



Базовая статистика

ТИП ПЕРЕМЕННЫХ ОПРЕДЕЛЯЕТ НАБОР СТАТИСТИЧЕСКИХ МЕТОДОВ АНАЛИЗА

ПЕРЕМЕННЫЕ

КОЛИЧЕСТВЕННЫЕ

ДИСКРЕТНЫЕ
(DISCRETE)

НЕПРЕРЫВНЫЕ
(CONTINUOUS)

ЧИСЛОВОЕ ОБОЗНАЧЕНИЕ ИМЕЕТ
СОДЕРЖАТЕЛЬНЫЙ СМЫСЛ

КАТЕГОРИАЛЬНЫЕ


ПОРЯДКОВЫЕ
(ORDINAL)

НОМИНАЛЬНЫЕ
(NOMINAL)

ЧИСЛОВОЕ ОБОЗНАЧЕНИЕ НЕ ИМЕЕТ
СОДЕРЖАТЕЛЬНЫЙ СМЫСЛ



Пакет rstatix



```
install.packages(c("rstatix", "ggpubr"))  
library(rstatix)  
library(ggpubr)
```

<https://rpkgs.datanovia.com/rstatix/>

Описательная статистика



Описательная статистика

```
data_summary <- data %>%  
  group_by(Group) %>%  
  get_summary_stats()
```

```
type = c("full", "common", "robust", "five_number",  
"mean_sd", "mean_se", "mean_ci", "median_iqr",  
"median_mad", "quantile", "mean", "median", "min", "max")
```


Нормальное распределение

```
data_norm <- data %>%  
  group_by(Group) %>%  
  shapiro_test(  
    Gene1_expression,  
    Gene2_expression  
  )
```

Figure 3: Boxplot of Normally Distributed Data

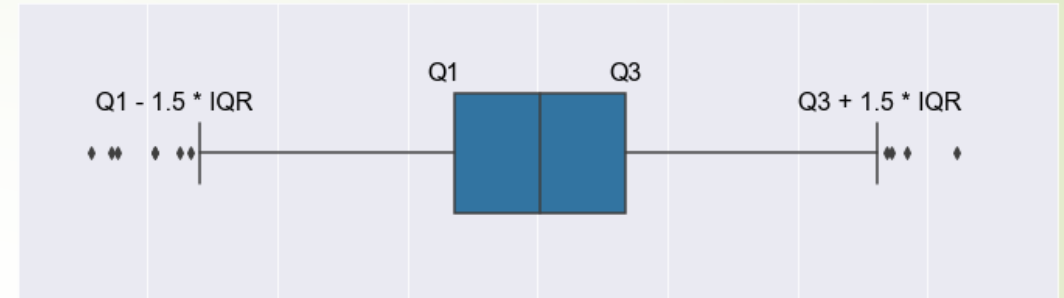


Figure 4: IQR Rule

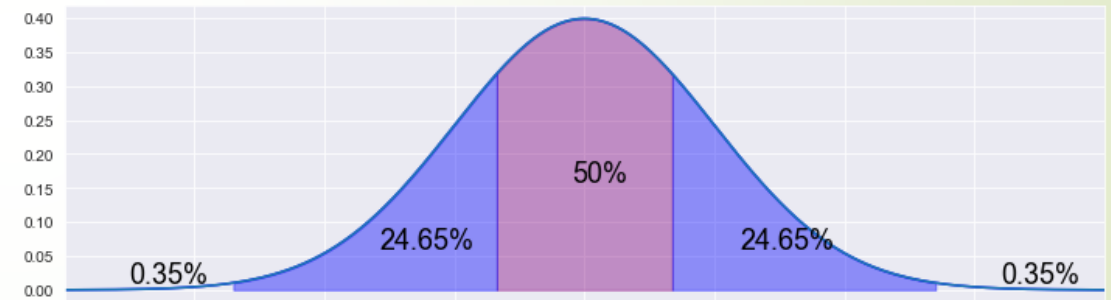
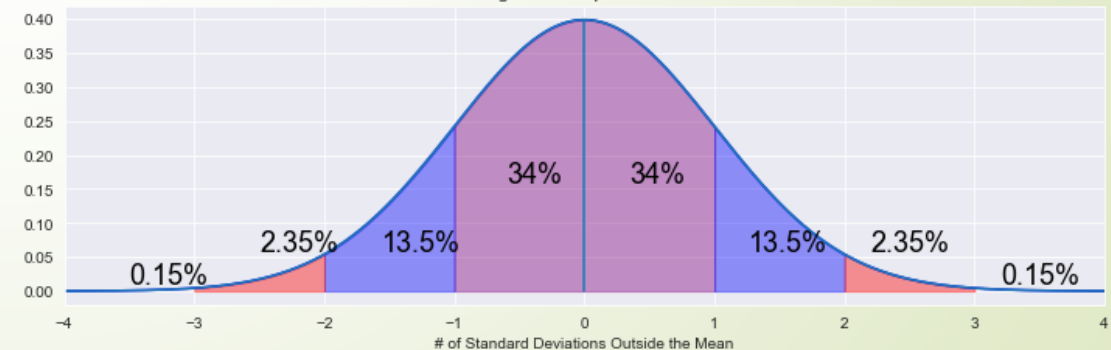


Figure 5: Empirical Rule



Как выбрать статистический тест

Сравнение с теоретическим
средним

НР: Т-тест

ННР: тест Уилкоксона

Сравнение 2 групп

Не парное сравнение

НР: непарный Т-тест

ННР: тест Манна-Уитни

Парное сравнение

НР: парный Т-тест

ННР: тест Уилкоксона

Сравнение 3 и более групп

Не повторные измерения

НР: однофакторный дисперсионный
анализ +post-hoc
ННР: тест Крускала-Уоллиса +post-hoc

Повторные измерения

НР: дисперсионный анализ с
повторными измерениями +post-hoc

ННР: тест Фридмана +post-hoc

Корреляция

НР: корреляция Пирсона

ННР: корреляция Спирмана

Предсказание

На основе 1 переменной

НР: линейная регрессия

ННР: непараметрическая регрессия

На основе нескольких
переменных

Множественная линейная и
нелинейная регрессии

НР – нормальное
распределение
ННР – не
нормальное
распределение

Сравнение с теоретическим средним - HP

```
data_exp <- data %>%  
  filter(Group == "Experiment")
```

```
bxp <- ggboxplot(  
  data_exp$Gene1_expression, width =  
  0.5, add = c("mean", "jitter"),  
  ylab = "Gene", xlab = FALSE  
)  
bxp
```

```
ggqqplot(data_exp, x =  
  "Gene1_expression")
```

```
ttest_theor <- data_exp %>%  
  t_test(Gene1_expression ~ 1, mu = 3)
```

```
bxp + labs(  
  subtitle = get_test_label(ttest_theor,  
    detailed = TRUE)  
)
```


Сравнение 2 групп - НР

```
bxp <- ggboxplot(  
  data, x = "Group", y =  
  "Gene1_expression",  
  ylab = "Gene 1", xlab = "Groups", add =  
  "jitter"  
)  
bxp
```

```
ggqqplot(data, x = "Gene1_expression",  
  facet.by = "Group")
```

```
ttest_2g <- data %>%  
  t_test(Gene1_expression ~ Group) %>%  
  add_significance()
```

```
ttest_2g <- ttest_2g %>%  
  add_xy_position(x = "Group")  
bxp +  
  stat_pvalue_manual(ttest_2g, tip.length  
  = 0) +  
  labs(subtitle = get_test_label(ttest_2g,  
  detailed = TRUE))
```

Сравнение с теоретическим средним - ННР

```
bxp3 <- ggboxplot(  
  data = data_exp, y =  
  "Gene2_expression", width = 0.5, add =  
  c("mean", "jitter"),  
  ylab = "Gene 2", xlab = FALSE  
)  
bxp3
```

```
wilcox_test_theor <- data_exp %>%  
  wilcox_test(Gene2_expression ~ 1,  
  mu = 0.9)
```

```
bxp3 +  
  labs(subtitle =  
  get_test_label(wilcox_test_theor,  
  detailed = TRUE))
```

Сравнение 2 групп - ННР

```
bxp4 <- ggboxplot(  
  data, x = "Group", y = "Gene2_expression",  
  ylab = "Gene 2", xlab = "Groups", add =  
  "jitter"  
)  
bxp4
```

```
wilcox_test_2g <- data %>%  
  wilcox_test(Gene2_expression ~ Group)  
  %>%  
  add_significance()
```

```
wilcox_test_2g <- wilcox_test_2g %>%  
  add_xy_position(x = "Group")  
  
bxp4 +  
  stat_pvalue_manual(wilcox_test_2g,  
    tip.length = 0) +  
  labs(subtitle =  
    get_test_label(wilcox_test_2g, detailed =  
    TRUE))
```

Сравнение 3 и более групп - HP

```
ggboxplot(data, x = "Category", y =  
"Gene2_expression")
```

```
ggqqplot(data, "Gene2_expression",  
facet.by = "Category")
```

```
anova_1 <- data %>%  
anova_test(Gene2_expression ~  
Category)
```

Не достоверно!

```
pwc1 <- data %>%  
tukey_hsd(Gene2_expression ~  
Category)
```

```
pwc1 <- pwc1 %>% add_xy_position(x =  
"Category")
```

```
ggboxplot(data, x = "Category", y =  
"Gene2_expression") +
```

```
  stat_pvalue_manual(pwc1, hide.ns =  
TRUE) +
```

```
  labs(
```

```
    subtitle = get_test_label(anova_1,  
detailed = TRUE),
```

```
    caption = get_pwc_label(pwc1)
```

```
)
```

Сравнение 3 и более групп - ННР

```
ggboxplot(data, x = "Category", y =  
"Gene1_expression")
```

```
ggqqplot(data, "Gene1_expression",  
facet.by = "Category")
```

```
kruskal <- data %>%  
kruskal_test(Gene1_expression ~  
Category)
```

Не достоверно!

```
pwc2 <- data %>%  
  dunn_test(Gene1_expression ~ Category,  
p.adjust.method = "bonferroni")
```

```
pwc2 <- pwc2 %>% add_xy_position(x =  
"Category")
```

```
ggboxplot(data, x = "Category", y =  
"Gene1_expression") +
```

```
  stat_pvalue_manual(pwc2, hide.ns =  
TRUE) +
```

```
  labs(  
    subtitle = get_test_label(kruskal, detailed  
= TRUE),  
    caption = get_pwc_label(pwc2)  
  )
```