   17.1176   15.4706   16.7059   18.5882   14.1176
```

```
function newstat = fun(stat)
n = 17;
ind = randi(numel(stat),1,n); %// n random
newstat = stat(ind);
end
```

```
    9    14    19     8    15    21    11    23    20

>> nboot = 10;
>> bootstat = bootstrp(nboot, @(stat) fun(stat), stat)

bootstat =

     4    11    16    23    20     4     4     4     4     4    15     4    11    19    11    23    13
    14    16     4     3    19    11     4    14    14    19     4     3    14     4     4    11    19
    15     8    13     8    19    15    21    21    31    14     4    15    31     8     8    21     9
    31    19    15    20     8    31    15     8     9    23     8    23    31     8    23    31    20
    20    15    20    31     4    14    20    15    14    15    20    15     9     4     4    19    14
     3    14    16    16    23    15     4     3    16    16     3     3    15    15     4     3    23
     4     9    31     4    31    20    21    20     4    19     3    31     9    31    11    19    19
```

У нас получилось $n = 1000$ значений выборочных средних

Их выборочное среднее $\bar{x} = 15.4173$
Выборочная дисперсия $\tilde{\sigma}^2 = 5.6369$
Выборочное среднеквадратичное отклонение $\tilde{\sigma} = 2.3742$

Настоящие матожидание и дисперсию мы не знаем.
Как и в пункте А, будем считать интервал с помощью центральной статистики:

$(\bar{x} - x)\sqrt{n-1}/\tilde{\sigma} \sim t(n-1)$ - центральная статистика, имеет распределение Стьюдента с 999 степенями свободы - по теореме Фишера

$P(t(n-1)(\alpha/2) < (\bar{x} - x)\sqrt{n-1}/\tilde{\sigma} < t(n-1)(1-\alpha/2)) = \ldots$
$= P(15.4173 - t(999)(1-\alpha/2) * 2.3742/\sqrt{999} < x < 15.4173 + t(999)(1-\alpha/2) * 2.3742/\sqrt{999}) = 1-\alpha$

$1-\alpha = 0.9$ $\alpha = 0.1$ $1 - \alpha/2 = 0.95$ - уровень надёжности и соотв. уровень квантили

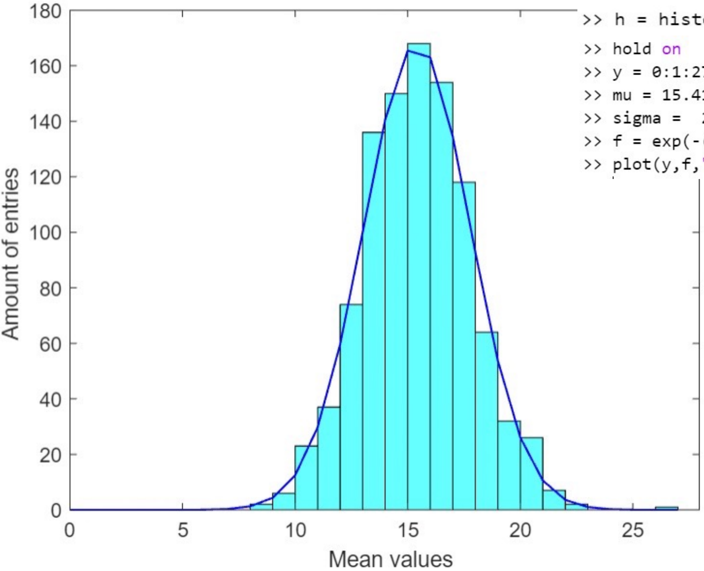
Найдём квантиль: $t(999)(0.95) = 1.6464$ - из питонового скрипта

Доверительный интервал уровня 0.9 для среднего по выборке из бутстрапа:
(15.29363; 15.54097)

В то время как выборочное матожидание почти не изменилось по сравнению с полученным ранее, дисперсия существенно уменьшилась, а доверительный интервал сузился - это произошло потому, что мы сделали новую выборку из средних значений, разброс которых намного меньше, также количество наблюдений увеличилось.

Г) Для каждой перевыборки полученной с помощью бутстрапа на предыдущем этапе рассчитаем выборочное среднее. Мы получили 1000 значений

Построим на их основе гистограмму:




```
>> h = histogram(means)
>> hold on
>> y = 0:1:27;
>> mu = 15.4173;
>> sigma =  2.3742 ;
>> f = exp(-(y-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi))*1000;
>> plot(y,f,"LineWidth",1, "Color",'blue')
```

Как можно видеть, гистограмма очень хорошо соотносится с графиком соответствующего гауссовского распределения.
Можем в очередной раз предположить, что продолжительность периода грудного вскармливания подчиняется закону нормального распределения.

(Послесловие и комментарий: математические вычисления сделаны в матлабе)