

Учебные материалы тренинга:

**Статистический дизайн и анализ данных полевых экспериментов
в странах Центральной Азии и Южного Кавказа
1 – 12 декабря, 2014
Ташкент, Узбекистан**

Курс охватывает следующие области:

- Обобщение данных по одной переменной, используя основные описательные статистики и отображения распределения
- Оценка и испытания гипотез основной значимой, дисперсий и пропорций
- Ассоциация и статистические связи между переменными, используя простую линейную регрессию и множественную линейную регрессию.
- Проектирование и анализ сортоиспытания: конструкций в рандомизированных полных блоках (RCB -) и неполных блоках (альфа дизайн), р-REP конструкций, пространственного анализа полевых испытаний, сборный анализ данных полученных из разных условий окружающей среды для оценки взаимодействия GxE, анализа устойчивости, оценка наследуемости.
- Проектирование и анализ агрономических полевых испытаний: эксперименты многофакторной в RCB, разнофакторный эксперимент в RCB, линейный эксперимент в RCB
- Методы многомерного такие как кластерного анализа и анализа главных компонент
- Введение в программное обеспечение Genstat для Windows, и его использование для управления данными
- Онлайн объект Био компьютеризации ИКАРДА
- Идентификация статистических подходов для анализа данных, анализ базы данных участника тренинга и интерпретации результатов
- Использование некоторых статистических программ, закодированных в программном обеспечении Genstat

Автор

М. Сингх

Старший эксперт по Биометрии
Управление Заместителя Генерального Директора (Исследования)



**Международный Центр по сельскохозяйственным исследованиям в
засушливых регионах (ICARDA)**

Основные Статистические Модули

Содержание:

1. Основные статистические Модули
2. Наборы данных
3. Основная статистика одной величины
 - 3.1 . Совокупность
 - 3.2. Случайная выборка
 - 3.3 . Описательная Статистика
 - 3.4 . Описательная Статистика - Вычисление
 - 3.5 . Пример
 - 3.6 . Применение Меню Genstat
 - 3.7 . Табулирование данных
4. Вывод на совокупных параметрах
 - 4.1. Оценка среднего значения и различия
 - 4.2. Испытание гипотезы / Испытание значения
 - 4.2.1 Испытание гипотезы равенства средств двух нормальных совокупностей
 - 4.2.2 Испытание гипотезы равенства различий двух нормальных совокупностей
5. Корреляция и Регресс
 - 5.1. Мера ассоциации
 - 5.2. Корреляция
 - 5.3. Линейная регрессия
 - 5.4 Множественная линейная регрессия

1. Основные Статистические Модули

Эти модули представляют концепции и методы для проведения суммирования данных по одной величине:

- путем использования основной описательной статистики, и демонстрации распределения
- методом оценки и выполнения испытаний гипотез по средствам, различиям и пропорциям
- путем оценки ассоциированных и статистических отношений между двумя переменными, используя простой линейный регресс, и разработки модели предсказания, используя многократный линейный регресс

Для обсуждения вышеупомянутых понятий, было бы целесообразно начать с некоторых реальных наборов данных от нескольких отобранных областей исследований:

2. Наборы данных

1. Высота культуры пшеницы (Источник: Г-н А. А. Ньян)

Был проведен эксперимент для оценки ООС (ООС) (отчетливости,

однородности и стабильности) вариантов пшеницы в Тел Хадя, в

течение 2004г. Для иллюстрации методов, рассматриваются данные по

высоте растения (см), то есть 30 растений в этих двух сортах, 4-сорта и

5-сорта в Ячейке-1:

Генотип-4:

91	95	94	93	92	91	91	91	89	89
89	90	90	91	97	99	99	98	89	88
85	79	79	81	92	91	100	100	108	90

Генотип-5:

115	112	118	118	114	112	115	118	110	117
111	110	110	115	116	115	119	118	113	116
118	115	116	118	117	116	113	115	115	116

Мы стремимся просуммировать вышеупомянутые наборы данных, и сравнивать эти два Генотипа для их роста, и изменчивости в их пределах.

1. Нут (данные по числу стручков)

Здесь представлен пример по числу потомств от $N = 86$ растений ILC 482, линия нута (источник: Г-н. Н. А. Хак, 1990, FLIP, ICARDA)

113	91	84	76	73	64	60	56	50	50	47	45	44	43	42	42	41
39	36	35	35	33												
	32	32	31	31	31	31	30	30	29	28	28	28	28	27	27	26
26	26	25	22	22	22											
	22	22	21	21	21	19	19	19	18	18	17	17	17	16	16	16
16	15	15	15	14												
	14	13	13	13	12	12	12	10	10	10	10	10	9	8	8	7
7:																

Задачей является суммирование распределения потомства в вышеуказанной линии нута.

2. Данные по нуту (24 генотипа в 9 условиях, 6 признаках, данные по средствам для повторений)

Envt	GENO	BYLD	GYLD	HSW	DFLR	DMAT	PTHT
1	1	1079	220.6	23.55	68.5	96.5	30.5
1	2	1397	420.6	38.8	52.5	90.5	23.5
1	3	1127	220.6	26.4	62.5	92.5	26.5
1	4	1127	284.1	23.45	57.5	91.5	26.5
1	5	1333	287.3	25.8	57.5	92	27
1	6	1540	396.8	30.3	57.5	91.5	25.5
1	7	1333	350.8	21.95	58.5	91.5	25
1	8	1254	284.1	25.65	53.5	93	24
1	9	1000	42.9	23.1	66.5	96.5	29
1	10	1206	311.1	22.45	55.5	92	27.5
1	11	1190	234.9	25.85	55.5	94.5	28.5
1	12	1413	374.6	23.95	58.5	91	22.5
1	13	1190	173	22.9	60	92.5	28.5
1	14	1254	368.3	22.85	52	91	23
1	15	1381	282.5	23.55	56	92.5	27
1	16	1222	311.1	30.3	50	89.5	24.5
1	17	1190	346	27.4	52.5	91	22.5
1	18	1270	533.3	28.8	47.5	86.5	23.5
1	19	1238	542.9	23.9	44	83.5	18.5
1	20	1302	479.4	29.45	47.5	86	21.5
1	21	1254	338.1	29.7	48.5	92	22.5
1	22	1127	219.1	26.4	61	96	24
1	23	1317	415.9	27.85	57.5	91	21.5
1	24	1397	465.1	29.95	54	88	23
2	1	1206	27	18	69	96.5	30
2	2	1127	17.5	23.6	68	96.5	28.5
2	3	1222	165.1	28.1	60	94.5	23.5
2	4	1333	330.1	19.65	59.5	94	24.5
.	.	милле	.	.	.	.	.

Семинар: Загрузите эти комплекты данных путем использования Меню Genstat.

Основная статистика одной величины

3.1. Совокупность

- Список участков под урожаем в данном поле является Совокупностью. Список предела текучести этих участков - статистическая Совокупность.
- Список растений в данном участке является Совокупностью. Список значений высоты этих растений - статистическая Совокупность.
- Рассмотрите область (данную деревню, город, и т.д.) и список всех маленьких жвачных, скажем, овец в ней.
 - Список овец является Совокупностью.
 - Отчеты урожаев молока и шерсти от каждого животного - статистическая Совокупность.
- В случае, если доступны Генотипы всей совокупности, возникает потребность в описании совокупности с точки зрения ее основных особенностей, таких как, основные тенденции урожаев участка, или урожаев молока, распространение тех значений, и других аспектов распределения этих значений.

Совокупные единицы $U_1 \quad U_2, \quad \dots \quad U_N$

Значения величин $Y_1 \quad Y_2, \quad \dots \quad Y_N$

3.2. Случайная выборка

- Другой путь состоял бы в том, чтобы собирать информацию относительно только части совокупности.
- Процедура отбора единиц, названных осуществлением выборки, должна обеспечить образец (комплект совокупности), который, как ожидается, представит совокупность в целом.

Имеется несколько способов по выбору репрезентативной пробы.

- Процедура, которая прилагает вероятность, отличную от нуля отбора каждой единицы совокупности, называется случайным осуществлением выборки
- Если эти возможности являются равными, в таком случае, мы имеем простое случайное осуществление выборки.
- У осуществления выборки имеется несколько преимуществ, включая, низкие затраты, большую скорость, больший масштаб, и т.д.

3.3. Описательная Статистика

Особенности, такие как

- средняя величина совокупности (для того, чтобы представить единые характерные величины основной тенденции совокупных ценностей основной тенденции величин совокупности),
- различие (чтобы дать общее представление или измерения о величине дисперсии от величины совокупности),
- коэффициент перекоса (чтобы представить индикаторы отхода от симметричности его распределения), и
- коэффициент эксцесса (в симметричных распределениях, чтобы представить идею о том, как величины в совокупности обычно распределяются).

Совокупность:

Презентация в графическом виде:

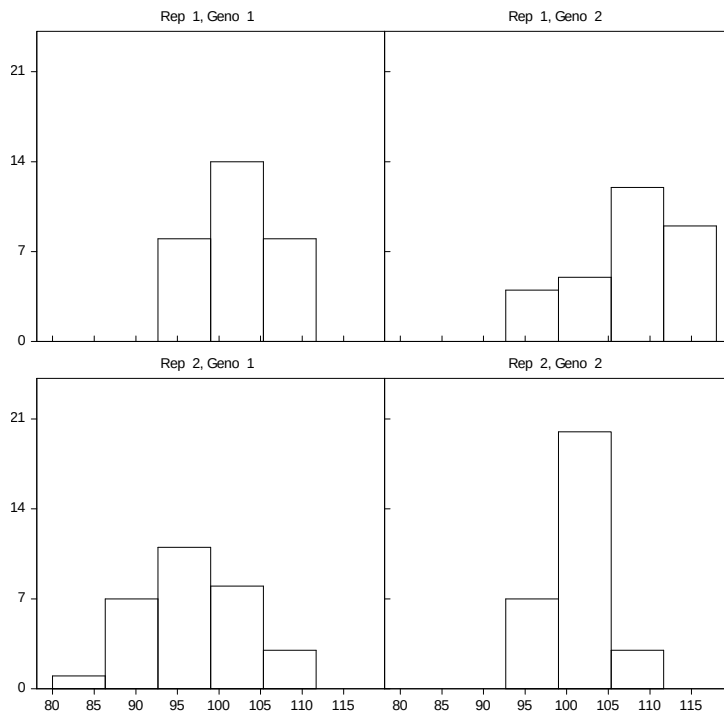
Высотные отметки растения (см) двух сортов пшеницы (ООС 2003-04)

Необходимо учесть, что каждый участок имел всего 30 растений. Эти записи представляют совокупность.

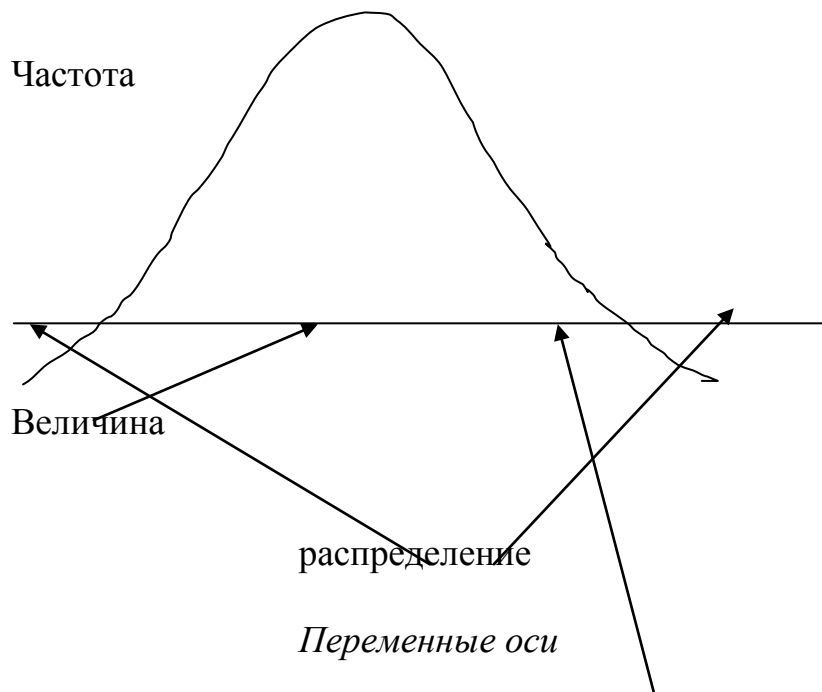
Geno1		Geno2	
Block 1	Block 2	Block 1	Block 2
105	98	113	102
93	110	97	105
106	95	113	107
108	100	111	102
95	105	110	104
98	102	105	101
95	98	108	107
109	98	109	105
110	93	113	103
107	91	112	104
103	94	113	101
108	91	98	103
102	90	99	100
105	80	104	99
106	91	103	97
100	91	98	101
97	99	101	102
104	104	107	102
103	108	103	100
99	91	118	95
102	91	109	100
106	99	115	96
102	99	107	98
105	104	110	107
99	108	111	102
103	100	111	101
102	104	110	100

105	103	115	99
100	98	112	99
99	97	109	100

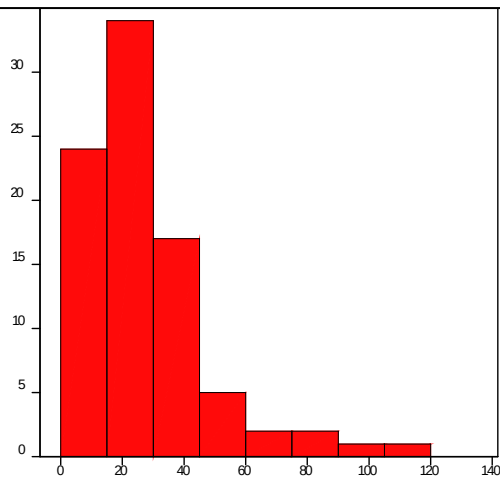
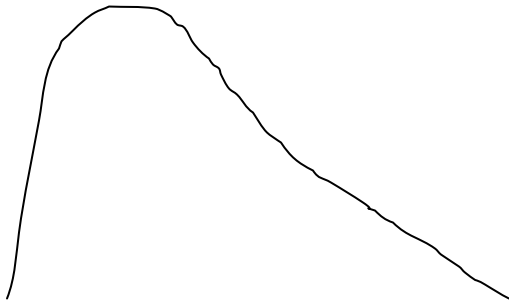
Гистограмма:



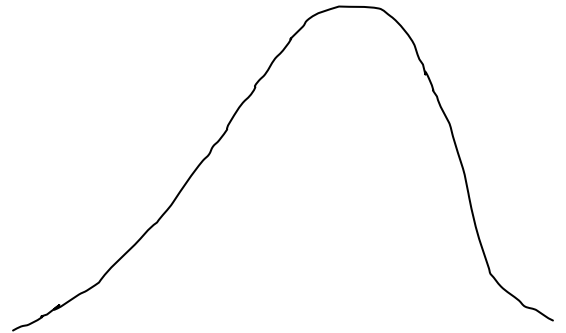
Симметричное распределение



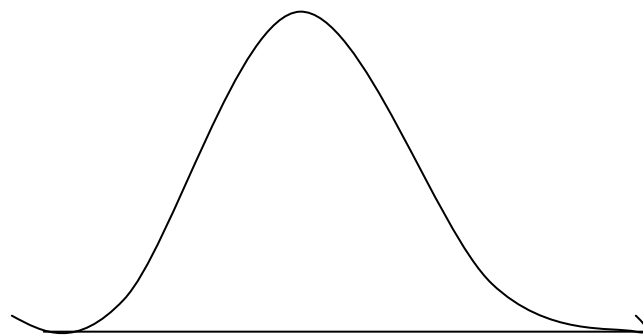
Асимметричное распределение
(Асимметрия в право)



Асимметричное распределение
(Асимметрия влево)



Нормальная совокупность
Симметричная, нулевая кurtosis (нормальный пиковый)



3.4 Описательная Статистика - Вычисление

3.4.1 Версии совокупности

i. Величина

$$\bar{Y} = (Y_1 + Y_2 + Y_3 + \dots + Y_N) / N$$

Или

$$\bar{Y} = \sum_{i=1}^{i=N} Y_i / N$$

ii. Дисперсия

$$\sigma^2 = \frac{(Y_1 - \bar{Y})^2 + (Y_2 - \bar{Y})^2 + (Y_3 - \bar{Y})^2 + \dots + (Y_N - \bar{Y})^2}{N}$$

или

$$\sigma^2 = \frac{(Y_1^2 + Y_2^2 + Y_3^2 + \dots + Y_N^2)}{N} - \bar{Y}^2$$

или,

$$\sigma^2 = \sum_{i=1}^{i=N} (Y_i - \bar{Y})^2 / N$$

Стандартное отклонение (CO) = σ
iii. Коэффициент вариации

$$CV = \sigma / \mu$$

Или,

$$CV\% = 100 * \sigma / \mu$$

iv. Коэффициент асимметрии

$$\gamma_1 = \frac{\sum_{i=1}^{i=N} (Y_i - \bar{Y})^3 / N}{\sigma^3}$$

v. Коэффициент Куртосиса

$$\gamma_2 = \frac{\sum_{i=1}^{i=N} (Y_i - \bar{Y})^4 / N}{\sigma^4} - 3$$

3.4.2 Вариант получения образца

Если мы наугад возьмем 10 растений, отобранных из 30 растений, тогда 10 величин из взятых наугад образцов могут быть использованы для расчета различных параметров, как величина, дисперсия, асимметрия и Куртосис.

Размер образца = n

величина $\bar{y} = (y_1 + y_2 + y_3 + \dots + y_n) / n$

дисперсия

$$s^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + \dots + (y_n - \bar{y})^2}{n - 1}$$

Или,

$$s^2 = \sum_{i=1}^{i=n} (y_i - \bar{y})^2 / (n - 1)$$

Коэффициент асимметрии

$$g_1 = \frac{\sum_{i=1}^{i=n} (y_i - \bar{y})^3 / n}{s^3}$$

Коэффициент Куртосиса

$$g_2 = \frac{\sum_{i=1}^{i=n} (y_i - \bar{y})^4 / n}{s^4} - 3$$

3.5 Примеры:

а) Данные по высоте растения

Для двух сортов пшеницы, записи по высоте растения могут быть просуммированы следующим образом:

Гистограмма:

```
50 Rest Height; Entry==1.and.Rep==1
51 Hist Height : Rest Height
```

Histogram of height

-	96	3	***
96	- 100	7	*****
100	- 104	8	*****
104	- 108	10	*****
108	-	2	**

Scale: 1 asterisk represents 1 unit.

```
52 Rest Height; Entry==1.and.Rep==2
53 Hist Height : Rest Height
```

Histogram of height

-	88	1	*
88	- 96	10	*****
96	- 104	15	*****
104	- 112	4	****
112	-	0	

Scale: 1 asterisk represents 1 unit.

Семинар: Проверьте следующее.

```
54 Rest Height; Entry==2.and.Rep==1
55 Hist Height : Rest Height
```

Histogram of height

-	100	4	****
100	- 105	5	*****
105	- 110	9	*****
110	- 115	11	*****
115	-	1	*

Scale: 1 asterisk represents 1 unit.

```
56 Rest Height; Entry==2.and.Rep==2
57 Hist Height : Rest Height
```

Histogram of height

-	97.5	3	***
97.5	- 100.0	9	*****
100.0	- 102.5	9	*****
102.5	- 105.0	6	*****
105.0	-	3	***

Scale: 1 asterisk represents 1 unit.

Basic statistics

```
108 Rest Height; Entry==1.and.Rep==1
109 DESCRIBE
[SELECTION=nval,nobs,nmv,mean,median,min,max,range,
var,sd,%cv,skew,seskew,\
110 kurtosis,sekurtosis] Height
```

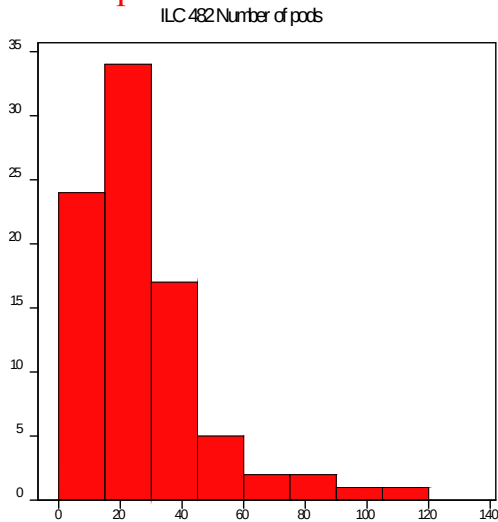
Краткое изложение статистики для Высоты

№ величин = 30
Число наблюдений = 30
Число отсутствующих величин = 0
Величина = 102.5
Средняя = 103
Минимальная = 93
Максимальная = 110
Диапазон = 17
Стандартное отклонение = 4.345
Дисперсия = 18.88
Коэффициент вариации = 4.238
Асимметрии = -0.367
Стандартная ошибка Асимметрии = 0.427
Куртосиса = -0.574
Стандартная ошибка Куртосиса = 0.833

b) Distribution of number of pods in ILC 482

113 91 84 76 73 64 60 56 50 50 47 45 44 43 42 42 41
 39 36 35 35 33
 32 32 31 31 31 31 30 30 29 28 28 28 28 27 27 26 26
 26 25 22 22 22
 22 22 21 21 21 19 19 19 18 18 17 17 17 16 16 16 16
 16 15 15 15 14
 14 13 13 13 12 12 12 10 10 10 10 10 9 8 8 7 7
 7 7 7 :

Гистограмма



104 Hist NoPods

Histogram of NoPods

```

-----
          -   15   24 *****
        15 -   30   34
*****
        30 -   45   17 *****
        45 -   60    5 *****
        60 -   75    2 **
        75 -   90    2 **
        90 -  105    1 *
       105 -  120    1 *
  
```

120 - 0

Scale: 1 asterisk represents 1 unit.

Basic statistics

```
114 DESCRIBE  
[SELECTION=nval,nobs,nmv,mean,median,min,max,range,  
var,sd,%cv,skew,seskew,\  
115 kurtosis,sekurtosis] NoPods
```

Краткое изложение статистики для Числа стручков

Число величин = 86
Число наблюдений = 86
Число отсутствующих величин = 0
Величина = 27.91
Средняя = 22
Минимальная = 7
Максимальная = 113
Диапазон = 106
Стандартное отклонение = 20.0
Дисперсия = 400.1
Коэффициент вариации = 71.68
Асимметрии = 1.851
Стандартная Ошибка Асимметрии = 0.260
Куртосис = 4.033
Стандартная ошибка Куртосиса = 0.514

Используется преобразование: квадратный корень, логарифм

```
116 "Преобразование"  
117 Calc SqrtNoPods= sqrt(NoPods)  
118 Calc LogNoPods = log(NoPods)  
119 Hist SqrtNoPods  
104 Hist NoPods
```

Гистограмма NoPods

```
      -   15  24 *****  
15 -   30  34  
*****  
30 -   45  17 *****  
45 -   60   5 *****  
60 -   75   2 **  
75 -   90   2 **  
90 -  105   1 *  
105 - 120   1 *  
120 -      0
```

Значения: 1 звездочка представляет 1 единицу.

Гистограмма SqrtNoPods

```
      -   3   8 *****  
3 -   4  21 *****  
4 -   5  17 *****  
5 -   6  22 *****  
6 -   7   8 *****  
7 -   8   5 *****  
8 -   9   2 **  
9 -  10   2 **  
10 -      1 *
```

Значения: 1 звездочка представляет 1 единицу.

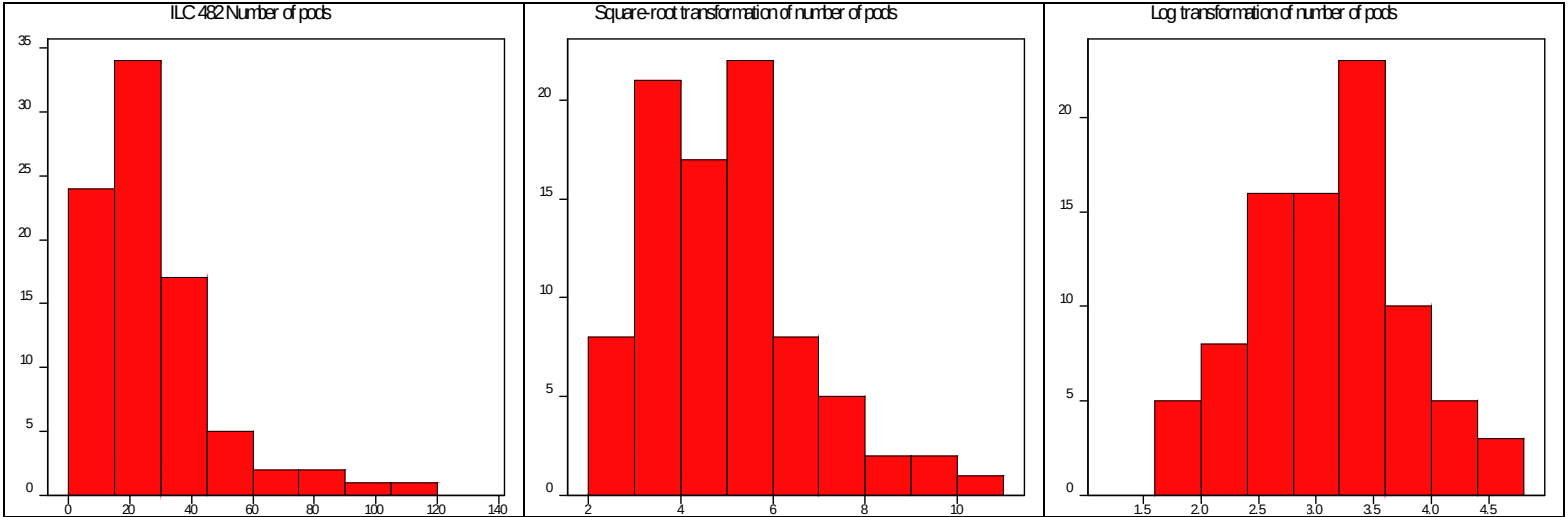
120

Hist LogNoPods

Histogram of LogNoPods

	- 1.6	0	
1.6	- 2.0	5	*****
2.0	- 2.4	8	*****
2.4	- 2.8	16	*****
2.8	- 3.2	16	*****
3.2	- 3.6	23	*****
3.6	- 4.0	10	*****
4.0	- 4.4	5	*****
4.4	-	3	***

Scale: 1 asterisk represents 1 unit.



Теоретические распределения:

Несколько обычно применяемых теоретических распределений для продолжительных переменных

- **Нормальное распределение**

Полученные распределения от нормального распределения

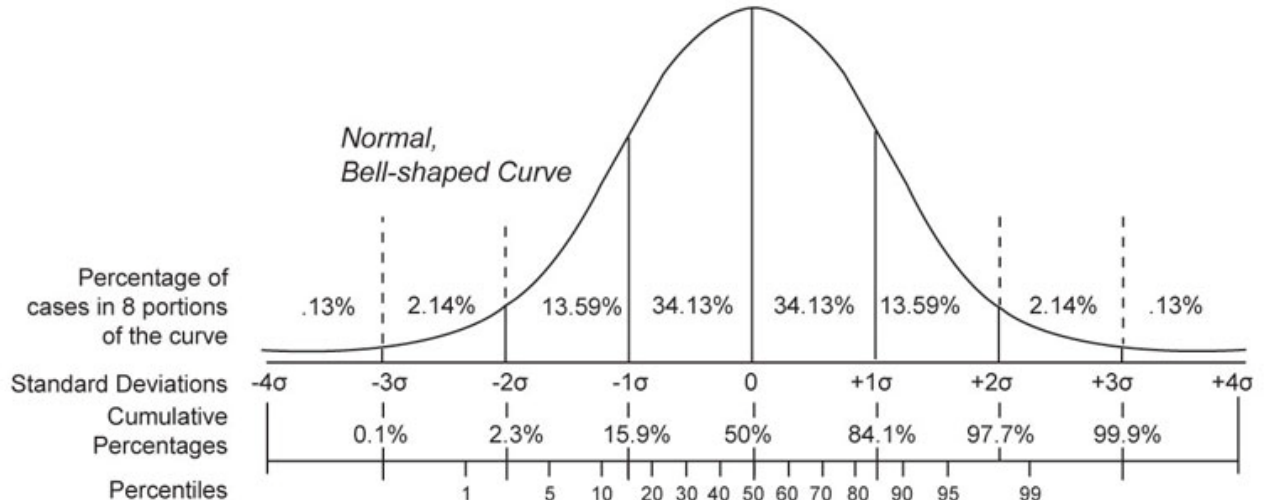
- **t-распределение**
- **распределение Хи-квадрат**
- **F-распределение**

1. Нормальное Распределение

Часто мы сталкиваемся с конечной Совокупностью с переменными величинами, измеренными в продолжительном масштабе (например, высота, вес, площадь листа, и т.д.). Имеются также примеры дискретных весов (таких как число стручков на растение, число сеялок, число выживших растений). Среди продолжительного распределения, в широком масштабе применяется распределение (или Совокупность), названное нормальным распределением, или Нормальной Совокупностью. Коэффициенты асимметрии и Куртосиса являются нулевыми для этого распределения. Распределение полностью описано значениями величины μ и дисперсионными σ^2 параметрами. Формы его функции, распределения функции плотности вероятности, выглядят следующим образом:(pdf)

$$P_r [Y \leq y] = \int_{-\infty}^y \text{pdf } dx = \int_{-\infty}^y (1/(2\pi\sigma^2)^{1/2}) e^{-(x-\mu)^2/(2\sigma^2)} dx \\ = \Phi\left(\frac{y-\mu}{\sigma}\right), \text{ where } \int_{-\infty}^z (1/\sqrt{2\pi}) e^{-(1/2)x^2} dx = \Phi(z)$$

Имеется таблица величин $\Phi(z)$ для величины z . Частота распределения (или функция плотности вероятности) нормального распределения будет в форме:



Источник: Wikipedia.org

Примечание

$$\text{Prob} [Y \leq \mu\sigma] = \text{Prob} [Y \geq \mu + \sigma] = 0.16$$

$$\text{Prob} [Y \leq \mu - 2\sigma] = \text{Prob} [Y \geq \mu + 2\sigma] = 0.023$$

$$\text{Prob} [Y \leq \mu - 1.96\sigma] = \text{Prob} [Y \geq \mu + 1.96\sigma] = 0.025$$

$$\text{Prob} [Y \leq \mu - 1.64\sigma] = \text{Prob} [Y \geq \mu + 1.64\sigma] = 0.05$$

$$\text{Prob} [\mu - \sigma \leq Y \leq \mu + \sigma] = 0.68$$

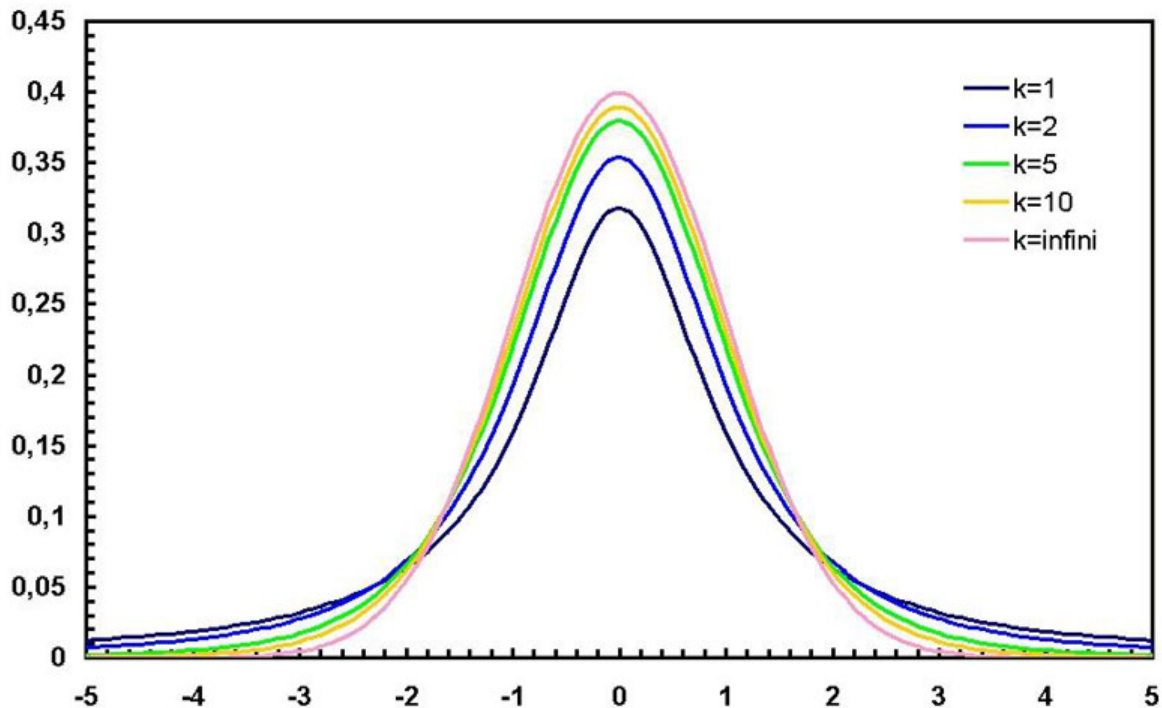
$$\text{Prob} [\mu - 2\sigma \leq Y \leq \mu + 2\sigma] = 0.954$$

$$\text{Prob} [\mu - 3\sigma \leq Y \leq \mu + 3\sigma] = 0.99$$

2. t-распределение

Данное распределение является симметричным, и имеет один параметр, называемый степенью свободы ($df = \nu$) вместе с pdf

$pdf(t, \nu) = \text{constant} (1 + t^2 / \nu)^{-(\nu+1)/2}$. Его формы для некоторых отобранных ν являются:



Источник: Wikipedia.org

Если $\bar{x}$ является величиной наугад принятого образца размера n от нормальной совокупности с величиной ноль, и s^2 -образец дисперсии ($n-1$ в знаменателе), тогда

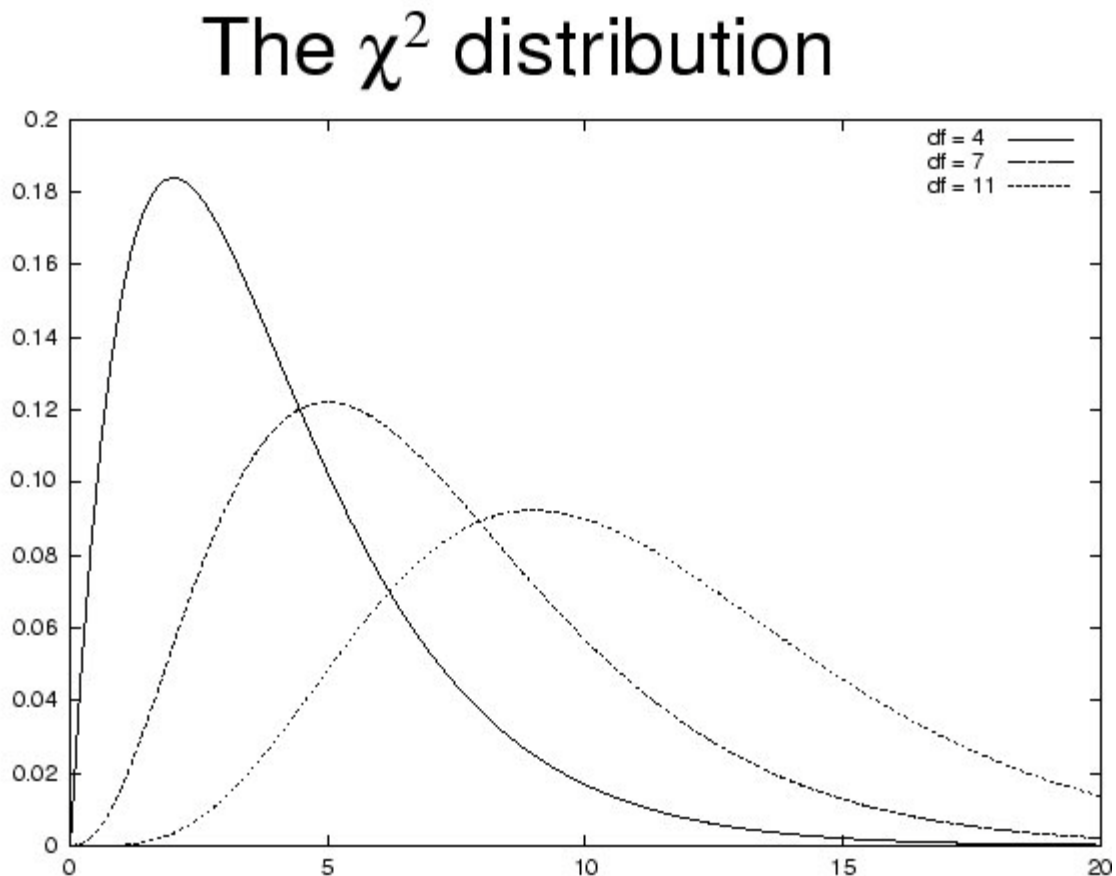
$$t = \bar{x} / (s / n^{1/2})$$

следует t-распределение с $\nu = n-1$ степенью свободы. Применение t-распределения возникает при сравнении единой величины, двух величин, соотношения коэффициента, и.т.д.

3 χ^2 - распределение (распределение хи-квадрата)

Это-асимметричное распределение параметра, названный df (ν) вместе с pdf:

$\text{pdf}(\chi^2, \nu) = \text{constant } \chi^{\nu-2} e^{-(1/2)\chi^2}$. Для величины df, его формы выглядят как:



Degrees of freedom: 4, 7 and 11

Источник: Wikipedia.org

Если s^2 образец дисперсии ($n-1$ в знаменателе), где образец был начертан от нормальной совокупности с дисперсией 1, тогда

$$\chi^2 = (n-1)s^2$$

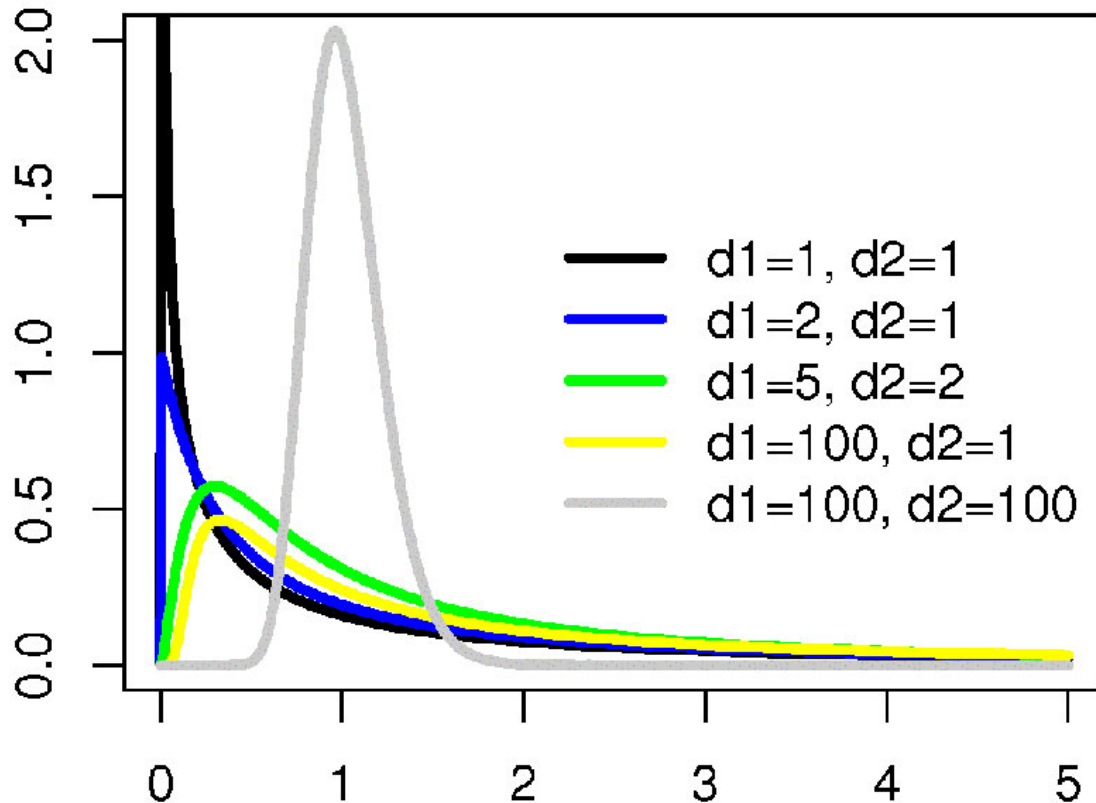
следует распределение хи-квадрата с $\nu = n-1$ степенью свободы.

4. F-распределение

Это-асимметричное распределение, и основано на два параметра, названные числителем уровни свободы (v_1) и уровнем знаменателя свободы (v_2) с pdf:

$$\text{pdf}(F, v_1, v_2) = \text{constant } F^{(v_1/2-1)} (v_1 F + v_2)^{-(v_1+v_2)/2}$$

Его форма:



Источник: Wikipedia.org

Представим s_1^2 и s_2^2 являются дисперсиями (размер образца-1 знаменатель) от образцов размеров n_1 и n_2 , и образцы были начертаны от двух нормальных совокупностей с дисперсиями σ_1^2 и σ_2^2 соответственно. Тогда

$$F = s_1^2 / s_2^2$$

следует за F-распределением с $v_1 = n_1 - 1$ и $v_2 = n_2 - 1$ степенью свободы соответственно.

3.6. Применение Меню Genstat

'Подведите итоги Статистик...>Просуммируйте содержание переменных"

3.7. Табулирование данных:

- Суммирование данных для каждого уровня фактора, или сочетаний уровней многократных факторов.

В данном случае, мы обсуждаем итоги переменных для каждого индивидуального уровня (или комбинацию уровней) фактора (многократных факторов).

Семинар: Используя данные из файла (файлов) по высоте растений, просуммируйте данные по высоте растений для Генотипа 2 для каждой из двух повторений

3.7.1 Пример:

а..) Составить таблицу основной статистики данных ООС 2003-04 для каждой комбинации генотипа и повтора

```
44  TABULATE  
[PRINT=means,variances,minima,maxima,sd;  
CLASSIFICATION=Rep,Entry; MARGINS=no] Height
```

Entry 1

	Mean	Minimum	Maximum	Variance	s.d.
Rep					
1	102.5	93	110	18.88	4.345
2	97.7	80	110	44.27	6.654

Entry 2

	Mean	Minimum	Maximum	Variance	s.d.
Rep					
1	108.1	97	118	31.02	5.569
2	101.4	95	107	9.28	3.047

4. Генотип параметров по совокупности:

- Расчет
- Испытание гипотезы / Критерия значимости

4.1 Расчет величины и дисперсии

Проблема:

- Вычисление параметров величины и дисперсии нормальной совокупности, от взятых наугад образцов размера n .

Совокупность:

$$N(\mu, \sigma^2)$$

Образец:

$$y_1, y_2, y_3, \dots, y_n$$

Подсчет величины

$$\mu$$

$$\bar{y} = \sum_{i=1}^{i=n} y_i / n$$

Вычисление дисперсии

$$\sigma^2$$

:

$$s^2 = \sum_{i=1}^{i=n} (y_i - \bar{y})^2 / (n - 1)$$

Стандартные ошибки

Вычисление Стандартных ошибок Вычисленной величины:

SEm =

$$s / \sqrt{n}$$

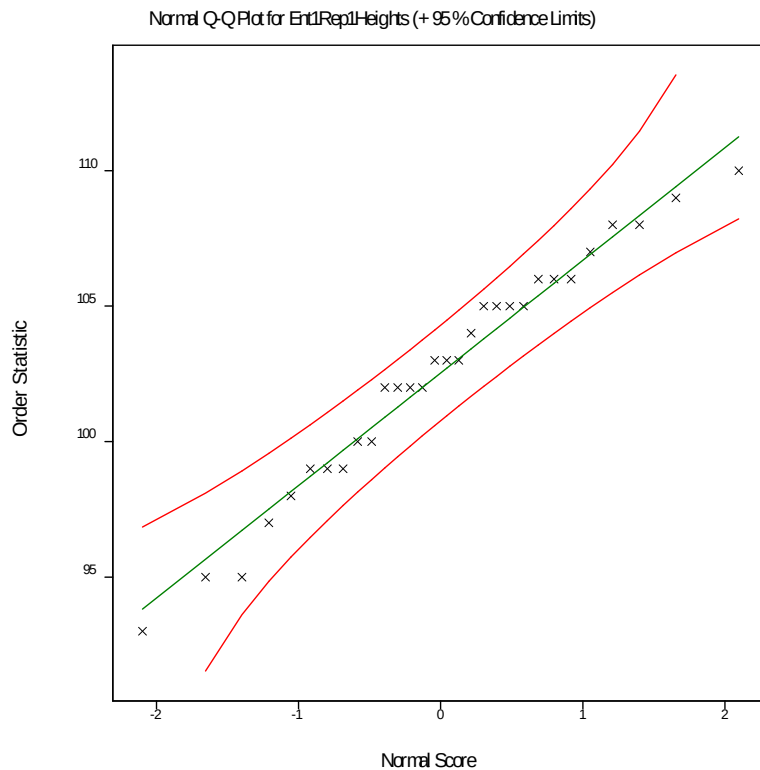
Границы доверительного интервала (интервал) совокупности величины:

$$(\bar{y} - t_{n-1; \alpha/2} s / \sqrt{n}, \quad \bar{y} + t_{n-1; \alpha/2} s / \sqrt{n})$$

Пример: ООС 2003-04, Генотип-1.

- Вычисление параметра величины и дисперсии нормальной совокупности от взятого наугад образца размера $n = 30$.

Нормальный участок вероятности для Генотипа 1, повтор-1 Высота



Суммарная статистика для Ent1Rep1Height

Число наблюдений = 30

Величина = 102.5

Стандартное отклонение = 4.345

Дисперсия = 18.88

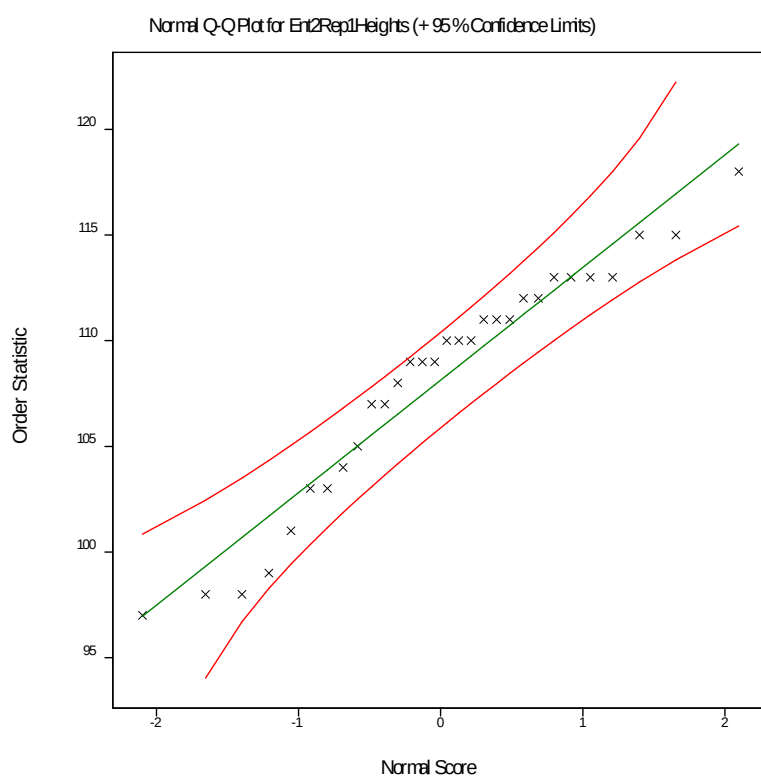
95% доверительный интервал для величины: (100.9, 104.2)

Суммарная статистика для Ent1Rep2Height

Число наблюдений = 30
Величина = 97.73
Стандартное отклонение = 6.654
Дисперсия = 44.27

95% доверительный интервал для величины: (95.25, 100.2)

Семинар: Проверьте следующее для Ent2Rep1Height



Суммарная статистика для Ent2Rep1Height

Число наблюдений = 30
Величина = 108.1
Стандартное отклонение = 5.569
Дисперсия = 31.02

95% доверительный интервал для величины: (106.1, 110.2)

Суммарная статистика для Ent2Rep2Height

Число наблюдений = 30

Величина = 101.4

Стандартная отклонение = 3.047

Дисперсия = 9.283

95% доверительный интервал для величины: (100.3, 102.5)

4.2 Испытание гипотезы / Испытание важности

- Испытание гипотез величин и дисперсий двух совокупностей.
- Сравнение величин и дисперсий двух образцов
- Нормальные совокупности

Процедуры для проведения Испытаний:

Состоят из:

- Гипотез – Нулевые и Альтернативные, оба
- Величина Образца/данных
- Статистика Испытания (функция величин образца)
- Предварительно определенные уровни значимости (Тип I норма ошибки: вероятность отказа нулевой гипотезы, когда она в действительности правильная)

$$\circ \alpha = 0.05, 0.01, 0.001 \text{ i.e. } 5\%, 1\%, 0.1\%$$

- Критическая область
- Правило по принятию решения
- Заключение

4.2.1 Испытание гипотезы равенства средств двух нормальных совокупностей

С учетом двух совокупностей:

$$N(\mu_1, \sigma_1^2)$$

$$N(\mu_2, \sigma_2^2)$$

Примеры:

- 1) С учетом того, что их дисперсии являются равными, Испытание для равенности величин
- 2) С учетом того, что их дисперсии являются неравными, Испытание для равенности величин
- 3) Однородность дисперсий

Пример 1: Испытание для равенства величин, когда их дисперсии являются равными

С учетом,

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

Проведите испытание для:

- Нулевой гипотезы:

$$H_0 : \mu_1 = \mu_2$$

По сравнению с

- Двухсторонними альтернативами:

$$H_1 : \mu_1 \neq \mu_2$$

Или

- Односторонняя альтернатива (правый хвост):

$$H_1 : \mu_1 > \mu_2$$

или

- Односторонняя альтернатива (левый хвост)

$$H_1 : \mu_1 < \mu_2$$

Данные Образца:

Взятый наугад образец от:

а) Совокупности-1:

$$X_1, X_2, X_3, \dots, X_{n_1}$$

б) совокупности-2

$$Y_1, Y_2, Y_3, \dots, Y_{n_2}$$

- Испытание статистики
-

$$t_{calc} = \frac{(\bar{x} - \bar{y})}{s\sqrt{(1/n_1) + (1/n_2)}}$$

где,

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$\nu = n_1 + n_2 - 2$$

Испытания Критической области:

- Отказаться от нулевой гипотезы при $100 \alpha \%$ если

1)

$$Absolute(t_{calc}) > t_{v;\alpha/2}$$

В пользу двухсторонней альтернативы.

2)

$$t_{calc} > t_{v;\alpha}$$

В пользу правосторонней альтернативы,

3)

$$t_{calc} < t_{v;\alpha}$$

В пользу левосторонней альтернативы.

- Обычно, программное обеспечение распечатает таким образом

$$Pr obability(| t_v | \geq t_{calc})$$

or

$$Pr obability(t_v \geq t_{calc})$$

or

$$Pr obability(t_v \leq t_{calc})$$

- 100(1- α)% границы доверительного интервала для разницы между величинами

$$Lower\ limit = \bar{x} - \bar{y} - t_{v;\alpha} s \sqrt{(1/n_1) + (1/n_2)}$$

$$Upper\ limit = \bar{x} - \bar{y} + t_{v;\alpha} s \sqrt{(1/n_1) + (1/n_2)}$$

$$(\bar{x} - \bar{y} - t_{v;\alpha} s \sqrt{(1/n_1) + (1/n_2)}, \bar{x} - \bar{y} + t_{v;\alpha} s \sqrt{(1/n_1) + (1/n_2)})$$

Пример: ООС 2003-04, Генотип 1 и 2 (повтор 1) образец n=30

- **Рассмотрите, являются ли величины высоты растений совокупностей двух Генотипов равными; взятый наугад образец размера n =30**
- **Проверьте предположение: являются ли дисперсии гомогенными?**

141 VHOMOGENEITY [PRINT=test,variance; GROUPS=Entry]
Rep1Height

Variances and degrees of freedom

Entry	Var_	d_f
1	18.88	29
2	31.02	29

Bartlett's test for homogeneity of variances

Chi-square 1.74 on 1 degrees of freedom: probability 0.187

Верхнее Испытание Барлетта является приблизительным испытанием для сравнения двух дисперсий. F-Испытание является точным испытанием, и представлен следующим образом.

Испытание по равенности величин высоты растений двух Генотипов по пшенице

```
144 TTEST [PRINT=summary,test,confidence,variance; METHOD=twosided;  
GROUPS=Entry; CIPROB=0.95;\n145 VMETHOD=automatic; NTIMES=4999; SEED=0] Y1=Rep1Height
```

Two-sample t-test

Variate: Rep1Height
Group factor: Entry

Test for equality of sample variances

Test statistic F = 1.64 on 29 and 29 d.f.

Probability (under null hypothesis of equal variances) = 0.19

Summary

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
1	30	102.5	18.88	4.345	0.7933
2	30	108.1	31.02	5.569	1.0168

Difference of means: -5.600
Standard error of difference: 1.290

95% confidence interval for difference in means: (-8.181, -3.019)

Test of null hypothesis that mean of Rep1Heights with Entry = 1 is equal to mean with Entry = 2

Test statistic t = -4.34 on 58 d.f.

Probability < 0.001

Или

```
186 TTEST [PRINT=summary,test,confidence,variance; METHOD=twosided; CIPROB=0.95;  
VMETHOD=automatic;\n187 NTIMES=4999; SEED=0] Y1=Ent1Rep1Heights; Y2=Ent2Rep1Height
```

Two-sample t-test

Variates: Ent1Rep1Height, Ent2Rep1Height.

Test for equality of sample variances

Test statistic F = 1.64 on 29 and 29 d.f.

Probability (under null hypothesis of equal variances) = 0.19

Summary

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
Ent1Rep1Height	30	102.5	18.88	4.345	0.7933
Ent2Rep1Height	30	108.1	31.02	5.569	1.0168

Difference of means: -5.600

Standard error of difference: 1.290

95% confidence interval for difference in means: (-8.181, -3.019)

Test of null hypothesis that mean of Ent1Rep1Height is equal to mean of Ent2Rep1Height

Test statistic t = -4.34 on 58 d.f.

Probability < 0.001

Решение:

Применение Меню Genstat

....>Stats...>Statistical tests....> One and two sample t-tests

Пример 2: Испытание на равенство величин, где их дисперсии являются неровными

С учетом,

$$\sigma_1^2 \neq \sigma_2^2$$

Проведите Испытание для:

- Нулевой гипотезы,

$$H_0 : \mu_1 = \mu_2$$

По сравнению с

- Двусторонней альтернативой:

$$H_1 : \mu_1 \neq \mu_2$$

Или

- Односторонней альтернативой (правый хвост):

$$H_1 : \mu_1 > \mu_2$$

или

- Односторонней альтернативой (левый хвост)

$$H_1 : \mu_1 < \mu_2$$

Данные Образца:

Взятый наугад образец от

б) Совокупности-1:

$$X_1, X_2, X_3, \dots, X_{n_1}$$

б) совокупности-2

$$y_1, y_2, y_3, \dots, y_{n_2}$$

Процедуры по проведению Испытаний:

- Статистика Испытания

$$t_{calc} = \frac{(\bar{x} - \bar{y})}{\sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}}$$

$$\nu = \frac{(s_1^2 / n_1 + s_2^2 / n_2)^2}{\frac{(s_1^2 / n_1)^2}{n_1 - 1} + \frac{(s_2^2 / n_2)^2}{n_2 - 1}}$$

(Аппроксимация Satterthwaite для степени свободы)

Критические области и Испытания:

- Отказ от нулевой гипотезы при $100\alpha\%$ если

2)

$$Absolute(t_{calc}) > t_{v;\alpha/2}$$

В пользу двусторонних альтернатив.

2)

$$t_{calc} > t_{v;\alpha}$$

В пользу правосторонних альтернатив,

3)

$$t_{calc} < t_{v;\alpha}$$

В пользу левосторонних альтернатив.

Probability($|t_v| \geq t_{calc}$)

or

Probability($t_v \geq t_{calc}$)

or

• *Probability*($t_v \leq t_{calc}$)

- 100(1- α)% границы доверительного интервала для разницы между величинами

$$\text{Lower limit} = \bar{x} - \bar{y} - t_{v;\alpha} \sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}$$

$$\text{Upper limit} = \bar{x} - \bar{y} + t_{v;\alpha} \sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}$$

$$(\bar{x} - \bar{y} - t_{v;\alpha} \sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)}, \bar{x} - \bar{y} + t_{v;\alpha} \sqrt{(s_1^2 / n_1) + (s_2^2 / n_2)})$$

Семинар: Данные по Ent1Rep2Height, и Ent2Rep2. Рассмотрите, происходили ли два образца из совокупностей с такой же величиной.

Решение : Примените Меню Genstat

....>Stats...>Statistical tests....> One and two sample t-tests –

Пример 2: Предположим, что дисперсии являются неравными.

Испытание на равенство величин высот в in Rep1 and Rep 2 для Entry 2.

```
188 TTEST [PRINT=summary,test,confidence,variance; METHOD=twosided; CIPROB=0.95;  
          VMETHOD=automatic;\n189 NTIMES=4999; SEED=0] Y1=Ent2Rep1Height; Y2=Ent2Rep2Height
```

Two-sample t-test

Variates: Ent2Rep1Height, Ent2Rep2Height.

Test for equality of sample variances

Test statistic $F = 3.34$ on 29 and 29 d.f.

Probability (under null hypothesis of equal variances) = 0.00

Note: strong evidence of unequal sample variances -
variances estimated separately for each group.

Summary

Sample	Size	Mean	Variance	Standard deviation	Standard error of mean
Ent2Rep1Height	30	108.1	31.02	5.569	1.0168
Ent2Rep2Height	30	101.4	9.28	3.047	0.5563

Difference of means: 6.733

Standard error of difference: 1.159

95% confidence interval for difference in means: (4.399, 9.068)

Test of null hypothesis that mean of Ent2Rep1Height is equal to mean of Ent2Rep2Height

Test statistic $t = 5.81$ on approximately 44.93 d.f.

Probability < 0.001

4.2.2 Испытание для равенности дисперсий

Пример 3: Гомогенность/равность дисперсии

Подготовьте испытание для

- Нулевой гипотезы

$$H_0 : \sigma_1^2 = \sigma_2^2$$

- Альтернативной гипотезы

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Статистика Испытания:

$$F_{calc} = s_1^2 / s_2^2$$

Где, число выше, чем знаменатель.

- Отклоните нулевые гипотезы при $100 \alpha \%$ в случае, если

$$F_{calc} > F_{n_1-1, n_2-1; \alpha}$$

где

$F_{n_1-1, n_2-1; \alpha}$ является табличное значение F-распределения на n_1-1, n_2-1 норма свободы при α выше нормы вероятности.

Пример:

141 VHOMOGENEITY [PRINT=test,variance; GROUPS=Entry] Rep1Heights

Variances and degrees of freedom

Entry	Var_	d_f
1	18.88	29
2	31.02	29

Bartlett's test for homogeneity of variances

Chi-square 1.74 on 1 degrees of freedom: probability 0.187

Проблема: Испытание гомогенности дисперсий данных Ent1 and Ent2 от Rep 2.

Расчеты и Испытание пропорции (ий)

5. Генеральная совокупность (π)

Представим n является числом семян данного генотипа пшеницы, находящийся в чашке Петри для изучения состояния покоя семян. Представим m пророс в течении приемлемого периода времени. Вычисление пропорции (вероятности) пророста и испытание роста данного генотипа составляет 80%.

Мы можем рассмотреть эту ситуацию следующим образом.

1. Начертите случайный образец размера n (семена), из которого m , был определен как желательный Тип (то есть проросший). Статистическое распределение проросших семян m может быть приближено двучленным распределением с вероятностью прорастания π (). В статистической терминологии m является числом 'успехов', наблюдаемых от двучленного распределения с n как 'Число испытаний' и π вероятность успеха. Таким образом,

Расчет:

Вычисление: $\hat{\pi} = \frac{m}{n}$,

Вычисление $se(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$

Испытание гипотезы:

Испытание $H_0: \pi = \pi_0$

$H_1: \pi \neq \pi_0$

2. Рассчитайте Статистику испытания

$$Z = \left| \frac{\hat{\pi} - \pi_0}{\left(\frac{\hat{\pi}(1 - \hat{\pi})}{n} \right)^{1/2}} \right|$$

3. Величина таблицы

$$Z_{tab} = Z_{\alpha/2}$$

4. Если $Z > Z_{tab}$ отказаться от H_0 .

Пример: Представим $n = 50$, $m=41$.

191 BNTEST [PRINT=summary,test,confidence; METHOD=twosided; TEST=exact; NULL=.8;
CIPROB=0.95] R1=41; N1=50

One-sample binomial test

Summary

Sample Size	Successes	Proportion
50	41	0.82

Test of null hypothesis that proportion is equal to 0.8000

Exact probability = 0.860

95% confidence interval for proportion: (0.6856, 0.9142)

Семинар: Представим $n = 50$, $m=41$ для Генотипа А и $n= 75$, $m = 53$ для Генотипа В. Провести Испытание, имеют ли два Генотипа одинаковые нормы пророста

194 BNTEST [PRINT=summary,test,confidence; METHOD=twosided; TEST=normal; CIPROB=0.95] R1=41; N1=50; R2=53; n2=75

Two-sample binomial test

Summary

Sample	Size	Successes	Proportion
1	50	41	0.820
2	75	53	0.707

Approx s.e. of difference between proportions: 0.0756

Test of null hypothesis that proportion 1 is equal to proportion 2

Normal Approximation = 1.437
Probability = 0.151

95% confidence interval for difference between proportions: (-0.03485, 0.2615)

Решение:

Примените Меню Genstat

....>Stats...>Statistical tests....> One and two sample t-tests

5. Корреляция и Регресс

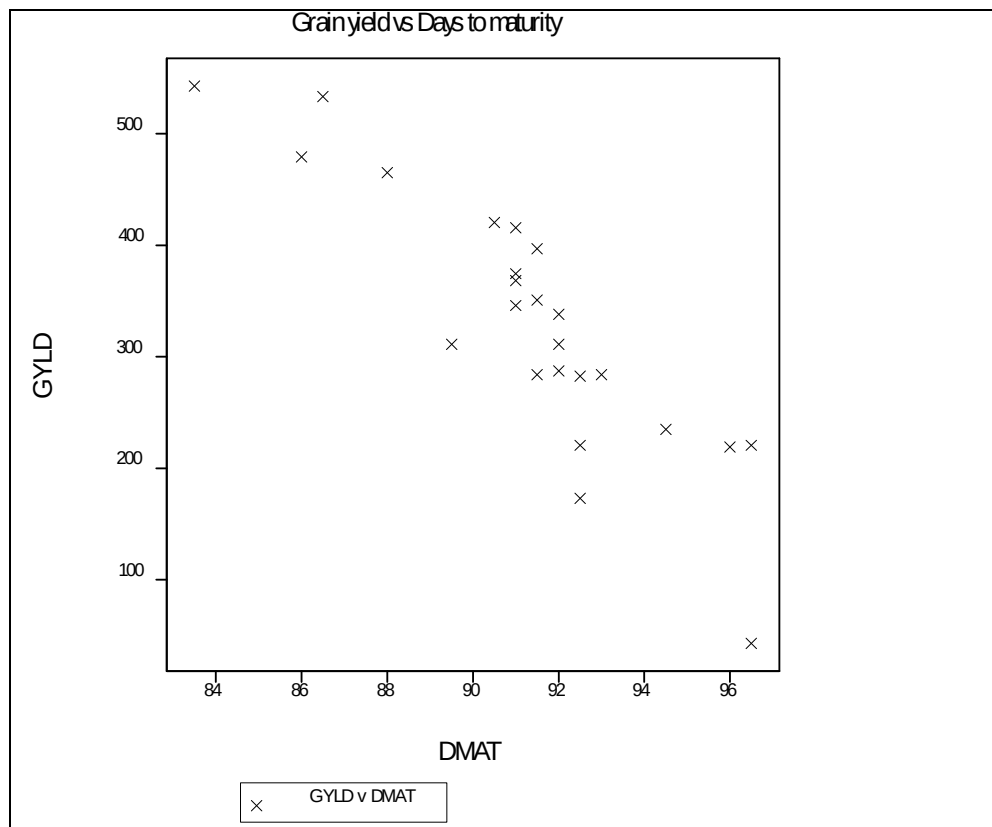
Часто, больше чем одна переменная, или черты, наблюдаются относительно каждой единицы а совокупности, или образца. Мы рассмотрим здесь ситуацию, когда две переменные X и Y , будут зарегистрированы на каждой из единиц N совокупности.

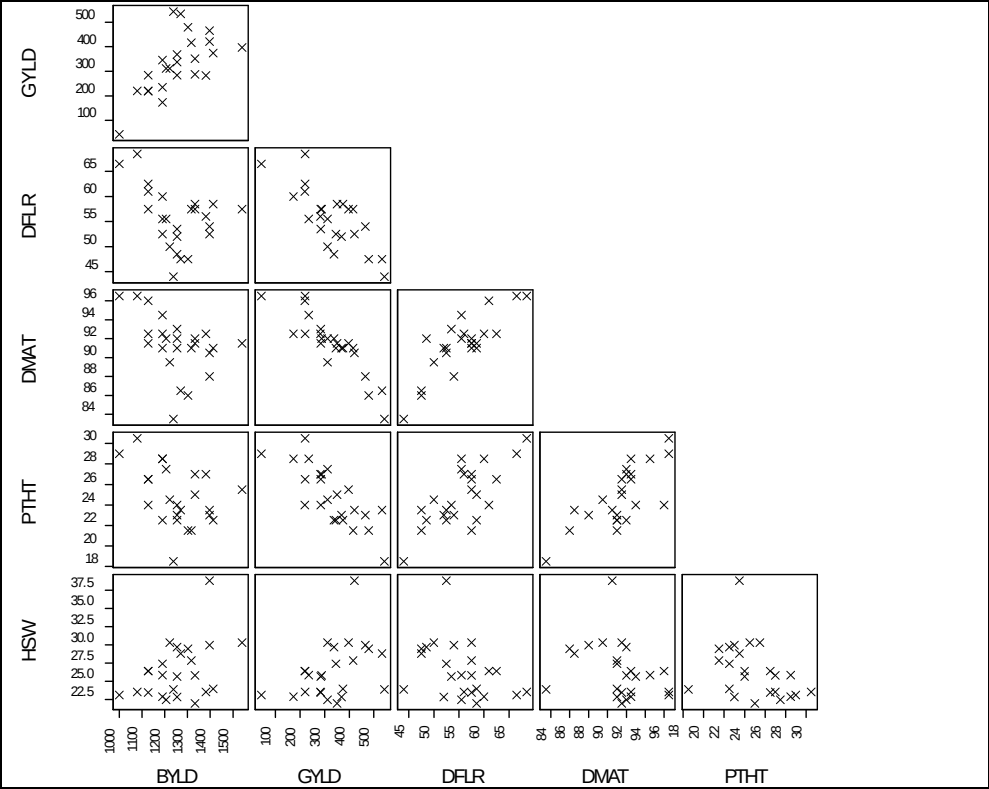
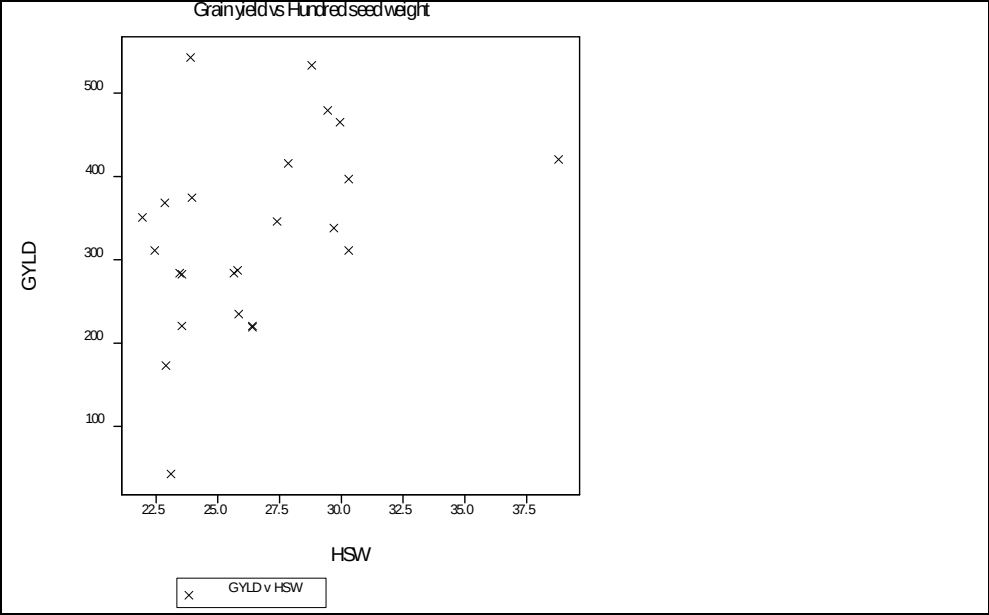
- Эти две переменные могут быть связаны, или показать некоторую степень зависимости, например, урожай зерна может быть связан со днями до цветения; урожай молока во второй лактации может быть связан с урожаем молока в первой лактации; и т.д.
- Помимо различных параметров совокупности для переменных, X и Y индивидуально являются очень полезными параметрами для совариаций между X и Y .
- Если мы имеем

Units	U_1	U_2	U_3	.	.	U_N
Y	Y_1	Y_2	Y_3	.	.	Y_N
X	X_1	X_2	X_3	.	.	X_N

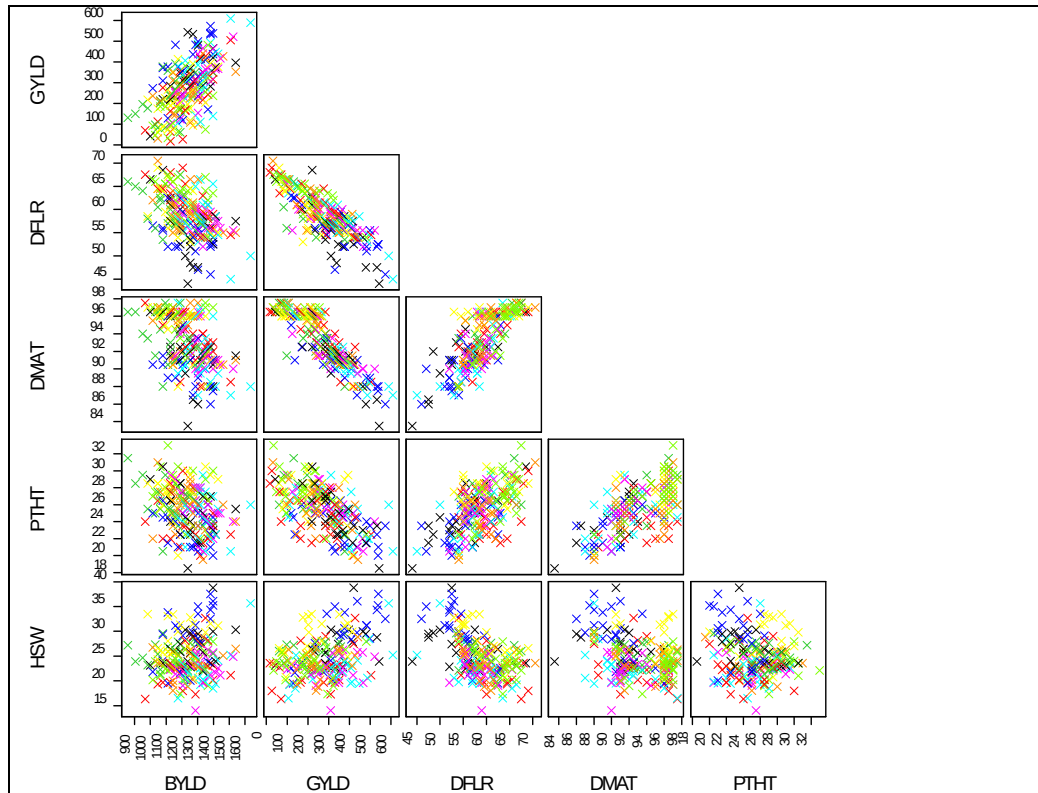
Пример: Рассмотрите данные по урожаю, и характерные черты генотипов нута (24), рассчитанные в 9 окружающих условиях. Смотрите следующие два графика.

Окружающая среда =1





Все виды Окружающей среды (1- 9):



BYLD vs DMAT или DMAT vs BYLD – плохая ассоциация

GYLD vs DMAT или DMAT vs GYLD– более сильная ассоциация;
- также негативная ассоциация(?)

GYLD vs HSW или HSW vs GYLD - сильная (?) позитивная
ассоциация

5.1 Мера ассоциации:

Для индивидуальных переменных X и Y , мы знаем их как

- Величины
- Дисперсии
- Нужны ли Совариации/ содисперсия?

В таком случае, содисперсия между X и Y указана σ_{xy}

$$\text{Cov}(X, Y) = \sigma_{xy}$$

$$= \sum_{i=1}^{i=n} (X_i - \bar{X}) (Y_i - \bar{Y}) / (n - 1)$$

- Параметры дисперсии и содисперсии могут быть использованы для описания функциональных форм отношений между X и Y переменными, особенно, когда форма является линейной.
- Можно получить меру степени, для которой форма является линейной, и
- Уравнение к прямой линии, подходящей лучше всего к графику рассеяния (X, Y) в некотором смысле.

5.2 Корреляция

- Корреляция является мерой степени, в отношении, которого соотношение между X и Y является линейной.
- Измеряется как стандартная содисперсия
- Обозначена и представлена как

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

- ρ 'rho' lies in $[-1, 1]$

Пример: Получите соотношения между каждой парой шести характеров
и проведите испытание значимости.

Окружающая среда =1

```
180 FCORRELATION [PRINT=correlations,test; METHOD=twosided]
BYLD,GYLD,HSW,DFLR,DMAT,PTHT
```

Correlations

BYLD	1	-					
GYLD	2	0.6284	-				
HSW	3	0.4229	0.4196	-			
DFLR	4	-0.3731	-0.7681	-0.3837	-		
DMAT	5	-0.4385	-0.8801	-0.2931	0.8247	-	
PTHT	6	-0.4059	-0.7878	-0.3560	0.7458	0.7455	-
		1	2	3	4	5	6

Number of observations: 24

Two-sided test of correlations different from zero

BYLD	1	-					
GYLD	2	0.0010	-				
HSW	3	0.0395	0.0413	-			
DFLR	4	0.0725	<0.001	0.0642	-		
DMAT	5	0.0321	<0.001	0.1645	<0.001	-	
PTHT	6	0.0491	<0.001	0.0878	<0.001	<0.001	-
		1	2	3	4	5	6
PTHT	6	-					
		6					

Окружающая среда = 2

```
184 FCORRELATION [PRINT=correlations,test; METHOD=twosided]
BYLD, GYLD, HSW, DFLR, DMAT, PTHT
```

Correlations

BYLD	1	-						
GYLD	2	0.7256	-					
HSW	3	0.3664	0.3343	-				
DFLR	4	-0.6984	-0.8732	-0.6162	-			
DMAT	5	-0.6671	-0.8330	-0.3213	0.7987	-		
PTHT	6	-0.2040	-0.5544	0.0177	0.3545	0.5151	-	
		1	2	3	4	5	6	

Number of observations: 24

Two-sided test of correlations different from zero

BYLD	1	-						
GYLD	2	<0.001	-					
HSW	3	0.0782	0.1103	-				
DFLR	4	<0.001	<0.001	0.0013	-			
DMAT	5	<0.001	<0.001	0.1258	<0.001	-		
PTHT	6	0.3390	0.0049	0.9346	0.0892	0.0100	-	
		1	2	3	4	5	6	
PTHT	6	-						
		6						

Solution:

Use Genstat Menu> Stats> Summary Statistics .>Correlations.

5.3 Линейная Регрессия

Установив, что эти две переменные X и Y ассоциированные, или особенно линейно связаны, может возникнуть вопрос: каковы функциональные отношения между ними?

- Линейный регресс (или регресс в целом) является функциональным отношением ожидаемых величин Y , данного X с точки зрения величин X .
- Y : зависимая переменная
- X : независимый или регрессор

Предостережение: Прежде, чем начать поиск функционального отношения, мы должны знать причину и переменные воздействия. Предположите, X является каузальной величиной, и Y переменной воздействия.

$X \text{-----} > Y$

то есть. Ожидаемая (Y для данного X) = функция (X)

При линейной регрессии (простая регрессия)

$$Y = \alpha + \beta X$$

- Вычислите изменение в Y для единицы изменения в X

$$Change = [\alpha + \beta (X + 1)] - (\alpha + \beta X)$$

$$= \beta$$

- Откос Y на X
- Коэффициент регрессии Y на X

- Вычислите Y при $X=0$

$$Y = \alpha$$

- Пересечение

Также

- Применяя 'наилучшим образом подходящую' линию
- Квадрат корреляций может быть интерпретирован как

$$\rho^2 = \frac{\text{Variation}(\alpha + \beta X)}{\text{Variation}(Y)}$$

- Пропорция вариации при Y объясняется линейной регрессией Y при X.
- Мера аппроксимации является процентом дисперсии, расцененная для

- $\bar{R}^2 \% = 100(1 - \frac{\text{residual mean square}}{\text{total mean square}})$

- откорректированная R^2 .

- **Вычисление линии регрессии**

Пример: Данные по нуту из Окружающей среды- 1.

- GYLD on DFL
- GYLD on HSW
- GYLD on DFL

```
186 Rest BYLD,GYLD,HSW,DFLR,DMAT,PTHT; Env1==1
187 "Simple Linear Regression"
188 MODEL GYLD
189 TERMS DFLR
190 FIT [PRINT=model,summary,estimates; CONSTANT=estimate;
FPROB=yes; TPROB=yes] DFLR
```

Regression analysis

Response variate: GYLD
Fitted terms: Constant, DFLR

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	1	184571.	184571.	31.66	<.001
Residual	22	128269.	5830.		
Total	23	312840.	13602.		

Percentage variance accounted for 57.1
Standard error of observations is estimated to be 76.4.

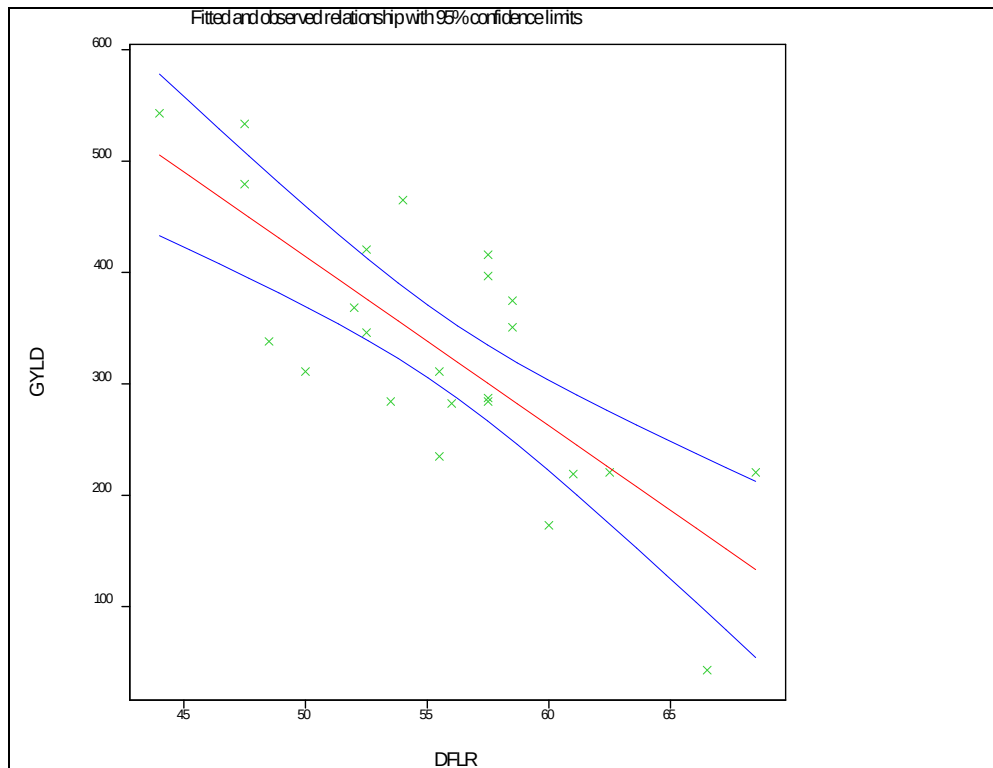
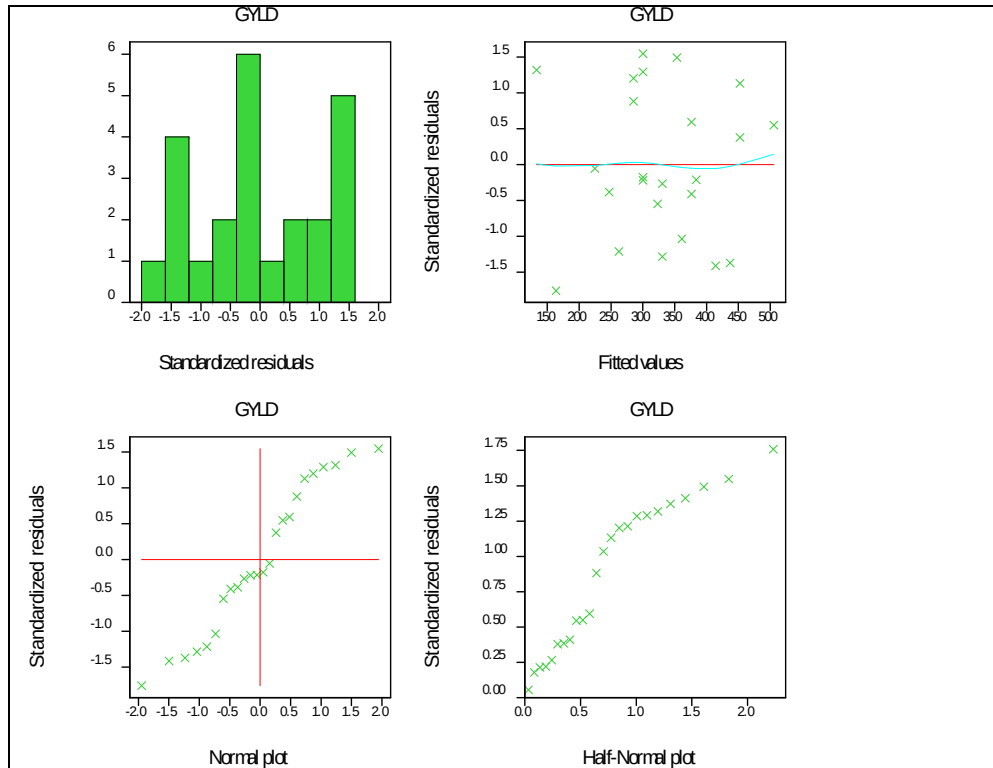
Message: the following units have high leverage.

Unit	Response	Leverage
1	220.6	0.250
9	42.9	0.190
19	542.9	0.210

Estimates of parameters

Parameter	estimate	s.e.	t(22)	t pr.
Constant	1174.	151.	7.78	<.001
DFLR	-15.19	2.70	-5.63	<.001

```
191 RCHECK [RMETHOD=deviance] residual; composite
192 RGRAPH [CIPLLOT=yes]
```



- GYLD on HSW
-

```

193 "Simple Linear Regression"
194 MODEL GYLD
195 TERMS HSW
196 FIT [PRINT=model,summary,estimates; CONSTANT=estimate;
FPROB=yes; TPROB=yes] HSW

```

Regression analysis

Response variate: GYLD
Fitted terms: Constant, HSW

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	1	55071.	55071.	4.70	0.041
Residual	22	257769.	11717.		
Total	23	312840.	13602.		

Percentage variance accounted for 13.9
Standard error of observations is estimated to be 108.

Message: the following units have large standardized residuals.

Unit	Response	Residual
9	43.	-2.34
19	543.	2.35

Message: the following units have high leverage.

Unit	Response	Leverage
2	421.	0.498

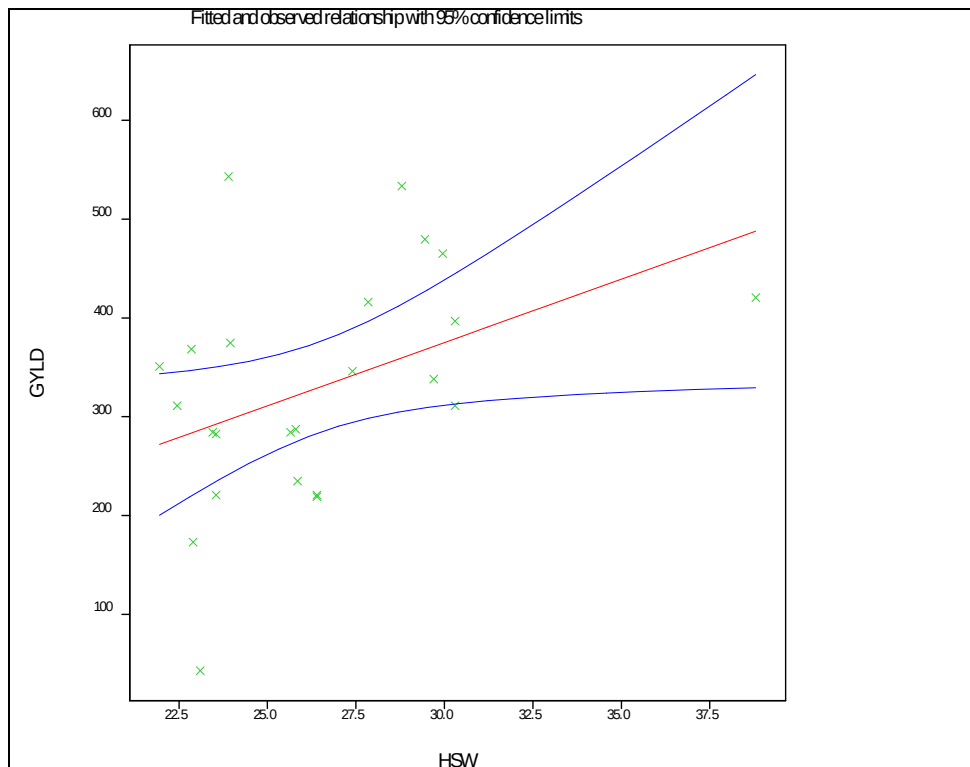
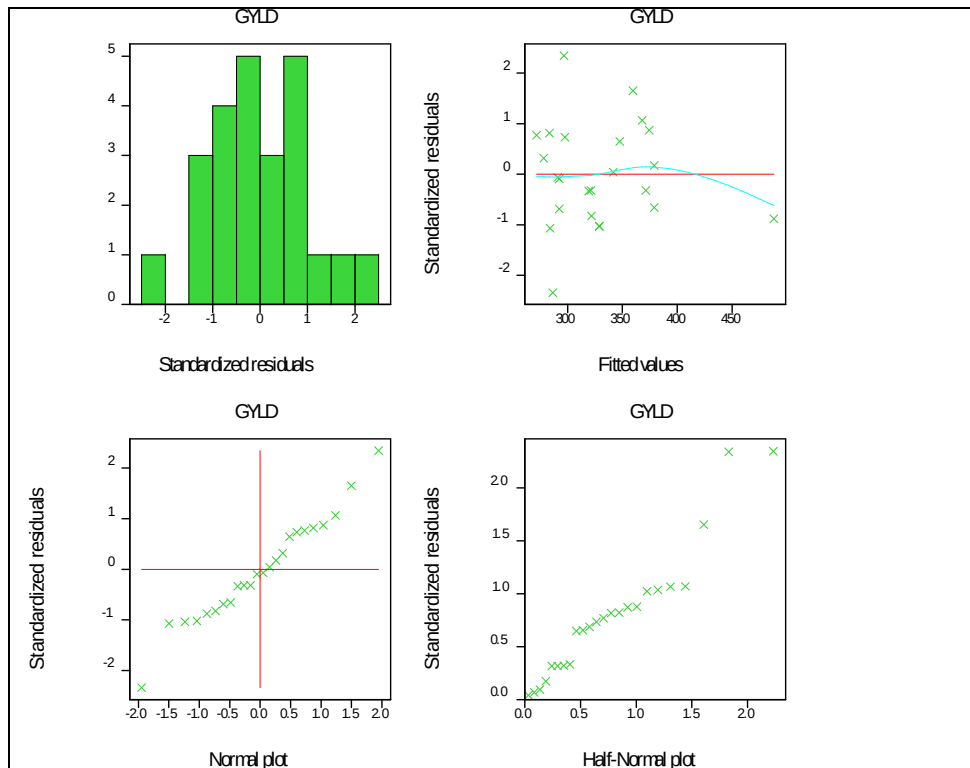
Estimates of parameters

Parameter	estimate	s.e.	t(22)	t pr.
Constant	-9.	158.	-0.06	0.953
HSW	12.81	5.91	2.17	0.041

```

197 RCHECK [RMETHOD=deviance] residual; composite
198 RGRAPH [CIPLLOT=yes]

```



Семинар. Получите линии регрессии GYLD on PTHT.

Solution: Use Genstat Menu>Stats....>Regression Analysis.....>Linear

models....>Simple regression

5.4 Множественная регрессия

- В предыдущем разделе, мы рассматривали только один объяснительный (или независимый) переменный для выражения зависимую переменную, используя линейные формы.
- Однако, на практике, зависимая переменная может быть ассоциирована с более чем одной независимой переменной.
- Например
 - Урожай зерна может быть ассоциирован с генетическими характеристиками, типа дни до цветения, и дни до зрелости и управления средств производства, типа объем удобрений и числа орошения;
 - При опытах с животными, скажем, сохранение качества образцов молока, может быть ассоциировано с, и может, зависит от различных переменных (напр. Сокращение времени метилового голубого, и логарифмов Стандартного чашечного метода подсчёта) и.т.д.
 - Весь ягненок может быть ассоциирован с возрастом и весом овцематки.
- Соотношение между зависимой переменной (Y) и различными независимыми переменными ($X_1, \dots, X_p$) может быть представлено линейным модулем в переменных p следующим образом

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p$$

где

- α пересечение, и
- β_i ($i=1,2, \dots, p$) являются тангенсом угла наклона, или (частично)

Коэффициентом регрессии Y на X .

Если у нас будут наблюдаемые величины ($Y, X_1, \dots, X_p$) на экспериментальных и образцовых единицах n , тогда мы можем вычислить параметры пересечения и тангенсы угла наклона (Коэффициенты).

- Используя метод вычисления для линейной модели, скажем, относящийся к методу наименьших квадратов, можно получить
 - Вычисление пересечений, α ; его Стандартные ошибки
 - Вычисление Коэффициентов регрессии β_i и их Стандартные ошибки
 - Остаточный средний квадрат, и его степени свободы

- Мера степени согласия линейной модели.
- Одна мера является квадратом из многократной корреляции (также, известный как коэффициент определения). Эта мера представлена как:

$$R^2 \% = 100(1 - \frac{\textit{residual sum of square}}{\textit{total sum of square}})$$

- Другой мерой является процент дисперсии, рассчитанный для (также, известный, как отрегулированный R^2). Эта мера представлена как:

$$\bar{R}^2 \% = 100(1 - \frac{\textit{residual mean square}}{\textit{total mean square}})$$

Пример

- GYLD по компонентам урожая: DFLR, DMAT, PHT, HSW

```
211 "Multiple Linear Regression"
212 MODEL GYLD
213 TERMS [FACT=9] DFLR,PTHT,DMAT,HSW
214 FIT [PRINT=model,summary,estimates,accumulated; CONSTANT=estimate;
FPROB=yes; TPROB=yes;\
215 FACT=9] DFLR,PTHT,DMAT,HSW
```

Regression analysis

Response variate: GYLD

Fitted terms: Constant, DFLR, PTHT, DMAT, HSW

Summary of analysis

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	4	259896.	64974.	23.32	<.001
Residual	19	52944.	2787.		
Total	23	312840.	13602.		

Percentage variance accounted for 79.5

Standard error of observations is estimated to be 52.8.

Message: the following units have large standardized residuals.

Unit	Response	Residual
9	42.9	-2.23

Message: the following units have high leverage.

Unit	Response	Leverage
2	420.6	0.53
19	542.9	0.44

Estimates of parameters

Parameter	estimate	s.e.	t(19)	t pr.
Constant	2746.	460.	5.97	<.001
DFLR	0.78	3.60	0.22	0.831
PTHT	-10.72	6.20	-1.73	0.100
DMAT	-25.28	6.70	-3.77	0.001
HSW	4.37	3.15	1.39	0.182

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ DFLR	1	184571.	184571.	66.24	<.001
+ PTHT	1	32568.	32568.	11.69	0.003
+ DMAT	1	37399.	37399.	13.42	0.002
+ HSW	1	5358.	5358.	1.92	0.182
Residual	19	52944.	2787.		

Total	23	312840.	13602.
216 RCHECK [RMETHOD=deviance] residual; composite			

```

217 "Multiple Linear Regression"
218 MODEL GYLD
219 TERMS [FACT=9] DFLR,PTHT,HSW,DMAT
220 FIT [PRINT=model,summary,estimates,accumulated; CONSTANT=estimate;
FPROB=yes; TPROB=yes;\
221 FACT=9] DFLR,PTHT,HSW,DMAT

```

Regression analysis

Response variate: GYLD

Fitted terms: Constant, DFLR, PTHT, HSW, DMAT

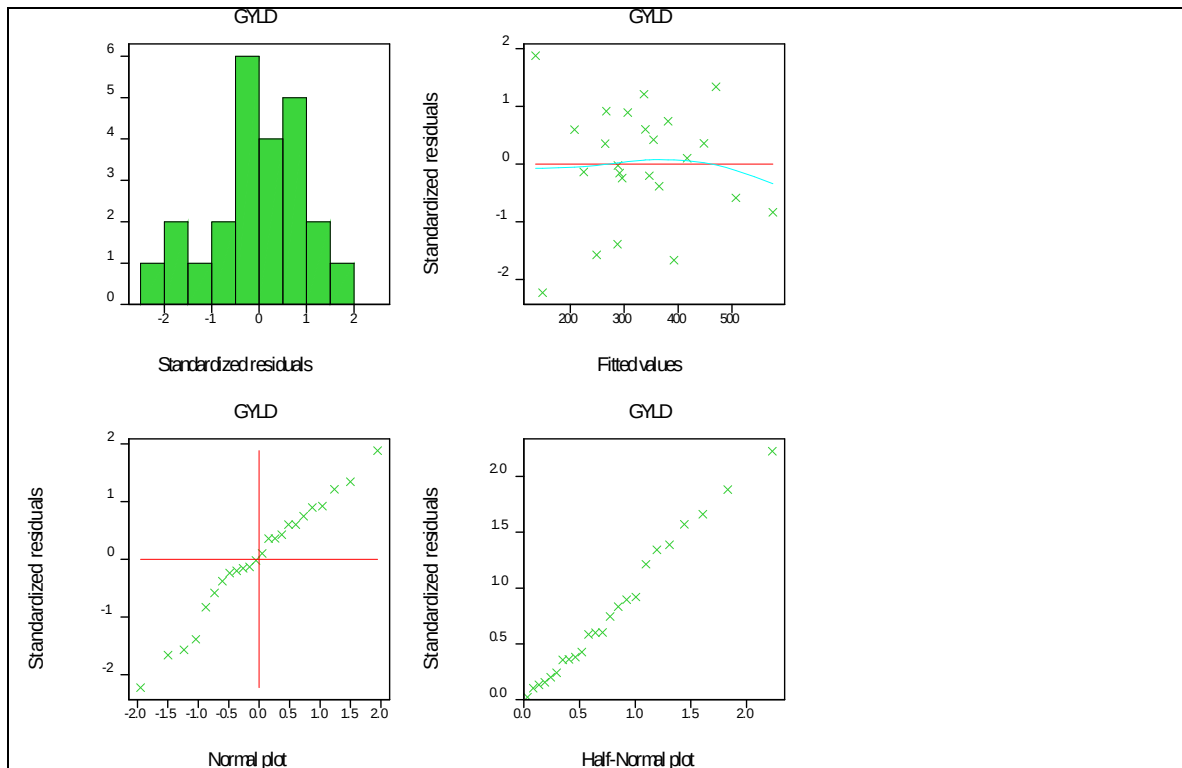
Summary of analysis.....

Estimates of parameters

Parameter	estimate	s.e.	t(19)	t pr.
Constant	2746.	460.	5.97	<.001
DFLR	0.78	3.60	0.22	0.831
PTHT	-10.72	6.20	-1.73	0.100
HSW	4.37	3.15	1.39	0.182
DMAT	-25.28	6.70	-3.77	0.001

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F pr.
+ DFLR	1	184571.	184571.	66.24	<.001
+ PTHT	1	32568.	32568.	11.69	0.003
+ HSW	1	3079.	3079.	1.10	0.306
+ DMAT	1	39679.	39679.	14.24	0.001
Residual	19	52944.	2787.		
Total	23	312840.	13602.		



Семинар. BYLD по компонентам урожая DFLR, DMAT, PHT, HSW

Solution: Use Genstat Menu>Stats....>Regression Analysis.....>Linear models....>Multiple Linear Rgression

Усовершенствованный Статистический Анализ Испытаний Сорта в различных местах

Часть-II. Анализ Стабильности

3 Эксплуатация Данных Г x Е: Анализ Стабильности

- Производительность генотипов меняется за счет экологических давлений, или стрессов (из-за популяционной разнородности, или промежуточного преобразования (буферизации) популяции, и изменений в генетическом составе, которые происходят в течении поколений), и из-за различия в их способности приспособиться к факторам стресса (краткосрочную акклиматизацию).
- В литературе были обсуждены много статистических моделей по изучению генотипной адаптации, основанные на фенотипичную производительность.
- Бит и Мангомери (1981г.) обсуждали следующие три концепции: *Стабильность*, *адаптируемость*, и *предсказуемость*.

Фенотипичная *стабильность* относится к способности генотипа в поддержании близкого постоянного фенотипа для изучения характера переменной окружающей среды. Такой генотип был бы расценен, как владеющий способностью *широкой адаптации*. Но определенные генотипы могут также, очевидно, показать превосходящую производительность в специфических типах окружающей среды, указывающей, что широкая адаптация неизбежно вовлекает жертву производительности в определенной окружающей среде. Таким образом, стратегии по улучшению растения для широкой адаптации (ограничивая взаимодействие Генотипа и Окружающей среды-далее G x E), и определенной адаптации (подчеркивая благоприятное

взаимодействие) находятся в прямом конфликте.

Предсказуемость относится к той степени, когда реакция является систематической.

Реактивность – способность генотипа реагировать в специфической манере на общее изменение в экологическом потенциале.

Чувствительность (также *стабильность*) относится к степени непредсказуемого изменения в реакции. Некоторые исследователи связывают стабильность с изменчивостью производительности в течение времени (временное изменение) в одном месте, в то время, как адаптируемость к изменчивости в производительности, происходит во всех местах (пространственное изменение).

Стабильность, используя обычно практикуемые статистические меры

- Два подхода, обсужденные Лин и другими. (1986г.).

3.1 Параметрический Подход

Статистические данные стабильности получены (вычислены) для каждого генотипа от двухсторонних таблиц данных генотипа, и окружающей среды. Эти статистические данные основаны на любой из следующих трех типов понятий стабильности.

Генотип, как считается, будет устойчивым, если
- среди различной окружающей среды, он является маленьким (Тип I стабильности),

- его реакция в окружающей среде параллельна к средним реакциям всех испытываемых генотипов (Тип II стабильности)

- остаточный средний квадрат от модели регресса на индексе окружающей среды, является маленьким (Тип III стабильности).

Чтобы перечислить различную статистику, мы будем использовать следующие примечания. Позволим y_{ij} обозначать среднее значение генотипа i в среде j ($i=1,2, \dots, p, j=1,2,\dots,q$).

Genotype	Environments						mean
	1	2	3	.	.	q	
1	y_{11}	y_{12}	y_{13}	.	.	y_{1q}	$\dot{y}_{1.}$
2	y_{21}	y_{22}	y_{23}	.	.	y_{2q}	$\dot{y}_{2.}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
p	y_{p1}	y_{p2}	y_{p3}	.	.	y_{pq}	$\dot{y}_{p.}$
mean	$\dot{y}_{.1}$	$\dot{y}_{.2}$	.	.	.	$\dot{y}_{.q}$	

Позволим

$$\dot{y}_{i.} = \sum_j y_{ij} / q; \quad \dot{y}_{.j} = \sum_i y_{ij} / p, \quad \dot{y}_{..} = \sum_i \sum_j y_{ij} / (pq)$$

представлять соответственно, среднюю величину генотипа - i , на окружающей среде- j , и общую среднюю величину.

Взаимодействие:

$$I_{ij} = y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..}$$

Девять статистических данных, и еще одна в настоящее время применяемая статистика, кратко описаны следующим образом:

1. Изменчивость генотипа в различной окружающей среде

$$S_i^2 = \sum_{j=1}^q (y_{ij} - \dot{y}_{i.})^2 / (q-1),$$

2. Коэффициент вариации

$$CV_i = S_i / \dot{y}_{i.}.$$

Франсис и Канненберг (1978г.) использовали традиционную КВ-CV (коэффициент вариации) % для каждого генотипа, в качестве меры стабильности.

3. Компонент средней изменчивости Плейстада и Петерсона (1959г.) для пары - нормального взаимодействия G x E ($\dot{\theta}_i$)

$$\dot{\theta}_i = (p \sum_{j=1}^q (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2 + \sum \sum (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2) / (2(p-1)(q-1)).$$

Среднее значение вычисленных компонентов изменчивости взаимодействия G x E для всех пар генотипов, которые включают генотип- I, является мерой стабильности генотипа-і.

4. Компонент расхождения Плейстеда (1960г.) для взаимодействия G x E ($\dot{\theta}_{(i)}$)

$$\theta_{(i)} = (-p \sum_{j=1}^q (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2 / (p-1) + \sum \sum (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2) / ((p-2)(q-1)).$$

Один генотип-і исключен из общего списка данных, и расхождение взаимодействий G x E из данного комплекта, является индексом стабильности для генотипа-і.

5. Экологическая валентность Рика (1962г.) (w_i^2)

$$w_i^2 = \sum_{j=1}^q (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2.$$

Данное взаимодействие G x E воздействует на генотип-і, согласованное и суммированное по всем видам окружающей среды, и является мерой стабильности для генотипа і.

6. Изменение стабильности по методу Шукла (1972г.) (σ_i^2)

$$\sigma_i^2 = (p \sum (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2 - \sum \sum (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2) / (p-1) / ((p-2)(q-1)).$$

На основе остатков в двусторонней классификации, мерой по стабильности является различие генотипа по всем типам окружающей среды.

7. Коэффициенты регрессии Финлей и Вилкинсона (1963г.) (b_i)

$$b_i = \sum_j (y_{ij} - \dot{y}_{i.})(\dot{y}_{.j} - \dot{y}_{..}) / \sum_j (\dot{y}_{.j} - \dot{y}_{..})^2$$

Наблюдаемые величины возвращены на основе индикаторов окружающей среды, и на основе общих срединных величин. Коэффициент регресса каждого генотипа, был взят как его параметр стабильности.

8. Коэффициент регрессии Перкинса и Джинкса (1968г.) (β_i)

$$\beta_i = \sum_j (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..}) (\dot{y}_{.j} - \dot{y}_{..}) / \sum_j (\dot{y}_{.j} - \dot{y}_{..})^2.$$

Подобно (7), кроме, наблюдаемых значений, данные были отрегулированы с учетом воздействия окружающей среды, с целью расчета коэффициента регрессии ($\beta_i = b_i - 1$).

9. Параметр отклонения Эберхарта и Рассела (1966г.) (δ_i^2)

$$\delta_i^2 = (\sum_j (y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..})^2 - \beta_i^2 \sum_j (\dot{y}_{.j} - \dot{y}_{..})^2) / (q - 2)$$

Это - остаточный средний квадрат (MS) отклонения от регресса, определенного в (7), или (8), что является мерой стабильности.

10. Различие генотипов по всем видам окружающей среды, в отношении коэффициента урожая, на средства окружающей среды

Яу (1972г.) представил другую статистику, обозначенную здесь как ξ_i в качестве различия по всем видам окружающей среды, в отношении коэффициента урожая к средним величинам, при соответствующей окружающей среде

$$\xi_i = \sum_{j=1}^q (r_{ij} - \dot{r}_{i.})^2 / (q-1),$$

где $\dot{r}_{i.} = \sum_{j=1}^q r_{ij} / q$, and $r_{ij} = y_{ij} / \dot{y}_{.j}$

Эта статистическая величина, как можно увидеть, измеряет Тип-II стабильности.

Группировка индексов и их схож ест ь

Первые девять статистических данных, основаны на отклонение от среднего воздействия генотипа (ВГ) = $y_{ij} - \dot{y}_{i.}$ или на условия взаимодействия GxE $I_{ij} = y_{ij} - \dot{y}_{i.} - \dot{y}_{.j} + \dot{y}_{..}$ (в форме их сумм квадратов СК, коэффициент регресса или отклонение от регресса), и были классифицированы в четыре группы (A, B, C, D):

Группа A: ВГ, СК: (S_i^2, CV_i) -Тип-I стабильности

Группа B: GE, СК: $(\theta_i, \theta_{(i)}, w_i^2, \sigma_i^2)$ -тип-II

стабильности

Группа С: ВГ или коэффициент регресса GE:

(b_i, β_i) -тип-II стабильности

Группа D: ВГ или отклонение регресса GE

(δ_i^2) Тип -III стабильности

Лин и другие (1986) отмечали:

(i) Так как переменная (регистрированная(y)) $\square$ переменная(y)/(среднее значение(y))² = (CV(y))², поэтому две статистические группы А являются эквивалентными, кроме преобразования данных.

(ii) Четыре статистических данных в группе В эквивалентны для занимающих место генотипам. σ_i^2 , также объективной оценке различия генотипа i. Приблизительный тест на однородность σ_i^2 был предоставлен Шуклой (1972b).

(iii) Начиная с $\beta_i = b_i - 1$, два статистических данных в группе С являются эквивалентными. Точно так же, статистические данные группы D, являются эквивалентными.

(iv) Когда изменчивость в реакции может быть удовлетворительно выражена моделью регресса, коэффициент регресса (группы С) может служить в качестве параметра стабильности, и может быть предпочтителен мерам по изменчивости (группа В), так как, они (меры Группы С) предоставляют информацию о виде реакции, наряду с ее изменением.

Индексы стабильности и меры по стабильности

- Статистика в Группе А измеряет стабильность Типа-I; принадлежащие Группе В и индексу Яу статистики измеряют Тип-II, в то время как таковые из Группы D, измеряют Тип III.
- Статистические данные Группы С имеют измерения по стабильности Типа I, или Типа II, в зависимости от природы устойчивого генотипа.
- Если устойчивый генотип определен при наличии $b_i=1$ ($\beta_i=0$), то подразумевается Тип II ; но если генотипы определены как $b_i=0$ ($\beta_i=-1$), то подразумевается Тип I.

Стабильность Типа-I:

- Тип-I указывает наличие гомеостаза, биологическое понятие (Becker 1981г.).
- Этот тип отличается от агрономического понятия стабильности, представленный Типом II.
- Хотя Тип I является теоретически зрелым, но селекционер часто не использует этот тип, поскольку селекционер хотел бы выбрать культурные сорта растения с высоким урожаем, помимо наличия стабильности Типа I.
- Стабильность Типа-I связана с относительно плохим урожаем в окружающей среде, которая сильно уступает другим культурным сортам растения. Кроме того, b_i урожаи положительно коррелированы (Финли и Вилкинсон, 1963г.). Хотя широкая (повсеместная) адаптация может быть желательной, но ее трудно достичь практически. Все же более удобный путь состоял бы в том, чтобы вывести культурные сорта растения с (определенной) адаптацией к различной окружающей среде, чтобы увеличить производство.

- Так как стабильность Типа-I не зависит от наличия других генотипов, у нее имеется широкая основа, для того чтобы сделать вывод. Однако, она не предоставляет информацию о структуре реакции.

Стабильность Типа-II:

- Выводы, сделанные по измерениям стабильности Типа II, являются относительными по генотипам, включенных в тест.
- Например, генотип А может быть оценен устойчивым, а В непостоянным, если генотип А схож с большинством генотипов в комплекте более близко, чем это делает генотип В.
- В другом наборе генотипов, если В напоминает большинство генотипов ближе, чем, это делает генотип А, то генотип В устойчив и непостоянен.
- Эта мера полезна для сравнения определенного набора генотипов, и таким образом, не имеет широкой основы для того, чтобы сделать вывод по общей оценки.

Стабильность Типа III

- • Эберхарт и Рассел (1966) предложили другую меру стабильности, основанную на среднем квадрате отклонения σ_i^2).
- Таким образом, имеются две меры стабильности, (b_i, σ_i^2) для единого характера.
- Использование σ_i^2 было защищено Бризом (1969г.), поскольку он полагал, что 'стабильность' должна отнести непредсказуемое изменение (неисправности) как реакцию на окружающую среду.
- Изменчивость реакции на окружающую среду может быть разделена на предсказуемое изменение (представленные средними квадратами регресса) и непредсказуемое изменение (измеренный, отклонением MS, σ_i^2).
- Этот аргумент является здравым, но мера стабильности отклонением MS является несоответствующей, поскольку она представляет совершенство пригодности модели, которую мы выбираем.
- Чтобы поддержать аргумент практически, нужно измерить независимые переменные, объясняя окружающую среду и модель предсказания, которые будут созданы вместе с ними.
- Экологический индекс (основанный на средствах всех генотипов) не может обеспечить независимое измерение потенциала окружающей среды.
- Таким образом нижнее значение различия процента составляло высоко σ_i^2 или гетерогенная MS просто указывает, что модель регресса не адекватна для стабильности, и должны быть исследованы некоторые другие методы.
- Тип III будет полезным только тогда, когда модель предсказания будет рассмотрена, и основана на независимо

взвешенных экологических переменных.

3.2 Непараметрический Подход

В продолжение к нашему предыдущему обсуждению, мы теперь полагаем, что непараметрический подход изучает генотип x и взаимодействие окружающей среды. Этот подход используется для поиска образца в генотипах, и или в окружающей среде.

Статистические данные при параметрическом подходе выражают многомерную информацию (реакции на многократную окружающую среду, считаются многомерными) с точки зрения одномерного, и измеряют только отдельные аспекты (Типы I, II, или III) стабильности. Возможно, достигнуть противоречия, то есть генотип может быть найден устойчивым для одного типа измерения по стабильности, но может быть найден непостоянным для другой меры (мер). Они не обеспечивают взаимосвязи среди генотипа, эксплуатирующего образцы реакции от этой (общей) окружающей среды. Классификация генотипов в количественно гомогенные комплекты стабильности, основанные на подобии их реакции на окружающую среду, является другой линией для осмысления, чтобы оценить взаимосвязи среди генотипов, и такой подход, как полагают, является непараметрическим. Метод классификации имеет преимущество в том смысле, что, хотя генотипы сгруппированы на основе определенного набора данных, относительные отношения среди генотипов могут быть независимыми от них, или любого определенного набора данных. Например, два генотипа, скажем A и B, с несходными (несвязанными) образцами реакции, могут всегда сгруппироваться в два различных набора стабильности, независимо от присутствия генотипов, напоминающих A и, или B. Теперь мы будем рассматривать методы для классификации.

3.2.1 Анализ Кластеров

В ссылках в конце этого материала доступны процитированные несколько обычно используемых методов для группирования генотипов (или окружающую среду), основанные на общность особенностей реакции. Каждая техника объединения в кластеры имеет два расчета. (i) определение матрицы общности, (ii) стратегия для того, чтобы сгруппироваться. Возникают два случая. В одном случае, общность основана на генетическом воздействии, и взаимодействии $G \times E$ (средства таблицы $p \times q$), в то время как в другом случае, общность основана только на взаимодействии $G \times E$. Стратегии для того, чтобы сгруппироваться могли быть, например, в виде возрастающих сумм квадратов (ВСК) стратегии синтеза, и среднего числа группы (GA) стратегии синтеза (см. Кормак, 1971г.). Дело обстоит так в отношении иерархического объединения в кластеры.

Ограничения методов объединения в кластеры

Специфические выборы матриц общности и кластерных стратегий дают начало различным кластерным группам, и это может привести к проблемам предпочтения одного метода кластера, по сравнению с другим. Другой критикой кластерного метода является то, что он может также вызвать нежелательную структуру при получении набора данных, предлагая вводящие в заблуждение результаты.

3.2.2 Неиерархическое объединение в кластеры

В неиерархическом объединении в кластеры, цель состоит в том, чтобы сгруппировать единицы (генотипы или местоположения) во многих несвязных классах, выбранных, заранее используя на них информацию в ряде переменных. Как ожидается, единицы с классом будут гомогенными на основе некоторого критерия. В иерархическом объединении в кластеры, можно срезать древовидную диаграмму на уровне схожести, чтобы обеспечить отобранное число групп, но статистические свойства такого группирования, пока еще определены не совсем ясно. В неиерархическом объединении в кластеры, группы единиц были получены, путем оптимизации отобранного критерия. Некоторые из них включают:

- i. максимизацию между - суммой групп квадратов
- ii. максимальную прогнозируемую классификацию,
- iii. уменьшение детерминанта объединенной матрицы дисперсии в пределах класса,
- iv. максимизирование полного согласованного расстояния "Махаланобиса" между группами.

3.2.3 Методы расположения

Методы расположения используются для упрощения многомерных данных для группы индивидуумов, суммируя отношения среди популяций или среди признаков, путем их описания. Это было сделано, производя простое визуальное представление индивидуумов как пункты, которые могут быть подготовлены, чтобы изобразить их отношения, приемлемые, как свободные от искажения. Методы расположения пытаются эффективно снизить размерность многомерных систем, чтобы сохранить отношения среди индивидуумов в максимально возможной степени, но чтобы обеспечить упрощенное представление тех отношений в меньшем количестве измерений, чем это указывается при оригинальных переменных. Имеются два метода расположения.

3.2.4 Анализы Принципиальных Компонентов (АПК).

АПК рассматривает обнаружение нового набора координационных осей, которые более эффективно подходят для вариаций среди единиц, чем те, основанные на оригинальных переменных. АПК представляют преобразование данных от одного набора координат, до другого. Возможно, это не обязательно приводит к сокращению размерности. Однако, когда только (сначала) несколько основных компонентов составляют большую часть изменения, тогда это становится эффективно полезным. Алгебраически, основные оси определены скрытыми векторами от матрицы исправленных сумм квадратов, и продуктов среди переменных. Элементы каждого вектора определяют, что линейные комбинации оригинальных переменных, необходимых для представления соответствующих ПК, и связанного скрытого корня, дают изменение, относящееся к компоненту. АПК может также быть применен на среду тем же самым способом, как это могло быть сделано в отношении генотипов. Мандел (1969г.) рассматривал это для эффекта взаимодействия $G \times E$.

3.2.5 Генотипы Анализов Принципиальных Координат. (ГАПК)

ГАПК анализи требуют обнаружения ряда прямоугольных координационных осей, которые, как считается, должны быть настолько эффективными, насколько это возможно для изменения среди индивидуумов, и могут впоследствии привести к сокращению размерности для упрощения. Эти цели подобны АПК, но ГАПК основаны на большем количестве общего подхода. ГАПК не предполагает автоматически, что оригинальные переменные определяют многомерное пространство Эвклидия, где отношения между парами индивидуумов обозначены расстоянием Эвклидия. Могли быть использованы много мер схожести (например, коэффициенты корреляции), или мер по несходству (расстояние). ГАПК вовлекает два этапа для вычисления.

1) представление комплекта индивидуумов в качестве пунктов в координационном месте произошло из оригинальной матрицы измерений. Гауэр (1966г.) показал, что разделение знаками препинания расстояний Эвклидия в этом месте, является простой функцией оригинальных мер отношений между индивидуумами. Значение этого метода состоит в том, что он отсылает индивидуумов к осям координат Эвклидия, даже, когда начальная координационная структура недоступна, и метод представляет оригинальные меры отношений, как расстояния Эвклидия, даже если они не являются Эвклидийскими.

2) выполнение АПК по данным, полученным поэтапно (1).

Гауэр представил два этапа ГАПК, объединенные в один (1967г.). Единственным требованием, чтобы гарантировать представление ГАПК без искажений является то, что

оригинальная матрица мер должна быть симметричной (так, чтобы никакие отрицательные скрытые корни не были бы получены). В целом, основные оси не будут в виде линейной комбинации оригинальных переменных, как в обычном АПК. Однако, возможно изучить отношения оригинальных переменных к каждой основной оси, коррелируя набор основного координатного множества для каждой оси, с каждой из оригинальных переменных. Корреляция большой величины для конкретной переменной подразумевает, что она сильно отражена в затронутой оси. Гауэр (1966г.) также показал, что АПК является особым случаем ГАПК, когда меры, используемые в ГАПК, представляют согласованные расстояния Эвклидия.

4. Разделение взаимодействия GxE

Мы можем представить

а) Результаты анализа кластера, используемого для того, чтобы зонировать окружающую среду, и сгруппировать генотипы.

б) Разделение взаимодействия GxE, используя эти группировки.

Необходимо уделить внимание оправданию групп, возникающих из метода, с точки зрения числа групп, и природы местоположений, и генотипов в их пределах. Рекомендуется, чтобы группирования были изучены с учетом некоторых других (независимых) переменных, отражающих физические свойства окружающей среды, и фенологические и морфологические свойства генотипов.

Полные иерархии должны быть представлены с помощью древовидной диаграммы, когда используются скапливающиеся методы формирующихся групп. Представим, что n^e и n^g являются числами групп окружающей среды, и соответственно, числами групп генотипов. Также представим, что n_i^g является числом генотипов в группе генотипа i-th ($i=1... n^g$) и n_j^e будет числом окружающей среды в j-th группе

окружающей среды ($j=1... n^e$). Отметим, что $\sum_{j=1}^{n^e} n_j^e = q$ и

$\sum_{i=1}^{n^g} n_i^g = p$. Далее, с уменьшенным G x E, можно представить матрицу данных:

(i) Разделение изменения, связанного с группировкой модели.

Скелет анализа различия.

Источник	d.f.	СК	MS
Окружающая среда (E)	q-1		
Среди групп E	$n^e - 1$		
В пределах групп E	$\Sigma(n_j^e - 1)$		
Генотипы (G)	p-1		
Среди групп Г	$n^g - 1$		
В пределах групп Г	$\Sigma(n_i^g - 1)$		
G x E	(q-1) (p-1)		
Среди групп Г x среди групп E	-		
Среди групп Г x в пределах групп E	-		
В пределах групп Г x среди групп E	-		
В пределах групп Г x в пределах групп E	-		
Остаточный	-		

(ii) Участки с групповой производительностью и

(iii) Образцы (G x E) взаимодействия на сгруппированных

комплектах.

5. Стохастическое Верховенство Сортот

Данная процедура подчеркивает рискованность (нового) генотипа, или сорта. Новые варианты сельскохозяйственных культур (или, в целом новые технологии) могут часто расцениваться фермерами, как более опасные, чем традиционные. Поэтому, риск может иметь тенденцию действовать как препятствие к принятию новых сортов культур. Улучшенные варианты, которые были бы предпочтены "несклонными к риску" фермерами, могут быть определены путем стохастической процедуры верховенства, согласно определенным предположениям. Андерсон (1974г.) использовал эту процедуру для анализа данных, полученных от Урожая Яровой Пшеницы Шестого Международного Питомника, которыми управляет CIMMYT. Он сделал следующие три предположения.

- (i) имеет смысл говорить о (или более региональной) вероятности распределения урожая пшеницы,
- (ii) выбор участков, сотрудников, полей и методов возделывания, и условий болезней, является представительным для соответствующего места производства в мире (или в регионе), и
- (iii) урожай, по существу, предоставляет разумную замену для аргумента сервисной функции среднего фермера.

Менз (1980г.) использовал кластерный анализ Бита и др. (1976г.), чтобы проанализировать Международный Питомник Урожая Яровой Пшеницы CIMMYT в течении, более чем пяти лет, и также использовал стохастическое верховенство. Он нашел значительную степень соответствия в результатах, на основе этих двух методов.

6. Совокупные Основные Воздействия и Мультипликативная Модель Взаимодействия

Другим аспектом эксплуатации взаимодействия GxE является поиск образцов в меньших измерениях. Имеются несколько моделей, которые могут быть изучены для этой цели.

1. Модель CBMM

Модель CBMM ♦ представляет совокупные основные воздействия, и мультипликативные модели взаимодействия.

- ♦ Данные по GxE приспособлены, используя
 - основные воздействия генотипов и окружающей среды,
 - взаимодействие GxE приспособлено как сумма мультипликативного множества АПК для генотипа, и окружающей среды.

Модель:

$$y_{ij} = \mu + \tau_i + \theta_j + \phi_{ij} [= (\tau\theta)_{ij}] + \bar{\varepsilon}_{ij}$$

Где, индексы $i = 1, 2, 3$. Г: число генотипов

$j = 1., 2., E$: число окружающей среды

y_{ij} = среднее для i -th генотипа в j -th окружающей среде (оно может быть приспособлено в среднем к неполным делянкам, и считаться как любая пространственным образом используемая структура),

τ_i = воздействия i -th генотипа

θ_j = воздействия j-th окружающей среды

ϕ_{ij} [= $(\tau\theta)_{ij}$] = взаимодействие между i-th генотипом и j-th окружающей средой

$\bar{\varepsilon}_{ij}$ = экспериментальная ошибка предполагаемого среднего y_{ij}

СВММ облегчает экспертизу образца взаимодействия GxE, как выражено общим числом основных компонентов.

Дальнейшие детали доступны в недавних сериях статей, включая Гауча (1988г.), и Гауча и (1988г.), Абдуллы и др. (1997г.), Корнелиуса & Кросса (1997), и других.

Рассматривая представления основных компонентов k, k = 1, 2.. мин. (G-1, E-1).

$$y_{ij} = \mu + \tau_i + \theta_j + \sum_{k=1}^m \lambda_k \pi_{ik}^g \pi_{jk}^e + \bar{\varepsilon}_{ij}$$

λ_k (k=1 ... m), собственные значения (корни) сингулярного разложения значений (Г x E) матрицы

$$GEI = \phi_{ij}$$

π_{ik}^g = i-th элемент вектора, связанного с генотипом [Г x м.] в

сингулярном разложении. Имеются $k=1,2..m$ векторы, называемые как k th основные компоненты (ОК), и генотипы могут быть графически показаны на оси k -th.

π_{jk}^e = элемент вектора j -th, связанный с генотипом $[E \times m]$ в сингулярном разложении значения.

- ◆ Собственное значение представляет вклад изменения во взаимодействии $G \times E$ вдоль выбранной оси (или ОК), и в порядке убывания.
- ◆ Был изучен совокупный вклад первых немногих компонентов.
- ◆ Вклад каждого из этих ОК мог бы быть проверен на предмет его значения.
- ◆ Если более поздние ОК найдены незначимыми, то GEI может быть выражен как образец (или с точки зрения) нескольких ОК, таким образом, представляя низкое размерное GEI.
- ◆ Низко размерное GEI может использоваться, чтобы предсказать GEI и следовательно работу генотипов, в данной окружающей среде.
- ◆ Это может также быть графически показано в низкоразмерном месте, например, в 1- измерении (первый ОК, ОК1) или с 2- измерениями (сначала два ОК, ОК1, ОК2).

Пример:

На линиях СВММ имеются также другие модели.

2. CBMM: Модель регресса участка.

Это - также модель GGE, где Генотип и взаимодействие GxE совместно могли бы быть представлены в более низком измерении.

- ◆ Данные приспособлены, используя
 - основные воздействия на окружающую среду (участок)
 - объединенное воздействие генотипа и взаимодействия GxE приспособлено как сумма мультипликативного множества АПК для генотипа, и окружающей среды.

Модель:

$$y_{ij} = \mu + \theta_j + \phi_{ij} + \bar{\varepsilon}_{ij}$$

Его версия ОК является

$$y_{ij} = \mu + \theta_j + \sum_{k=1}^m \lambda_k \pi_{ik}^g \pi_{jk}^e + \bar{\varepsilon}_{ij}$$

3. МРГ: Модель регресса генотипа.

- ◆ Данные приспособлены, используя
 - основные воздействия генотипов
 - объединенное воздействие окружающей среды и взаимодействия GxE приспособлено как сумма мультипликативного множества АПК для генотипа и окружающей среды.

Модель:

$$y_{ij} = \mu + \tau_i + \phi_{ij} + \bar{\varepsilon}_{ij}$$

Его версия ОК является

$$y_{ij} = \mu + \tau_i + \sum_{k=1}^m \lambda_k \pi_{ik}^g \pi_{jk}^e + \bar{\varepsilon}_{ij}$$

4. SHMM: Перемещенная Мультипликативная модель

- ◆ Данные приспособлены, используя
 - Постоянное среднее (изменение)
 - объединенное воздействие генотипа, окружающей среды

и взаимодействия $G \times E$ (ϕ_{ij}) приспособлены как сумма мультипликативного множества АПК для генотипа и окружающей среды.

Модель:

$$y_{ij} = \mu + \phi_{ij} + \bar{\varepsilon}_{ij}$$

Его версия ОК является

$$y_{ij} = \mu + \sum_{k=1}^m \lambda_k \pi_{ik}^g \pi_{jk}^e + \bar{\varepsilon}_{ij}$$

5. ПММ: Полностью Мультипликативная модель

♦ Данные y_{ij} как таковые, приспособлены, используя сумму мультипликативного множества АПК для генотипа и окружающей среды.

Его версия ОК

$$y_{ij} = \sum_{k=1}^m \lambda_k \pi_{ik}^g \pi_{jk}^e + \bar{\varepsilon}_{ij}$$

Вычисление

Эти ОК могут быть получены, используя директивы Генстат и процедуры:

SVD

CBMM

BIPLLOT

Примеры:

Данные:

Окружающая среда = 5

Проект: РППД : 25 генотипов, 3 повтора

Мес Пов. Ген. Урожай

1 1 20 600

1 1 15 507

1 1 6 433

1 1 23 660

1 1 12 320

1 1 24 367

1 1 19 453

1 1 4 393

1 1 17 673

1 1 8 553

1 1 22 347

.

.

.

5 3 22 378

5 3 25 650

5 3 24 628

5 3 1 694

5 3 18 703

5 3 12 541

:

1. Модель СВММ и GE в совмещенных двух точек в участках (далее би-участки-biplots)

1. Основные Компоненты:

	Корни	Процент	Включая процент
1.000	1783.5	45.98	45.98
2.000	1242.0	32.02	77.99
3.000	508.3	13.10	91.09
4.000	345.5	8.91	100.00
5.000	0.0	0.00	100.00

2. Таблица ANOVA для модели СВММ

Source	df	SS	MS	F	F_prob
Total	374	257270453	687889	*	*
Treatments	124	247471371	1995737	56.68	0.00000
Genotypes	24	3541544	147564	4.19	0.00000
Environments	4	228626204	57156551	423.63	0.00000
Block	10	1349216	134922	3.83	0.00008
Interactions	96	15303623	159413	4.53	0.00000
IPCA	27	9542845	353439	10.04	0.00000
IPCA	25	4627540	185102	5.26	0.00000
Residuals	44	1133238	25755	0.73	0.89313
Error	240	8449866	35208	*	*

3. AMMI-estimates per environment

Environment	1 AMMI-estimates	Ranked
G1	500.5	G13 722.6
G2	616.2	G2 616.2
G3	448.5	G16 557.9
G4	465.6	G17 553.2
G5	541.5	G5 541.5
G6	392.8	G25 540.0
G7	441.2	G19 535.3
G8	528.0	G15 531.8
G9	512.6	G8 528.0
G10	525.7	G10 525.7
G11	403.5	G9 512.6
G12	406.6	G21 508.7
G13	722.6	G23 505.8
G14	465.9	G1 500.5
G15	531.8	G20 487.7
G16	557.9	G14 465.9
G17	553.2	G4 465.6

G18	410.5	G3	448.5
G19	535.3	G24	445.9
G20	487.7	G7	441.2
G21	508.7	G22	417.9
G22	417.9	G18	410.5
G23	505.8	G12	406.6
G24	445.9	G11	403.5
G25	540.0	G6	392.8

Environment	2 AMMI-estimates	Ranked
G1	754.0	G13 895.2
G2	883.9	G2 883.9
G3	704.3	G8 842.4
G4	778.2	G19 831.3
G5	772.8	G16 830.5
G6	739.2	G17 821.1
G7	772.6	G15 792.9
G8	842.4	G23 788.8
G9	784.9	G9 784.9
G10	784.0	G10 784.0
G11	678.8	G21 782.0
G12	639.2	G4 778.2
G13	895.2	G5 772.8
G14	735.4	G7 772.6
G15	792.9	G1 754.0
G16	830.5	G20 746.7
G17	821.1	G25 739.7
G18	689.7	G6 739.2
G19	831.3	G14 735.4
G20	746.7	G22 732.8
G21	782.0	G24 724.1
G22	732.8	G3 704.3
G23	788.8	G18 689.7
G24	724.1	G11 678.8
G25	739.7	G12 639.2

Environment	3 AMMI-estimates	Ranked
G1	1162	G13 2915
G2	1711	G12 2193
G3	1606	G11 2070
G4	1264	G5 1941
G5	1941	G9 1799
G6	1154	G22 1780
G7	1540	G19 1741
G8	1577	G17 1732
G9	1799	G15 1715
G10	1623	G2 1711
G11	2070	G10 1623
G12	2193	G21 1620
G13	2915	G3 1606
G14	1584	G14 1584
G15	1715	G8 1577

G16	1493	G25	1561
G17	1732	G7	1540
G18	1468	G24	1507
G19	1741	G16	1493
G20	1405	G18	1468
G21	1620	G23	1431
G22	1780	G20	1405
G23	1431	G4	1264
G24	1507	G1	1162
G25	1561	G6	1154

Environment	4 AMMI-estimates		Ranked
G1	2143	G7	3183
G2	2656	G22	3125
G3	2392	G6	3112
G4	2833	G8	3055
G5	2351	G19	2948
G6	3112	G11	2844
G7	3183	G4	2833
G8	3055	G9	2710
G9	2710	G2	2656
G10	2464	G17	2642
G11	2844	G21	2619
G12	2445	G23	2619
G13	2329	G24	2583
G14	2538	G16	2564
G15	2549	G18	2555
G16	2564	G15	2549
G17	2642	G14	2538
G18	2555	G10	2464
G19	2948	G12	2445
G20	2333	G3	2392
G21	2619	G5	2351
G22	3125	G20	2333
G23	2619	G13	2329
G24	2583	G1	2143
G25	1793	G25	1793

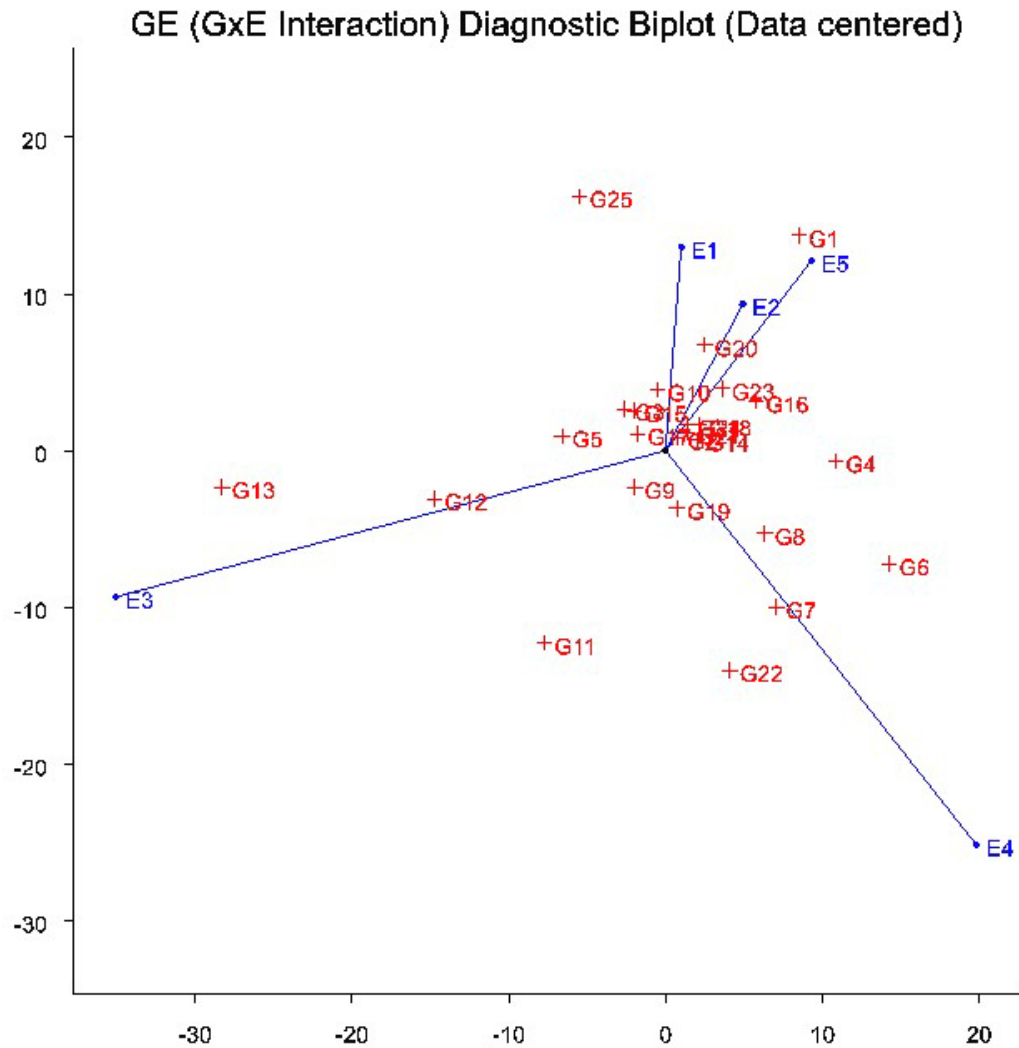
Environment	5 AMMI-estimatesRanked		
G1	826.5	G2	904.3
G2	904.3	G16	869.8
G3	717.4	G8	867.3
G4	833.1	G19	837.8
G5	757.5	G4	833.1
G6	797.8	G17	831.4
G7	791.1	G23	829.1
G8	867.3	G1	826.5
G9	782.2	G10	804.3
G10	804.3	G15	802.8
G11	630.6	G21	800.3
G12	577.6	G6	797.8
G13	786.4	G7	791.1

G14	752.9	G20	788.5
G15	802.8	G13	786.4
G16	869.8	G9	782.2
G17	831.4	G25	770.3
G18	714.3	G5	757.5
G19	837.8	G14	752.9
G20	788.5	G24	748.2
G21	800.3	G22	720.1
G22	720.1	G3	717.4
G23	829.1	G18	714.3
G24	748.2	G11	630.6
G25	770.3	G12	577.6

Таблица первых четырех отборов СВММ на каждую окружающую среду

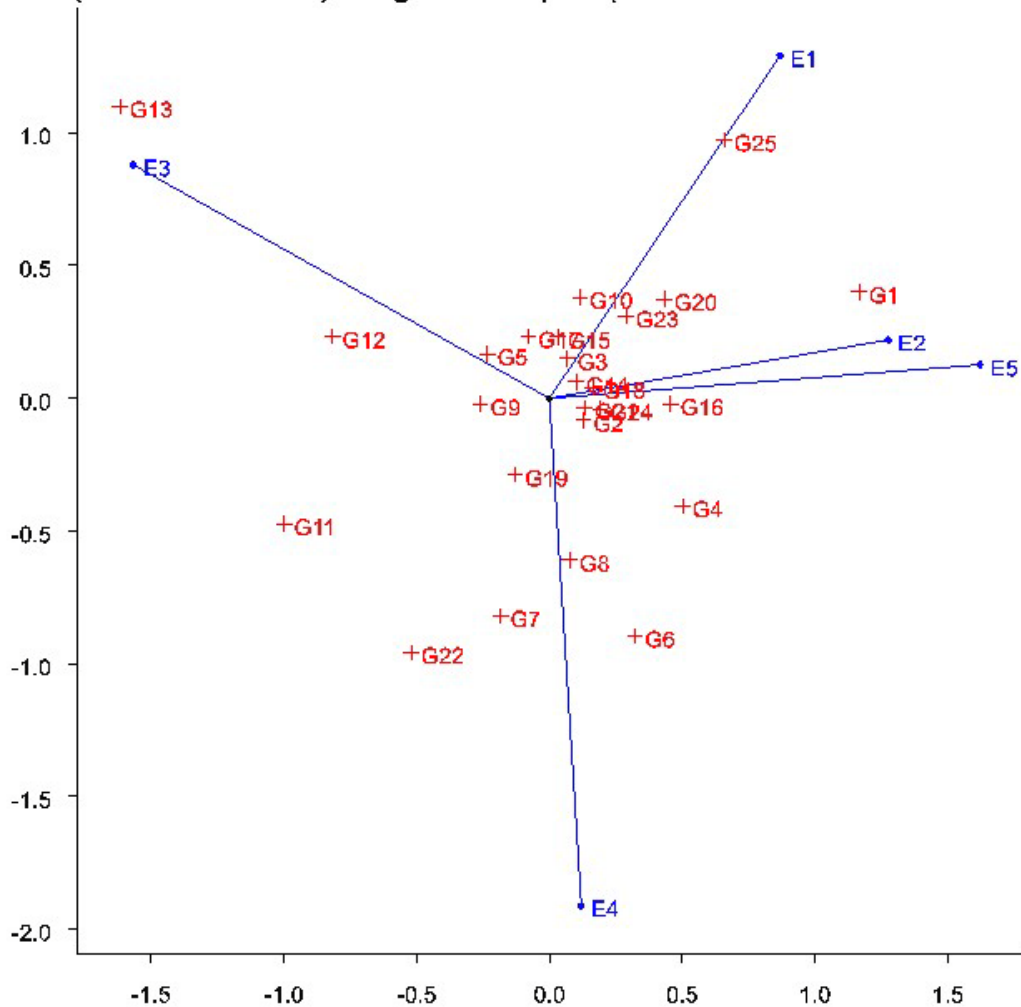
Number	Environment	Mean	Score	1	2	3	4
4	E4	2615	27.16	G7	G22	G6	G8
5	E5	782	5.22	G2	G16	G8	G19
2	E2	770	1.79	G13	G2	G8	G19
1	E1	499	-2.39	G13	G2	G16	G17
3	E3	1664	-31.78	G13	G12	G11	G5

Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении)



Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении, и для их
нормализации (в единицу суммы квадратов):

GE (GxE Interaction) Diagnostic Biplot (Data centered and normalized)



2. СВММ и Биучастки GGE

..... Процент Взаимодействия G x E, вычисленный....

	Roots	Percent	CumPercent
1.000	1810.4	38.45	38.45
2.000	1531.9	32.54	70.99
3.000	602.7	12.80	83.79
4.000	418.9	8.90	92.69
5.000	344.2	7.31	100.00

Расчеты GGE для генотипов

----- Окружающая среда- E 1

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	467.2	G13	650.1
G2	502.5	G12	563.5
G3	502.7	G5	541.8
G4	444.4	G11	530.7
G5	541.8	G25	527.8
G6	418.4	G9	509.0
G7	457.2	G15	507.7
G8	467.9	G17	505.3
G9	509.0	G3	502.7
G10	502.0	G2	502.5
G11	530.7	G10	502.0
G12	563.5	G21	493.8
G13	650.1	G14	493.6
G14	493.6	G19	490.9
G15	507.7	G22	485.8
G16	483.1	G20	484.5
G17	505.3	G16	483.1
G18	479.8	G24	482.8
G19	490.9	G18	479.8
G20	484.5	G23	473.5
G21	493.8	G8	467.9
G22	485.8	G1	467.2
G23	473.5	G7	457.2
G24	482.8	G4	444.4
G25	527.8	G6	418.4

----- Окружающая среда-

E 2

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	755.6	G13	795.4
G2	771.1	G11	780.0
G3	766.9	G12	780.0
G4	762.5	G22	775.7
G5	774.0	G5	774.0
G6	762.0	G19	773.7
G7	770.9	G9	773.5
G8	770.8	G17	771.6
G9	773.5	G2	771.1
G10	768.0	G7	770.9
G11	780.0	G8	770.8
G12	780.0	G15	770.5
G13	795.4	G21	768.9
G14	767.5	G10	768.0
G15	770.5	G14	767.5
G16	765.8	G3	766.9
G17	771.6	G24	766.1
G18	765.1	G16	765.8
G19	773.7	G18	765.1
G20	762.3	G23	764.9
G21	768.9	G4	762.5
G22	775.7	G20	762.3
G23	764.9	G6	762.0
G24	766.1	G25	761.9
G25	761.9	G1	755.6

----- Окружающая среда-

E 3

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	1163	G13	2922
G2	1714	G12	2181
G3	1600	G11	2061
G4	1265	G5	1939

G5	1939	G9	1802
G6	1153	G22	1775
G7	1540	G19	1741
G8	1578	G17	1737
G9	1802	G15	1717
G10	1627	G2	1714
G11	2061	G10	1627
G12	2181	G21	1621
G13	2922	G3	1600
G14	1583	G14	1583
G15	1717	G8	1578
G16	1497	G25	1559
G17	1737	G7	1540
G18	1466	G24	1504
G19	1741	G16	1497
G20	1409	G18	1466
G21	1621	G23	1438
G22	1775	G20	1409
G23	1438	G4	1265
G24	1504	G1	1163
G25	1559	G6	1153

----- Environment is E 4

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	2141	G7	3184
G2	2650	G22	3126
G3	2390	G6	3113
G4	2832	G8	3051
G5	2349	G19	2945
G6	3113	G11	2845
G7	3184	G4	2832
G8	3051	G9	2714
G9	2714	G2	2650
G10	2466	G17	2643
G11	2845	G23	2623
G12	2445	G21	2620
G13	2332	G24	2582
G14	2539	G16	2560
G15	2548	G18	2559
G16	2560	G15	2548
G17	2643	G14	2539

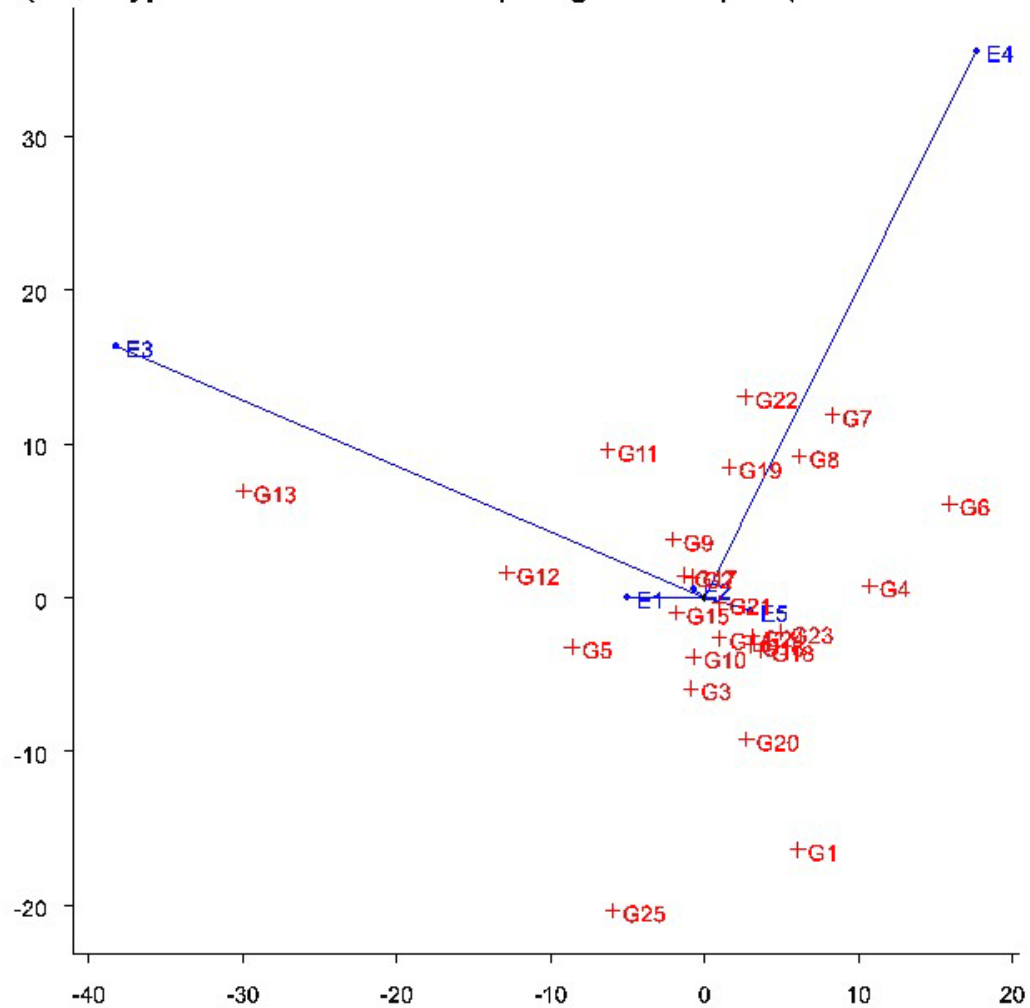
G18	2559	G10	2466
G19	2945	G12	2445
G20	2336	G3	2390
G21	2620	G5	2349
G22	3126	G20	2336
G23	2623	G13	2332
G24	2582	G1	2141
G25	1787	G25	1787

----- Environment is E 5

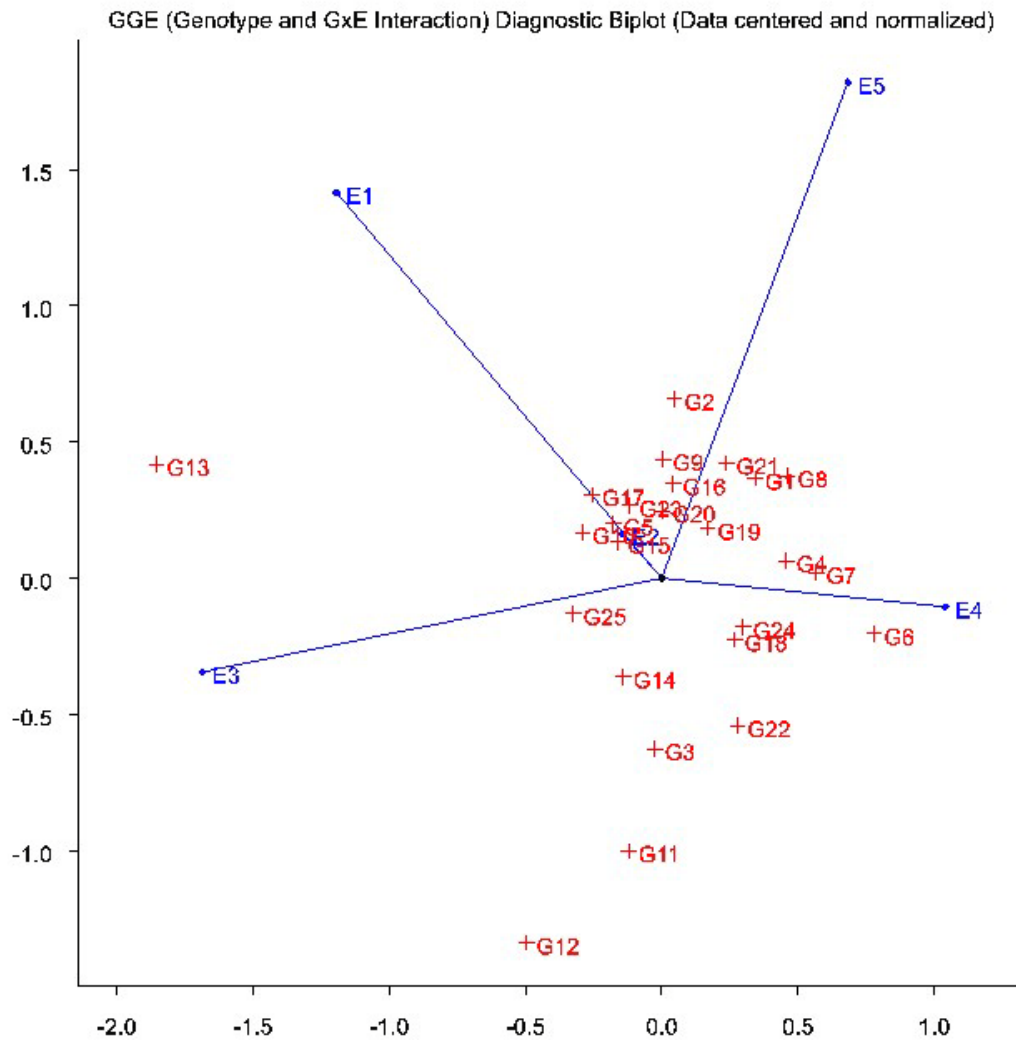
GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	813.1	G6	824.4
G2	778.3	G4	813.2
G3	783.9	G1	813.1
G4	813.2	G23	798.3
G5	758.6	G20	797.3
G6	824.4	G7	796.9
G7	796.9	G18	795.5
G8	792.7	G16	793.3
G9	772.6	G24	793.0
G10	782.7	G8	792.7
G11	755.1	G14	786.7
G12	742.0	G21	784.8
G13	686.6	G3	783.9
G14	786.7	G10	782.7
G15	777.1	G25	780.4
G16	793.3	G19	779.6
G17	776.6	G22	779.0
G18	795.5	G2	778.3
G19	779.6	G15	777.1
G20	797.3	G17	776.6
G21	784.8	G9	772.6
G22	779.0	G5	758.6
G23	798.3	G11	755.1
G24	793.0	G12	742.0
G25	780.4	G13	686.6

Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении)

GGE (Genotype and GxE Interaction) Diagnostic Biplot (Data centered at or



Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении, и для их
нормализации (в единицу суммы квадратов):



Пример 2:

Число окружающей среды: 5

Опытный проект: тройная решетка, 25 генотипов.

Env	Rep	Blk	Geno	GYld"
1	1	1	20	600
1	1	1	15	507
1	1	1	6	433
1	1	1	23	660
1	1	1	12	320
1	1	2	24	367
1	1	2	19	453
.				
.				
.				
5	3	5	25	650
5	3	5	24	628
5	3	5	1	694
5	3	5	18	703
5	3	5	12	541

1. Модель СВММ и би участки GE

Разделение взаимодействия GxE:

DFPCA FvalPCA ProbPCA

27.00 12.432 0.0000

25.00 5.621 0.0000

44.00 0.931 0.5988

Предлагая: 2 ОК

Корни для всех основных компонентов:

	Roots	Percent	CumPercent
1.000	1735.1	46.91	46.91
2.000	1122.6	30.35	77.26
3.000	503.7	13.62	90.88
4.000	337.3	9.12	100.00
5.000	0.0	0.00	100.00

Расчеты СВММ:

----- Окружающая среда-

E 1

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	490.4	G13	737.9
G2	613.3	G2	613.3
G3	451.6	G17	556.2
G4	468.6	G16	548.7
G5	541.4	G15	543.4
G6	401.8	G25	543.3
G7	436.9	G5	541.4
G8	532.2	G8	532.2
G9	524.8	G19	531.6
G10	522.9	G9	524.8
G11	385.4	G10	522.9
G12	418.3	G23	514.2
G13	737.9	G21	511.5
G14	445.4	G20	499.3
G15	543.4	G1	490.4
G16	548.7	G4	468.6
G17	556.2	G3	451.6

G18	394.7	G24	448.1
G19	531.6	G14	445.4
G20	499.3	G7	436.9
G21	511.5	G12	418.3
G22	404.3	G22	404.3
G23	514.2	G6	401.8
G24	448.1	G18	394.7
G25	543.3	G11	385.4

----- Environment is E 2

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	744.9	G13	906.7
G2	884.5	G2	884.5
G3	702.8	G8	847.5
G4	784.8	G16	831.0
G5	783.3	G19	819.0
G6	755.3	G17	816.5
G7	772.1	G15	797.4
G8	847.5	G9	796.7
G9	796.7	G23	785.0
G10	778.0	G4	784.8
G11	671.0	G21	784.0
G12	642.8	G5	783.3
G13	906.7	G10	778.0
G14	721.9	G7	772.1
G15	797.4	G20	755.7
G16	831.0	G6	755.3
G17	816.5	G1	744.9
G18	668.2	G22	742.7
G19	819.0	G25	733.9
G20	755.7	G14	721.9
G21	784.0	G24	719.0
G22	742.7	G3	702.8
G23	785.0	G11	671.0
G24	719.0	G18	668.2
G25	733.9	G12	642.8

----- Environment is E 3

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	1042	G13	2975
G2	1732	G12	2183
G3	1651	G11	2102
G4	1256	G5	1923
G5	1923	G9	1813
G6	1213	G17	1761
G7	1570	G19	1751
G8	1586	G22	1734
G9	1813	G2	1732
G10	1619	G15	1724
G11	2102	G3	1651
G12	2183	G10	1619
G13	2975	G21	1601
G14	1523	G8	1586
G15	1724	G7	1570
G16	1435	G25	1543
G17	1761	G24	1524
G18	1446	G14	1523
G19	1751	G23	1461
G20	1423	G18	1446
G21	1601	G16	1435
G22	1734	G20	1423
G23	1461	G4	1256
G24	1524	G6	1213
G25	1543	G1	1042

----- Environment is E 4

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	2244	G22	3133
G2	2714	G7	3067
G3	2418	G6	3063
G4	2815	G8	2970
G5	2498	G11	2825
G6	3063	G4	2815
G7	3067	G19	2802
G8	2970	G2	2714
G9	2694	G9	2694
G10	2482	G16	2654
G11	2825	G21	2611

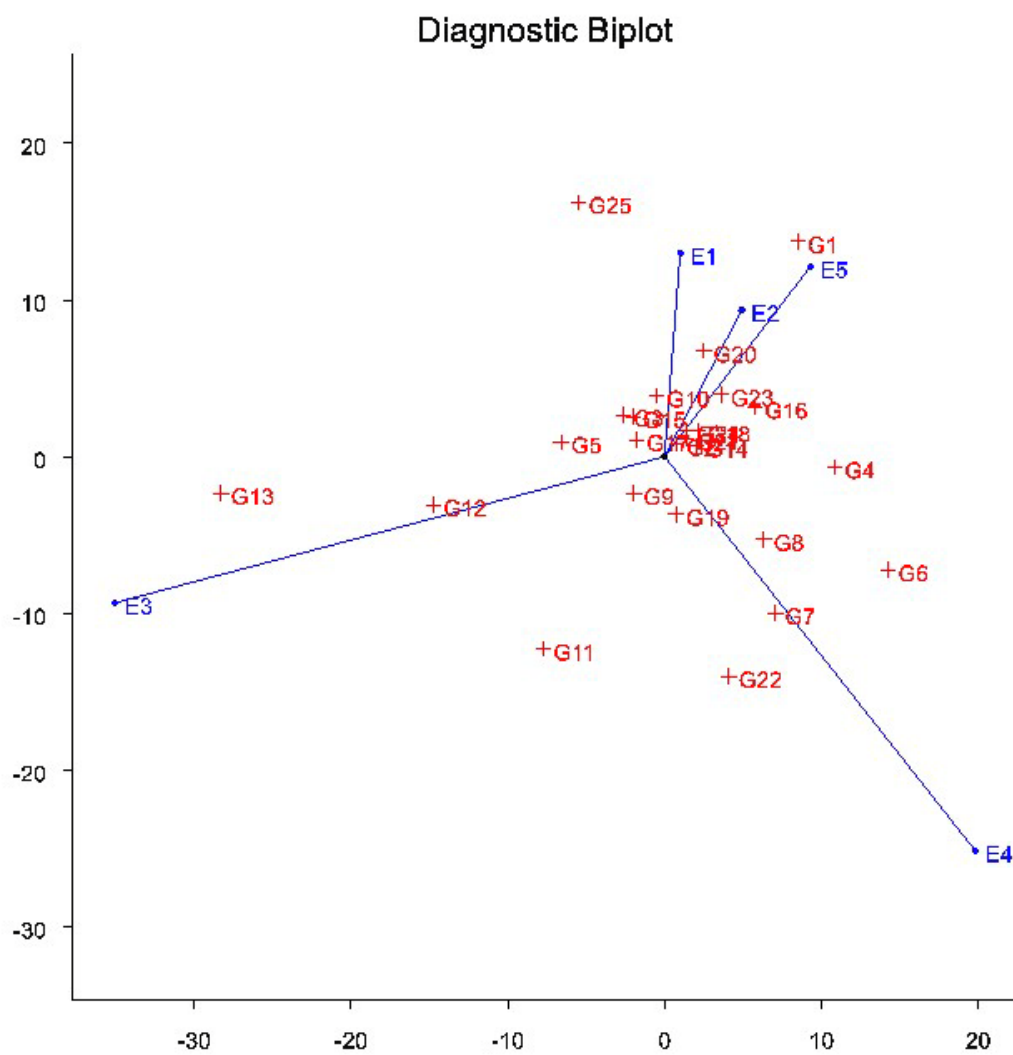
G12	2374	G17	2598
G13	2411	G14	2573
G14	2573	G23	2548
G15	2525	G24	2529
G16	2654	G15	2525
G17	2598	G5	2498
G18	2488	G18	2488
G19	2802	G10	2482
G20	2406	G3	2418
G21	2611	G13	2411
G22	3133	G20	2406
G23	2548	G12	2374
G24	2529	G1	2244
G25	1938	G25	1938

----- Environment is E 5

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	832.3	G2	902.0
G2	902.0	G16	876.8
G3	710.7	G8	872.5
G4	842.5	G4	842.5
G5	769.0	G1	832.3
G6	809.6	G19	824.0
G7	787.2	G23	823.7
G8	872.5	G17	823.6
G9	793.5	G6	809.6
G10	798.5	G15	807.6
G11	615.0	G21	805.1
G12	581.6	G10	798.5
G13	788.2	G20	797.3
G14	744.4	G9	793.5
G15	807.6	G13	788.2
G16	876.8	G7	787.2
G17	823.6	G5	769.0
G18	694.0	G25	766.9
G19	824.0	G14	744.4
G20	797.3	G24	741.9
G21	805.1	G22	733.6
G22	733.6	G3	710.7
G23	823.7	G18	694.0
G24	741.9	G11	615.0

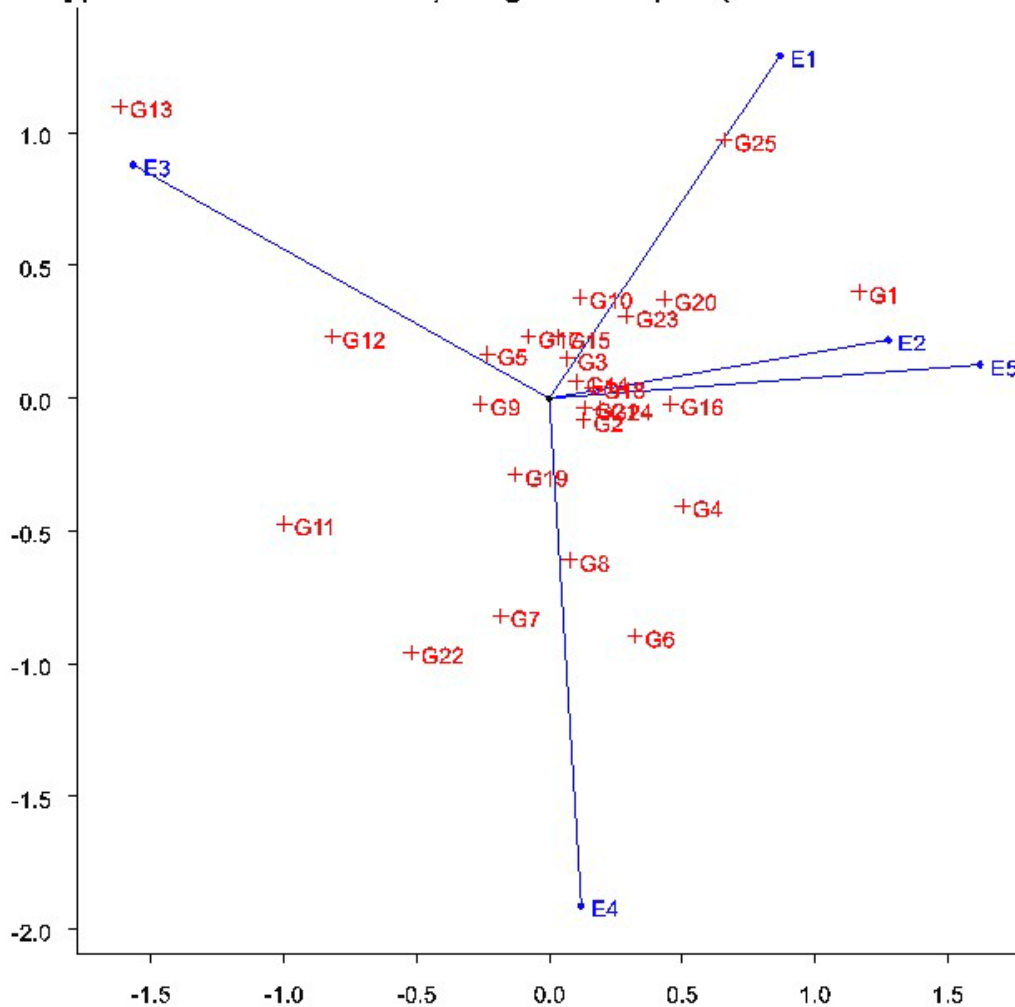
G25 766.9 G12 581.6

Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении)



Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении, и для их
нормализации (в единицу суммы квадратов):

⋮ (Genotype and GxE Interaction) Diagnostic Biplot (Data centered and norm



2. CBMM и GGE биучастки

Процент взаимодействия GE вычисленный по корням

		Roots Percent		CumPercent
1.000	1847.6	40.55	40.55	
2.000	1332.5	29.24	69.79	
3.000	603.0	13.23	83.02	
4.000	439.0	9.63	92.65	
5.000	334.8	7.35	100.00	

GGE estimates

----- Окружающая среда E 1

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	448.0	G13	655.8
G2	502.1	G12	567.5
G3	506.6	G11	536.1
G4	443.4	G5	533.4
G5	533.4	G25	519.1
G6	425.9	G9	511.5
G7	465.5	G17	510.7
G8	472.7	G15	510.1
G9	511.5	G3	506.6
G10	500.5	G2	502.1
G11	536.1	G10	500.5
G12	567.5	G19	499.3
G13	655.8	G21	492.0
G14	484.8	G24	487.0
G15	510.1	G14	484.8
G16	471.8	G20	482.4
G17	510.7	G18	480.1
G18	480.1	G22	480.1
G19	499.3	G23	479.7
G20	482.4	G8	472.7
G21	492.0	G16	471.8
G22	480.1	G7	465.5
G23	479.7	G1	448.0
G24	487.0	G4	443.4
G25	519.1	G6	425.9

----- Окружающая среда

E 2

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	738.7	G13	806.7
G2	775.1	G11	789.7
G3	763.1	G22	787.2
G4	762.6	G7	780.2
G5	774.6	G12	778.8
G6	768.4	G19	778.2
G7	780.2	G8	777.8
G8	777.8	G9	777.2
G9	777.2	G2	775.1
G10	764.7	G5	774.6
G11	789.7	G17	772.7
G12	778.8	G15	769.1
G13	806.7	G6	768.4
G14	764.0	G21	767.7
G15	769.1	G10	764.7
G16	763.7	G14	764.0
G17	772.7	G16	763.7
G18	758.9	G3	763.1
G19	778.2	G4	762.6
G20	756.0	G24	762.5
G21	767.7	G23	761.6
G22	787.2	G18	758.9
G23	761.6	G20	756.0
G24	762.5	G25	745.5
G25	745.5	G1	738.7

----- Окружающая среда

E 3

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	1045	G13	2983
G2	1738	G12	2169
G3	1645	G11	2090
G4	1258	G5	1922
G5	1922	G9	1815
G6	1211	G17	1766
G7	1568	G19	1752

G8	1588	G2	1738
G9	1815	G22	1727
G10	1623	G15	1726
G11	2090	G3	1645
G12	2169	G10	1623
G13	2983	G21	1602
G14	1522	G8	1588
G15	1726	G7	1568
G16	1440	G25	1542
G17	1766	G14	1522
G18	1442	G24	1520
G19	1752	G23	1468
G20	1427	G18	1442
G21	1602	G16	1440
G22	1727	G20	1427
G23	1468	G4	1258
G24	1520	G6	1211
G25	1542	G1	1045

----- Окружающая среда

E 4

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	2242	G22	3132
G2	2712	G7	3069
G3	2412	G6	3064
G4	2816	G8	2968
G5	2496	G11	2822
G6	3064	G4	2816
G7	3069	G19	2801
G8	2968	G2	2712
G9	2698	G9	2698
G10	2487	G16	2654
G11	2822	G21	2612
G12	2366	G17	2603
G13	2420	G14	2575
G14	2575	G23	2555
G15	2525	G24	2526
G16	2654	G15	2525
G17	2603	G5	2496
G18	2489	G18	2489
G19	2801	G10	2487
G20	2409	G13	2420

G21	2612	G3	2412
G22	3132	G20	2409
G23	2555	G12	2366
G24	2526	G1	2242
G25	1930	G25	1930

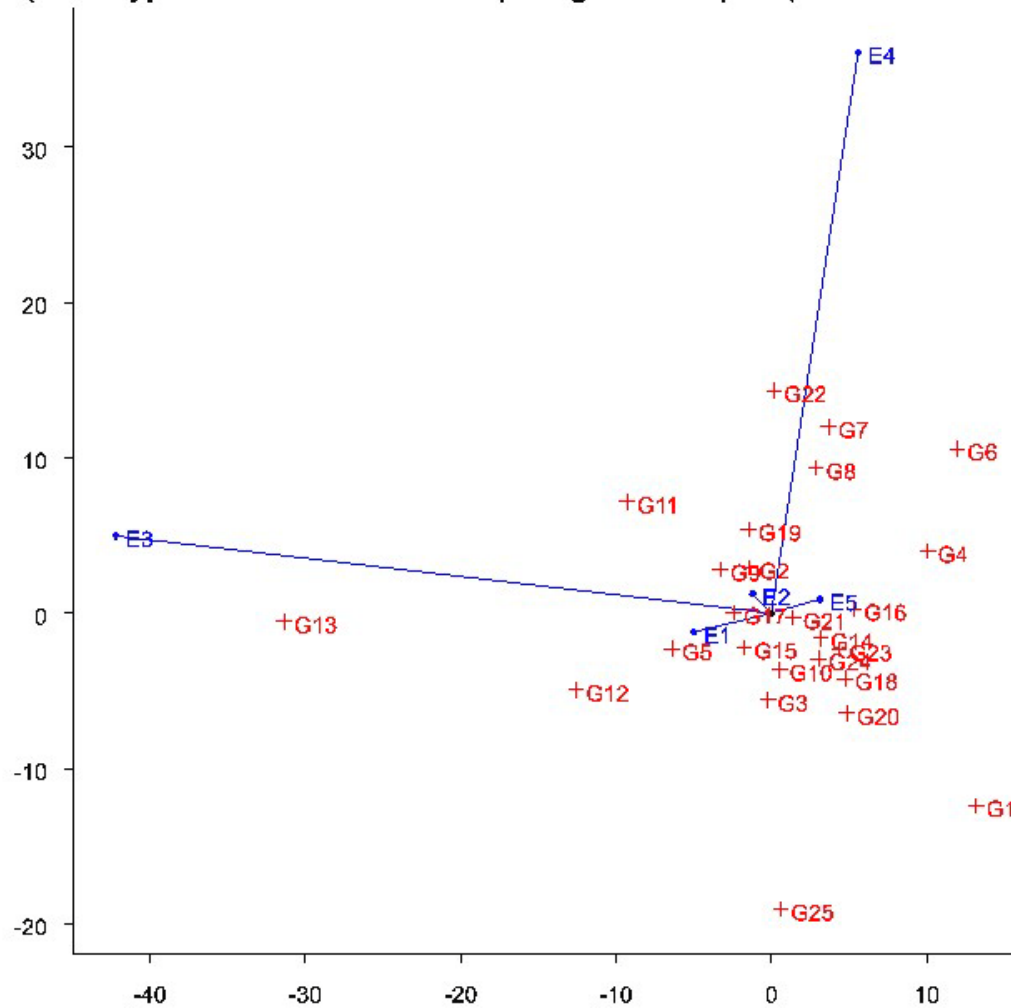
----- Окружающая среда

E 5

GenoLabel	GEFitted	GenoLabelR	GEFitRanked
G1	811.9	G6	828.4
G2	779.8	G4	816.7
G3	776.0	G1	811.9
G4	816.7	G7	803.8
G5	759.6	G8	798.9
G6	828.4	G16	798.5
G7	803.8	G22	794.9
G8	798.9	G23	793.2
G9	773.9	G18	792.8
G10	780.1	G20	791.1
G11	759.0	G14	790.1
G12	738.0	G24	788.6
G13	683.2	G21	785.8
G14	790.1	G19	781.8
G15	774.2	G10	780.1
G16	798.5	G2	779.8
G17	774.1	G3	776.0
G18	792.8	G15	774.2
G19	781.8	G17	774.1
G20	791.1	G9	773.9
G21	785.8	G25	766.7
G22	794.9	G5	759.6
G23	793.2	G11	759.0
G24	788.6	G12	738.0
G25	766.7	G13	683.2

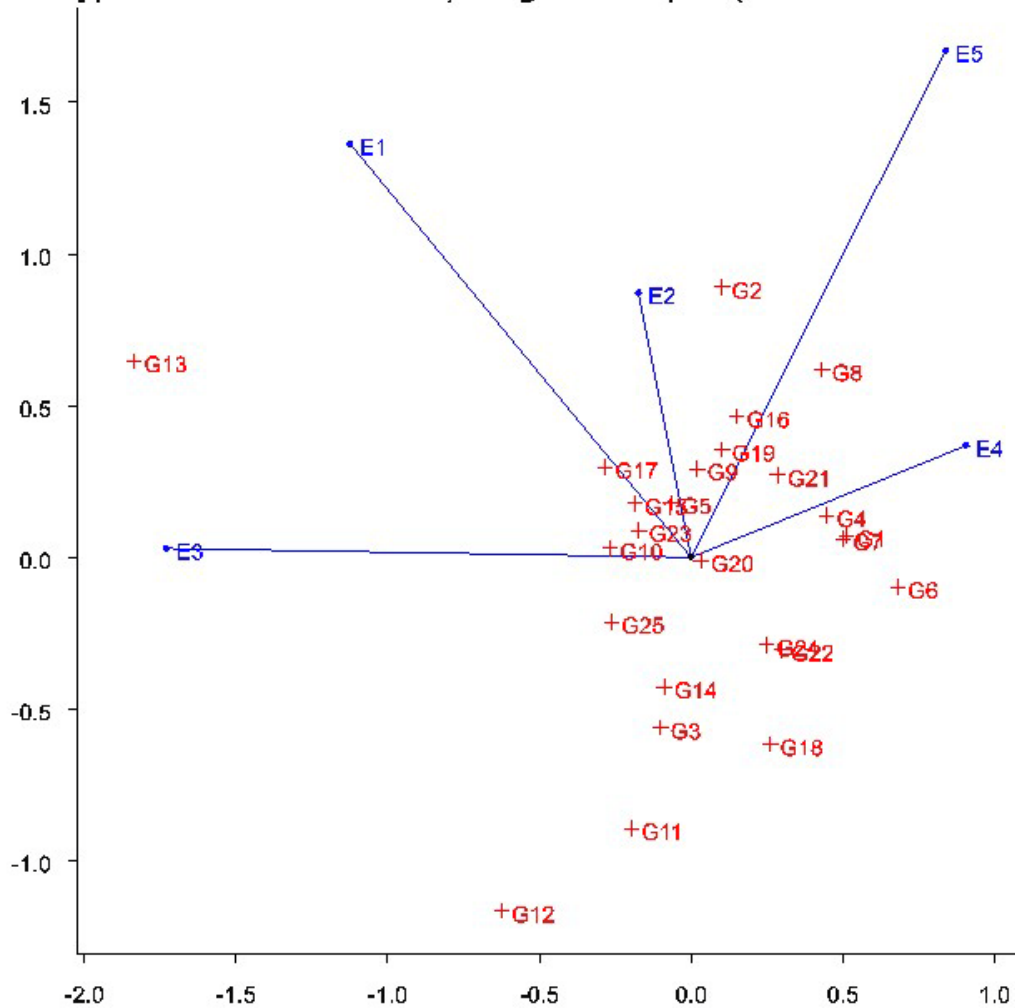
Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении)

GGE (Genotype and GxE Interaction) Diagnostic Biplot (Data centered at or



Би участки (данные, стандартизированные, чтобы
сосредоточить конфигурации в происхождении, и для их
нормализации (в единицу суммы квадратов):

E (Genotype and GxE Interaction) Diagnostic Biplot (Data centered and non



7. Переносимость Сортів Сельськогосподарських Культур Внутрі участка

- Развитие сортов и их оценка, часто происходит на ограниченном диапазоне окружающей среды (например, экспериментальные станции), но сорта фактически предназначены для производства в намного разнообразней окружающей среде (например, поля фермеров).
- Поэтому, переносимость является реакцией сорта на новое место, что является важным аспектом для рекомендации сорта.
- Сингх и др. (1996г.) представил статистическое измерение переносимости сорта, используя данные, полученные из различных участков.
- Подход выглядит следующим. Для данного сорта, скажем i , его реакция на окружающую среду может быть смоделирована, как линейная регрессия на экологическом индексе (часто рассматриваемая, как здоровой биологической мерой, и принятая, как средняя величина из всех генотипов в том месте).
- Для оценки переносимости реакции генотипа на окружающую среду, скажем j -th, вставьте линейную регрессию урожая на экологический индекс, используя данные по (реакция, индекс) парам для всех мест, кроме места j -th, и вычисляя различие в реакции наблюдаемого урожая, и предсказанную реакцию в месте j -th, используя вышеупомянутую линейную регрессию.
- Такое различие было названо, как остатки из участка (Вуд и Кади, 1981г.) и предсказуемым остатком (Кук и Висберг, 1982г.). Такие различия могут быть получены, оставляя одно место за один раз. Их (взвешенная) сумма квадратов дает остаточную сумму квадратов внутри участка.

- Для проверки остатков внутри участка, мы можем рассмотреть остатки в участке, как и остатки в пределах поля.
- Мерой (P) переносимости для генотипа (на рассмотрении -i), в таком случае, является коэффициент суммы квадрата передаточного остатка внутри участка, в отношении к остаточной сумме квадратов в пределах участка, взвешенных с повторениями.
- Были отработаны статистические распределения линейных функций P, когда ошибочные различия по местам, являлись гомогенными / гетерогенными. Сингхом и другими были представлены шесть испытаний в нескольких местах для ячменя и пшеницы, изменяющихся от 16-53 и сорта, изменяющегося от 21-23. (1996г.).

8. Иллюстрация

Мы перечисляем в следующей распечатке от GENSTAT 5 программ, написанных для анализа данных, полученных от испытаний сортов в различных участках, проведенных в рандомизированных полных деляночных конструкциях.

Окружающая среда: 10

Проект: РППД в 15 генотипах и 3 повторениях

Data:

Loc	Rep	Geno	Yield
1	1	1	567
1	1	2	800
1	1	3	1300
1	1	4	1700
1	1	5	1333
1	1	6	1600
1	1	7	1333
1	1	8	1350
1	1	9	1683
1	1	10	1600
1	1	11	1600
1	1	12	967
1	1	13	1567
.			
.			
.			
10	3	13	3160
10	3	2	4880
10	3	15	5080
10	3	1	4280
10	3	8	4720
10	3	4	3840
:			

Section 1. Analysis of data from individual environments

***** Environment is = 1 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	27563.	13781.	0.23	

Rep.Geno stratum					
Geno	14	666583.	47613.	0.78	0.682
Residual	28	1711141.	61112.		
Total	44	2405287.			

***** Environment is = 2 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	165032.	82516.	0.18	
Rep.Geno stratum					
Geno	14	5153189.	368085.	0.79	0.672
Residual	28	13060300.	466439.		
Total	44	18378521.			

***** Environment is = 3 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	3598694.	1799347.	2.94	
Rep.Geno stratum					
Geno	14	19108528.	1364895.	2.23	0.034
Residual	28	17115056.	611252.		
Total	44	39822278.			

***** Environment is = 4 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	10000371.	5000185.	2.91	
Rep.Geno stratum					
Geno	14	27233159.	1945226.	1.13	0.375
Residual	28	48085069.	1717324.		
Total	44	85318599.			

***** Environment is = 5 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	8853142.	4426571.	6.48	
Rep.Geno stratum					
Geno	14	27816546.	1986896.	2.91	0.008
Residual	28	19133779.	683349.		
Total	44	55803467.			

***** Environment is = 6 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	16804.	8402.	0.02	
Rep.Geno stratum					
Geno	14	4419494.	315678.	0.88	0.585
Residual	28	10018841.	357816.		
Total	44	14455139.			

***** Environment is = 7 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.(m.v.)	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	848920.	424460.	2.90	
Rep.Geno stratum					
Geno	14	4911812.	350844.	2.40	0.025
Residual	27(1)	3954364.	146458.		
Total	43(1)	9666992.			

***** Environment is = 8 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	200333.	100167.	13.87	
Rep.Geno stratum					
Geno	14	461750.	32982.	4.57	<.001
Residual	28	202167.	7220.		
Total	44	864250.			

***** Environment is = 9 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	1606671.	803336.	0.66	
Rep.Geno stratum					
Geno	14	40950364.	2925026.	2.39	0.024
Residual	28	34268396.	1223871.		
Total	44	76825431.			

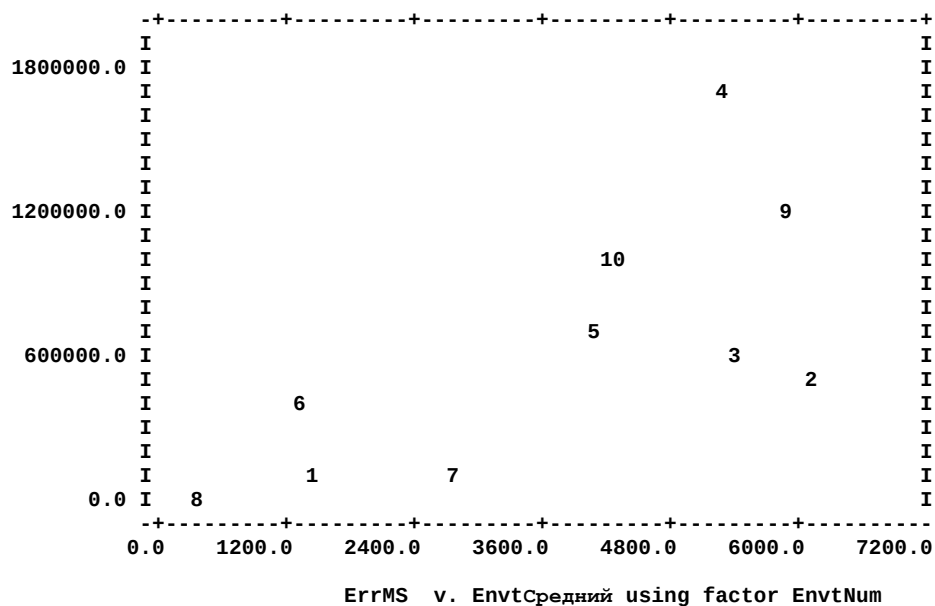
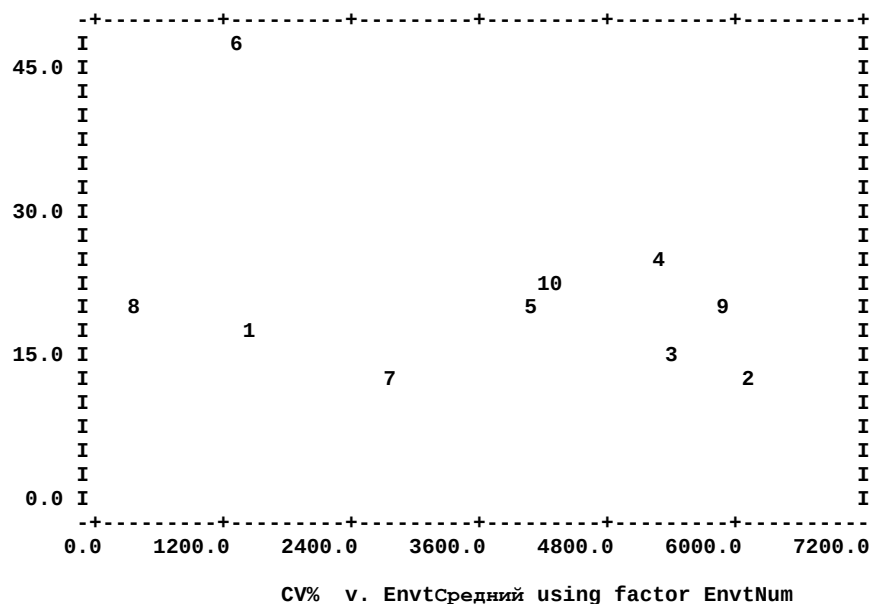
***** Environment is = 10 *****

Analysis of variance

Variate: GYield

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Rep stratum	2	933760.	466880.	0.47	
Rep.Geno stratum					
Geno	14	9569067.	683505.	0.69	0.769
Residual	28	27909973.	996785.		
Total	44	38412800.			

EnvNum	EnvMean	CV%	SEM	ErrMS	ErrDF
1	1409	17.54	142.7	61112	28.00
2	6069	11.25	394.3	466439	28.00
3	5417	14.43	451.4	611252	28.00
4	5324	24.61	756.6	1717324	28.00
5	4086	20.23	477.3	683349	28.00
6	1264	47.34	345.4	357816	28.00
7	2816	13.59	221.0	146458	27.00
8	415	20.48	49.1	7220	28.00
9	5872	18.84	638.7	1223871	28.00
10	4227	23.62	576.4	996785	28.00



Раздел 2. Тест Барлета на однородность ошибочных различий

Объединенный Средний квадрат Ошибки =
628886

DF для Объединенного Среднего квадрата ошибки. =
279.0

Вероятность большего **chi**-квадрата (для того, чтобы
проверить однородность ошибочных различий) =
0.000000

Ошибочные различия являются гетерогенными в 0.05000
вероятностях

**Раздел 3: Объединенный анализ данных для
взаимодействия GXE**

3.1 ANOVA под гомогенными ошибочными различиями

Дисперсионный анализ

Variate: GYield

Source of variation	d.f.(m.v.)	s.s.	m.s.	v.r.	F pr.
Env.t.Rep stratum					
Envt	9	1.759E+09	1.955E+08	148.92	<.001
Residual	20	2.625E+07	1.313E+06	2.09	
Envt.Rep.Geno stratum					
Geno	14	2.097E+07	1.498E+06	2.38	0.004
Envt.Geno	126	1.193E+08	9.470E+05	1.51	0.003
Residual	279(1)	1.755E+08	6.289E+05		
Total	448(1)	2.101E+09			

***** Environment means *****

Genotype	1	2	3	4	5	6	7	8	9	10
1	1167	6178	4000	5167	4026	660	2107	441.7	3567	4360
2	1333	6044	4933	6635	3173	1457	2308	391.7	6300	4933
3	1341	6556	6100	5967	3423	1050	2611	316.7	5813	4267
4	1411	5700	5292	6245	4314	699	2785	283.3	6958	3680
5	1464	5511	5867	5617	5551	1437	2833	333.3	7558	4013
6	1511	5595	6050	5489	3699	1168	3063	275.0	6003	3133
7	1411	5933	5350	5737	4987	1451	2740	358.3	5428	4440
8	1495	5867	4700	5445	3070	987	3181	375.0	6837	3880
9	1516	6389	5133	5700	3897	1899	2596	491.7	5387	4120
10	1576	5611	5167	5056	3423	1565	3061	408.3	6200	4280
11	1511	6189	5400	3833	5135	1254	3048	575.0	5637	4360
12	1161	6455	6600	5958	5019	1391	3501	483.3	6280	4373
13	1357	6389	6117	4278	4750	1469	2845	641.7	6360	3893
14	1533	6456	4800	4481	3494	1264	2816	366.7	5200	5013
15	1354	6167	5750	4259	3333	1203	2817	483.3	4558	4653
Mean	1409	6069	5417	5324	4086	1264	2821	415.0	5872	4227

3.2 Взвешенный ANOVA при гетерогенных изменениях ошибки

Дисперсионный анализ

Variate: GEData

Weight variate: AllWet

Source of variation	d.f.	s.s.	m.s.	v.r.	F pr.
Env1	9	9442.689	1049.188		
Geno1	14	61.356	4.383		
Env1.Geno1	126	201.326	1.598		
Total	149	9705.370			

3.2.1 Tests of significance for Genotype and GxE interaction

Genotype

DF = 14 Weighted Sum of Squares = 61.356 Prob > Chisq = 0.00000

G x E Interaction

DF = 126 Weighted Sum of Squares = 201.326 Prob > Chisq = 0.00002

Раздел 4: Наследуемость

4.1 Environment-wise heritability estimates

EnvNum	EnvtMean	RCBHerit	RCBBias	RCBSeh2	RCBSGg2	RCBSGe2
1	1409	0.0000	0.0000	0.00000	0	54666
2	6069	0.0000	0.0000	0.00000	0	417694
3	5417	0.2913	0.1184	0.17307	251214	611252
4	5324	0.0424	0.6086	0.16029	75967	1717324
5	4086	0.3887	0.0924	0.16765	434516	683349
6	1264	0.0000	0.0000	0.00000	0	328526
7	2821	0.3238	0.1101	0.17348	69756	145689
8	415	0.5432	0.0645	0.14705	8587	7220
9	5872	0.3252	0.1068	0.16992	576397	1195835
10	4227	0.0000	0.0000	0.00000	0	873018

4.2 Корреляции между средним, наследуемым, генотипным различием и ошибочное различие

EnvMean	1.0000			
RCBHerit	-0.1629	1.0000		
RCBSGg2	0.4837	0.4898	1.0000	
RCBSGe2	0.7009	-0.1992	0.4167	1.0000
	EnvMean	RCBHerit	RCBSGg2	RCBSGe2

4.3 Heritability estimate in presence of GxE interaction

heritability = 0.02429 Bias = 0.02664 SError= 0.02532

Раздел 5: Тесты на параллельность линий регрессий

5.1 Разделение GXE Int на гетерогенность линий регрессий при гомогенных изменениях ошибок

Дисперсионный анализ

Variate: GYield

Source of variation pr.	d.f.(m.v.)	s.s.	m.s.	v.r.	F
Envt.Rep stratum					
Envt	9	1.759E+09	1.955E+08	148.92	
<.001					
Lin	1	1.759E+09	1.759E+09	1340.28	
<.001					
Deviations	8	5.352E-22	6.690E-23	0.00	
1.000					
Residual	20	2.625E+07	1.313E+06	2.09	
Envt.Rep.Geno stratum					
Geno	14	2.097E+07	1.498E+06	2.38	
0.004					
Envt.Geno	126	1.193E+08	9.470E+05	1.51	
0.003					
Lin.Geno	14	1.147E+07	8.192E+05	1.30	
0.205					
Deviations	112	1.079E+08	9.630E+05	1.53	
0.003					
Residual	279(1)	1.755E+08	6.289E+05		
Total	448(1)	2.101E+09			

Анализы регрессии

Дисперсия отклика GEData

Фиксированные условия: Константа + Все виды окр.среды + генотип-1 +
Все виды окр.среды + генотип-1

Итоги анализов

Source	d.f.	s.s.	m.s.	v.r.	F pr.
Regression	29	597215036.	20593622.	68.74	<.001
Residual	120	35950869.	299591.		
Total	149	633165905.	4249436.		

Percentage variance accounted for 92.9

Standard error of observations is estimated to be 547..

Accumulated analysis of variance

Change	d.f.	s.s.	m.s.	v.r.	F
pr.					
+ AllEnvt	1	586402407.	586402407.	1957.35	
<.001					
+ Geno1	14	6989756.	499268.	1.67	
0.072					
+ AllEnvt.Geno1	14	3822873.	273062.	0.91	
0.549					
Residual	120	35950869.	299591.		
Total	149	633165905.	4249436		

5.2 Разделение GXE Int на гетерогенность линейных регрессий при гетерогенных изменениях ошибок

Анализы регрессий

Дисперсия отклика: GEData

Дисперсия веса: AllWet

Фиксированные темины: Константа + Все виды окр.среды + генотип-1+ Все
виды окр.среды + генотип-1

Итоги анализов

Source	d.f.	s.s.	m.s.	v.r.	chi pr
Regression	29	9539.8	328.959	328.96	<.001
Residual	120	165.6	1.380		
Total	149	9705.4	65.137		

Процент дисперсии для 97.9

Стандартная ошибка наблюдений установлена на 1.00.

* MESSAGE: коэффициенты отклонения основаны на параметре дисперсии с ценностью 1.

Накопленный дисперсионный анализ

Change	d.f.	s.s.	m.s.	v.r. chi
pr				
+ AllEnv	1	9442.689	9442.689	9442.69
<.001				
+ Geno1	14	61.356	4.383	4.38
<.001				
+ AllEnv.Geno1	14	35.762	2.554	2.55
0.001				
Residual	120	165.563	1.380	
Total	149	9705.370	65.137	

* MESSAGE: коэффициенты отклонения основаны на параметре дисперсии с ценностью 1

.....
.....

Раздел 6: Общая статистика стабильности

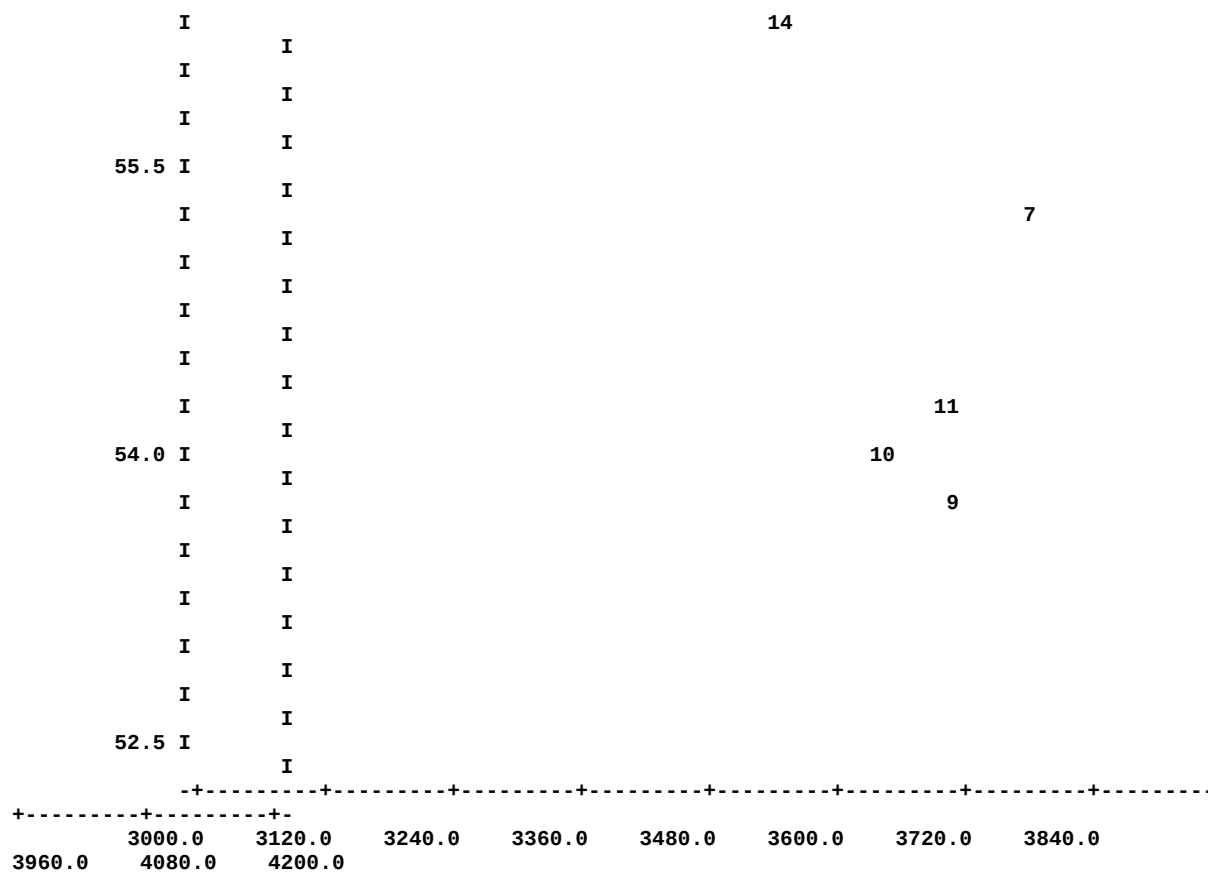
6.1 Индикаторы стабильности при гомогенных изменений ошибок

GenoNum	GenoMean	Slope	SeSlop	Probb1	DeviMS	ProbDev	RSq%	GenoCV
YauH								
1	3167	0.875	0.12606	0.1757	621262	0.0034	83.99	62.19
0.03663								
2	3751	1.062	0.10712	0.2893	448611	0.0323	91.53	61.36
0.02380								
3	3744	1.107	0.06156	0.0597	148129	0.6857	97.29	62.39
0.01585								
4	3737	1.113	0.08666	0.1149	293599	0.1959	94.79	63.55
0.03950								
5	4018	1.097	0.11631	0.2150	528832	0.0116	90.71	59.38
0.02798								
6	3599	1.001	0.08054	0.4966	253568	0.2931	94.46	59.44
0.02250								
7	3784	0.988	0.06195	0.4260	150046	0.6776	96.57	55.27
0.01113								
8	3584	1.007	0.09379	0.4698	343905	0.1132	92.70	60.58
0.01921								
9	3713	0.945	0.05525	0.1744	119352	0.8028	97.01	53.77
0.03144								
10	3635	0.929	0.05189	0.1041	105270	0.8543	97.26	53.93
0.01285								
11	3694	0.920	0.10187	0.2266	405679	0.0549	89.94	54.37
0.03123								
12	4122	1.122	0.05947	0.0370	138238	0.7270	97.53	57.37
0.01526								
13	3810	1.007	0.08812	0.4703	303553	0.1764	93.50	56.74
0.03940								
14	3542	0.925	0.08468	0.2013	280336	0.2248	92.93	56.23
0.01360								
15	3458	0.902	0.09509	0.1671	353476	0.1015	90.82	56.75
0.01741								

GenoNum	GenoMean	Wricke	Pla_Pet	Plaisted	Shukla
1	3167	5578639	489894	288860	690928
2	3751	3739272	380408	305704	455112
3	3744	1635091	255159	324973	185345
4	3737	2844787	327165	313895	340434
5	4018	4596131	431412	297857	564966
6	3599	2028564	278580	321370	235790
7	3784	1205943	229615	328903	130326
8	3584	2753349	321722	314733	328712
9	3713	1073016	221702	330120	113284
10	3635	1039503	219707	330427	108988
11	3694	3497666	366027	307917	424137
12	4122	1688955	258365	324480	192251
13	3810	2430222	302488	317692	287285
14	3542	2461612	304357	317404	291309
15	3458	3200992	348368	310633	386102

[illegible]

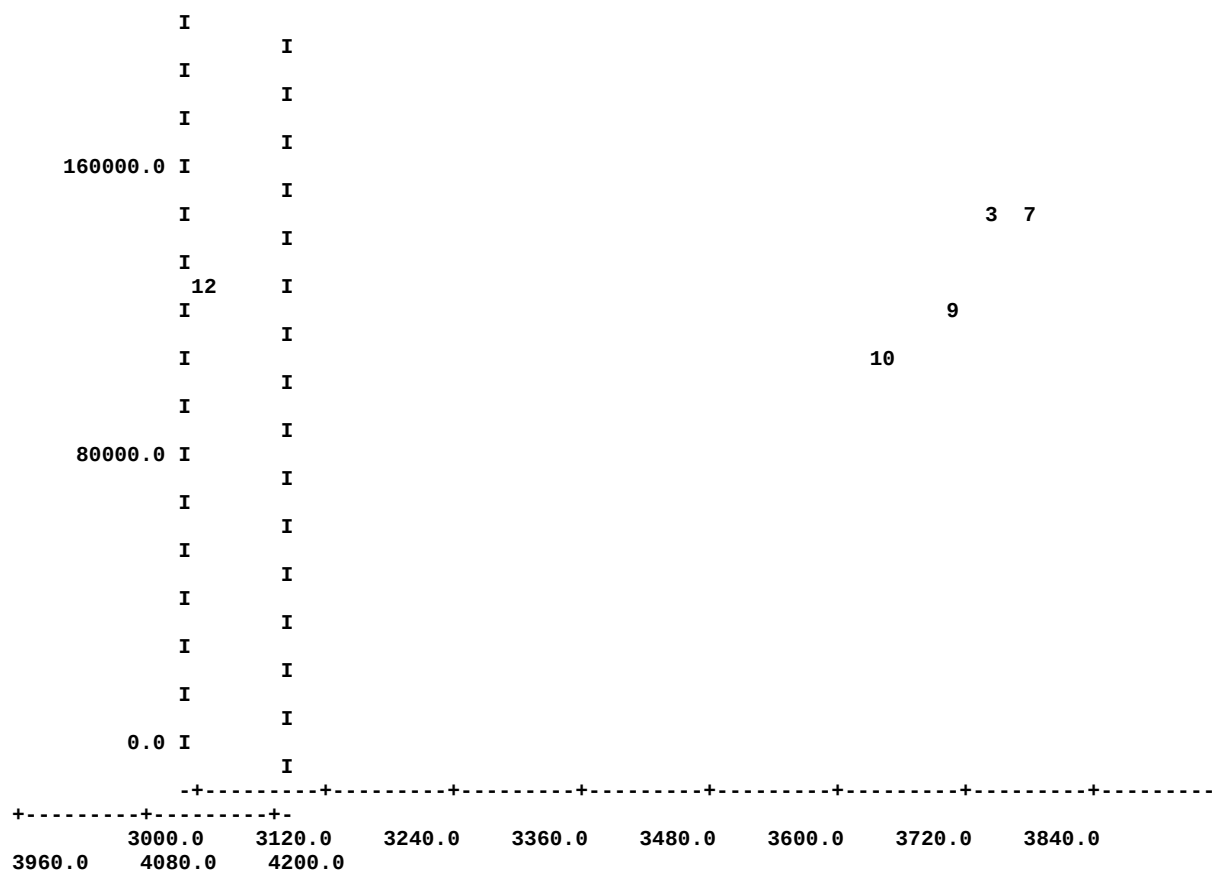
[illegible]



Генотип CV v. Средний генотип с применением фактора №

генотипа

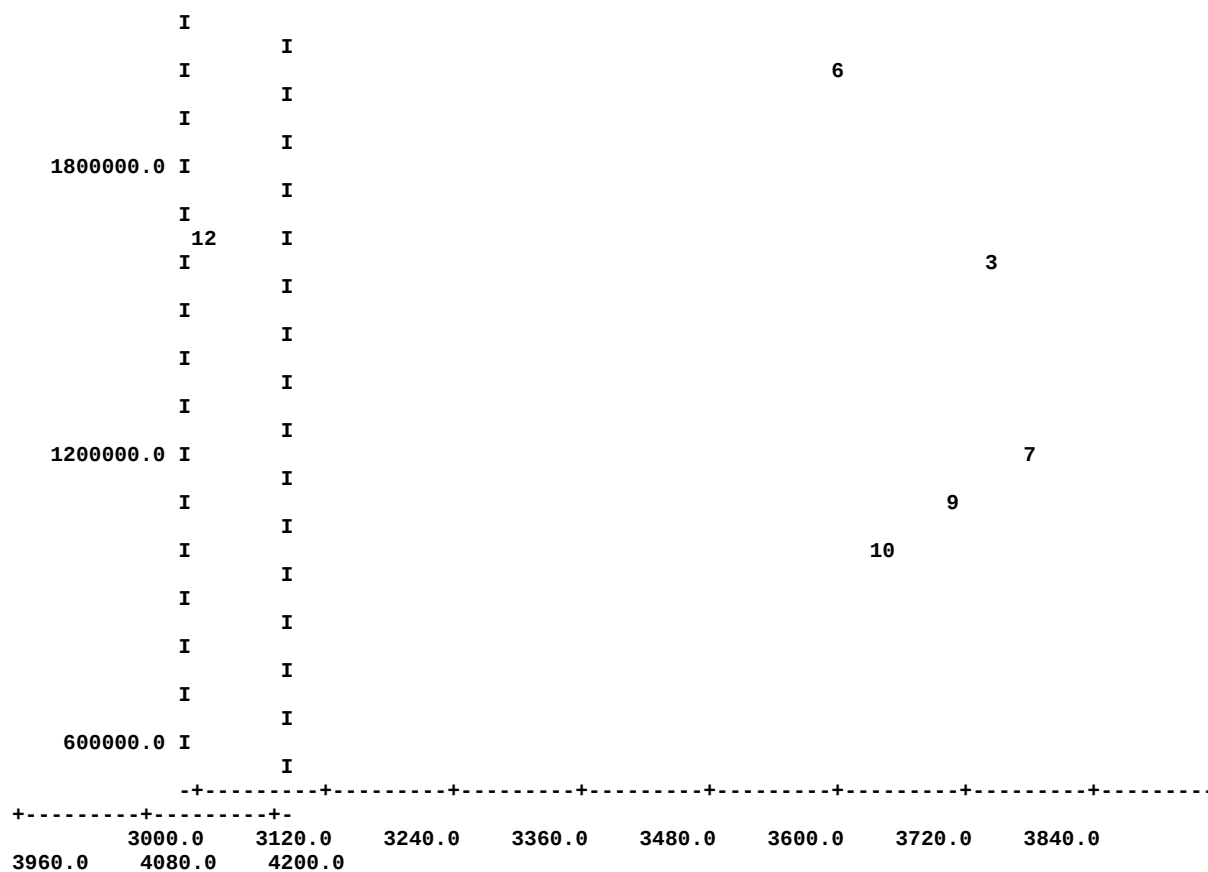
[illegible]



ОтклонениеMS v. Средние генотипа с применением фактора

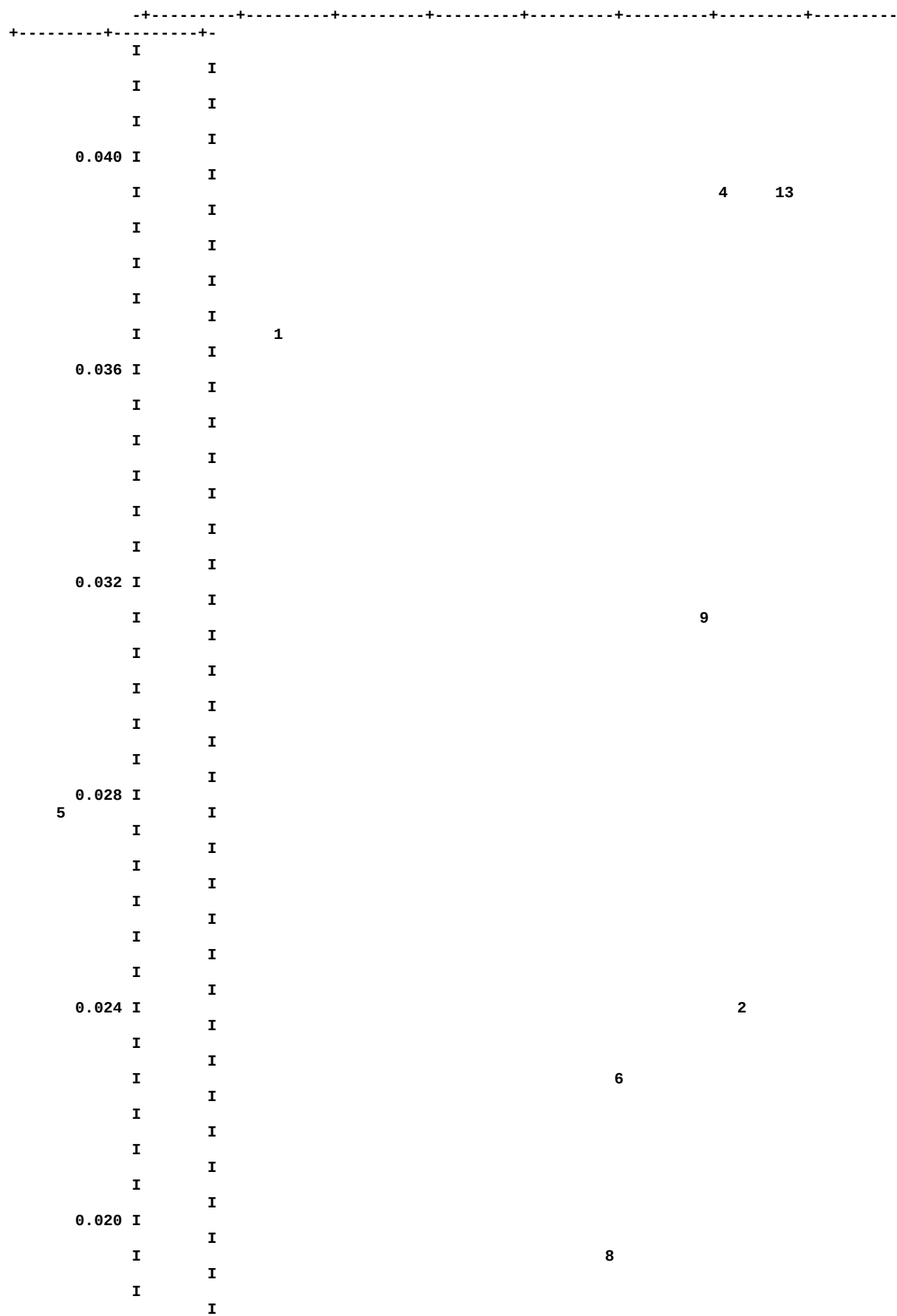
№ Генотипа

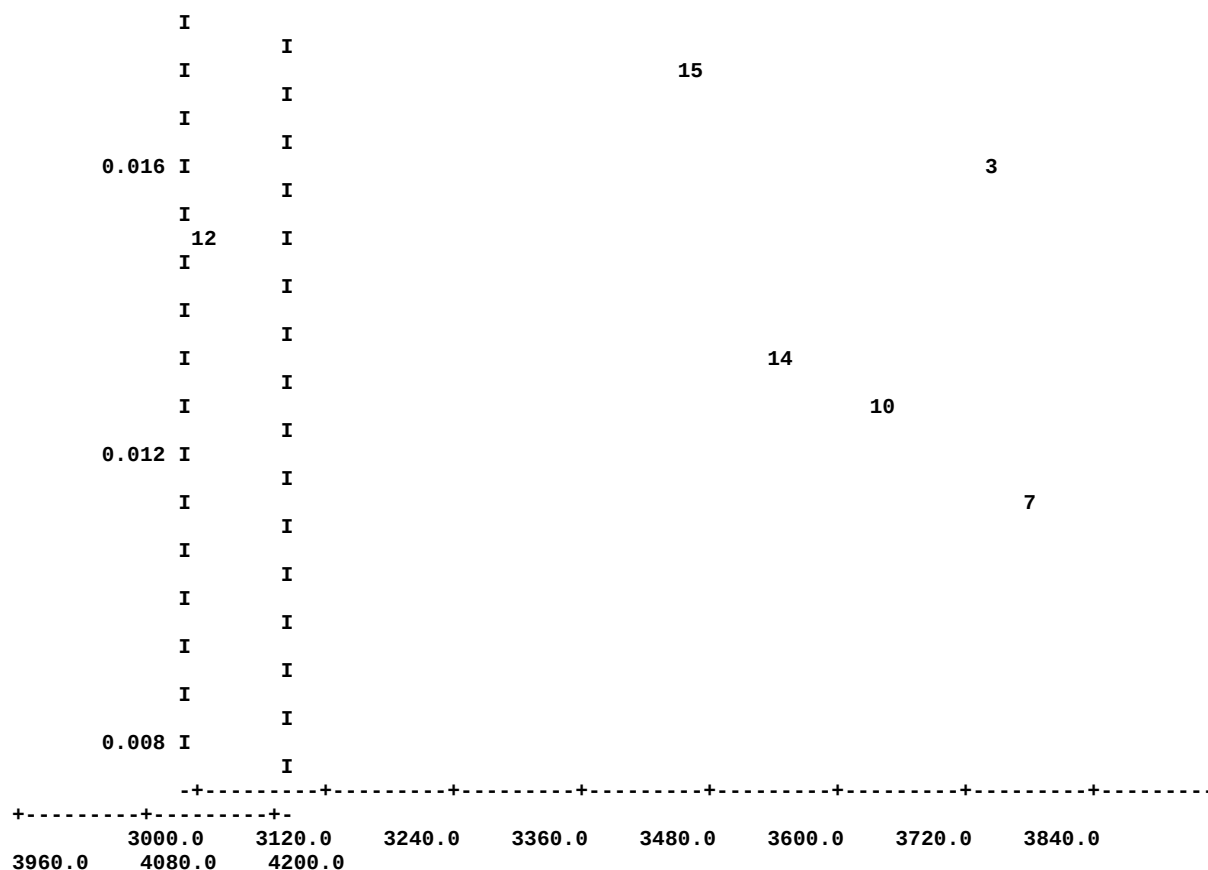
[illegible]



Рик V. Средние генотипа с применением фактора №

Генотипа





Яу V. Средние генотипа с применением фактора №

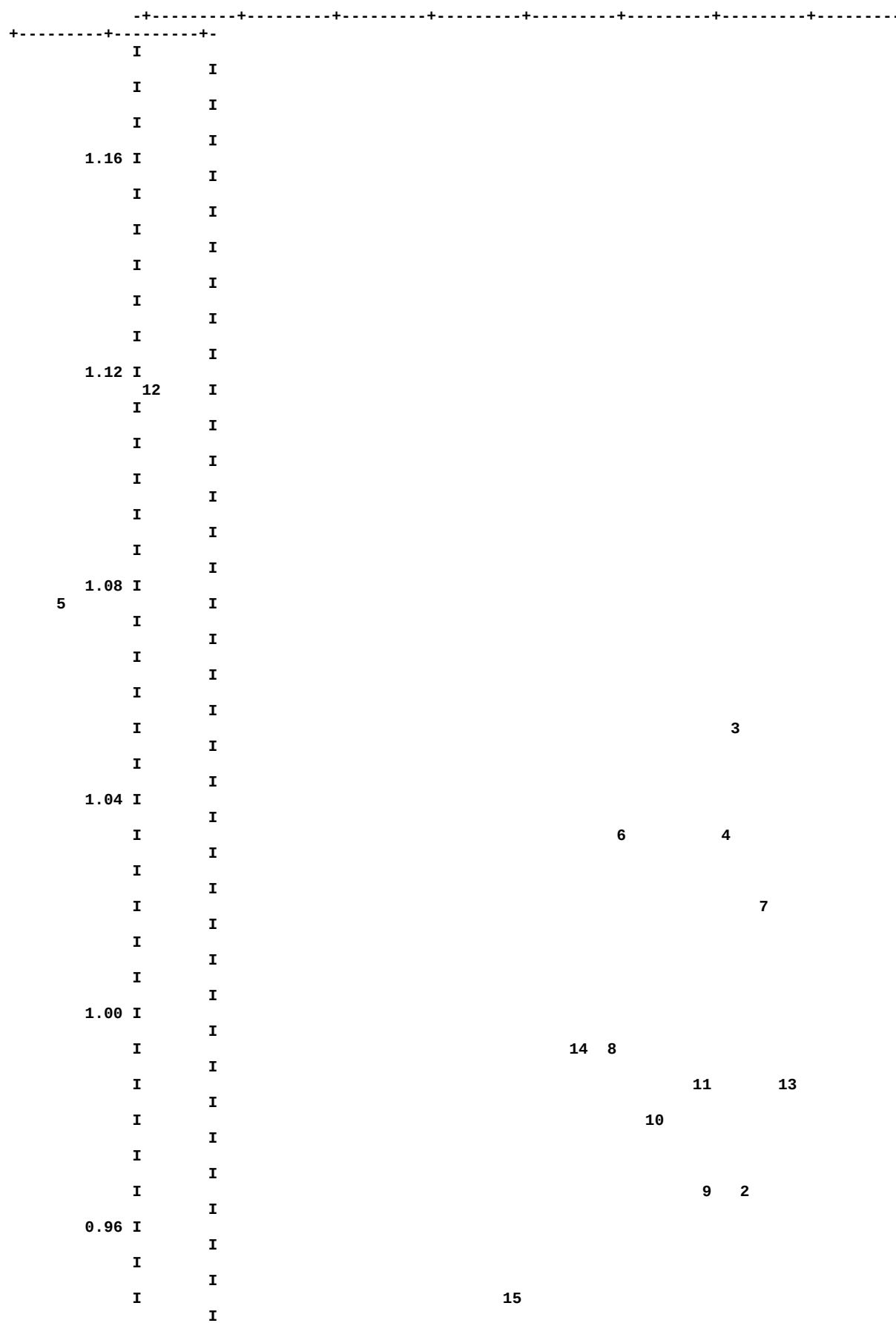
Генотипа

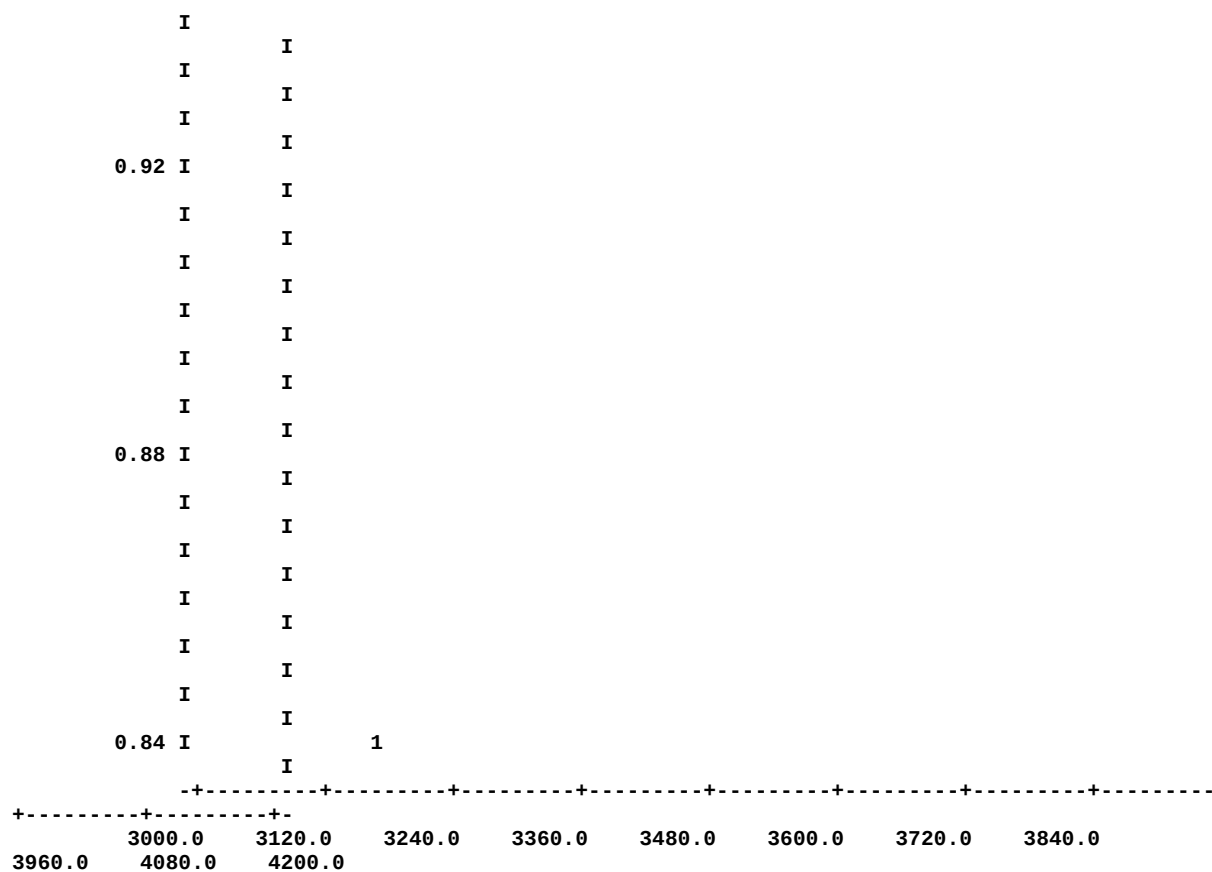
Точки совпадающие с 9

11

6.2 Индикаторы стабильности при гетерогенных дисперсий ошибок

GenoNum	GenoMean	SlopeW	SeSlopW	Probb1W	DeviSSW	ProbDevW	RSqW%
1	3167	0.841	0.06573	0.0209	21.76	0.0054	94.76
2	3751	0.968	0.05322	0.2809	14.26	0.0751	97.34
3	3744	1.052	0.03527	0.0890	6.27	0.6175	99.00
4	3737	1.036	0.04158	0.2080	8.71	0.3677	98.57
5	4018	1.079	0.05987	0.1105	18.05	0.0208	97.30
6	3599	1.034	0.04771	0.2508	11.46	0.1768	98.11
7	3784	1.022	0.03232	0.2558	5.26	0.7294	99.11
8	3584	0.992	0.05240	0.4402	13.83	0.0863	97.54
9	3713	0.969	0.03524	0.2012	6.26	0.6187	98.82
10	3635	0.982	0.03735	0.3218	7.03	0.5338	98.71
11	3694	0.984	0.04245	0.3550	9.08	0.3359	98.35
12	4122	1.117	0.05394	0.0308	14.65	0.0663	97.94
13	3810	0.986	0.04448	0.3816	9.97	0.2674	98.20
14	3542	0.993	0.04376	0.4367	9.65	0.2908	98.28
15	3458	0.946	0.04308	0.1227	9.34	0.3141	98.16





Уклон W V. Средние генотипа с применением фактора №

Генотипа

6.3 Корреляции между различными индексами

GenoMean	1.0000						
Slope	0.7633	1.0000					
DeviMS	-0.3673	-0.1752	1.0000				
GenoCV	-0.1419	0.5082	0.4421	1.0000			
Wricke	-0.3515	-0.1330	0.9875	0.4842	1.0000		
Pla_Pet	-0.3515	-0.1330	0.9875	0.4842	1.0000		
1.0000							
Plaisted	0.3515	0.1330	-0.9875	-0.4842	-1.0000	-	
1.0000	1.0000						
Shukla	-0.3515	-0.1330	0.9875	0.4842	1.0000		
1.0000	-1.0000	1.0000					
YauH	-0.1174	0.0041	0.5005	0.2849	0.4971		
0.4971	-0.4971	0.4971					
SlopeW	0.8758	0.7940	-0.4722	0.0054	-0.4434	-	
0.4434	0.4434	-0.4434					
DeviSSW	-0.1513	0.0485	0.7885	0.4775	0.8104		
0.8104	-0.8104	0.8104					
	GenoMean	Slope	DeviMS	GenoCV	Wricke		
Pla_Pet	Plaisted	Shukla					
YauH	1.0000						
SlopeW	-0.3188	1.0000					
DeviSSW	0.3063	-0.2228	1.0000				
	YauH	SlopeW	DeviSSW				

RGenoMn	1.0000					
RSlope	0.7393	1.0000				
RDeviMS	-0.1893	-0.2071	1.0000			
RGenoCV	-0.0143	0.4857	0.4036	1.0000		
RWricke	-0.2000	-0.1357	0.9679	0.4750	1.0000	
RPla_Pet	-0.2000	-0.1357	0.9679	0.4750	1.0000	
1.0000						
RPlaist	0.2000	0.1357	-0.9679	-0.4750	-1.0000	-
1.0000	1.0000					
RShukla	-0.2000	-0.1357	0.9679	0.4750	1.0000	
1.0000	-1.0000	1.0000				
RYauH	0.0179	0.0464	0.4929	0.3036	0.5071	
0.5071	-0.5071	0.5071				
	RGenoMn	RSlope	RDeviMS	RGenoCV	RWricke	
RPla_Pet	RPlaist	RShukla				
RYauH	1.0000					
	RYauH					

Раздел 7: Иерархическое объединение в кластеры генотипов

7.1: Иерархическое объединение в кластеры генотипов с использованием

a) Матрицы общности/расстояния, основанную на: расстояние Эвклида

b) Метода объединения в кластеры:
Скапливающийся метод, основанный на
Функцию связи: Среднее соединение (также **UPGMA**: невзвешенный метод группы пары, используя арифметические средние числа)
Здесь, общность между кластером и двумя слитыми кластерами
- среднее число общих черт
кластеров с каждым из двух.

Среднее соединение анализов кластера

Иерархические кластеры

Level 95.0

8 10

11 13

Ungrouped

1 2 3 9 14 15 4 6 5 7
12

Level 90.0

2 3

9 14 15

4 6 8 10

5 7

11 13

Ungrouped

1 12

Level 85.0

2 3 9 14 15

4 6 8 10 5 7

11 13 12

Ungrouped

1

Level 80.0

2	3	9	14	15	4	6	8	10	5
7	11	13	12						

Ungrouped
1

Level 75.0

2	3	9	14	15	4	6	8	10	5
7	11	13	12						

Ungrouped
1

Level 70.0

1	2	3	9	14	15	4	6	8	10
5	7	11	13	12					

Dendrogram

** Levels 100.0 90.0 80.0 70.0

```

1 .....
2 ..... )
3 ..... )
9 ..... )
14 ..... )
15 ..... )
4 ..... )
6 ..... )
8 .. )
10 .. )
5 ..... )
7 ..... )
11 .. )
13 .. )
12 ..... )

```

7.2: Иерархическое объединение в кластеры генотипов с использованием

а) Матрицы общности/расстояния, основанную на: расстояние Эвклида

б) Метода объединения в кластеры:

Скапливающийся метод, основанный на

Функцию связи: Среднее соединение (также UPGMA: невзвешенный метод группы пары, используя арифметические средние числа)

Здесь, общность между двумя кластерами - среднее число принятое выше всех единиц в двух слитых кластерах.

Среднее соединение анализов кластера

Иерархические кластеры

Level 95.0

8 10

11 13

Ungrouped

1 2 3 9 14 15 4 6 5 7
12

Level 90.0

2 3

9 14 15

4 6 8 10

5 7

11 13

Ungrouped

1 12

Level 85.0

2 3 9 14 15

4 6 8 10 5 7

11 13 12

Ungrouped

1

Level 80.0

2 3 9 14 15 4 6 8 10 5
7 11 13 12

Ungrouped

1

Level 75.0

1 2 3 9 14 15 4 6 8 10
5 7 11 13 12

Dendrogram

**** Levels 100.0 90.0 80.0**

```

1 .....
2 ..... )
3 ..... )
9 ..... )
14 ..... )
15 ..... )
4 ..... )
6 ..... )
8 .. )
10 .. )
5 ..... )
7 ..... )
11 .. )
13 .. )
12 ..... )

```

7.3: Иерархическое объединение в кластеры генотипов с использованием

а) Матрицы общности/расстояния, основанную на: расстояние
Эвклида

б) Метода объединения в кластеры:

Скапливающийся метод, основанный на Функцию связи: Полное соединение
(также самые отдаленный сосед)

Здесь, расстояние между двумя кластерами было взято как
максимальное расстояние между любыми из этих двух единиц в этих кластерах.

Среднее соединение анализов кластера

Иерархические кластеры

Level 95.0

8 10

11 13

Ungrouped

1 2 3 9 14 15 4 6 5 7

12

Level 90.0

2 3

14 15

4 6

5 7

8 10

11 13

Ungrouped

1 9 12

Level 85.0

2 3 9 14 15

4 6 5 7

8 10

11 13

Ungrouped

1 12

Level 80.0

2 3 9 14 15

4 6 5 7 8 10

11 13 12

Ungrouped

1

Level 75.0

1 2 3 9 14 15

4 6 5 7 8 10

11 13 12

Level 70.0

1 2 3 9 14 15

4 6 5 7 8 10 11 13 12

Level 65.0

1	2	3	9	14	15				
4	6	5	7	8	10	11	13	12	

Level 60.0

1	2	3	9	14	15	4	6	5	7
8	10	11	13	12					

Древовидная схема

** Levels 100.0 90.0 80.0 70.0 60.0

1					
2		)			
3	)	..	)		
9	)		)		
14		)	)		
15	)	..)	)		
4				)	
6	)	..		)	
5		)		)	
7	)	..)	..	)	
8	..	)		)	
10	..)	)		)	
11	..		)	)	
13	..)		)	)	
12	)	)	)		

7.4: Иерархическое объединение в кластеры генотипов с использованием

а) Матрицы общности/расстояния, основанную на: расстояние Эвклида

б) Метода объединения в кластеры:

Скапливающийся метод, основанный на Функцию связи: единое соединение (также самый отдаленный сосед)

Здесь, расстояние между двумя кластерами было взято как наименьшее расстояние между любыми из этих двух образцов/единиц из различных кластеров.

Единое соединение анализов кластера

Минимум покрывающего дерева

Similarity matrix: Simi

```

1..... 15..... 3..... 2
  86.1 ( 90.8 91.5
      (
      (..... 11..... 13..... 12
      ( 91.0 95.2 88.9
      (
      (..... 14..... 9..... 7..... 5..... 4..... 6
        94.3 91.3 92.4 93.8 91.7 ( 93.9
                                   (
                                   (.....
8..... 10
                                   93.4

```

95.3

Total length: 1289.7

Hierarchical clusters

Level 95.0

8 10

11 13

Ungrouped

1 15 14 9 7 5 4 6 3 2
12

Level 90.0

15 14 9 7 5 4 6 8 10 11
13 3 2

Ungrouped

1 12

Level 85.0

1	15	14	9	7	5	4	6	8	10
11	13	3	2	12					

Dendrogram

**** Levels 100.0 90.0**

```

1 .....
15 ..... )
14 ..... )
9 ..... )
7 ..... )
5 ..... )
4 ..... )
6 ..... )
8 .. )
10 ..)..) )
11 .. )
13 ..)..) )
3 ..... )
2 .....)..)
12 .....).....

```

Раздел 8: Иерархическое объединение в кластеры окружающей среды
 8.1: Иерархическое объединение в кластеры генотипов с использованием

а) Матрицы общности/расстояния, основанную на: расстояние Эвклида

б) Метода объединения в кластеры:

Скапливающийся метод, основанный на

функцию связи: Средняя связь (также НМПГИАС (UPGMA): невзвешенный метод пары-группы, используя арифметические средние)

Здесь, схожесть между кластерами и двумя слитыми кластерами в целом сходиться с каждым из двух.

Анализы кластеров среднего соединения

Иерархические кластеры

Level 95.0

1 6 8

2 3 9 4

5 10

Ungrouped

7

Level 90.0

1 6 8

2 3 9 4

5 10 7

Level 85.0

1 6 8

2 3 9 4

5 10 7

Level 80.0

1 6 8

2 3 9 4 5 10 7

Level 75.0

1 6 8

```

      2      3      9      4      5      10      7

Level      70.0

      1      6      8

      2      3      9      4      5      10      7

Level      65.0

      1      6      8

      2      3      9      4      5      10      7

Level      60.0

      1      6      8

      2      3      9      4      5      10      7

Level      55.0

      1      6      8      2      3      9      4      5      10      7

```

Dendrogram

```

** Levels  100.0  90.0  80.0  70.0  60.0

      1  ..
      6  ..)
      8  ..).....)
      2  ..                               )
      3  ..)                             )
      9  ..)                             )
      4  ..).....)
      5  ..                               )
     10  ..).....)
      7  .....).....).....).....).....

```

8.2: Иерархическое объединение в кластеры генотипов с использованием

а) Матрицы общности/расстояния, основанную на: расстояние Эвклида

б) Метода объединения в кластеры:

Скапливающийся метод, основанный на

функцию связи: Средняя группы

Здесь, схожесть между двумя кластерами является средним, взятым выше всех единиц в двух слитых кластерах.

Групповые анализы средних кластеров

Иерархические кластеры

Level 95.0

1 6 8

2 3 9 4

5 10

Ungrouped

7

Level 90.0

1 6 8

2 3 9 4

5 10 7

Level 85.0

1 6 8

2 3 9 4 5 10 7

Level 80.0

1 6 8

2 3 9 4 5 10 7

Level 75.0

1 6 8

2 3 9 4 5 10 7

Level 70.0

1 6 8

2 3 9 4 5 10 7

Level 65.0

1 6 8

2 3 9 4 5 10 7

Level 60.0

1 6 8

2 3 9 4 5 10 7

Level 55.0

1 6 8 2 3 9 4 5 10 7

Dendrogram

** Levels 100.0 90.0 80.0 70.0 60.0

```
1  ..
6  ..)
8  ..).....)
2  ..).....)
3  ..).....)
9  ..).....)
4  ..).....)
5  ..).....)
10 ..).....)
7  .....).....)
```


8.3: Иерархическое объединение в кластеры генотипов с использованием

a) Матрицы общности/расстояния, основанную на: расстоянии Эвклида

b) Метода объединения в кластеры:

Скапливающийся метод, основанный на Функцию связи: Полные связи (также самый отдаленный сосед)

Здесь, расстояние между двумя кластерами принято как максимальное расстояние между любыми двумя единицами в этих кластерах.

Полные анализы связей кластеров

Иерархические кластеры

Level 95.0

1	6	8	
2	3	9	4
5	10		

**Ungrouped
7**

Level 90.0

1	6	8	
2	3	9	4
5	10	7	

Level 85.0

1	6	8	
2	3	9	4
5	10	7	

Level 80.0

1	6	8	
2	3	9	4
5	10	7	

Level 75.0

1	6	8	
2	3	9	4

	5	10	7				
Level		70.0					
1	6	8					
2	3	9	4				
5	10	7					
Level		65.0					
1	6	8					
2	3	9	4	5	10	7	
Level		60.0					
1	6	8					
2	3	9	4	5	10	7	
Level		55.0					
1	6	8					
2	3	9	4	5	10	7	
Level		50.0					
1	6	8					
2	3	9	4	5	10	7	
Level		45.0					
1	6	8					
2	3	9	4	5	10	7	
Level		40.0					
1	6	8					
2	3	9	4	5	10	7	
Level		35.0					
1	6	8					
2	3	9	4	5	10	7	
Level		30.0					

Минимум покрывающего дерева

Similarity matrix: Simi

```
1..... 6..... 8
( 99.7      97.8
(
(..... 7..... 5..... 10..... 4..... 3..... 2
 94.2      94.0      97.5      94.5      97.1 ( 97.4
(
(..... 9
 97.5
```

Total length: 869.7

Hierarchical clusters

Level 95.0

1 6 8

5 10

4 3 9 2

Ungrouped

7

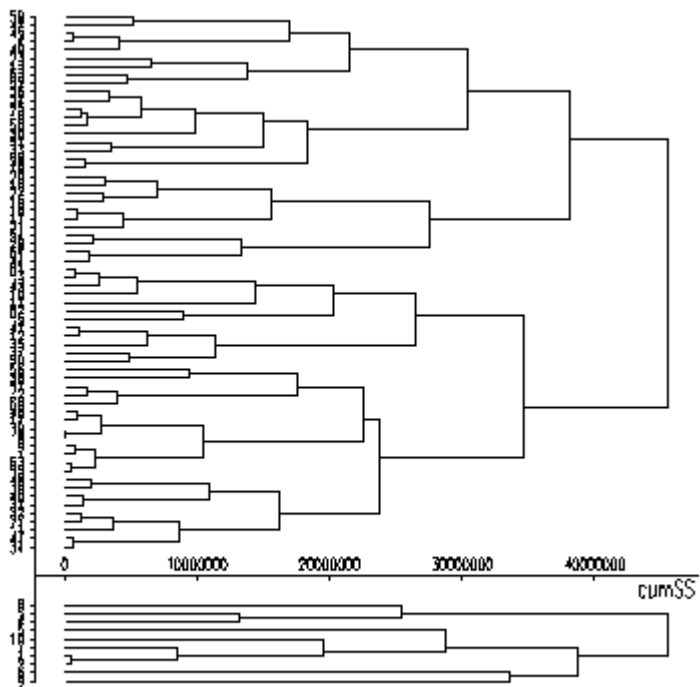
Level 90.0

1 6 8 7 5 10 4 3 9 2

Dendrogram

** Levels 100.0 90.0

```
1 ..
6 ..)
8 ..)..
7 .....)
5 .. )
10 ..)..)
4 .. )
3 ..) )
9 ..) )
2 ..)..).....
```



***** Principal components analysis *****

*** Latent Roots ***

	1	2	3
	19652260	11339087	5721755
*** Percentage variation ***			
	1	2	3
	42.02	24.25	12.24

*** Trace ***

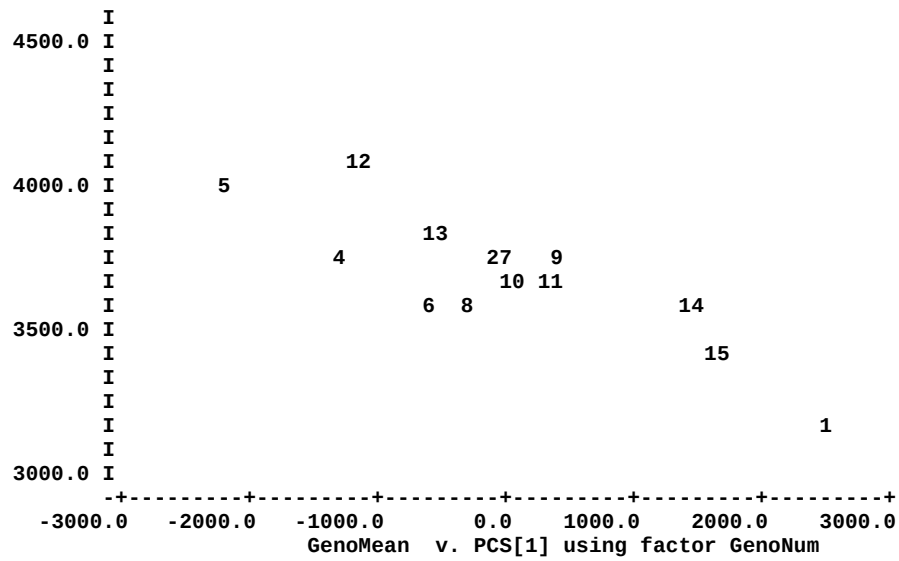
46763498

*** Latent Vectors (Loadings) ***

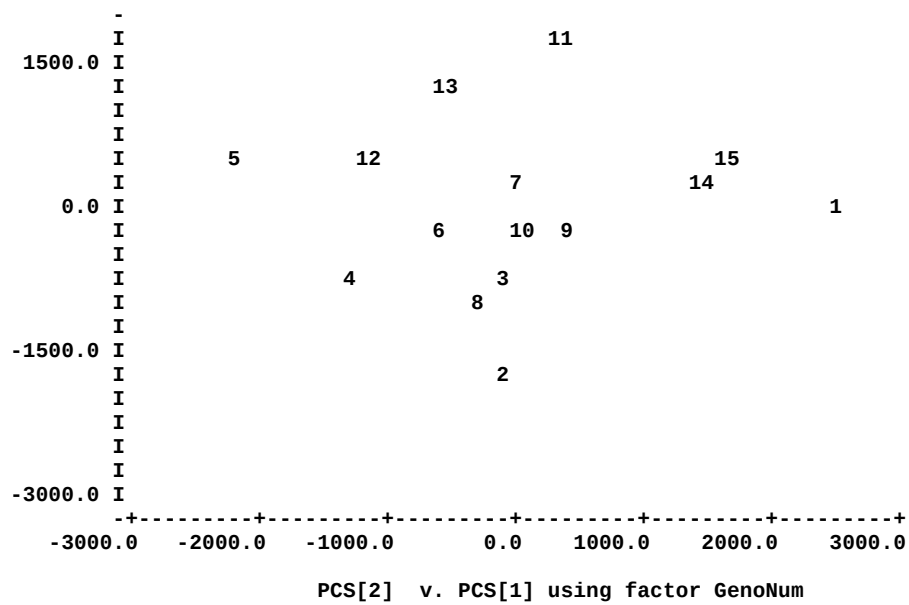
	1	2	3
Mean[1]	-0.01595	-0.00287	0.11513
Mean[2]	0.12483	0.08662	-0.06460
Mean[3]	-0.34907	0.30239	0.10301
Mean[4]	-0.29299	-0.66406	-0.60624
Mean[5]	-0.31794	0.63791	-0.63810
Mean[6]	-0.05731	0.06721	0.02572
Mean[7]	-0.14832	0.13131	0.24152
Mean[8]	0.02140	0.07354	0.01506
Mean[9]	-0.78072	-0.16108	0.34760
Mean[10]	0.20030	-0.00525	-0.13124

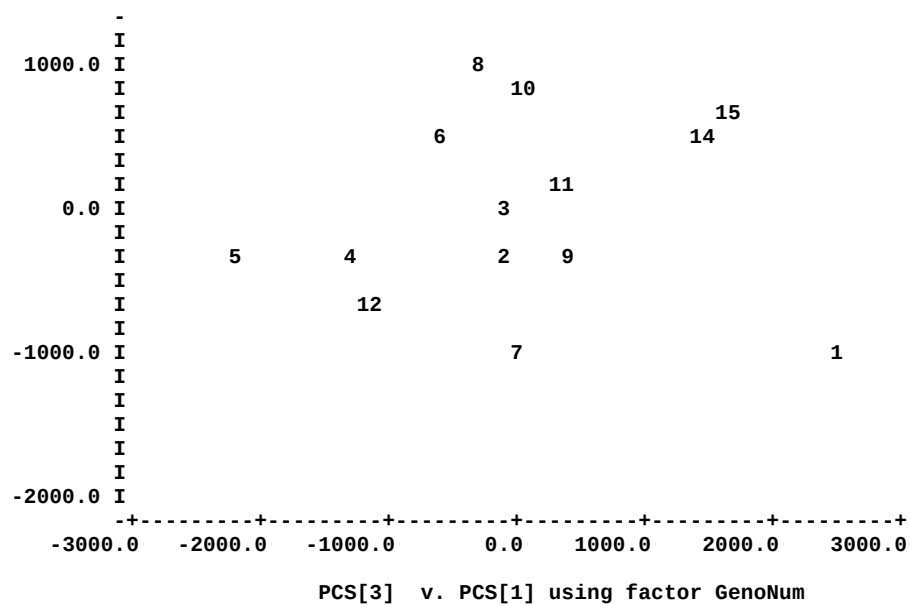
*** Significance tests for equality of final K roots ***

No. (K) Roots	Chi squared	df
2	11.51	2
3	32.02	5
4	37.26	9
5	42.67	14
6	59.62	20
7	73.35	27
8	84.70	35
9	105.01	44
10	129.82	54

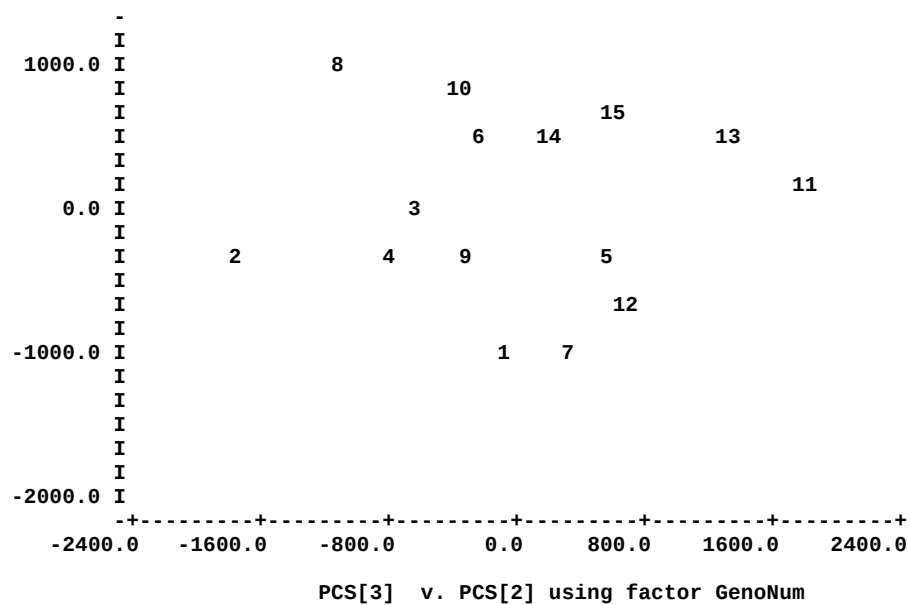


Points coinciding with 2
3





Points coinciding with 6
13



12. Principal component analysis for environments "

***** Principal components analysis *****

*** Latent Roots ***

1	2	3
590271617	13246608	10841600

*** Percentage variation ***

1	2	3
94.27	2.12	1.73

*** Trace ***

626176149

*** Latent Vectors (Loadings) ***

	1	2	3
GMean[1]	-0.22466	-0.52192	0.03860
GMean[2]	-0.27343	-0.05691	0.53187
GMean[3]	-0.28493	-0.10886	0.17753
GMean[4]	-0.28683	0.32571	0.16908
GMean[5]	-0.28277	0.46215	-0.20718
GMean[6]	-0.25788	0.23251	0.01987
GMean[7]	-0.25422	-0.07736	-0.08669
GMean[8]	-0.25951	0.20937	0.29682
GMean[9]	-0.24309	-0.13974	0.13184
GMean[10]	-0.23917	0.08706	0.11551
GMean[11]	-0.23638	-0.07437	-0.51634
GMean[12]	-0.28886	0.05121	-0.17524
GMean[13]	-0.25907	0.09627	-0.40939
GMean[14]	-0.23759	-0.33745	-0.00923
GMean[15]	-0.23171	-0.36133	-0.14441

*** Significance tests for equality of final K roots ***

The last 6 latent roots are (effectively) zero
and have been excluded from the tests.

No. (K)	Chi	
Roots	squared	df
2	0.92	2
3	2.40	5
4	4.48	9
5	11.95	14
6	17.98	20
7	27.14	27
8	33.37	35
9	135.61	44

-
I
4000.0 I
I
I

