



Санкт-Петербургский
ИНСТИТУТ ПСИХОЛОГИИ
И АКМЕОЛОГИИ



Лекция (2 часа).

Математическая статистика

Введение в дисциплину. Основные определения. Шкалы. Совокупность исходных данных. Распределения. Графическое представление распределений.

Некоторые основные определения

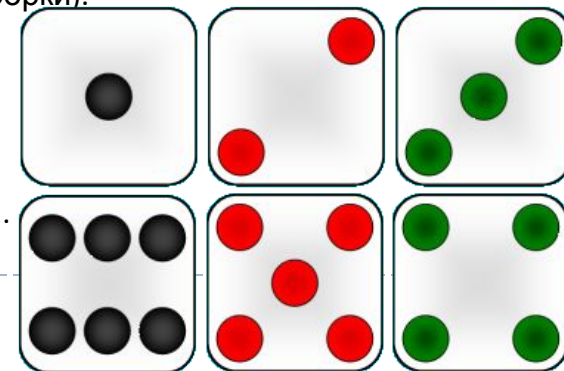


- ▣ **Статистика** – это прежде всего **способ мышления**, и для ее применения нужно лишь иметь немного здравого смысла и знать основы математики (Мак-Коннелл). Статистика оперирует понятием «случайный» или «стохастический»
- ▣ Слово **стохастический** (от греч. $\sigma\tau\omicron\chi\alpha\sigma\tau\iota\kappa\acute{o}\varsigma$ – «умеющий угадывать») используется во многих терминах из разных областей науки, и в общем означает неопределённость, случайность чего-либо.
- ▣ **Случайная величина** – это величина, которая принимает в результате опыта одно из множества значений, причём появление того или иного значения этой величины до её измерения нельзя точно предсказать.
- ▣ **Стохастический процесс** – это процесс, поведение которого не является детерминированным, и последующее состояние такой системы описывается как величинами, которые могут быть предсказаны, так и случайными. Однако (М. Как, Э. Нельсон) любое развитие процесса во времени (неважно, детерминированное или вероятностное) при анализе в терминах вероятностей будет стохастическим процессом (иными словами, все процессы, имеющие развитие во времени, с точки зрения теории вероятностей, стохастические)
- ▣ **Описательная статистика** – описывает, подытоживает и воспроизводит в виде таблиц или графиков данные того или иного **распределения**, вычисляет **среднее** для данного распределения, его **размах** и **дисперсию**.
- ▣ **Индуктивная статистика** – проверяет, можно ли распространить результаты, полученные на данной **выборке**, на всю **популяцию**, из которой взята эта выборка. Иными словами - позволяет выяснить, до какой степени можно путем индукции обобщить на большее число объектов ту или иную закономерность, обнаруженную при изучении их ограниченной группы в ходе какого-либо наблюдения или эксперимента.
- ▣ **Популяция** («генеральная совокупность») в статистике не обязательно означает какую-либо группу людей или естественное сообщество; этот термин относится ко всем объектам, образующим общую изучаемую совокупность
- ▣ **Выборка** («выборочная совокупность») – это n -ное количество элементов, отобранных с помощью научных методов так, чтобы она была **репрезентативной**, т.е. отражала популяцию в целом (n – объем выборки).

▣ **Данные** в статистике – это основные элементы, подлежащие анализу.

▣ Выделяют 2 группы и 4 шкалы данных:

- ▣ **качественные** – номинальные и порядковые
- ▣ **количественные** – интервальные (относительные) и отношений (абсолютные).



Шкалы представления данных



- **Номинальная шкала** обозначает принадлежность к группе. Например, М – мужчины, Ж – женщины; КГ – контрольная группа, ЭГ – экспериментальная группа.
- **Порядковая шкала** – ранжирование объектов по какому-либо признаку. Например, старший ребенок в семье получает номер 1, следующий за ним – номер 2 и т.д.; или лучший спортсмен получает 1-е место, следующий за ним – 2-е место и т. д.
- **Интервальная (относительная) шкала** – данные ранжированы, но интервалы между всеми последовательными величинами равны; здесь обходятся без такой величины, как «истинный» нуль. Например, температурная шкала Цельсия, возраст, календарная хронология. Так нулевой отметке термометра соответствует определенная температура; когда мы появляемся на свет, наш возраст не равен нулю; а эти строки писались в 2012 г. от Рождества Христова, в 7520 г. от сотворения мира и примерно через 12 млрд. лет после образования Солнечной Системы.
- **Шкала отношений (абсолютная)** – включает в себя «истинный» нуль и обладает всеми свойствами других шкал измерения, а также при желании могут быть преобразованы в интервальные, порядковые и номинальные данные. Например, расстояние, высота и промежуток времени. Отметка «нуль сантиметров» указывает на отсутствие расстояния или высоты, а нулевая отметка времени – на то, что еще ничего не произошло.



Совокупность исходных данных



- Как правило, **совокупность исходных данных** изображают в виде таблицы «объекты-признаки», в которой переменные (признаки) располагают в столбцах (графах), а данные индивидуального участника (объекты) – в строках (иногда называемых «записями»).

Номер	Пол	Порядок рождения	Возраст	СБ осенний	СБ весенний	Отработанные часы	Удовлетворенность учебой	Уровень знаний
X1	X2	X3	X4	X5	X6	X7	X8	X9
1	1	3	60	3,9	3,9	38	1	Н
2	1	3	22	4,6	4	15	2	В
3	2	2	17	4,7	5	10	2	В
4	1	2	19	4,2	4	30	2	В
5	2	1	23	5	4,9	12	3	В
6	1	3	27	3,7	3,9	35	3	В
7	2	4	18	3,7	4,5	30	1	Н
8	2	2	32	4,4	4,9	30	2	Н
9	2	2	22	4,6	4,7	20	3	Н
10	2	1	20	4,9	5	10	1	В
11	1	2	29	5	5	20	3	В
12	2	2	18	4	4	35	3	Н
13	1	2	18	4,6	3,7	30	1	В
14	1	2	20	4,2	4,5	30	1	Н
15	2	3	19	4	4,1	35	2	Н
16	1	1	19	4,9	4,8	10	1	В
17	2	3	37	4,7	4,5	10	3	Н

ШКАЛЫ

качественные

количественные



номинальная

порядковая

интервальная (относительная)

Виды распределений



Binomial



Discrete



DUmiform



Geomet



HyperGeo



IntUniform



NegBin



Poisson



Beta



BetaGeneral



BetaSubj



ChiSq



Cumul



CumulD



Erf



Erlang



Expon



ExtValue



Gamma



General



Histogram



InvGauss



Logistic



LogLogistic



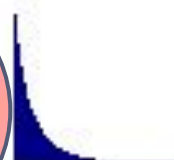
Lognorm



Lognorm2



Normal



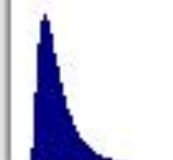
Pareto



Pareto2



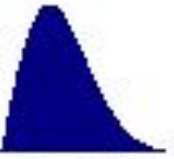
Pearson5



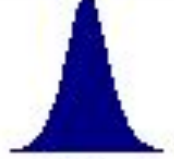
Pearson6



Pert



Rayleigh



Student



Triang



TriGen



Uniform



Weibull

□ Нормальное распределение.

□ *Другие названия:* закон Гаусса (1809); гауссовское распределение; второй закон Лапласа (1812); распределение Гаусса-Лапласа

Формула закона *нормального* распределения

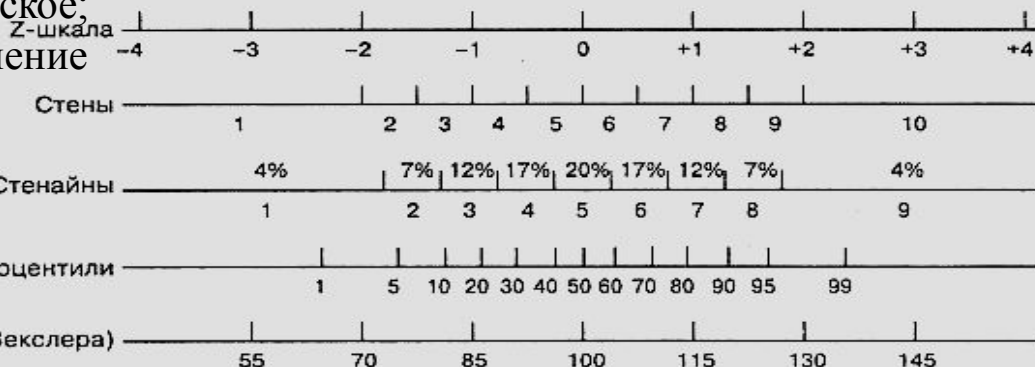
$$f(x_i) = \frac{1}{SD\sqrt{2\pi}} e^{-\frac{(x_i - M)^2}{2SD^2}}$$

где: $f(x_i)$ – высота подъема кривой (плотность вероятности для значения x_i);
 e – основание натурального логарифма (2,718);
 π – число «пи» (3,14159);
 M – среднее арифметическое;
 SD – стандартное отклонение

Иоганн Карл Фридрих Гаусс
 (30.04.1777 - 23.02.1855)



Пьер-Симон Лаплас
 (23.03.1749 - 05.03.1827)



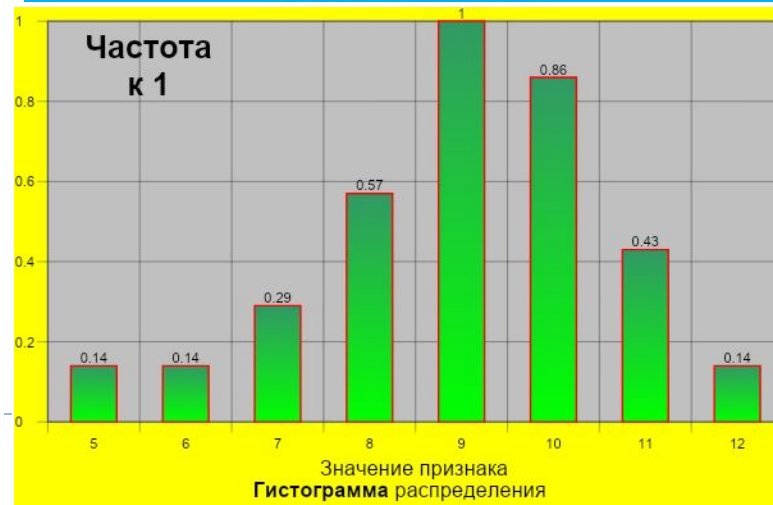
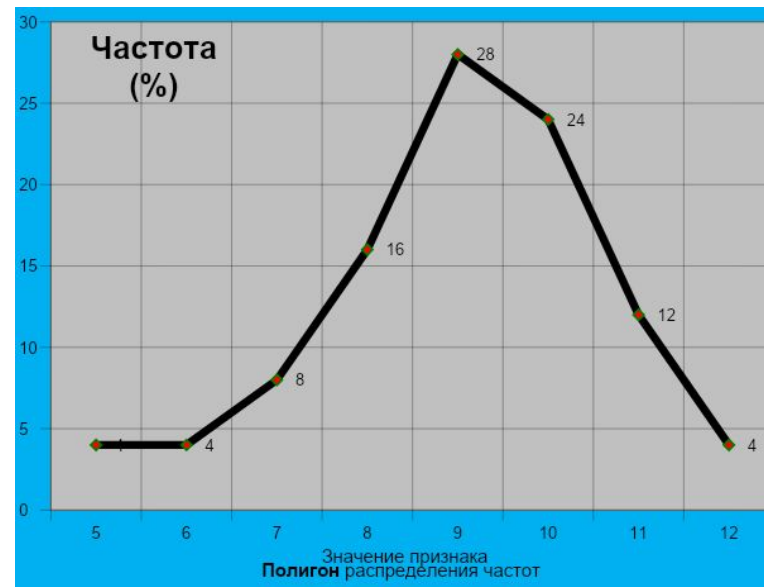
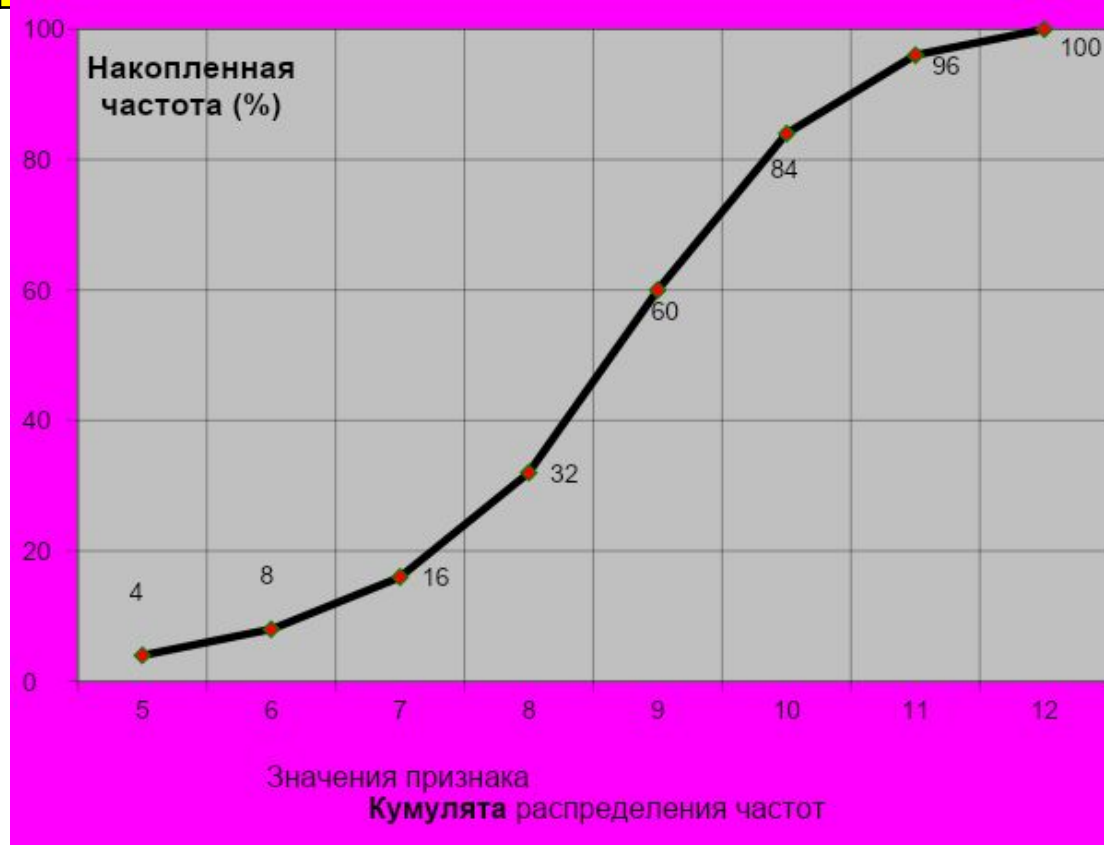


Графическое представление распределений

По 16-факторному опроснику Р. Кеттелла у 20 респондентов актеров измерялись показатели фактора MD – адекватность самооценки (диапазон от 0 до 14 баллов):

6, 9, 5, 10, 7, 9, 8, 10, 9, 10, 8, 11, 9, 12, 9, 10, 8, 10, 11, 9, 10, 9, 8, 7, 11.

Значение признака	5	6	7	8	9	10	11	12
Число вариант	1	1	2	4	7	6	3	1
Частота к 100%	4	4	8	16	28	24	12	4
Накопленная частота	4	8	16	32	60	84	96	100
Частота к 1	0.14	0.14	0.29	0.57	1.00	0.86	0.43	0.14



Самостоятельная работа студента (10 часов)



- Рекомендуется перед самостоятельным изучением дисциплины ознакомиться с содержанием брошюры (свободно распространяется в интернете) **Биркгофф Г. Математика и психология. –М.: Сов. Радио, 1977. – 96с.** (Garrett Birkhoff. Mathematics and Psychology. – SIAM Review, 1969, vol.11. №9, p.429-469.)



- Затем по учебному пособию **Воронов И.А. Эксперимент и методы обработки многомерных данных с применением SPSS: медико-биологические исследования, психология, физическая культура и спорт. СПб.: СПбГУТ им. Проф. М.А. Бонч-Бруевича, 2008.** (пособие в формате *.pdf свободно распространяется автором в интернете) ознакомиться с материалом, изложенным на **страницах: 4 – 5, 14 – 22, 31**, выполнить самостоятельно все предлагаемые **задачи от 2.1 до 2.4.**



- Примечание: Похожий материал можно найти во многих иных учебниках по математической статистике для психологов. Рекомендуем обратиться к изданиям **Наследов А.Д. Математические методы психологического исследования: анализ и интерпретация данных. СПб, 2004.** и **Наследов А.Д. SPSS15: профессиональный статистический анализ данных. СПб, 2008.**

