УДК 519.233.3:633/635
А. Т. Хохрякова, студентка 1 курса агрономического факультета
Научный руководитель: старший преподаватель кафедры высшей математики Е. Н. Соболева ФГБОУ ВО Ижевская ГСХА

**Проверка статистической гипотезы на примере растениеводства**

В статье методами математической статистики произведена обработка двух выборок: длины пророщенных ростков пшеницы без применения нанокремния и с применением нанокремния. Сформированы интервальные ряды распределения по данным выборкам, подсчитаны выборочные характеристики и выполнены их точечные оценки. А также проверена гипотеза о том, значимо ли различаются средние длины пророщенных ростков пшеницы.

Глобальной проблемой общества является огромное количество информации и данных в повседневной жизни. Данная проблема позволила учёным открыть такую науку, как математическая статистика. Что такое математическая статистика? Какие существуют методы обработки данных и информации? Математическая статистика – это наука, которая изучает математические методы обработки и использования статистических данных для научных и экспериментальных выводов. В современном мире существует много разных методов. Например, корреляционный анализ, регрессионный анализ, статистическое оценивание параметров распределения и проверка статистических гипотез. Можно выделить такой актуальный метод, как проверка статистических гипотез. Данный метод основан на том, чтобы проверить согласуются ли выборочные данные с выдвинутой гипотезой при исследовании какого-либо признака.[1-7] В нашем эксперименте были измерены длины пророщенных ростков пшеницы без применения нанокремния и с применением нанокремния.

*Цель эксперимента* заключается в том, чтобы определить значимо или незначимо различаются длины пророщенных ростков пшеницы.

*Задачи эксперимента:*

1. Изучить основные понятия темы «Проверка статистических гипотез»;
2. Обработать данные по выборкам объёмом 50:
   а) Построить интервальные ряды распределения по выборкам;
   б) Вычислить выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение;
3. На основе полученных данных проверить гипотезу о равенстве двух средних генеральных совокупностей (независимые выборки) при уровне значимости 0,01;
4. Сделать практические выводы.

*Материалы и методы.* Эксперимент проводился на злаковом растении пшеница (лат. *Tríticum L.*) в лабораторных условиях. Были пророщены семена пшеницы с применением нанокремния и без применения нанокремния (рис.1, рис.2)

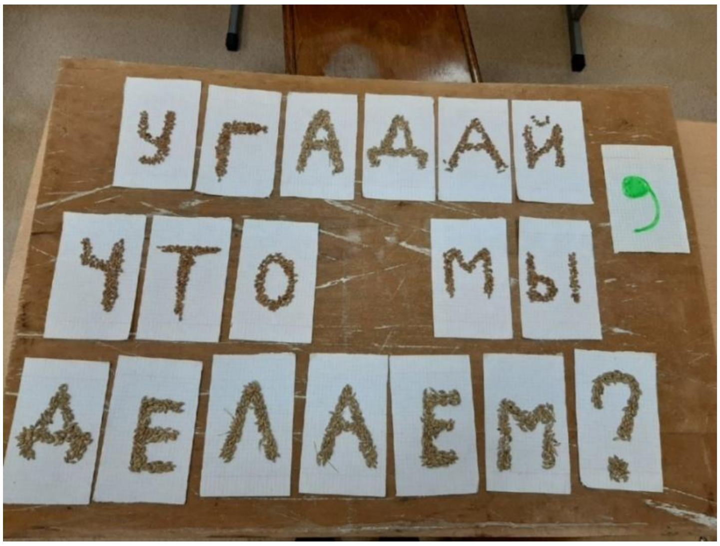


Рисунок 1 – **семена злаковых (ячмень и пшеница) до прорастания и обработки нанокремнием**

Рисунок 2 – **пророщенные семена пшеницы в виде ростков без обработки и с обработкой нанокремнием**

У ростков пшеницы (лат. *Tríticum L.*) измерены длины, которые представлены выборками в таблице 1. Объём выборки: n=50.

Таблица 1 **- Протокол измерений длин ростков пшеницы**

| **Длины необработанных семян пшеницы, см** | **Длины обработанных семян пшеницы, см** |
|---|---|
| Ростки | Ростки |
| 12,50 | 8,30 |
| 9,50 | 9,20 |
| 19,10 | 12,30 |
| 3,70 | 9,90 |
| 4,30 | 11,60 |
| 0,00 | 11,20 |
| 5,70 | 10,60 |
| 8,00 | 10,30 |
| 5,80 | 11,40 |
| 5,90 | 10,10 |
| 19,60 | 9,60 |
| 7,50 | 12,00 |
| 10,50 | 11,80 |
| 12,80 | 13,00 |
| 17,60 | 12,80 |
| 6,50 | 11,00 |
| 19,90 | 11,70 |
| 8,20 | 13,40 |
| 19,50 | 12,00 |
| 7,30 | 12,50 |
| 17,40 | 10,00 |
| 7,20 | 14,00 |
| 6,40 | 13,90 |
| 11,10 | 14,70 |
| 22,30 | 12,30 |
| 4,00 | 12,10 |
| 5,40 | 13,00 |
| 24,60 | 11,40 |
| 18,40 | 11,30 |
| 8,40 | 12,30 |
| 7,60 | 12,90 |
| 4,00 | 12,00 |
| 5,80 | 13,00 |
| 10,00 | 12,70 |
| 4,50 | 12,60 |
| 21,50 | 12,40 |
| 15,50 | 10,50 |
| 19,40 | 11,60 |
| 15,60 | 12,00 |
| 26,20 | 11,90 |
| 17,30 | 12,10 |
| 19,50 | 10,10 |
| 18,10 | 11,70 |
| 14,30 | 11,20 |

| | |
|---|---|
| 22,00 | 12,80 |
| 22,00 | 10,90 |
| 4,70 | 11,10 |
| 8,50 | 13,70 |
| 9,20 | 14,10 |
| 13,10 | 13,00 |

Пусть выборка X – это длина пророщенных ростков пшеницы без применения нанокремния, а выборка Y – это длина пророщенных ростков пшеницы с применением нанокремния. На основе этих выборок необходимо составить интервальный ряд распределения.

Выборка X (таб.2):

1) определим размах вариации R= $x_{max}$ - $x_{min}$ = 26,2 – 0,0 = 26,2;
2) определим длину интервала, то есть шаг h=R/k, где k – количество интервалов; h = 26,2/6 = 4,4. Длина интервалов равна 6 в связи с тем, что объём выборки равен 50.
3) Формируем интервалы ($x_i$; $x_{i+1}$), следовательно,
   ($x_1$; $x_2$) = ($x_{min}$; $x_{min}$ + h); ($x_2$; $x_3$) = ($x_2$; $x_2$+h); ($x_3$; $x_4$) = ($x_3$; $x_3$+h);
   ($x_4$; $x_5$) = ($x_4$; $x_4$+h); ($x_5$; $x_6$) = ($x_5$; $x_5$+h); ($x_6$; $x_7$) = ($x_6$; $x_6$+h).

Таблица 2 – **Интервальный ряд выборки X**

| № | $(X_i;X_{i+1})$ | $n_i$ | $W_i$ | $X_{сер}$ | $n_h$ |
|---|---|---|---|---|---|
| 1 | (0; 4,4) | 5 | 0,1 | 2,2 | 5 |
| 2 | (4,4; 8,8) | 17 | 0,34 | 6,6 | 22 |
| 3 | (8,8; 13,2) | 8 | 0,16 | 11 | 30 |
| 4 | (13,2; 17,6) | 5 | 0,1 | 15,4 | 35 |
| 5 | (17,6; 22) | 10 | 0,2 | 19,8 | 45 |
| 6 | (22; 26,4) | 5 | 0,1 | 24,2 | 50 |

$n_i$ – локальная частота, которая показывает сколько раз встретилась соответствующая варианта; $W_i$ – относительная частота (частость появления i-ой варианты); $X_{сер}$ – середина интервала; $n_h$ – накопленная частота.

Выборка Y (таб. 3):

1) определим размах вариации R= $y_{max}$ -$y_{min}$ = 14,7 – 8,3 = 6,4;
2) определим длину интервала, то есть шаг h=R/k, где k – количество интервалов; h = 6,4/6 = 1,1. Длина интервалов равна 6 в связи с тем, что объём выборки равен 50.

3) Формируем интервалы $(y_i; y_{i+1})$, следовательно,
$(y_1; y_2) = (y_{min}; y_{min} + h)$; $(y_2; y_3) = (y_2; y_2+h)$; $(y_3; y_4) = (y_3; y_3+h)$;
$(y_4; y_5) = (y_4; y_4+h)$; $(y_5; y_6) = (y_5; y_5+h)$; $(y_6; y_7) = (y_6; y_6+h)$.

Таблица 3 – **Интервальный ряд выборки Y**

| **№** | **$(y_i; y_{i+1})$** | **$n_i$** | **$W_i$** | **$y_{сер}$** | **$n_h$** |
|---|---|---|---|---|---|
| 1 | (8,3; 9,4) | 2 | 0,04 | 8,85 | 2 |
| 2 | (9,4;10,5) | 6 | 0,12 | 9,95 | 8 |
| 3 | (10,5; 11,6) | 10 | 0,2 | 11,05 | 18 |
| 4 | (11,6; 12,7) | 18 | 0,36 | 12,15 | 36 |
| 5 | (12,7; 13,8) | 10 | 0,2 | 13,25 | 46 |
| 6 | (13,8; 14,9) | 4 | 0,08 | 14,35 | 50 |

$n_i$ – локальная частота, которая показывает сколько раз встретилась соответствующая варианта; $W_i$ – относительная частота (частость появления i-ой варианты); $y_{сер}$ – середина интервала; $n_h$ – накопленная частота.
По представленным в таблице 2 и таблице 3 данным, вычислим выборочное среднее, выборочную дисперсию (исправленную выборочную дисперсию), выборочное среднее квадратическое отклонение (исправленное выборочное среднее квадратическое отклонение):

1) $\overline{x_В} = \frac{1}{n}\sum_{i=1}^{k}\tilde{x}_i n_i \approx 12{,}1$; $\overline{y_В} = \frac{1}{n}\sum_{i=1}^{k}\tilde{y}_i n_i \approx 11{,}9$ - выборочное среднее

2) $D_В = \frac{1}{n}\sum_{i=1}^{k}\tilde{x}_i^2 n_i - (\bar{x}_В)^2 \approx 47{,}9$; $D_В = \frac{1}{n}\sum_{i=1}^{k}\tilde{y}_i^2 n_i - (\bar{y}_В)^2 \approx 1{,}8$ – выборочная дисперсия

$S_x{}^2 = \frac{n}{n-1}D_В \approx 48{,}9$; $S_y{}^2 = \frac{n}{n-1}D_В \approx 1{,}8$ - исправленная выборочная дисперсия

3) $\sigma = \sqrt{D_В} = \sqrt{47{,}9} \approx 6{,}9$; $\sigma = \sqrt{D_B} = \sqrt{1{,}8} \approx 1{,}3$ – выборочное среднее квадратическое отклонение

4) $S_x = \sqrt{S_x{}^2} = \sqrt{48{,}9} \approx 7{,}0$; $S_y = \sqrt{S_y{}^2} = \sqrt{1{,}8} \approx 1{,}3$ – исправленное выборочное среднее квадратическое отклонение

На основе данных вычислений по выборке X и по выборке Y необходимо проверить гипотезу о равенстве двух средних генеральных совокупностей (независимые выборки). Однако данная гипотеза используется при нормальном законе. Поэтому нужно доказать, что X и Y распределены по нормальному закону, это можно сделать с помощью правила трёх сигм.

Для выборки X:

$P(|x-a|<3\sigma)\approx 0{,}997$, практически все значения случайных величин $x\in N(a;\sigma)$ лежат в интервале $(a-3\sigma; a+3\sigma)$.[1,2]

Так как выборочное среднее $\overline{x_B}=12{,}1$, следовательно, математическое ожидание $M(x)=a=\overline{x_B}=12{,}1$; среднее квадратическое отклонение $\sigma_B\approx 6{,}9$ можем вычислить интервал: $(12{,}1\pm 3*6{,}9)=(12{,}1\pm 20{,}7)\Rightarrow(-8{,}6; 32{,}8)$. Все длины ростков пшеницы без обработки нанокремнием принадлежат интервалу (-8,6; 32,8), следовательно, X распределена по нормальному закону.

Для выборки Y:

$P(|y-a|<3\sigma)\approx 0{,}997$, практически все значения случайных величин $x\in N(a;\sigma)$ лежат в интервале $(a-3\sigma; a+3\sigma)$.[1,2]

Так как выборочное среднее $\overline{x_B}=11{,}9$ , следовательно, математическое ожидание $M(x)=a=\overline{x_B}=11{,}9$; среднее квадратическое отклонение $\sigma_B\approx 1{,}4$ можем вычислить интервал: $(11{,}9\pm 3*1{,}4)=(11{,}9\pm 4{,}2)\Rightarrow(7{,}7; 16{,}1)$. Все длины ростков пшеницы с применением нанокремния принадлежат интервалу (7,7; 16,1), следовательно, Y распределена по нормальному закону.

Генеральные совокупности X и Y распределены нормально, причём известны их исправленные дисперсии. По независимым выборкам X и Y, объёмы которых соответственно равны 50, найдены выборочные средние. Если учесть, что выборочные средние являются несмещёнными оценками генеральных средних, то нулевую гипотезу можно записать так $H_0: \overline{x_Г}=\overline{y_Г}$ (нанокремний незначительно повлияет на длины ростков пшеницы), при этом уровень значимости $\alpha=0{,}01$. А конкурирующую гипотезу можно записать так $H_1: \overline{x_Г}<\overline{y_Г}$ (нанокремний значительно увеличит длину ростков пшеницы).

Необходимо вычислить наблюдаемое значение критерия по формуле:

$T_{набл}=\frac{\overline{x_B}-\overline{y_B}}{\sqrt{\frac{s_x^2}{n_1}+\frac{s_y^2}{n_2}}}$ , где $S^2$ – исправленная дисперсия, n – объём выборки, $\bar{x}_B$ и $\bar{y}_B$ – выборочные средние, тогда $T_{набл}=\frac{12{,}1-11{,}9}{\sqrt{\frac{48{,}9}{50}+\frac{1{,}8}{50}}}\approx 0{,}2$ .

Конкурирующая гипотеза имеет вид $H_1: \overline{x_Г}<\overline{y_Г}$, поэтому критическая область левосторонняя. По таблице функции Лапласа находим $T_{кр.пр.}$, используя неравенство $\phi(T_{кр.пр.})=0{,}5-\alpha=0{,}5-0{,}01=0{,}49\Rightarrow T_{кр.пр.}=2{,}33$. Так как $T_{кр.пр.}=-T_{кр.л.}$, следовательно, $T_{кр.л.}=-2{,}33$.

Строим левостороннюю критическую область (рис.3):

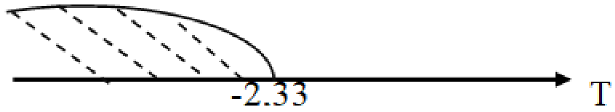


Рисунок 3 – **критическая область**

У нас, наблюдаемое значение признака $T_{набл} = 0,2$ не принадлежит критической области, то есть интервалу $(-\infty; -2,33)$, $T_{набл} > T_{кр.}$. Тем самым нет оснований отвергнуть нулевую гипотезу. То есть раствор нанокремния незначительно повлияет на длину ростков пшеницы.

**Заключение.** В результате исследования была принята нулевая гипотеза, анализ которой позволил заключить, что в сельском хозяйстве использование раствора нанокремения не приведёт к большой урожайности пшеницы.

Список литературы

1) Гмурман В. Е. Теория вероятностей и математическая статистика: учебное пособие для вузов / В. Е. Гмурман. – Москва : Высшая школа, 2003. – 281-303 С.

2) Теория вероятностей и математическая статистика : учебное пособие/ В. С. Мхитарян, Е. В. Астафьева, Ю. Н.Миронкина, Л. И. Трошкин; под ред. В. С. Мхитаряна . – 2-е изд., перера. и доп. – М.: Московский финансово-промышленный университет «Синергия», 2013. – 173 С.

3) Живетин В. Б. Высшая математика. Т.1. Лекции : учебное пособие / В. Б. Живетин . – Москва, 2005. - 535 С.

4) Рыжков И. Б. Основы научных исследований и изобретательства / И. Б. Рыжков. – учебное пособие. – 2-е изд., стер. – СПб. : «Лань», 2013. – 124 С.

5) Математическая статистика : практикум для студентов, обучающихся по направлениям бакалавриата / сост.: С. Я. Пономарёва, Е. Н. Соболева, Т. Р. Галлямова. – Ижевск,2015.

6) Кузнецова О.В., Соболева Е.Н. Нужна ли математика будущему агроному? // Научное и кадровое обеспечение АПК для продовольственного импортозамещения : материалы Всероссийской научно-практической конференции, 16-19 февраля 2016 г. / Ижевская ГСХА. – Ижевск, 2016. – Т. 3. – С. 190-196.

7) Соболева Е.Н., Применение математики при решении прикладных задач в сельскохозяйственном вузе // Е.Н. Соболева / Инновационные технологии для реализации программы научно-технического развития сельского хозяйства: материалы Международной научно-практической конференции 13-16 февраля 2018 года, г. Ижевск. В 3 т. – Ижевск: ФГБОУ ВО Ижевская ГСХА, 2018. – Т. 3. – С. 264-269.