

Ярылгасимов Султан
Домашнее задание №6

2. Сгенерировать необходимый набор данных, отвечающих условиям применимости статистических тестов, и продемонстрировать их применение для проверки различных гипотез и различных доверительных уровней (0.9, 0.95, 0.99) с использованием критериев:

2.а) **Стьюдента**, включая односторонние варианты, когда проверяемая нулевая гипотеза заключается в том, что одно из сравниваемых средних значений больше (или меньше) другого: использовать `t.test()` в сочетании с аргументом «alternative» (может принимать одно из трех значений: «two.sided» (по умолчанию), «greater» или «less»). Реализовать оценку мощности критериев при заданном объеме выборки или определения объема выборки для достижения заданной мощности;

Код:

```
data("morley")
?morley
head(morley)
morley
attach(morley)
# будем работать по expt #1, 2 и #3, 4
morley[Expt==2,]$Expt = 1 # перегруппируем
morley[Expt==3,]$Expt = 2
morley[Expt==4,]$Expt = 2
morley = morley[Expt != 5,]
attach(morley)
# чтобы применить тест Стьюдента
# 1. не должны быть коррелирующими
first = morley[Expt == 1,]$Speed
second = morley[Expt == 2,]$Speed
cor.test(first, second)
# p-value > 0.05 -> не коррелируют
# 2. соответствуют нормальному распределению
shapiro.test(first)
shapiro.test(second)
# p-value > 0.05 -> нормальные
var.test(first, second)
# p-value > 0.05 -> не совпадают дисперсии
t.test(Speed ~ Expt, var.equal=F)
# p-value < 0.05 -> средние не равны
t.test(first, second, alternative = 'less')
# p-value > 0.05 -> разница в средних меньше 0
power.t.test(n = length(first), delta = mean(first) - mean(second),
             sd = mean(sd(first), sd(second)), sig.level = 0.05)
# power == 0.7, что недостаточно много
# при каком объеме можно увеличить до 0.9?
power.t.test(delta = mean(first) - mean(second),
             sd = mean(sd(first), sd(second)), sig.level = 0.05, power = 0.9)
# n = 69
```

Комментарий: в качестве данных взяты данные с эксперимента по измерению скорости света. Стоит учесть, что фактически данные менее разбросаны, т. к. в данном датасете вычтено 299000. В комментариях указаны результаты тестов и выполненные шаги, включая проверку применимости теста.

2.b) Уилкоксона-Манна-Уитни (ранговые);

Код:

```
data("randu")
?randu
attach(randu)
# будем работать по x и y
shapiro.test(x)
shapiro.test(y)
# p-value < 0.05 -> распределение не нормальное
cor.test(x, y)
# p-value > 0.05 -> корреляции нет
wilcox.test(x, y)
# p-value > 0.05 -> средние совпадают
```

Комментарий: в качестве данных взяты случайные числа, сгенерированные RANDU.

2.c) Фишера, Левене, Бартлетта, Флигнера-Килина (проверка гипотез об однородности дисперсий).

Код:

```
library(car)
var.test(first, second) # Фишер
# p-value 0.15
leveneTest(first, second) # Левене
# p-value 0.15
bartlett.test(morley$Speed ~ morley$Expt, data=morley) # Бартлетт
# p-value 0.15
fligner.test(morley$Speed ~ morley$Expt) # Флигнер
# p-value 0.10
```

Комментарий: «допроверяем утверждение» об однородности дисперсий из первого пункта и убеждаемся еще раз в правильности.

3. Провести статистический анализ из п. 2 для собственных данных при выполнении условий применимости.

Код:

```
gds # from last hw
gds=sort(gds)
hist(gds, freq=F)
lines(density(gds))

shapiro.test(gds)$p.value
# p-value < 0.05 -> не нормальное распределение
mean(gds)
# 0
wilcox.test(gds)
# p-value > 0.05 -> среднее близко к 0
```

Комментарий: на прошлом занятии, мы пришли к выводу, что данные не являются нормально распределенными, и на этом занятии, используя этот факт, можем применить тест Уилкоксона-Манна-Уитни, и сделать вывод, что среднее действительно равно 0.