

Ярылгасимов Султан
Домашняя работа №7

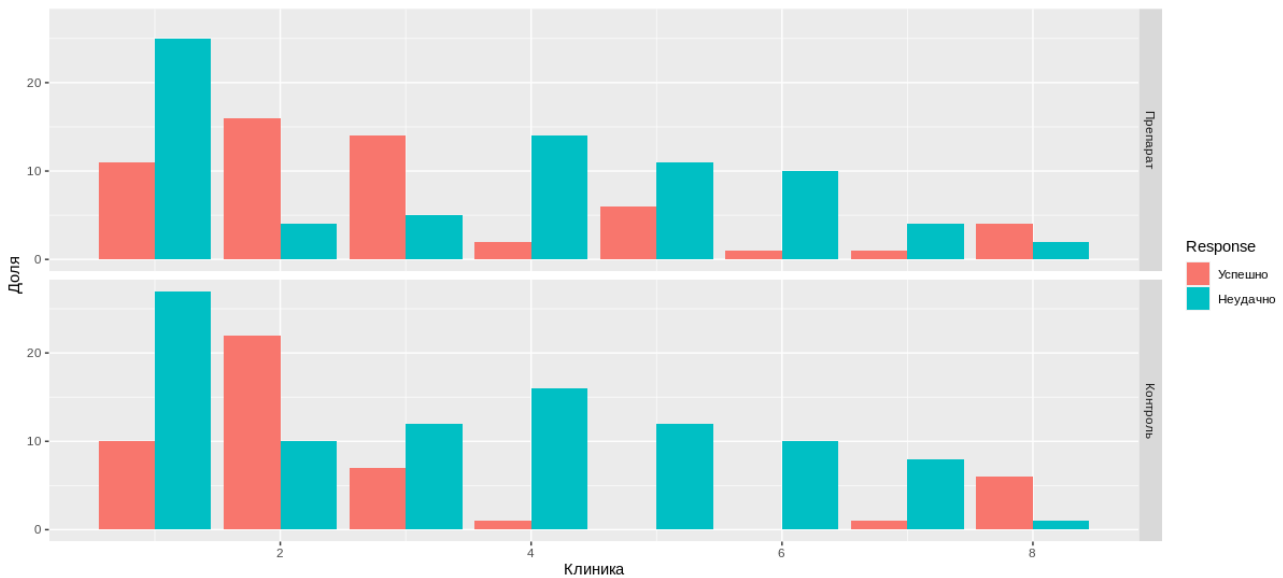
Задание 1. Выполнить задания из комментариев к программному коду занятия, включая изучение теоретических свойств рассмотренных объектов.

1.1. Преобразовать график к "парным" столбцам

Код:

```
p <- ggplot(data = drug.df, aes(x = Center, y = value, fill = Response))  
+ xlab("Клиника") + ylab("Доля")  
p + geom_bar(stat = "identity", position = "dodge") + facet_grid(Group~.)
```

Вывод:



Комментарий: требуемый результат достигается изменением одного параметра position с «fill» на «dodge».

Задание 2. Исследовать корреляционные взаимосвязи в собственных данных с помощью коэффициентов корреляции Пирсона, Спирмена и Кендалла, проверив условия применимости.

Код:

```
data("women")  
?women  
attach(women)  
# проверим нормальность  
shapiro.test(height)  
shapiro.test(weight)  
# данные нормальны  
# вычисляем коэф. корреляции Пирсона  
cor.test(height, weight)  
# cor = 1, значит есть прямая зависимость между величинами  
cor.test(height, weight, method = "spearman", alternative="less")  
# тест Спирмана с проверкой обратной зависимости выдает результат p-value =  
1  
# т.е. обратной зависимости точно нет  
cor.test(height, weight, method = "kendall", alternative = "greater")  
# тест Кендалла с гипотезой о том, что коэф. кор > 0  
# p-value < 0.05 -> коэф. корреляции больше 1
```

Комментарий: выводы и ход работы указаны в комментариях к коду. Стоит добавить, что данные тесты оказались неприменимы к моим данным, т. к. ни один параметр не оказался

нормально распределенным. Дополнительно об использованных данных: данные представляют информацию о росте и весе женщин. Заранее, можно предположить, что данные цифра имеют прямую зависимость и тесты проведенные выше это доказывают.

Задание 3. На собственных данных проверить использование методов хи-квадрат, точного теста Фишера, теста МакНемара, Кохрана-Мантеля-Хензеля.

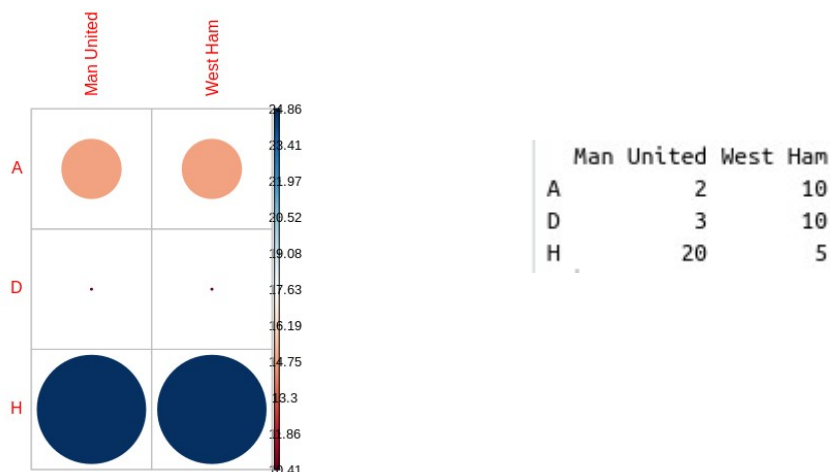
Код:

```
attach(match_results)
team_names <- c('Man United', 'West Ham')
home_mu = match_results[HomeTeam == 'Man United' & AwayTeam == 'West Ham',
                        c("HomeTeam", "AwayTeam", "FTR")]
home_whu = match_results[HomeTeam == 'West Ham' & AwayTeam == 'Man United',
                        c('HomeTeam', 'AwayTeam', 'FTR')]
mu_res = as.vector(table(home_mu$FTR))
table(home_mu$FTR)
whu_res = as.vector(table(home_whu$FTR))
table(home_whu$FTR)
mu_res
whu_res
results.data <- matrix(c(mu_res, whu_res), nrow = 3, byrow=F,
                        dimnames = list(c('A', 'D', 'H'), team_names))
results.data

chisq <- chisq.test(results.data)
# p<0.05
contrib <- 100*chisq$residuals^2/chisq$statistic
library(corrplot)
corrplot(contrib, is.cor = F)

fisher.test(results.data)
# p < 0.05
```

Вывод:



Комментарий: В качестве данных взяты результаты игр между двумя клубами за все время существования турнира. В результате тестов, можно прийти к выводу, что между результатами есть зависимость, т. е. одна команда побеждает чаще другой. При проверке на клубах, которые играют одинаково, тест не давал таких результатов.