Отчет по домашнему заданию

Фахртдинов Т. А.

8 апреля 2020 г.

Первая задача. Описательная статистика и проверка гипотезы согласия.

3 вариант.

1) Моделируем величину выигрыша в 50-ти кратном повторении эксперимента, с начальными данными:

$a = 10$, $b = 10$, $d = 100$, $p = 0.5$, $N = 50$, $NN = 30$.

```
n <- 50
nn <- 30
d <- 100
for (i in 1 : n) {
  x <- sample(c(10,-10), n, replace=T, prob=c(0.5, 0.5))
  res <- c(res, sum(x))
}
res <- as.numeric(res)
res

##  [1]  100   40   60   40   20   40  100  -40   80  -20   40  -40  100 -100  -60
## [16]  -80  -40   60   20   20   20  -80    0  -20  -40  -20    0  -80   60  -40
## [31]  -60   20   40  -40 -100  -20   80  -60  120 -100 -140  -80 -120  -20    0
## [46]  -40   80 -100  -20  -60
```

Берем выборку:

```
res_sample <- res[1:nn]
```

2) С помощью встроенных функций считаем выборочное среднее (mean) и выборочную дисперсию (var).

Среднее = 0

Выборочное среднее =

```
## [1] 4.666667
```

Дисперсия = 5000

Выборочная дисперсия =

```
## [1] 3329.195
```

3) Посчитать вероятность того, что проигрыш больше d. Посчитать вероятность того, что выигрыш больше d. Посчитать вероятность того, что выигрыш в промежутке от (-d/2; d/2).

Считаем нужные нам вероятности при помощи встроенных функций:

```
prop_loss_gtd <- pnorm(-d, mean = mean_res, sd = sd_res_sample)
prop_win_gtd <- 1 - pnorm(d, mean = mean_res, sd = sd_res_sample)
prop_prof_between <- pnorm(d/2, mean = mean_res, sd = sd_res_sample) -
pnorm(-d/2, mean = mean_res, sd = sd_res_sample)
```

Получаем вероятность того, что проигрыш больше d =

```
## [1] 0.05858696
```

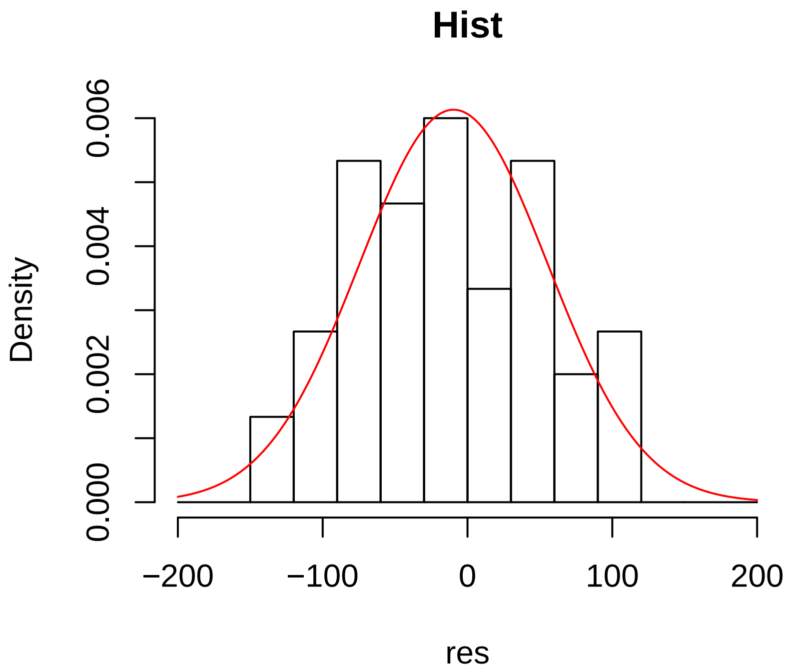
Получаем вероятность того, что выигрыш больше d =

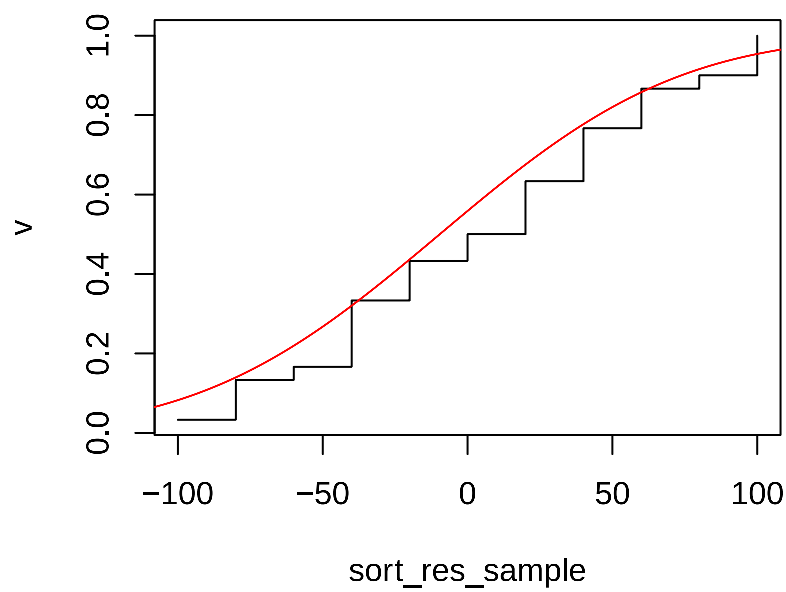
```
## [1] 0.02874892
```

Получаем вероятность того, что выигрыш в промежутке от (-d/2; d/2) =

```
## [1] 0.6072784
```

4) Строим гистограмму и функцию выборочного распределения:





5) С помощью статистики Пирсона проверяем гипотезу согласия эмпирического распределения с нормальным. Математическое ожидание моей случайной величины, вычисленное теоретически = 0, дисперсия = 5000.

H_0: Эмпирическое распределение совпадает с распределением $N(0, 5000)$

Возьмем r = 8. Для пункта a) (параметры распределения вычисленны теоретически) посчитаем вектор вероятностей p.

```
p = numeric()
p <-  c(p, pnorm(-90, mean = 0, sd = sqrt(5000)) -
    pnorm(-9000, mean = 0, sd = sqrt(5000)))
p <- c(p, pnorm(-55, mean = 0, sd = sqrt(5000)) -
           pnorm(-90, mean = 0, sd = sqrt(5000)))
p <- c(p, pnorm(-30, mean = 0, sd = sqrt(5000)) -
           pnorm(-55, mean = 0, sd = sqrt(5000)))
p <- c(p, pnorm(0, mean = 0, sd = sqrt(5000)) -
           pnorm(-30, mean = 0, sd = sqrt(5000)))
p <- c(p, pnorm(30, mean = 0, sd = sqrt(5000)) -
           pnorm(0, mean = 0, sd = sqrt(5000)))
p <- c(p, pnorm(55, mean = 0, sd = sqrt(5000)) -
           pnorm(30, mean = 0, sd = sqrt(5000)))
p <- c(p, pnorm(90, mean = 0, sd = sqrt(5000)) -
           pnorm(55, mean = 0, sd = sqrt(5000)))
p <- c(p, pnorm(9000, mean = 0, sd = sqrt(5000)) -
    pnorm(90, mean = 0, sd = sqrt(5000)))
```

Убедимся, что $n * p_i > 5$, а сумма $\sum_i p_i = 1$:

```
n*p

## [1] 5.077295 5.839621 5.867415 8.215669 8.215669 5.867415 5.839621 5.077295

sum(p)

## [1] 1

br <- c(-9000, -90, -55, -30, 0, 30, 55, 90, 9000)
resHIS <- hist(res, breaks = br, plot = FALSE)
chiSQ <- (resHIS$counts - n * p)^2
for (i in 1:length(p)) {
  chiSQ[i] = chiSQ[i] / (n * p[i])
}
chiSQsumm <- sum(chiSQ)
pValPir <- 1 - pchisq(chiSQsumm, 7)
```

В переменной chiSQsumm записанно значение статистики, а переменной pValPir записанно p - value, величины соответсвенно равны:

```
chiSQsumm

## [1] 2.880276

pValPir

## [1] 0.8958524
```

Убедимся, что результат совпадает со значение встроенной функции:

```
chisq.test(resHIS$counts, p = p)

##
##  Chi-squared test for given probabilities
##
## data:  resHIS$counts
## X-squared = 2.8803, df = 7, p-value = 0.8959
```

По полученным данным нет причин отклонить гипотезу.

В пункте б) (параметры распределения оцененны по выборке), воспользуемся функцией pearson.test, пакета nortest, чтобы посчитать значение статистики:

```
pearson.test(res, n.classes = 11)

##
##  Pearson chi-square normality test
##
## data:  res
## P = 9.4, p-value = 0.3097
```

По полученным данным нет причин отклонить гипотезу.

6) Считаем статистики: Медиана =

```
median_res_sample <- median(res_sample)
median_res_sample

## [1] 10
```

Стандартное отклонение =

```
sd_res_sample = sd(res_sample)
sd_res_sample

## [1] 57.69918
```

Коэффицент вариации =

```
sd_res_sample / mean_res_sample

## [1] 12.36411
```

Рассеяние =

```
var_res_sample * mean_res_sample

## [1] 15536.25
```

Асимметрия =

```
skew_res_sample <- skewness(res_sample)
skew_res_sample

## [1] -0.0004933734
```

Эксцесс =

```
kurt_res_sample <- kurtosis(res_sample)
kurt_res_sample

## [1] 2.018807
```

95% доверительный интервал =

```
dovInt <- MeanCI(res_sample)
#Левая граница
dovInt[2]

##    lwr.ci
## -16.87856

#Правая границы
dovInt[3]

##   upr.ci
## 26.21189
```