

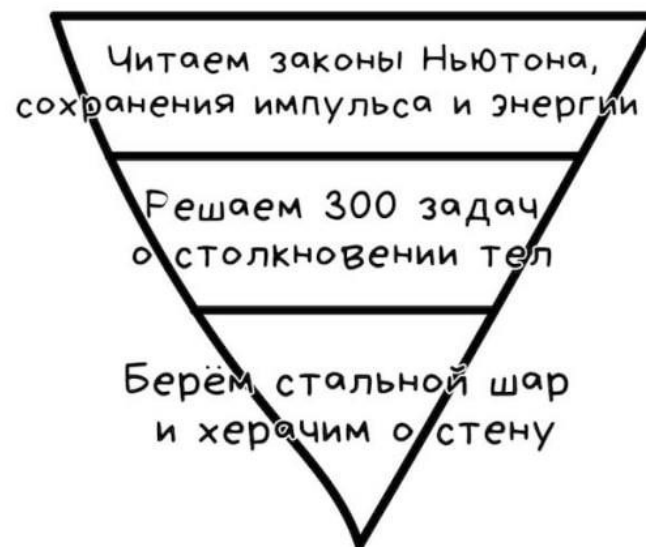
Ремесленный метод (практический)

- Хочу знать физику!



Академический метод (теоретический)

- Хочу знать физику!



Метод сломанной пирамиды



О важном

«Главным источником сведений о биометрии для многих поколений отечественных селекционеров остается классический труд Б.А. Доспехова «Методика полевого опыта», выдержавший несколько переизданий. **Спустя десятилетия после создания «Методики» сама процедура статистического анализа данных совершенно изменилась.** Исчезла необходимость в кропотливых многостадийных расчетах вручную. Появились компьютерные программы, выполняющие сотни видов анализа (например, IBM SPSS, STATISTICA и многие другие – **например, R**).

Несмотря на кажущуюся простоту машинной обработки результатов эксперимента, вопрос применимости тех или иных методов не стал менее существенным; возможно, напротив, оказался острее, потому что существенно расширился выбор из множества на первый взгляд сходных приемов.»

Из статьи «Статистические ошибки и как их избегают»

<https://drive.google.com/drive/folders/1UgKVVKBrBtLwy5bZPYT1VEMkk4WwSifv>

План встречи

1. Демонстрация возможностей R для анализа данных селекционных опытов на примере факторного опыта
2. Разбор простых примеров работы в R

Пример:

- 3 сорта озимой пшеницы
- 3 варианта срока сева
- 4 варианта густоты сева (2,5 млн., 3,5 млн., 4,5 млн., 5,5 млн.)

В 4-х повторностях

Всего 36 вариантов

144 деланки

Факторные опыты

- Обычная ситуация в селекционных опытах – исследуется один фактор - «генотип»
- Некоторые эксперименты изучают влияние >1 фактора одновременно, например, генотип и срок сева (и/ или густота сева)
- Факторные опыты оценивают взаимодействие сорта и факторов влияющих на урожайность сорта
 - Дают больше идей для понимания урожайности сорта
 - Идентификация интенсивных / экстенсивных сортов
 - Обычно проводятся на поздних стадиях селекционных программ
- Самый популярный способ размещения вариантов дизайн эксперимента для факторных опытов: split plot design (он же метод расщепленных делянок)

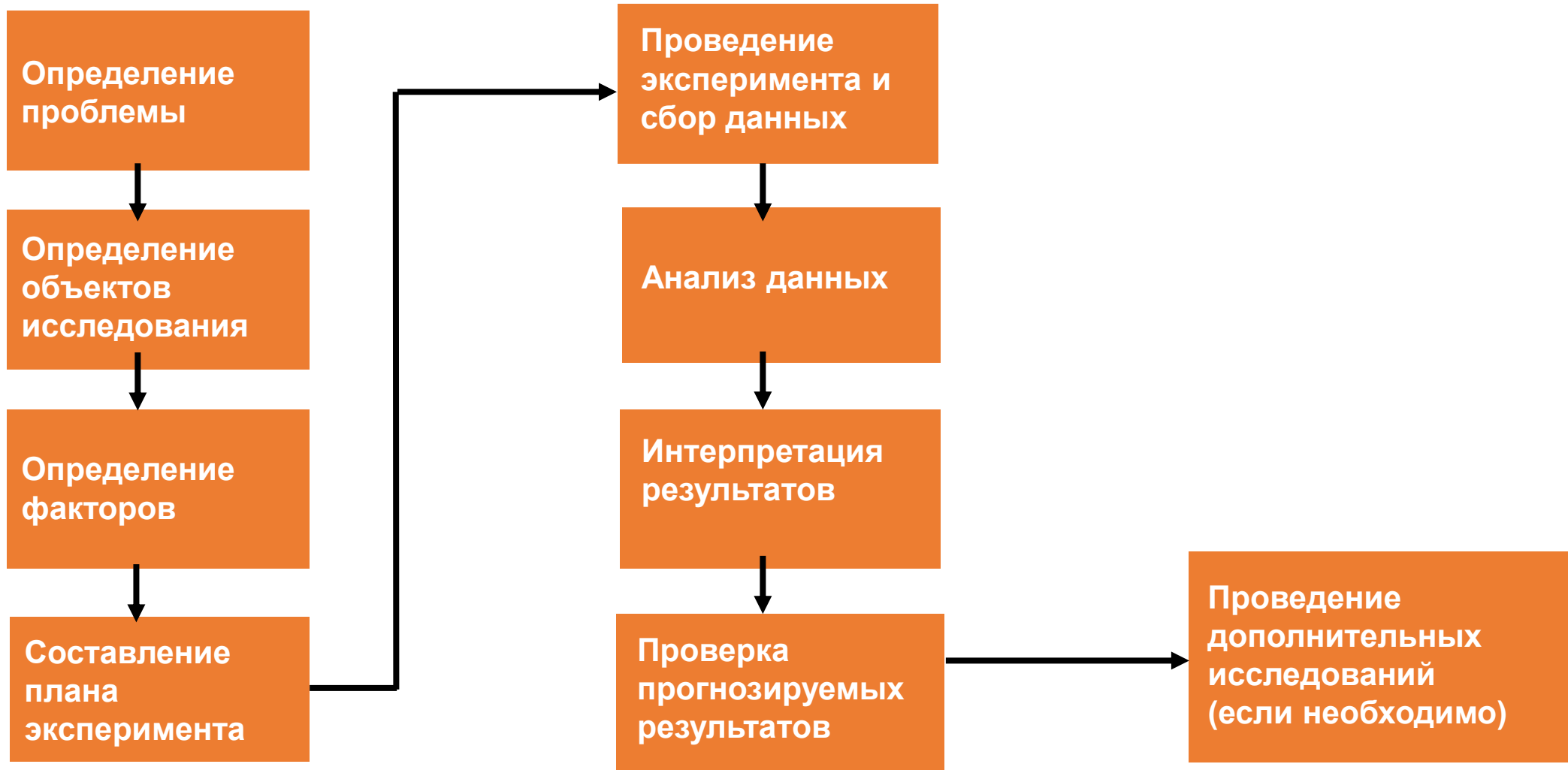
*О планировании и рандомизации факторных опытов – поговорим на других встречах

Как выглядел опыт, если бы выбрали split-plot в качестве дизайна эксперимента

Full Factorial Design (CRD) 36X4

ROWS	2*1*0	2*2*0	2*1*3	2*0*1
	0*1*1	2*2*0	2*2*1	1*1*1
	1*2*2	2*0*2	0*0*3	1*2*3
	0*2*0	1*1*3	0*0*2	1*0*2
	0*1*3	0*0*3	1*0*0	0*0*0
	1*2*2	1*0*2	1*0*3	2*1*2
	1*2*2	2*0*0	0*1*0	1*1*3
	0*1*1	1*1*1	2*1*2	2*2*1
	2*1*3	1*2*0	2*2*2	2*2*3
	0*0*3	1*1*0	1*1*2	2*0*1
	1*2*1	2*2*0	0*1*3	2*1*2
	2*0*0	0*0*2	0*2*3	0*2*1
	0*2*2	1*2*1	2*1*0	2*2*3
	0*0*1	0*1*2	0*0*1	0*2*0
	1*0*1	0*0*2	2*0*2	2*0*0
	0*2*3	2*2*2	0*0*1	1*0*0
	1*0*3	1*0*3	0*2*2	1*2*3
	1*0*1	1*0*3	1*0*1	2*2*2
	2*2*2	0*0*1	1*1*0	2*1*0
	1*2*1	2*0*0	1*2*0	2*0*3
	2*2*3	2*2*1	2*0*3	1*2*2
	1*0*2	1*2*0	0*0*2	0*2*3
	2*1*3	1*0*0	0*1*0	0*2*2
	2*1*1	0*1*0	1*1*1	1*1*2
	0*2*1	0*2*0	0*1*1	1*2*0
	1*2*1	1*1*0	2*1*3	2*2*0
	0*1*3	1*2*3	1*2*3	2*1*1
	2*2*1	2*0*3	1*0*2	0*1*3
	0*2*1	0*1*2	2*2*3	0*1*0
	2*1*1	2*1*2	2*0*2	1*1*2
	1*1*3	0*2*0	1*0*1	0*0*0
	0*1*2	2*0*1	2*0*3	0*2*3
	1*0*0	0*1*1	1*1*1	2*1*0
	0*2*1	2*0*1	0*1*2	0*2*2
	0*0*0	1*1*0	0*0*0	1*1*2
	2*1*1	2*0*2	0*0*3	1*1*3
COLUMNS				

Продуманный (стандартизированный) план эксперимента – проще анализ и интерпретация данных!



Какие исследовательские вопросы можно рассмотреть?

- 1. Влияние густоты посева на урожайность:** Как изменение густоты посева влияет на урожайность каждого исследуемого сорта пшеницы?
- 2. Влияние срока сева на урожайность:** Как разные сроки сева влияют на урожайность? Можно анализировать, как ранний, средний и поздний сев влияют на разные сорта пшеницы.
- 3. Взаимодействие густоты посева и срока сева:** Как густота посева в сочетании со сроками сева влияет на урожайность? Включает в себя анализ комбинированного эффекта этих двух факторов.
- 4. Сравнение сортов пшеницы:** Какой сорт пшеницы показывает лучшие результаты по урожайности при различных сроках сева и густоте посева?
- 5. Адаптивность сортов к условиям выращивания:** Какой сорт пшеницы наиболее устойчив к изменениям в густоте посева и сроках сева? Это может быть измерено через изменение урожайности или качества зерна.
- 6. Статистическая значимость различий:** Насколько статистически значимы различия в урожайности между различными сортами, вариантами густоты сева и сроками сева?

Постановка гипотез

1. Гипотеза о влиянии густоты посева:

Нулевая гипотеза (H_0): Густота посева не влияет на урожайность пшеницы.

Альтернативная гипотеза (H_1): Густота посева влияет на урожайность пшеницы.

2. Гипотеза о влиянии срока сева:

H_0 : Срок сева не оказывает влияния на урожайность пшеницы.

H_1 : Срок сева оказывает влияние на урожайность пшеницы.

3. Гипотеза о взаимодействии густоты посева и срока сева:

H_0 : Нет взаимодействия между густотой посева и сроком сева в отношении урожайности пшеницы.

H_1 : Существует взаимодействие между густотой посева и сроком сева, влияющее на урожайность пшеницы.

4. Гипотеза о сравнении сортов пшеницы:

H_0 : Нет различий в урожайности между разными сортами пшеницы при различных условиях сева и густоты посева.

H_1 : Существуют значимые различия в урожайности между разными сортами пшеницы при различных условиях сева и густоты посева.

5. Гипотеза об адаптивности сортов к условиям выращивания:

H_0 : Все сорта пшеницы одинаково адаптируются к различным условиям сева и густоты посева.

H_1 : Некоторые сорта пшеницы лучше адаптируются к определенным условиям сева и густоты посева, чем другие.

Функция `lm()`

(линейная модель) предназначена в основном для регрессионного анализа. Она оценивает коэффициенты для каждой переменной (или их комбинации), указанных в модели, и предоставляет подробную статистику для каждого коэффициента.

Преимущества:

- 1. Детализация результатов:** Функция `lm()` предоставляет подробные результаты для каждого коэффициента в модели, включая оценки, стандартные ошибки, t-статистику и p-значения. Это позволяет глубже понять влияние каждой переменной (и их взаимодействий) на зависимую переменную, что особенно полезно при тестировании гипотез о взаимодействиях между факторами.
- 2. Гибкость модели:** `lm()` позволяет легко моделировать как основные эффекты, так и взаимодействия между параметрами. Это особенно важно при тестировании гипотез о взаимодействиях между различными факторами, например, такими как густота посева, срок сева и сорт пшеницы.
- 3. Универсальность:** Хотя функция `aov()` для дисперсионного анализа часто используется для анализа дизайнов экспериментов, функция `lm()` также подходит для этих целей. `lm()` может быть использована для анализа того же набора данных и проверки тех же гипотез, что и `aov()`, но с дополнительными деталями в выводе.

Демонстрация работы в
R...

#Гипотеза о влиянии густоты посева

Базовый уровень
для показателя
«густота» (в нашем
примере 2,5 млн)

Call:
lm(formula = YIELD ~ DENSITY, data = WT)

Residuals:

Min	1Q	Median	3Q	Max
-9.6705	-3.1813	0.0176	2.5022	11.0225

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	67.6879	0.6824	99.198	< 2e-16 ***
DENSITY3,5 млн	3.7433	0.9650	3.879	0.000161 ***
DENSITY4,5 млн	4.0960	0.9650	4.245	3.96e-05 ***
DENSITY5,5 млн	4.5462	0.9650	4.711	5.87e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

При базовом уровне
густоты - средняя
урожайность 67,68
ц/га

При густоте посева 5,5
млн. урожайность
увеличивается на 4,5
ц/га по сравнению с
базовым уровнем

Residual standard error: 4.094 on 140 degrees of freedom
Multiple R-squared: 0.1674, Adjusted R-squared: 0.1496
F-statistic: 9.384 on 3 and 140 DF, p-value: 1.081e-05

Базовый уровень - intercept (или категория отсчёта) для факторных переменных в линейной модели в R обычно выбирается автоматически. По умолчанию R выбирает первый уровень фактора в алфавитном порядке (но можно задать вручную) в качестве базового уровня. Этот выбор влияет на интерпретацию коэффициентов модели.

Все коэффициенты
статистически
значимы (значение
ниже 0.05)

16.74% изменчивости
в урожайности можно
объяснить с помощью
густоты посева

#Гипотеза о влиянии густоты посева

Нулевая гипотеза (H_0): Густота посева не влияет на урожайность пшеницы.

Альтернативная гипотеза (H_1): Густота посева влияет на урожайность пшеницы.

Результаты показали, что все коэффициенты для различных уровней густоты посева (2,5 млн, 3,5 млн, 4,5 млн и 5,5 млн) оказались **статистически значимыми** (р-значения меньше 0.05), следовательно, мы можем **отклонить нулевую гипотезу**.

Это означает, что **густота посева действительно оказывает влияние на урожайность пшеницы** (в нашем конкретном опыте!).

#Гипотеза о взаимодействии густоты посева и срока сева

Call:

```
lm(formula = YIELD ~ DENSITY * TERM, data = WT)
```

Residuals:

Min	1Q	Median	3Q	Max
-8.9824	-3.0007	-0.0485	2.6084	10.3232

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	67.8326	1.1924	56.889	<2e-16 ***
DENSITY3,5 млн	3.5732	1.6863	2.119	0.0360 *
DENSITY4,5 млн	3.9399	1.6863	2.337	0.0210 *
DENSITY5,5 млн	4.1121	1.6863	2.439	0.0161 *
TERM2 срок	1.1484	1.6863	0.681	0.4971
TERM3 срок	-1.5825	1.6863	-0.938	0.3497
DENSITY3,5 млн:TERM2 срок	-0.2617	2.3847	-0.110	0.9128
DENSITY4,5 млн:TERM2 срок	-0.4377	2.3847	-0.184	0.8547
DENSITY5,5 млн:TERM2 срок	0.1787	2.3847	0.075	0.9404
DENSITY3,5 млн:TERM3 срок	0.7722	2.3847	0.324	0.7466
DENSITY4,5 млн:TERM3 срок	0.9058	2.3847	0.380	0.7047
DENSITY5,5 млн:TERM3 срок	1.1234	2.3847	0.471	0.6384

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.13 on 132 degrees of freedom

Multiple R-squared: 0.201, Adjusted R-squared: 0.1344

F-statistic: 3.018 on 11 and 132 DF, p-value: 0.001274

Нулевая гипотеза о том, что нет взаимодействия между густотой посева и сроком сева в отношении урожайности, не может быть отвергнута, так как нет статистически значимых результатов, подтверждающих это взаимодействие.

Ни одно из взаимодействий между густотой посева и сроком сева не показывает статистической значимости (все р-значения выше 0.05), что указывает на отсутствие значимого взаимодействия между этими факторами в отношении урожайности.

#Гипотеза о сравнении сортов пшеницы

Модель установила
Гомер базовым
сортом, 2,4 млн –
базовым значением
густоты, Срок 1 –
базовым сроком

```
Call:
lm(formula = YIELD ~ VARIETY + TERM + DENSITY + TERM:DENSITY,
    data = WT)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.8780	-1.7979	0.0583	1.6278	6.1889

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	71.9669	0.7906	91.023	< 2e-16 ***
VARIETYльговская 4	-7.5915	0.5176	-14.667	< 2e-16 ***
VARIETYСобербаш	-4.8113	0.5176	-9.295	4.49e-16 ***
TERM2 срок	1.1484	1.0352	1.109	0.269339
TERM3 срок	-1.5825	1.0352	-1.529	0.128773
DENSITY3,5 млн	3.5732	1.0352	3.452	0.000752 ***
DENSITY4,5 млн	3.9399	1.0352	3.806	0.000217 ***
DENSITY5,5 млн	4.1121	1.0352	3.972	0.000117 ***
TERM2 срок:DENSITY3,5 млн	-0.2617	1.4640	-0.179	0.858427
TERM3 срок:DENSITY3,5 млн	0.7722	1.4640	0.527	0.598790
TERM2 срок:DENSITY4,5 млн	-0.4377	1.4640	-0.299	0.765449
TERM3 срок:DENSITY4,5 млн	0.9058	1.4640	0.619	0.537205
TERM2 срок:DENSITY5,5 млн	0.1787	1.4640	0.122	0.903026
TERM3 срок:DENSITY5,5 млн	1.1234	1.4640	0.767	0.444276

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.536 on 130 degrees of freedom
Multiple R-squared: 0.7034, Adjusted R-squared: 0.6738
F-statistic: 23.72 on 13 and 130 DF, p-value: < 2.2e-16

Сорта Льговская 4 и
Собербаш показывают
статистически значимое
снижение урожайности по
сравнению с базовым сортом

Все уровни густоты посева (3,5
млн, 4,5 млн, 5,5 млн)
статистически значимо
увеличивают урожайность по
сравнению с базовой густотой.

Ни одно из взаимодействий
между сроками сева и густотой
посева не показывает
статистической значимости.

Ни один из сроков
сева не оказывает
статистически
значимого влияния
на урожайность.

#Гипотеза об адаптивности сортов к условиям выращивания

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	70.9560	1.1526	61.564	< 2e-16	***
VARIETYЛьговская 4	-5.0003	1.6300	-3.068	0.002726	**
VARIETYСобербаш	-4.3698	1.6300	-2.681	0.008494	**
TERM2 срок	2.4650	1.6300	1.512	0.133375	
TERM3 срок	-3.8576	1.6300	-2.367	0.019729	*
DENSITY3,5 млн	3.1984	1.6300	1.962	0.052305	.
DENSITY4,5 млн	6.2487	1.6300	3.834	0.000213	***
DENSITY5,5 млн	3.6608	1.6300	2.246	0.026742	*
VARIETYЛьговская 4:TERM2 срок	-2.0326	2.3051	-0.882	0.379855	
VARIETYСобербаш:TERM2 срок	-1.9173	2.3051	-0.832	0.407370	
VARIETYЛьговская 4:TERM3 срок	1.2295	2.3051	0.533	0.594858	
VARIETYСобербаш:TERM3 срок	5.5959	2.3051	2.428	0.016853	*
VARIETYЛьговская 4:DENSITY3,5 млн	0.5857	2.3051	0.254	0.799924	
VARIETYСобербаш:DENSITY3,5 млн	0.5386	2.3051	0.234	0.815710	
VARIETYЛьговская 4:DENSITY4,5 млн	-3.9387	2.3051	-1.709	0.090382	.
VARIETYСобербаш:DENSITY4,5 млн	-2.9874	2.3051	-1.296	0.197745	
VARIETYЛьговская 4:DENSITY5,5 млн	-0.2927	2.3051	-0.127	0.899191	
VARIETYСобербаш:DENSITY5,5 млн	1.6466	2.3051	0.714	0.476574	
TERM2 срок:DENSITY3,5 млн	-0.8106	2.3051	-0.352	0.725798	
TERM3 срок:DENSITY3,5 млн	4.7047	2.3051	2.041	0.043693	*
TERM2 срок:DENSITY4,5 млн	-0.9690	2.3051	-0.420	0.675044	
TERM3 срок:DENSITY4,5 млн	4.0211	2.3051	1.744	0.083929	.
TERM2 срок:DENSITY5,5 млн	0.1147	2.3051	0.050	0.960407	
TERM3 срок:DENSITY5,5 млн	6.7362	2.3051	2.922	0.004233	**
VARIETYЛьговская 4:TERM2 срок:DENSITY3,5 млн	0.7910	3.2599	0.243	0.808739	
VARIETYСобербаш:TERM2 срок:DENSITY3,5 млн	0.8557	3.2599	0.262	0.793450	
VARIETYЛьговская 4:TERM3 срок:DENSITY3,5 млн	-5.7640	3.2599	-1.768	0.079863	.
VARIETYСобербаш:TERM3 срок:DENSITY3,5 млн	-6.0335	3.2599	-1.851	0.066929	.
VARIETYЛьговская 4:TERM2 срок:DENSITY4,5 млн	0.2158	3.2599	0.066	0.947348	
VARIETYСобербаш:TERM2 срок:DENSITY4,5 млн	1.3783	3.2599	0.423	0.673289	
VARIETYЛьговская 4:TERM3 срок:DENSITY4,5 млн	-4.3599	3.2599	-1.337	0.183894	
VARIETYСобербаш:TERM3 срок:DENSITY4,5 млн	-4.9862	3.2599	-1.530	0.129056	
VARIETYЛьговская 4:TERM2 срок:DENSITY5,5 млн	-0.9494	3.2599	-0.291	0.771433	
VARIETYСобербаш:TERM2 срок:DENSITY5,5 млн	1.1415	3.2599	0.350	0.726908	
VARIETYЛьговская 4:TERM3 срок:DENSITY5,5 млн	-6.8775	3.2599	-2.110	0.037194	*
VARIETYСобербаш:TERM3 срок:DENSITY5,5 млн	-9.9611	3.2599	-3.056	0.002830	**

Residuals:

Min	1Q	Median	3Q	Max
-1.12500	-0.32187	-0.02125	0.35562	1.08500

Некоторые взаимодействия между сортами, сроками сева и густотой посева оказываются статистически значимыми.

Эти комбинации значительно снижают урожайность, что может указывать на то, что определенные комбинации условий выращивания более или менее благоприятны для разных сортов.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.305 on 108 degrees of freedom
Multiple R-squared: 0.7964, Adjusted R-squared: 0.7304
F-statistic: 12.07 on 35 and 108 DF, p-value: < 2.2e-16

Гомер

Call:

```
lm(formula = YIELD ~ TERM * DENSITY, data = Gomer_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.382	-1.385	-0.468	1.321	4.106

Coefficients:

Изменения в урожайности
по сравнению с базовым
сроком (Срок 1)

Изменения в
урожайности по
сравнению с базовым
значением густоты
(2,5 млн)

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	70.9560	1.1465	61.888	< 2e-16	***
TERM2 срок	2.4650	1.6214	1.520	0.137178	
TERM3 срок	-3.8576	1.6214	-2.379	0.022776	*
DENSITY3,5 млн	3.1984	1.6214	1.973	0.056261	.
DENSITY4,5 млн	6.2487	1.6214	3.854	0.000461	***
DENSITY5,5 млн	3.6608	1.6214	2.258	0.030120	*
TERM2 срок:DENSITY3,5 млн	-0.8106	2.2931	-0.353	0.725787	
TERM3 срок:DENSITY3,5 млн	4.7047	2.2931	2.052	0.047527	*
TERM2 срок:DENSITY4,5 млн	-0.9690	2.2931	-0.423	0.675108	
TERM3 срок:DENSITY4,5 млн	4.0211	2.2931	1.754	0.088009	.
TERM2 срок:DENSITY5,5 млн	0.1147	2.2931	0.050	0.960383	
TERM3 срок:DENSITY5,5 млн	6.7362	2.2931	2.938	0.005737	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.293 on 36 degrees of freedom

Multiple R-squared: 0.7136, Adjusted R-squared: 0.6261

F-statistic: 8.156 on 11 and 36 DF, p-value: 6.476e-07

Модель статистически
значима

Сравнение результатов – табличный расчет и R

Гомер снижает урожайность при загущении более 4,5 млн и при поздних сроках сева. Лучшая норма 4,5

- | | |
|--|--|
| <ul style="list-style-type: none">Поздний срок сева привел к снижению урожайности при норме высева 2,5, а ранний - на норме 5,5 | <ul style="list-style-type: none">Срок 3 – густота 2,5 – ДАСрок 1 – густота 5,5 меньше чем срок 3, густота 5,5 |
| <ul style="list-style-type: none">При раннем сроке сева Гомер 4,5 млн дает максимальную урожайность.При среднем и позднем сроке сева снижение нормы после 4,5 приводит к существенному снижению урожайности, а повышение свыше не ведет к ее увеличению | <ul style="list-style-type: none">Максимальная урожайность срок 1 – густота 4,5 – ДА (+ 6,2 ц/га по сравнению с базовым значением густоты)Срок 2 и Срок 3, густота < 4,5 |
| <ul style="list-style-type: none">средний срок сева существенно лучше позднего и раннего | <ul style="list-style-type: none">Нет (+2,5 ц/га по сравнению с базовым сроком, но значение не статистически значимо) |
| <ul style="list-style-type: none">вне зависимости от срока сева норма 4,5 существенно превышает по урожайности прочие нормы | <ul style="list-style-type: none">Да! |

Вывод: Эффект влияния сроков сева недоказан, эффект влияния норм высева и взаимодействие факторов доказаны

- Густота посева оказывает значительное влияние на урожайность, а срок сева 3 влияет на урожайность различным образом в зависимости от густоты посева. Например, при густоте 5,5 млн срок 3 увеличивает урожайность более значительно, чем при других уровнях густоты.

Оценка статистической значимости (достоверности)

Доспехов

В большинстве отечественных работ в качестве оценки достоверности используется наименьшая существенная разность (НСР, least significant difference, LSD). Этот параметр указывает на границу возможных случайных отклонений в эксперименте, за пределами которой различия средних считаются значимыми на соответствующем уровне. На практике обычно используют 1%-ный и 5%-ный уровни значимости (НСР01 и НСР05).

НО расчет НСР обращается к t-критерию Стьюдента информативен не для любых распределений

Современный подход

При анализе данных с помощью специальных программ обычно автоматически рассчитывается уровень значимости, выраженный через

p-value