Теннисные турниры:

Помогут ли данные победить?

ВШЭстером

01

ВВЕДЕНИЕ

- Исследовательский вопрос
- Гипотеза
- Механизм

Исследовательский вопрос

Как **возраст** игроков и **длительность** матча влияют на его исход?

Гипотеза

В **длительных** матчах чаще побеждает более **молодой** игрок

Механизм

- Молодые люди более выносливы
- Им легче переносить физические нагрузки и активно играть на протяжении всего матча
- Из-за этого в продолжительных матчах более молодые игроки имеют преимущество

02

ДАННЫЕ

- Описание переменных
- Очистка выбросов
- Разведывательный анализ

Описание переменных

winner_name - имя игрока, выигравшего матч

loser_name - имя игрока, проигравшего матч

winner_age - возраст победителя

loser_age - возраст проигравшего

winner_rank_points - количество очков в ранге у победителя

loser_rank_points - количество очков в ранге у проигравшего

minutes - длительность матча

Чистка данных

Распределение продолжительности матчей
Количество матчей
400
350
300
250
200
150
100
50
0
27
42
49
56
63
70
77
84
91
98
105
112
119
126
133
140
147
154
161
168
175
182
189
196
203
210
217
231
Минуты

03

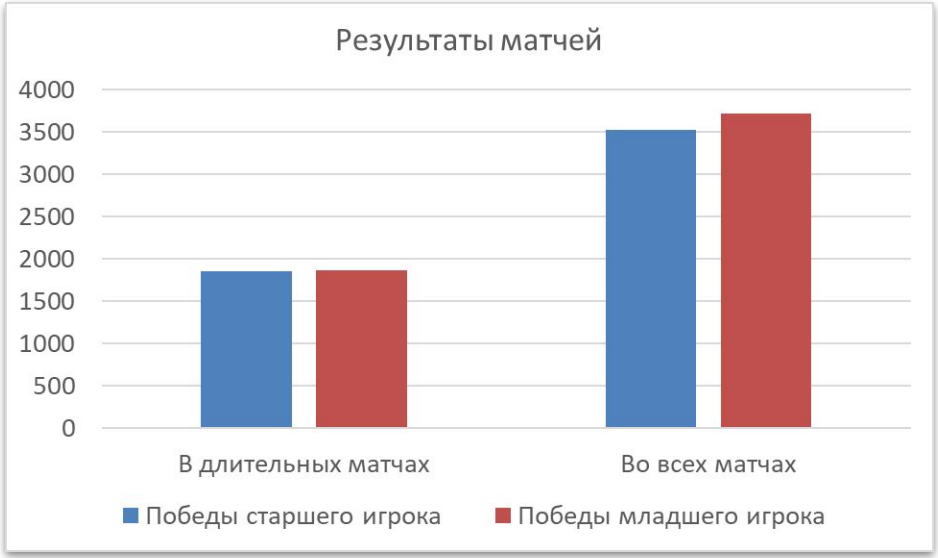
ПРОВЕРКА ГИПОТЕЗЫ

- Математическая модель
- Проверка устойчивости
- Интерпретация

Математическая модель

Используем биномиальный тест

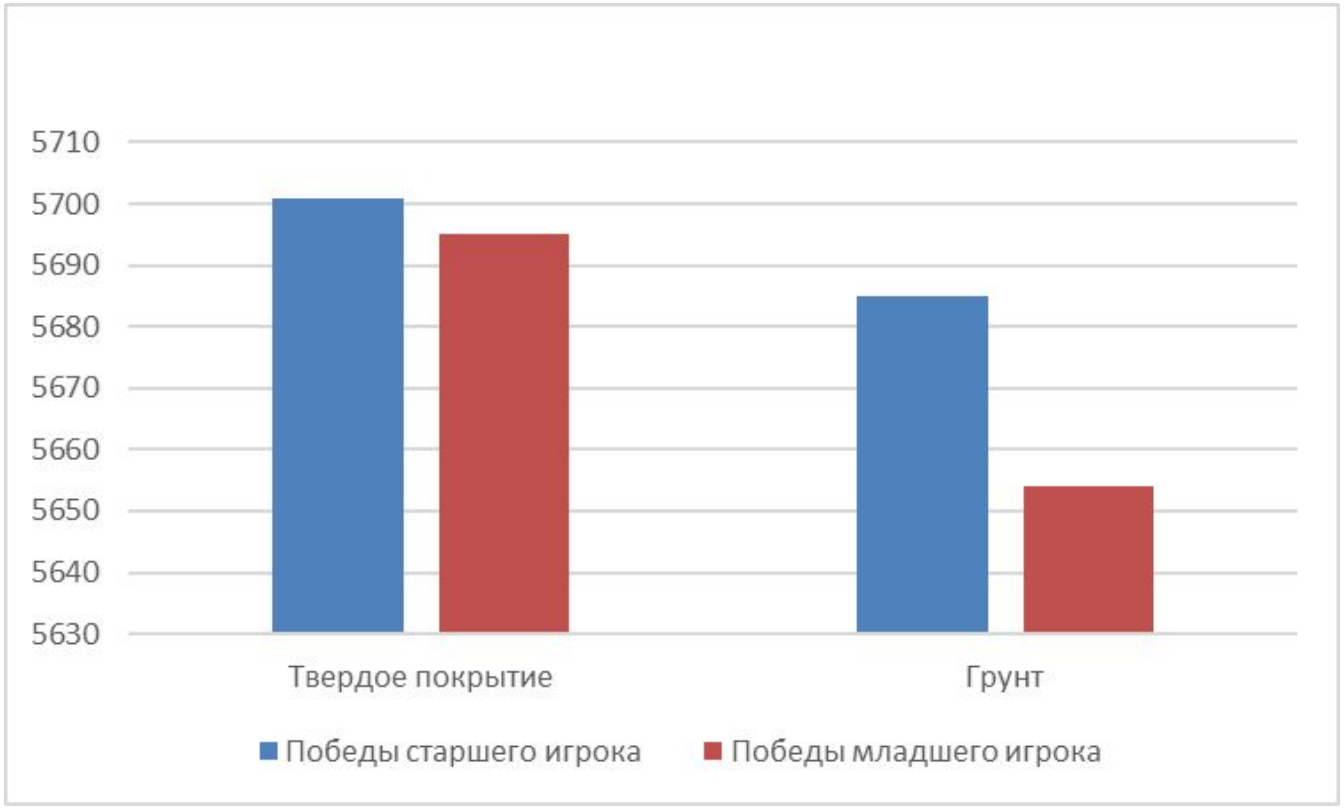
H0 - между долями побед нет значимых различий



P-value = 0.3

H0 **не отвергается**

Устойчивость



04

ВЫВОДЫ

- Перспективы
- Ограничения
- Policy Implication

ПЕРСПЕКТИВЫ

Сбор данных о женщинах

Выделить зависимость среди женщин, а не мужчин

Детализация хода матча

Проверить зависимость результата гейма от предыдущих

ОГРАНИЧЕНИЯ

Короткий период

Нельзя отследить тенденцию развития игрока в течении долгого времени

Много пустых значений в бд

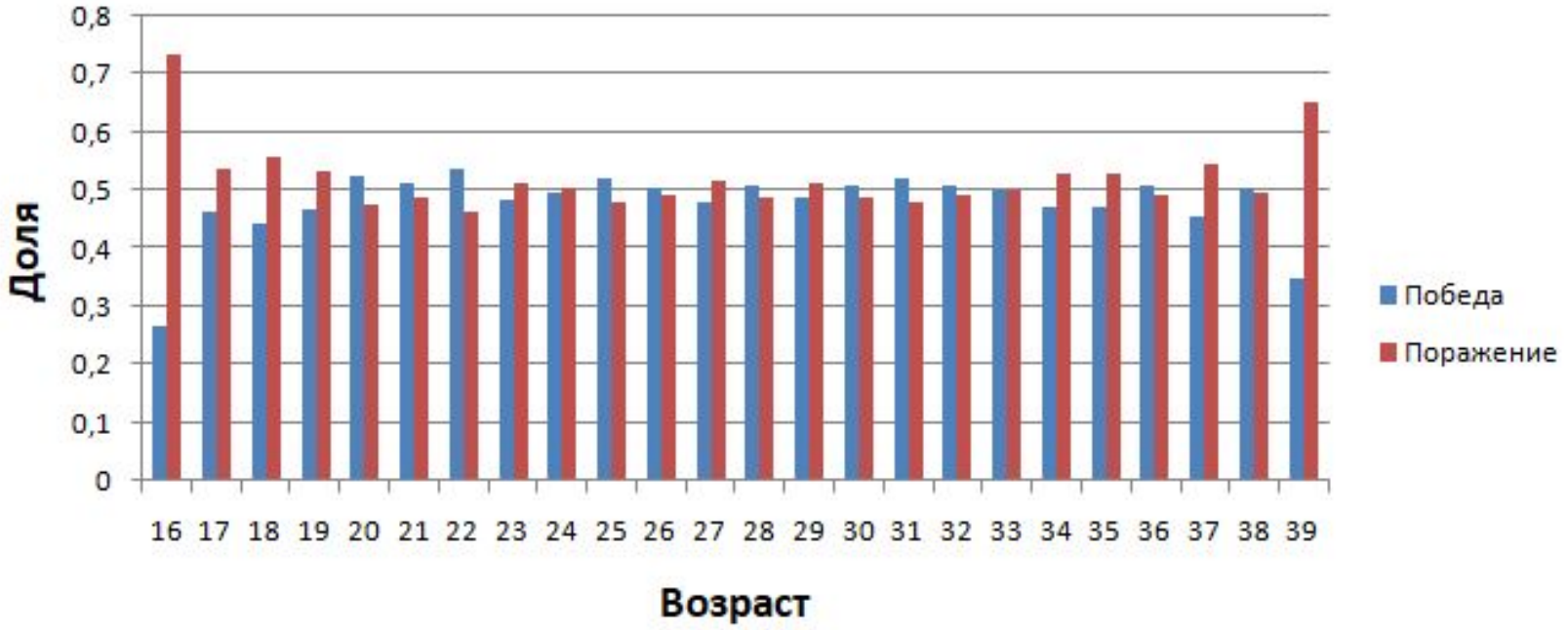
Сложность в построении боксплотов

-1

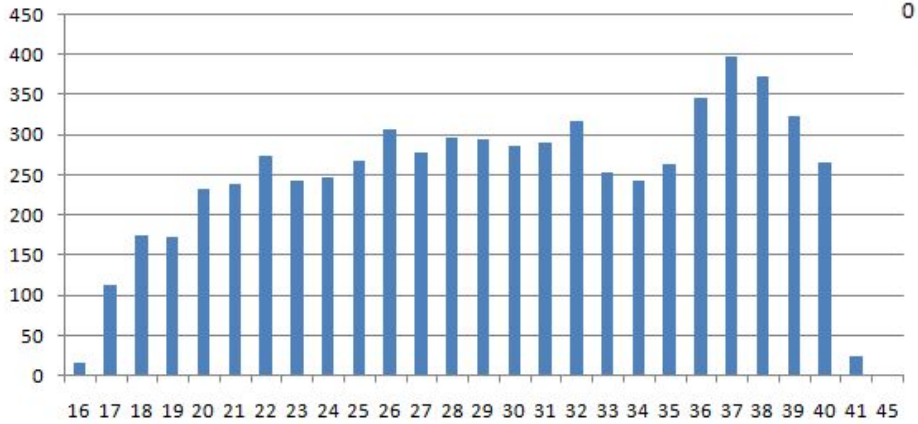
ПОДВАЛ

- Матрица корреляций
- Дополнительные вычисления

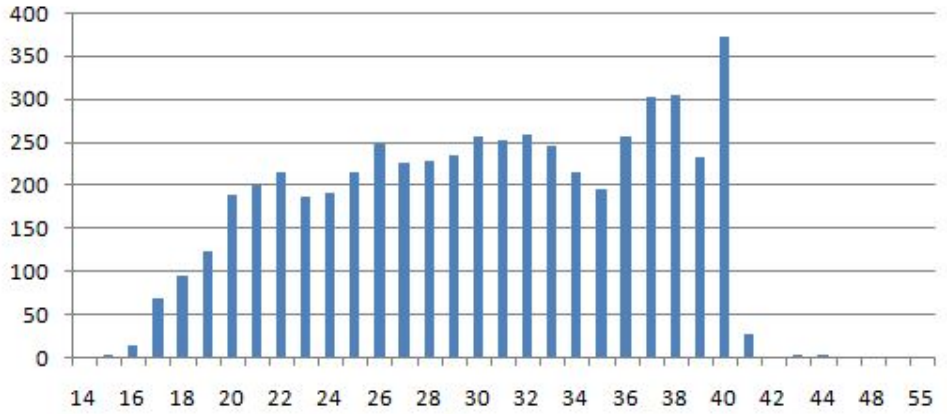
Количество игр, сыгранных на покрытии
Количество сыгранных матчей
12000
10000
8000
6000
4000
2000
0
Carpet
Clay
Grass
Hard
Тип покрытия


0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
Доля
16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
Возраст
Победа
Поражение

Усредненное распределение очков по возрастам победителей

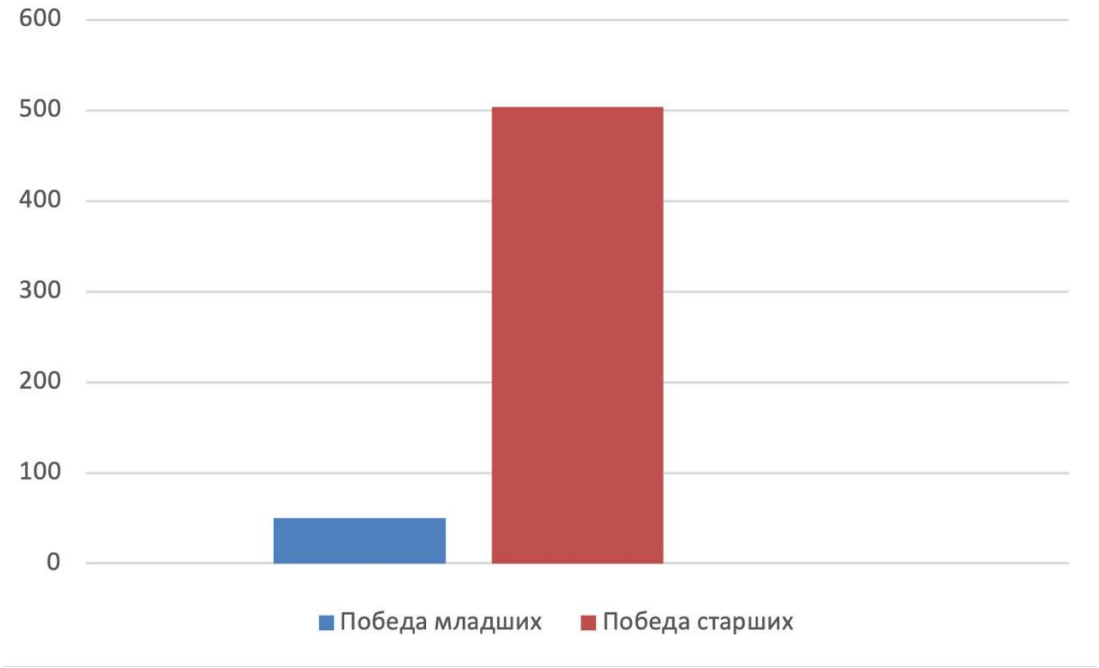


Усредненное распределение очков по возрастам проигравших



Результаты матча
Количество матчей
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Левша
Правша
Не определено
Победа
Поражение

Распределение побед среди младших и старших



Матрица Корреляций

	Количество эйсов у победителя	Количество двойных ошибок у победителя	Количество розыгрышей подачи у победителя	Количество первых подач в игре у победителя	Количество очков, выигранных на первой подаче у победителя	Количество очков, выигранных на второй подаче у победителя	Количество геймов на своей подаче, выигранных победителем	Количество брейк-пойнтов, спасенных победителем	Количество брейк-пойнтов, сыгранных победителем
Количество эйсов у проигравшего	0,308	0,117	0,382	0,361	0,452	0,343	0,388	0,074	0,056
Количество двойных ошибок у проигравшего	0,042	0,177	0,200	0,145	0,126	0,175	0,176	0,125	0,177
Количество розыгрышей подачи у проигравшего	0,244	0,344	0,819	0,754	0,743	0,587	0,807	0,344	0,478
Количество первых подач в игре у проигравшего	0,237	0,285	0,757	0,717	0,707	0,519	0,751	0,314	0,433
Количество очков, выигранных на первой подаче у проигравшего	0,338	0,287	0,788	0,743	0,786	0,587	0,788	0,285	0,360
Количество очков, выигранных на второй подаче у проигравшего	0,253	0,286	0,658	0,592	0,626	0,528	0,649	0,240	0,309
Количество геймов на своей подаче, выигранных проигравшим	0,249	0,334	0,824	0,768	0,759	0,582	0,976	0,353	0,483
Количество брейк-пойнтов, спасенных проигравшим	-0,019	0,115	0,197	0,164	0,120	0,132	0,200	0,109	0,185
Количество брейк-пойнтов, сыгранных проигравшим	-0,153	0,131	0,155	0,119	0,002	0,045	0,150	0,156	0,296

ВШЭстером

Николай Матвеев
Тимлид

Полина Жуйкова
Аналитик

Захарий Димидов
Программист

Денис Игнатов
Программист

Полина Мирошниченко
Аналитик

Андрей Чагин
Аналитик