ТЕОРЕТИЧЕСКИЕ РАСПРЕДЕЛЕНИЯ

1. Нормальное распределение
2. t – распределение Стьюдента
3. F – распределение Фишера
4. χ^2– распределение
5. Распределение Пуассона

Различают эмпирические и теоретические распределения частот совокупности результатов наблюдений.

Эмпирическое распределение – распределение результатов измерений, полученных при изучении выборки (длина колоса, масса клубня и т.п.). В основе эмпирического распределения лежат определенные математические закономерности, которые при большом числе наблюдений характеризуются некоторыми теоретическими распределениями.

1. ***Нормальное распределение***. Наиболее часто в исследовательской работе используют нормальное распределение, или закон нормального распределения. Нормальным, или гауссовым называют распределение случайной независимой переменной величины X в однородной совокупности, которое описывается уравнением Гаусса:

$$Y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

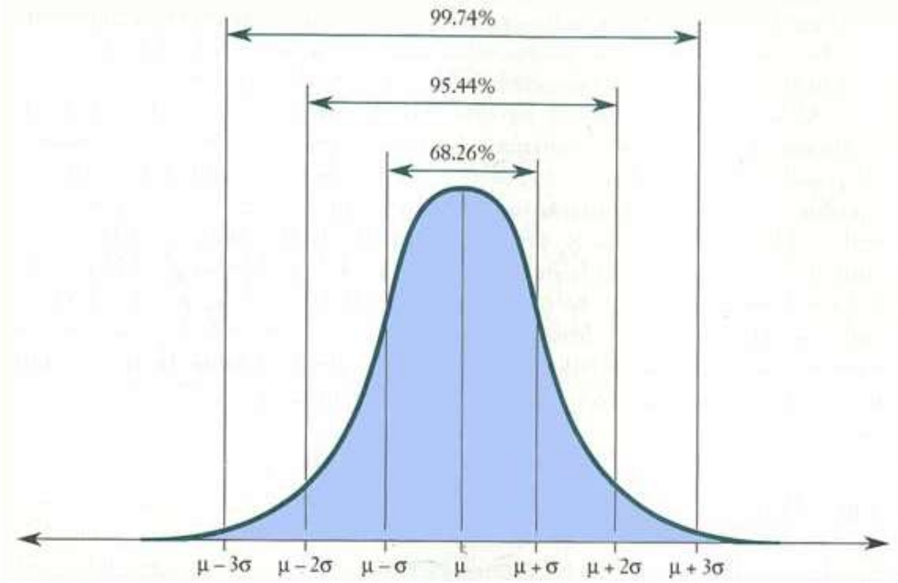
где Y – ордината кривой, или вероятность;

μ - генеральная средняя;

σ – стандартное отклонение генеральной совокупности;

π и e - константы (π=3,14; e=2,72).

Графически уравнение имеет следующий вид



Положение и форма кривой нормального распределения определяется двумя параметрами: генеральной средней μ, которая находится в центре распределения, и стандартным отклонением σ, которое измеряет вариацию отдельных наблюдений относительно средней арифметической.

Максимум, или центр нормального распределения лежит в точке X= μ; точки перегиба кривой находятся при X1= $\mu - \sigma$ и X2= $\mu + \sigma$. При X±∞ кривая достигает нулевого значения.

Кривые нормального распределения могут значительно различаться по форме. Вид кривой полностью соответствует степени варьирования изучаемого признака, т.е. от величины стандартного отклонения σ. Чем оно больше, тем более пологой будет вариационная кривая, при малых значениях σ она приобретает иглообразную форму.

Размах колебаний от μ вправо и влево зависит от величины стандартного отклонения и укладывается в основном в пределах 3 стандартных отклонений. Для нормального распределения характерны следующие закономерности:

- в пределах $\pm\sigma$ лежит 68% всех наблюдений;
- в пределах $\pm 2\sigma$ лежит 95% всех наблюдений;
- в пределах $\pm 3\sigma$ лежит 99% всех наблюдений.

Площадь под кривой, ограниченную на n стандартных отклонений и выраженную в процентах, называют ***уровнем вероятности*** P. Вероятность того, что значение варьирующего признака находится вне указанных пределов, называется ***уровнем значимости*** P_1. Уровень значимости указывает на величину ошибки наших выводов. Уровень значимости и уровень вероятности записывают в процентах или в долях целого.

Выбор уровня значимости или уровня вероятности для тех или иных исследований определяется практическими соображениями, ответственностью выводов и возможностями. В практике биологических и агрономических исследований наиболее часто используют уровень вероятности 95%, которому соответствует уровень значимости 5%.

Средняя арифметическая генеральной совокупности μ и стандартное отклонение σ являются параметрами генеральной совокупности, когда число наблюдений стремится к бесконечности. Получить оценки этих параметров позволяют выборочные наблюдения. Так, выборочная средняя $\bar{x}$ является оценкой генеральной средней μ, выборочное стандартное отклонение s – оценкой стандартного отклонения генеральной средней σ. Для выборок большого размера (n>20-30 и особенно n >100) закономерности нормального распределения, рассмотренные выше для генеральной совокупности, справедливы и для их оценок, т.е. в области x±1s лежит 68% всех наблюдений, внутри пределов, x±2s - 95 и в интервале x±3s - 99%.

Средняя арифметическая и стандартное отклонение – являются основными статистическими характеристиками, при помощи которых задается эмпирическое распределение частот. Этих двух характеристик достаточно, чтобы на основе знания закономерностей теоретических распределений построить эмпирическое распределение и воспроизвести в нем определенную закономерность.

Результаты различных наблюдений чаще всего располагаются приблизительно в соответствии с симметричной кривой нормального распределения, когда частоты вариант равно отстоящих от средней, равны между собой, т.е. симметричны. Но нередко некоторые признаки растений и животных дают распределения, значительно отличающиеся от нормального – ***асимметричные, или скошенные***.

Асимметрия может быть положительной, или правосторонней, когда увеличиваются частоты правой стороны, или левосторонней, когда увеличиваются частоты левой части вариационной кривой.

Причинами асимметричных распределений могут быть следующими:

- неправильно взята выборка, когда в нее вошло непропорционально много (или мало) представителей варианты с большим или меньшим значением;
- на признак действуют определенные факторы, сдвигающие частоту в ту или иную сторону от среднего значения.

Когда какие-либо причины благоприятствуют более частому появлению и средних и крайних значений признака, образуются так называемые ***положительные эксцессивные распределения,*** имеющие вид острой пирамиды с расширенным основанием, или ***отрицательные эксцессивные распределения,*** когда в центре кривой имеется не вершина, а впадина, и вариационная кривая становится двухуровневой.

Многовершинные и двухвершинные кривые в большинстве случаев указывают, что в выборку попали представители нескольких совокупностей с различными средними. В генетических работах двухвершинные кривые могут свидетельствовать о появлении объектов с новыми свойствами или признаками

Закон нормального распределения проявляется при большом числе наблюдений (n>20-30). Однако экспериментатор часто основывает свои выводы на выборках малого объема. При небольшом числе наблюдений результаты обычно близки и редко появляются большие отклонения, что соответствует закону нормального распределения: вероятность появления малых отклонений больше чем появления больших. Поэтому стандартное отклонение подсчитанное на основании малой выборки, в большинству случаев, будет меньше чем в генеральной совокупности, поэтому полностью полагаться на критерии нормального распределения в своих выводах нельзя. В этих случаях используют t- распределение Стьюдента, F- распределение Фишера и χ^2 – распределение Пирсона, которые являются частными случаями нормального распределения.

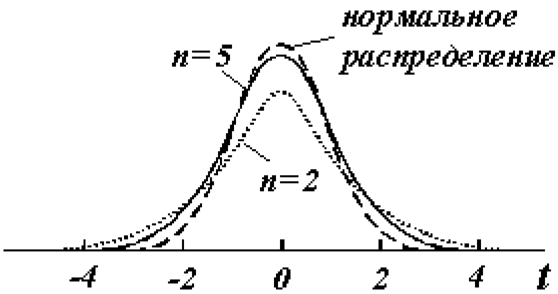
2. ***t-распределение Стьюдента***. С начала 20 века в математической статистике стало разрабатываться навое направление, которое получило название «статистика малых выборок». Наибольшее практическое значение для исследовательской работы имело открытое в 1908 году В. Госсетом t-распределение, получившее название распределение Стьюдента.

Распределение Стьюдента для выборочных средних определяется следующим равенством:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

Таким образом, величина t измеряется отклонением выборочной средней от средней генеральной совокупности, выраженным в долях ошибки выборки, принятой за единицу. Теоретическое распределение критерия Стьюдента дается в статистической таблице.

Графически распределение Стьюдента при n=2 и n=5 в сравнении с нормальным распределением имеет следующий вид:



Максимум частоты нормального и t распределения совпадают, но форма кривой t распределения полностью зависит от числа наблюдений. При очень малом количестве наблюдений она принимает вид плосковершинной кривой, причем площадь, ограниченная кривой, больше, чем при нормальном распределении. При увеличении числа наблюдений распределение t приближается к нормальному и переходит в него при $n \to \infty$.

Распределение Стьюдента позволяет определить доверительный интервал и проверить ту или иную гипотезу относительно генеральной совокупности. При этом нет необходимости знать ее среднюю арифметическую и стандартное отклонение, достаточно иметь их оценки, т.е. среднюю арифметическую выборки и стандартное отклонение для изучаемой выборки объемом n.

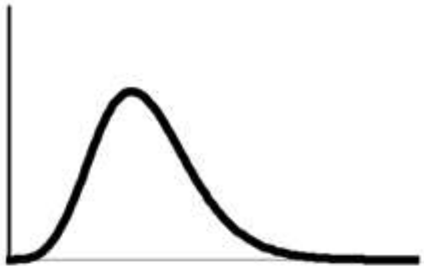
3. ***F- распределение Фишера***. Если из нормально распределенной совокупности взять 2 независимые выборки объемом n_1 и n_2 и подсчитать их дисперсии, то можно определить соотношение дисперсий, или критерий Фишера:

$$F = \frac{S_1^2}{S_2^2}$$

Отношение дисперсий рассчитывают таким образом, чтобы в числителе была большая дисперсия и поэтому $F \geq 1$.

Распределение Фишера зависит только от числа степеней свободы 2 выборок, т.е. от объема выборок. Когда 2 сравниваемые выборки являются случайными независимыми из общей совокупности с генеральной средней μ, то фактическое значение критерия Фишера не выйдет за определенные пределы и не превысит критическое для данных значений степеней свободы теоретическое значение критерия Фишера, т.е. $F_{факт} < F_{теор}$. Если генеральные параметры сравниваемых групп различны, то $F_{факт} \geq F_{теор}$. Теоретическое значение критерия Фишера берется из статистической таблицы для 1 или 5% уровня значимости.

Кривые F-распределения особенно при небольшом числе наблюдений имеют асимметричную форму - длинный хвост больших значений и большую концентрацию малых величин F:



$F_{10,25}$

4. χ^2 ***– распределение Пирсона***. χ^2 – распределение Пирсона используется для сравнения фактических и теоретических частот при качественной изменчивости при решении селекционно-генетических задач.

Для сравнения частот рассчитывают χ^2 – критерий, или критерий согласия по следующей формуле:

$$\chi^2 = \sum \frac{(f - F)^2}{F}$$

где, f – частота фактическая, F – теоретическая частота.

Фактическое значение критерия сравнивают с теоретическим, которое берут из статистических таблиц.

Если χ^2 факт. $\geq \chi^2$ теор., различия между частотами существенны, если χ^2факт$< \chi^2$ теор, различия между частотами несущественны.

Кривая распределения Пирсона в значительной степени зависит от числа степеней свободы, т.е. как и распределение Фишера, от объема выборки. Для малого числа степеней свободы кривая асимметрична, но с увеличением числа степеней свободы при n ->∞ кривая становится нормальной.

5. ***Распределение Пуассона***. Когда наступление некоторого события имеет очень маленькую вероятность, например, несколько раз на 1000 или 10000 обычных, то распределение случайной величины следует определенному закону редких событий, который выражается формулой Пуассона:

$$Px = \frac{a^{x} \cdot e^{-a}}{x}$$

где, Р – вероятность значения х;

х – число редких событий, происшедших в каждой большой группе (ч=0, 1, 2, 3 и т.д.);

а – среднее число редких событий на каждую большую группу;

х! – произведение чисел от 1 до х (факториал); считается что факториал 0=1;

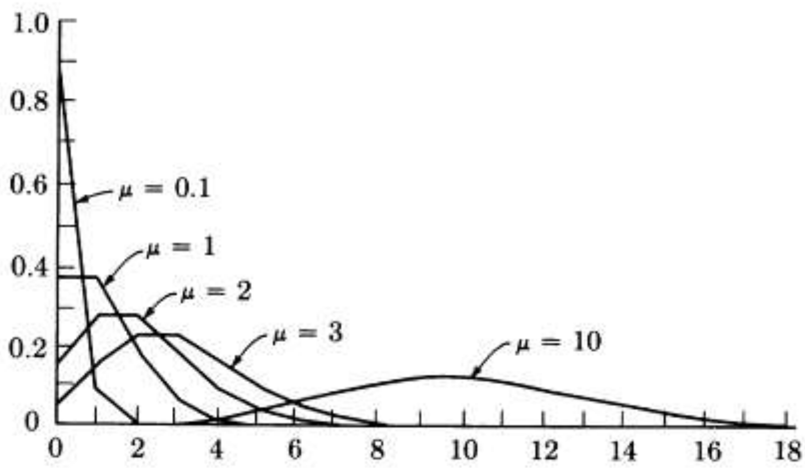
е – константа, основание натурального логарифма =2,718.

Если событие х подчинено распределению Пуассона со средней а, то вероятность значений х=0, 1, 2, 3 и т.д. будут соответственно равны:

$$P_{x=0} = \frac{a^{0} \cdot e^{-a}}{1} = e^{-a} \qquad P_{x=1} = a \cdot e^{-a}$$

$$P_{x=2} = \frac{a^{2} \cdot e^{-a}}{2} \qquad P_{x=3} = \frac{a^{3} \cdot e^{-a}}{6}$$

Графически распределение редких событий представляет асимметричную кривую, причем чем меньше вероятность события, тем сильнее асимметрия.



Примером распределения Пуассона может служить количество семян сорняков в семенном зерне, численность перезимовавших вредителей, частоты спонтанных мутаций у бактерий.