Лекция 5

ҚАЗАҚСТАННЫҢ АШЫҚ УНИВЕРСИТЕТІ

ОСНОВЫ СТАТИСТИКИ

Меры дисперсии

Здравствуйте!

Сегодня мы рассмотрим и научимся:

1. Понимать меру рассеивания и интерпретировать показатели
2. Вычислять и объяснять понятие размаха - R, межквартильного размаха (диапазона) – Q, стандартного отклонения – s и дисперсию – s2.
3. Выбирать подходящую меру дисперсии, правильно рассчитывать и интерпретировать статистику.
4. Описывать и объяснять математические характеристики стандартного отклонения.
5. Анализировать график «ящик с усами»
6. Используйте SPSS для получения стандартного отклонения и диапазона.

Статистические методы используются для описания изменчивости или разнообразия в наборе значений. Они могут быть использованы для описания:

- Вариаций в оценках. Некоторые студенты получают одинаковую оценку практически все время, в то время как другие варьируются от высоких до низких.
- Разнообразие образа жизни в разных социальных условиях. Большие города вообще поддерживают более широкий спектр образов жизни, чем в небольших городах.
- Различия в неравенстве доходов в зависимости от времени. У некоторых регионах есть огромные различия в богатстве и доходах от самых богатых до самых бедных членов, в то время как в других странах расстояние меньше.

Важность концепции дисперсии может быть легче понять при помощи примера. Предположим, что директор общественной безопасности хочет оценить две службы скорой помощи, которые заключили договор с городом на оказание экстренной медицинской помощи. В рамках исследования она собрала данные, свидетельствующие о том, что среднее время ответа на обращения за помощью составляет 7,4 минуты для Службы А и 7,6 минут для Службы Б.

Эти средние значения практически идентичны и не дают четкой основы для оценки одной услуги как более эффективной, чем другой. Меры дисперсии, однако, могут выявить существенные различия между комбинациями, даже когда меры центральной тенденции одинаковы. Например, рассмотрим графики на рисунке 5.1, которые отображают распределение времени отклика для двух служб. Обратите внимание, что линейный график для Службы Б гораздо более плоский, чем для Службы А. Это объясняется тем, что оценки для Службы Б более разбросаны или более разнообразны, чем оценки для Службы А. Другими словами, Служба Б была гораздо более переменный во времени отклика и имел больше баллов в высоких и низких диапазонах и меньше в середине. Служба А была более последовательной по времени отклика, а ее оценки более сгруппированы или сгруппированы по среднему значению. Оба распределения имеют, по существу, одинаковое среднее время отклика, но время отклика для службы Б значительно больше различий или дисперсий.

Если бы вы были директором службы общественной безопасности, вы бы с большей вероятностью выбрали службу скорой помощи, которая всегда находилась на месте возникновения чрезвычайной ситуации примерно в одно и то же время (служба А) или службу, которая иногда была очень медленной, а иногда и очень быстрой? (Служба Б)?

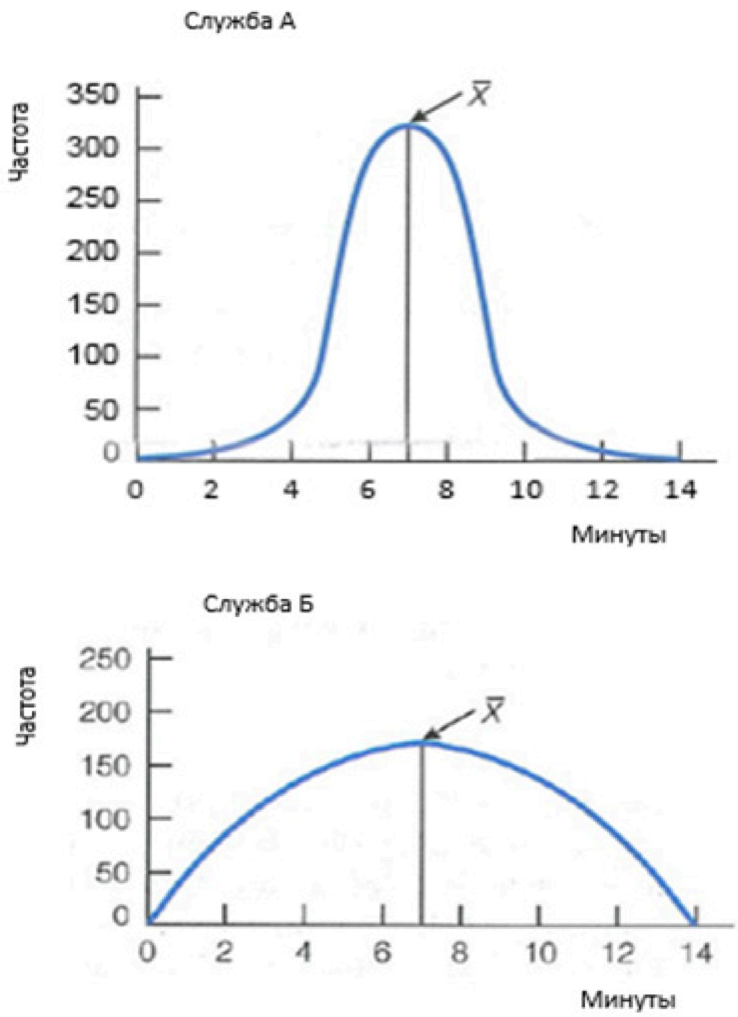

Рисунок 5.1. Время ответа для двух служб скорой помощи

Обратите внимание, что, если вы не рассматривали дисперсию, возможно, не заметили бы различие в работе двух служб скорой помощи.

Помните о двух формах на рисунке 5.1 как о визуальном представлении понятия дисперсии. Большая кластеризация баллов вокруг среднего значения в распределении для Службы А указывает на меньшую дисперсию, а более плоская кривая распределения для Службы Б указывает на большее разнообразие или дисперсию. Любая мера дисперсии будет уменьшаться по мере того, как оценки становятся менее изменчивыми, а распределение становится более пиковым (то есть, когда распределение становится все более похожим на службу А) и увеличивается, когда оценки становятся более изменчивыми, а распределение становится более плоским (что есть, так как дистрибутив все больше и больше напоминает Службу Б).

Эти идеи и рисунок 5.1 могут дать вам общее представление о том, что подразумевается под дисперсией, но эта концепцию нелегко описать одними словами: нам нужно рассмотреть некоторые статистические данные, которые предназначены для количественного определения вариаций в наборе оценок. Мы представим некоторые из наиболее распространенных меры дисперсии. Начнем с размаха и межквартильного размаха, но особое внимание уделим стандартному отклонению. Мы также рассмотрим визуальное представление дисперсии, называемое коробчатым графиком.

Размах (R) и межквартильный диапазон (Q) размах представляет собой расстояние между самой высокой и самой низкой оценками в распределении:

Размах = Высокий балл – Низкий балл

Размах легко вычисляется и, возможно, наиболее полезен в качестве быстрого и общего индикатора изменчивости и его легко интерпретировать: чем больше значение Размаха тем больше расстояние от высокого до низкого балла и тем больше дисперсия в распределении.

К сожалению, поскольку он основан только на самом высоком и самом низком баллах, размах имеет некоторые важные ограничения. Во-первых, почти любое значительное распределение будет содержать некоторые оценки, которые являются нетипично высокими и/или низкими (выбросами) по сравнению с большинством оценок. Таким образом, R может преувеличивать величину дисперсии для большинства оценок в распределении. Кроме того, R не дает информации о вариации оценок между самым высоким и самым низким оценками.

В статистике для анализа выборки довольно часто прибегают к другому показателю вариации – межквартильному размаху. Квартиль – это то значение, которые делит ранжированные (отсортированные) данные на части, кратные одной четверти, или 25%. Так, 1-й квартиль – это значение, ниже которого находится 25% совокупности. 2-й квартиль делит совокупность данным пополам (то есть медиана), ну и 3-й квартиль отделяет 25% наибольших значений. Так вот межквартильный размах – это разница между 3-м и 1-м квартилями. У данного показателя есть одно неоспоримое преимущество: он является робастным, т.е. не зависит от аномальных отклонений.

В таблице 5.1 представлены коэффициенты рождаемости (число рождений на 1000 человек населения) для группы из 40 стран. Каков размах и межквартильный размах этих данных? Обратите внимание, что оценки уже были упорядочены от высокой к низкой. Это облегчает расчет диапазона и является необходимым для определения межквартильного диапазона. Из этих 40 стран Нигер занял самое высокое место с коэффициентом рождаемости 50, а Германия и Япония заняли самое низкое место с коэффициентом рождаемости 8. Диапазон составляет 50-8 или 42 [R - 42).

Чтобы найти Q, мы должны найти первый и третий квартили (Q и Q3). Мы определяем эти точки в терминах баллов, связанных с определенными случаями, как мы это делали при поиске медианы. Q, определяется умножением N на 0,25. Поскольку (40) X (0,25) = 10, Q – это оценка, связанная с десятым случаем, начиная с самой низкой оценки. Десятый случай – Исландия с результатом 14. Таким образом, Q = 14.

Таблица 5.1. Уровни рождаемости (Количество рождений за 1000 населения) для 40 стран, 2013

Разряд	Страны	Уровень рождаемости	Разряд	Страны	Уровень рождаемости
40(высокого)	Нигерия	50	20	Ливия	22
39	Ангола	47	19	Индия	22
38	Уганды	45	18	Венесуэла	21
37	Мозамбик	44	17	Мексика	19
36	Нигерия	42	16	Колумбия	19
35	Маливия	40	15	Кувейт	19
34	Танзания	40	14	Австралия	18
33	Гвинея	38	13	Вьетнам	17
32	Сенегал	38	12	Ирландия	16
31	Того	37	11	Чили	15
30	Кения	36	10	Исландия	14
29	Мавритания	35	9	Соединенные Штаты	13
28	Эфиопия	34	8	Россия	13
27	Гана	33	7	Франция	13
26	Гватемала	32	6	Китай	12
25	Пакистан	30	5	Канада	11
24	Гаити	26	4	Испания	10
23	Камбоджия	25	3	Италия	9
22	Египт	25	2	Япония	8
21	Сирия	25	1(низких)	Германия	8

Случай, который лежит в третьем квартиле (Qz), получено умножением N на 0,75 и (40) X (0,75) = 30-й случай. 30-й случай, снова считающийся с самого низкого балла, - это Кения, набравший 36 баллов (Q3 = 36). Следовательно, Случай, который лежит в третьем квартиле (Q3), получено умножением N на 0,75 и (40) X (0,75) = 30-й случай. 30-й случай, снова считающийся с самого низкого балла, - это Кения, набравший 36 баллов (Q3 = 36). Следовательно:

Q= Q3- Q1

Q=36-14

Q=22

Стандартное отклонение и дисперсия

Основное ограничение как Q, так и R состоит в том, что они используют только две оценки и, следовательно, не используют всю доступную информацию. Кроме того, ни одна из статистических данных не предоставляет никакой информации о том, как далеко находятся оценки друг от друга или от какой-либо центральной точки, такой как среднее значение. Как мы можем разработать меру дисперсии, которая позволит избежать этих ограничений?

Одним из способов разработки статистики, соответствующей этим критериям, было бы начать с расстояний между каждой оценкой и средним значением или отклонений (Xi-X). Значение отклонений будет увеличиваться по мере увеличения различий между оценками и средним значением. Если оценки более сгруппированы вокруг среднего значения (помните график для Службы А на рисунке 5.1), отклонения будут небольшими. Если оценки будут более распределенными или более разнообразными (например, оценки для Службы Б на рисунке 5.1), отклонения будут больше. Как мы можем использовать отклонения для разработки полезной статистики?
Одним из способов действий будет использование суммы отклонений

$\sum (X_i - \overline{X})$ в качестве основы для статистики, но, как мы видели ранее, сумма отклонений всегда будет равна 0. Для иллюстрации рассмотрим распределение пяти баллов, представленных в таблице 5.2. Если мы суммируем отклонения любого набора баллов от их среднего значения, мы всегда получим итоговое значение 0, независимо от степени разнообразия в баллах.

Таблица 5.2. Демонстрация, что сумма отклонений очков вокруг среднего будет всегда составлять ноль

Баллы (X_i)	Отклонение ($X_i - \overline{X}$)
10	(10-30) = -20
20	(20-30) = -10
30	(30-30) = 0
40	(40-30) = 10
50	(50-30) = 20
$\sum X_i = 150$	$\sum(X_i - \overline{X}) = 0$
$\overline{X}$ = 150/50 = 30	

Таблица 5.3. Вычисление среднеквадратичного отклонения

Баллы (X_i)	Отклонение ($X_i - \bar{X}$)	Согласованные отклонение $(X_i - \bar{X})^2$
10	(10-30) = -20	$(-20)^2 = 400$
20	(20-30) = -10	$(-10)^2 = 100$
30	(30-30) = 0	$(0)^2 = 0$
40	(40-30) = 10	$(10)^2 = 100$
50	(50-30) = 20	$(20)^2 = 400$
$\sum X_i = 150$	$\sum(X_i - \bar{X}) = 0$	$\sum(X_i - \bar{X})^2 = 1000$

Тем не менее, сумма отклонений является логической основой для измерения дисперсии, и статистики разработали способ обойти тот факт, что положительные отклонения всегда равны отрицательным отклонениям. Если мы возведем в квадрат каждое из отклонений, все значения будут положительными, потому что отрицательное число, умноженное само по себе, становится положительным. Например, (-20) X (- 20) = + 400. Для баллов в таблице 5.2 сумма квадратов отклонений будет (400 + 100 + 0 + 100 + 400) или 1000 (см. Таблицу 5.3). , Таким образом, статистика, основанная на сумме квадратов отклонений, будет иметь свойства, которые мы хотим в хорошей мере дисперсии.

Прежде чем мы закончим разработку нашей меры дисперсии, мы должны решить еще одну проблему. Сумма квадратов отклонений будет увеличиваться с увеличением размера выборки: чем больше число баллов, тем больше значение меры. Это очень затруднит сравнение относительной изменчивости распределений на основе выборок N (размер выборки) и, таким образом, стандартизации для выборок разных размеров.
Эти процедуры дают статистику, известную как дисперсия, которая обозначается как S2. Дисперсия используется главным образом в логической статистике, хотя она является центральной концепцией при определении некоторых мер ассоциации. В целях описания дисперсии распределения обычно используется тесно связанная статистика, называемая стандартным отклонением (обозначаемая как S).

Дисперсия (S2) и стандартное отклонение (S)

$$s^2 = \frac{\Sigma(X_i - \bar{X})^2}{N}$$

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N}}$$

Строго говоря, формулы 5.3 и 5.4 предназначены для дисперсии и стандартного отклонения совокупности. Немного разные формулы, с N - 1 вместо N в знаменателе, должны использоваться, когда мы работаем со случайными выборками, а не целыми популяциями. Это важный момент, потому что многие из электронных калькуляторов и статистических программных пакетов, которые вы можете использовать (включая SPSS), используют N - 1 в знаменателе и, таким образом, дают результаты, которые будут быть хотя бы немного отличным от тех, которые приведены в формулах 5.3 и 5.4. Размер разницы будет уменьшаться по мере увеличения размера выборки, но многие проблемы и примеры в этой главе используют небольшие выборки, и различия в использовании N и N - 1 в знаменателе могут быть значительными в таких случаях. Некоторые калькуляторы позволяют пользователю выбирать N - 1 или N при расчете стандартного отклонения. Если у вас есть такой калькулятор, выберите «N», и ваши ответы должны совпадать с нашими.

Рекомендуется использовать таблицу, такую как Таблица 5.3, для вычисления стандартного отклонения. Пять баллов в нашем примере перечислены в левом столбце таблицы, отклонения - в среднем столбце, а квадратичные отклонения - в правом столбце.
Сумма последнего столбца в таблице 5.3 представляет собой сумму квадратов отклонений, которая подставляется непосредственно в формулу 5.4:

$$s = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N}}$$

$$s = \sqrt{\frac{\Sigma 1000}{5}}$$

$$s = \sqrt{200}$$

s= 14.14

Пример вычисление стандартного отклонения:

Исследователь сравнивает студенческие органы двух кампусов. Один колледж находится в небольшом городке, и почти все студенты проживают в кампусе. Другой находится в большом городе, и студенты почти все заняты на полставки. Исследователь хочет сравнить возраст студентов в двух кампусах и собрал информацию, представленную в таблице 5.4. Какой из учеников старше, а какой по возрасту более разнообразен? (Излишне говорить, что эти очень маленькие группы слишком малы для серьезных исследований и используются здесь только для упрощения вычислений).

Таблица 5.4 Вычисление стандартного отклонения для двух кампусов

Жилой кампус		
Баллы (X_i)	Отклонение ($X_i - \bar{X}$)	Согласованные отклонение $(X_i - \bar{X})^2$
18	(18-20) = -1	$(-1)^2 = 1$
19	(19-19) = 0	$(0)^2 = 0$
20	(20-19) =1	$(1)^2 = 1$
18	(18-19) = -1	$(-1)^2 = 1$
20	(20- 19) = 1	$(1)^2 = 1$
$\sum X_i = 95$	$\sum(X_i - \bar{X}) = 0$	$\sum(X_i - \bar{X})^2 = 4$
$\bar{X} = \frac{\Sigma X_i}{N} = \frac{95}{5} = 19$		
Городской кампус		
Баллы (X_i)	Отклонение ($X_i - \bar{X}$)	Согласованные отклонение $(X_i - \bar{X})^2$
20	(20-23) = -3	$(-3)^2 = 9$
22	(22-23) = -1	$(-1)^2 = 1$
18	(18-23) = -5	$(-5)^2 = 25$
25	(25-23) = 2	$(2)^2 = 4$
30	(30-23) = 7	$(7)^2 = 49$
$\sum X_i = 115$	$\sum(X_i - \bar{X}) = 0$	$\sum(X_i - \bar{X})^2 = 88$
$\bar{X} = \frac{\Sigma X_i}{N} = \frac{115}{5} = 23$		

Мы видим из средств, что студенты из жилого кампуса немного моложе, чем студенты из городского кампуса (19 против 23 лет). Какая группа более разнообразна по возрасту? Вычисление стандартного отклонения (Формула 5.4) ответит на этот вопрос:

Жилой кампус

$$S = \sqrt{\frac{\sum (X_i - \overline{X})^2}{N}} = \sqrt{\frac{4}{5}} = \sqrt{0.8} = 0.89$$

Городской кампус

$$S = \sqrt{\sum \frac{(X_i - \overline{X})^2}{N}} = \sqrt{\frac{88}{5}} = \sqrt{17.6} = 4.20$$

Более высокое значение стандартного отклонения для городского кампуса означает, что оно более разнообразно по возрасту. Как видно из результатов сканирования, возраст учащихся в жилом колледже находится в узком возрастном диапазоне (R= 20–18 = 2), в то время как студенты в городском кампусе более смешанные и включают студентов 25 и 30 лет (R = 30 - 18 = 12).

Визуализация дисперсии

График, называемый блокпостом или блоком или графиком усов, предоставляет полезный способ визуализации и анализа дисперсии и дает нам возможность применить некоторые из наших растущих массивов статистических инструментов. Блокплоты используют медиану, диапазон (R) и межквартильный диапазон (Q) для отображения как центральной тенденции, так и изменчивости. Они также отображают любые выбросы или экстремальные оценки, которые могут быть включены в распределение.
Давайте начнем с изучения частей коробочного графика, используя рисунок 5.2 в качестве примера. На этом графике представлены коэффициенты рождаемости (число рождений на 1000 человек населения) в 99 странах по различным уровням развития и достатка. Участок состоит из коробки и двух Т-образных линий, одна над и одна под коробкой. Коробка простирается от 3-го квартиля вверху (Q3,) до 1-го квартиля внизу (Q1,). Таким образом, высота коробки отражает значение R или межквартильный размах. Горизонтальная линия, проходящая через прямоугольник, проведена по медиане (Md).

Т-образные линии (или «усы») отражают диапазон баллов. Они простираются в 1,5 раза по высоте коробки или до высоких и низких баллов, в зависимости от того, что ближе к коробке. В этом случае самый высокий фактический коэффициент рождаемости составляет 51 рождаемость на 1000 человек населения, и именно здесь вырисовывается верхний усик. Нижний уровень усов определяется при уровне рождаемости 8, что является самым низким показателем рождаемости в этой группе наций.
На бокс-листе «выброс» определяется как любая оценка за пределами «усов» или любая оценка, превышающая в 1,5 раза высоту коробки в любом направлении. Результаты, которые

превышают высоту окна в три раза в любом направлении, называются «крайними выбросами». Нации, показанные на рисунке 5.2, не имеют экстремальных показателей, но вскоре мы увидим некоторые примеры.

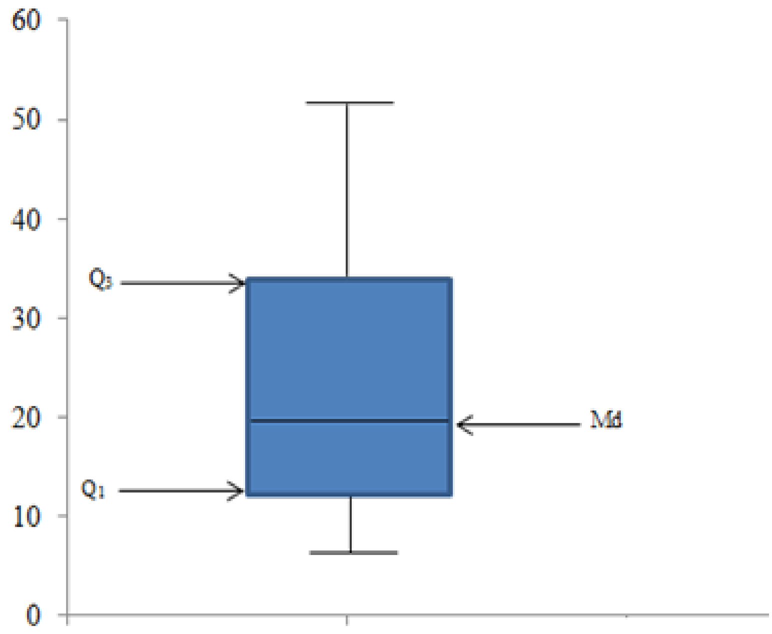

Рисунок 5.2 Boxplot для коэффициентов рождаемости.

Данный тип диаграмм особенно полезен, когда мы хотим сравнить распределение переменной по различным условиям или времени. На рисунке 5.3 представлены показатели рождаемости по уровню доходов для тех же 99 стран, но на этот раз страны были классифицированы по группам доходов. Сводная статистика по рождаемости по группам доходов представлена в таблице 5.5.

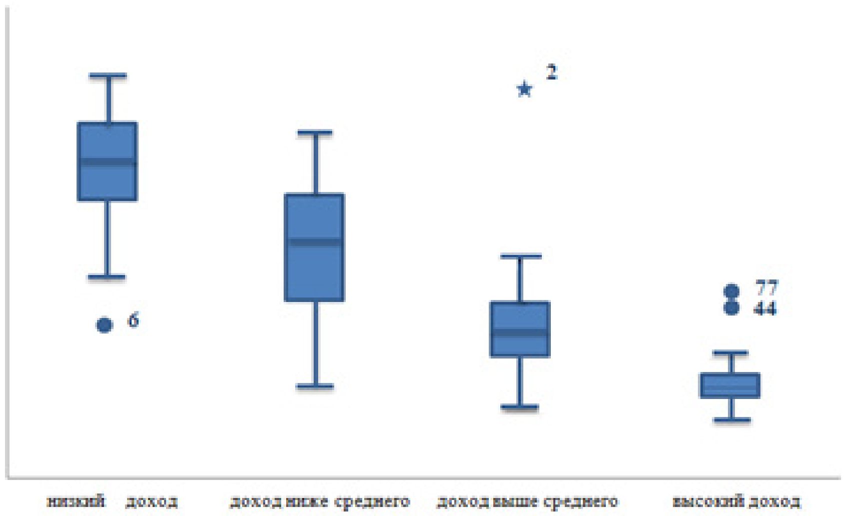

Рисунок 5.3 Бокcплот для коэффициентов рождаемости по уровню доходов

Статистика	Уровень дохода			
	Низкий доход	Более низкий средний доход	Верхний средний доход	Высокий доход
Низкий балл	31	11	9	8
Q_1	35.5	22	15.3	10
Медиана	40	30	19	11
Q_3	45	36	22	13
Высокий балл	51	44	47	22
Q	9.5	14	6.7	3
R	20	33	38	14
S	7.5	8.6	7.4	3.4
$\overline{X}$	39.2	29.1	19.7	11. 8
	N = 25	21	24	29

Таблица 5.5. Уровни рождаемости уровнями дохода для 99 стран, 2013

Группы доходов основаны на экономических критериях, разработанных Всемирным банком (http://data.wor1dbank.org/about/counny-classification/country-and-1ending- groups), и страны в каждой группе доходов имеют много социальных характеристик:
• Страны со средним уровнем дохода, как правило, находятся на ранних стадиях индустриализации и включают Египет, Боливию и Индию.
• Страны с доходом выше среднего имеют более развитую экономику и включают в себя Аргентину, Турцию и Китай.
• Почти все страны с высоким уровнем дохода высокоиндустриализированы и урбанизированы, и у них самый высокий уровень жизни. В этих странах воспитание детей обходится дороже, поэтому семьи меньше и рождаемость ниже. В эту категорию входят США, Канада, большинство стран Западной Европы, а также Япония, Австралия и Новая Зеландия.
Рассматривая сначала центральную тенденцию, мы видим, что средний уровень рождаемости (отмеченный горизонтальной линией в рамке) резко снижается с ростом уровня дохода, как и среднее значение (см. Таблицу 5.5). В странах с низким доходом средний уровень рождаемости намного выше, а в странах с высоким доходом самый низкий средний уровень рождаемости. Страны со средним уровнем ниже среднего и выше среднего имеют средний уровень рождаемости.

Обращаясь к дисперсии, мы видим несколько закономерностей. Во-первых, страны с доходами выше среднего имеют самый большой диапазон (R= 38) и, согласно этой статистике, являются наиболее изменчивыми группами. Тем не менее, диапазон для стран в этой категории завышен выбросом (нация № 2, которая является Анголой). Оценка Анголы является «экстремальным» выбросом (она имеет оценку более чем в 3 раза превышающую высоту поля) и отмечена * на графике. Диапазон, как вы помните, основан только на самом высоком и самом низком баллах и может быть не лучшим показателем дисперсии в этом случае.
Во-вторых, страны со средним уровнем дохода имеют самый высокий межквартильный размах (Q) и стандартное отклонение (S) и, по этим статистическим данным, наибольшая дисперсия.

Поскольку R раздувается для стран с доходами выше среднего, то выбросы в этой ситуации лучше измерять дисперсию, и мы, вероятно, пришли бы к выводу, что в странах с доходом ниже среднего уровень рождаемости наиболее изменчив.
В-третьих, страны с высоким уровнем дохода являются наименее изменчивыми на сегодняшний день. Это видно из их короткого коробочного графика и их гораздо более низкого межквартильного диапазона (3), диапазона (14) и стандартного отклонения (3.4). Эти случаи наиболее похожи друг на друга, несмотря на два выброса (№ 77 - Саудовская Аравия и № 44 - Израиль). Почему эти две нации так сильно отличаются от других стран с высоким уровнем дохода по уровню рождаемости?
Как для R, Q и S, блокпосты наиболее полезны при сравнении переменных в разных условиях.

Интерпретация стандартного отклонения

Вполне возможно, что значение стандартного отклонения (то есть, почему мы его вычисляем) не совсем очевидно для вас на данный момент. Вы можете спросить: «Как только у меня возникнет проблема расчета стандартного отклонения?» Значение этой меры дисперсии можно выразить тремя способами. Первая и самая важная включает нормальную кривую, и мы отложим эту интерпретацию до следующей главы.
Второй способ мышления о стандартном отклонении - это показатель дисперсии, значение которого увеличивается по мере того, как распределение становится более изменчивым. Другими словами, стандартное отклонение выше для более разнообразных распределений и ниже для менее разнообразных распределений. Наименьшее значение, которое может иметь стандартное отклонение, равно 0, и это будет иметь место для распределений без дисперсии (т. Е. Если бы каждый отдельный случай в выборке имел одинаковую оценку). Таким образом, 0 является наименьшим значением, допустимым для стандартного отклонения (хотя верхнего предела нет).
Третий способ понять значение стандартного отклонения состоит в сравнении одного распределения с другим. Мы уже сделали это при сравнении двух служб скорой помощи на рисунке 5.1 и жилых и городских кампусов в таблице 5.4. Вы также можете сделать это при сравнении одной группы с другой или одна и та же переменная в два разных времени. Например, предположим, что мы обнаружили, что возраст студентов в Кампусе был изменен с течением времени, о чем свидетельствует сводная статистика, представленная в таблице 5.6.

1990	2010
$\overline{X} = 21$	$\overline{X} = 25$
$s = 1$	$s = 4$

Таблица 5.6 Возраст студентов в Кампусе колледжа с течением времени

Ясно, что студенческий состав стал в среднем старше, и, согласно стандартное отклонение, оно также стало более разнообразным по возрасту. Более низкое стандартное отклонение для 1990 года указывает на то, что возрасты будут более сгруппированы по среднему значению в этом году (вспомните распределение для Службы А на рисунке 5.1).

Выводы:

1. Меры дисперсии суммируют информацию о разнородности или разнообразие, в распределении очков. Когда объединено с мерой центральной тенденции, они передают большой объем информации во всего нескольких числах. Меры центральной тенденции определяют местонахождение центральных точек распределения, меры дисперсии указывают на сумму разнообразия в распределении.

2. Размах (R) является расстоянием от самого высокого до самого низкого счета в распределении. Диапазон межквартиля (Q) расстояние от третьего до первого квартиля («диапазон» середины 50% процентилей). Эти два диапазона могут использоваться с переменными, измеренными или на порядковом уровне, или на уровне отношения интервала.

3. Стандартное отклонение (отклонения) - самый важная мера дисперсии из-за его центральной роли в намного более передовом статистическом применении. Стандартное отклонение имеет минимальное значение 0 и увеличивается в зависимости как изменяется с увеличением распределения.

4. Коробчатые диаграммы обеспечивают визуализацию мер.