Содержательный модуль 8.
Краткие сведения о зависимых случайных величинах и зависимых погрешностях

8.1. Виды зависимостей

На процесс геодезических измерений, как показано в п.п. 2.3 оказывает влияние множество взаимосвязанных между собой факторов. Поэтому результаты измерений и их погрешности, как правило, представляются случайными величинами, которые имеют между собой некоторую зависимость. Например, результаты измерений длины и погрешности измерений зависят от температуры и влажности окружающей среды, результаты измерения углов превышения зависит от точности используемого прибора и т.д.

В математической статистике выделяют три основные зависимости между парами двух случайных величин X и Y.

1. **Функциональная зависимость**.

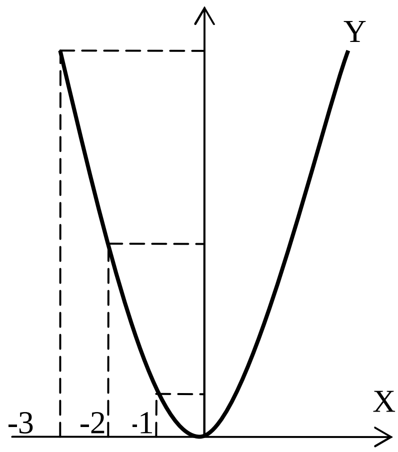
Строгая функциональная зависимость между случайными величинами в геодезической практике присутствует редко. Это связано с тем, что обе величины или одна из них подвержены действию случайных факторов. Более того, среди этих факторов могут быть и общие, т.е. воздействующие и на X, и на Y.

Это зависимость, когда значению X соответствует одно значение Y. Такая зависимость может быть выражена в виде функции $Y = f(x)$, представленной в виде графика или таблицы.

Так, например, функция $y = x^2$ может быть представлена в виде параболы или в виде таблицы (см. рис. 8.1).

2. Стохастическая зависимость

При стохастической зависимости, которая имеет место только для случайных величин, в том числе и для случайных погрешностей, одному значению *x* может соответствовать несколько или ни одного значения *y*. Такая зависимость может быть выражена только в виде таблицы или графика (см. рис. 8.2). Здесь иллюстрируется, так называемая эмпирическая зависимость, которая на рисунке показана совокупностью точек, т.е. результатов измерений. Сплошной линией на рисунке показана теоретическая зависимость, полученная в результате математической обработки эмпирической зависимости.



x	-3	-2	-1	0	1	2	3
y	9	4	1	0	1	4	9

Рис.8.1. Функциональная зависимость $y = x^2$

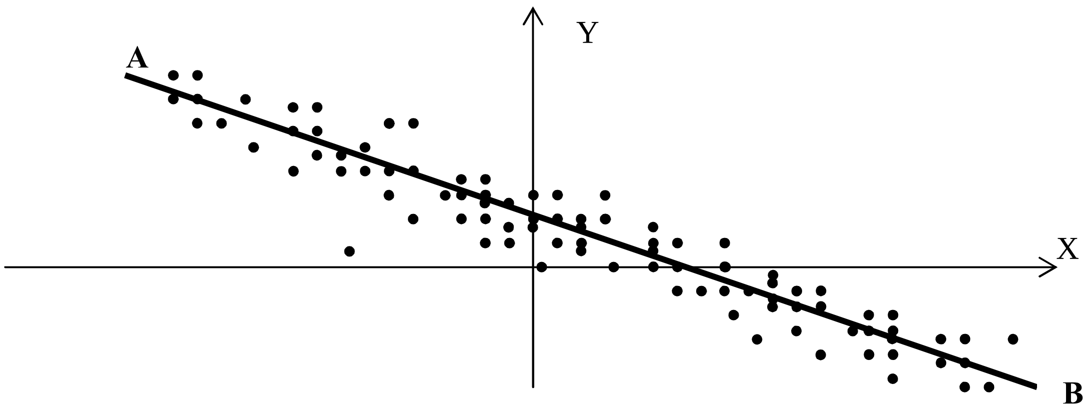
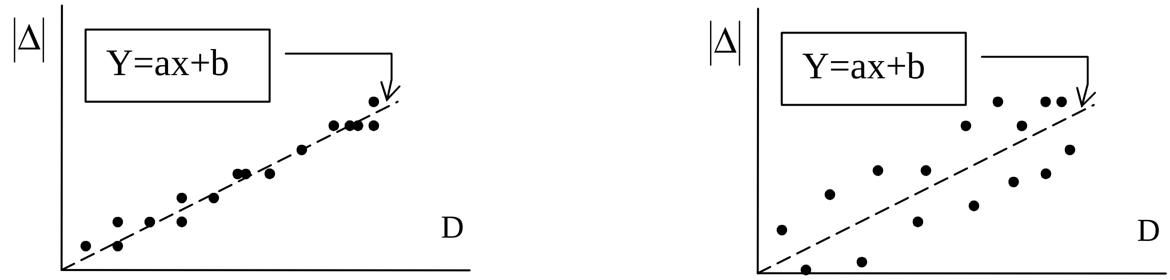


Рис. 8.2. Эмпирическая и теоретическая зависимости результатов измерения физической величины

На рис.8.2 наглядно показано, что результаты измерений расположены узкой полосой вдоль линии AB и имеют линейный стохастический (случайный) характер.

Пример 8.1. Одно и тоже расстояние D измерено двумя группами специалистов. Первую группу составили специалисты, имеющие большой опыт геодезических работ, которые проводили измерения высокоточными приборами. Вторая группа состояла из студентов, проходящих практику геодезических измерений с приборами невысокой точности.

Результаты измерений и зависимости их от погрешностей показаны на рис.8.3. Видно, что разброс эмпирических значений во второй группе больше, чем в первой группе. Поэтому в качестве математического аппарата для оценивания точности измерений зависимых случайных величин используется корреляционный и регрессионный анализ.



Результаты измерений 1-й группой   Результаты измерений 2-й группой

Рис.8.3. Коррелограммы и прямые регрессии

Существуют и более сложные стохастические зависимости, когда точки группируются вдоль узкой полосы, напоминающей кривую. Однако они здесь рассматриваться не будут. Заметим только, что многие нелинейные зависимости можно преобразовать в линейные.

3. Отсутствие зависимости

Имеет место только для случайных величин, в том числе и для случайных погрешностей. Как и в случае стохастической зависимости, такое множество можно представить в виде таблицы или показать графически (см. рис.8.3).

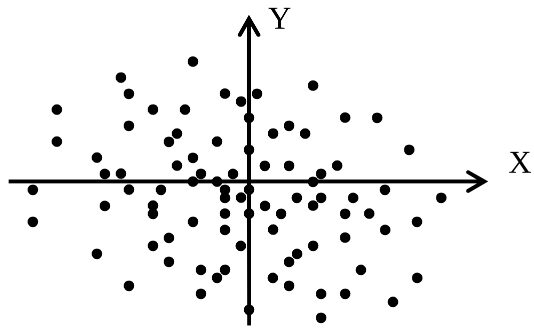


Рис.8.3. Иллюстрация отсутствия зависимости между результатами измерений

Основным признаком, подтверждающим отсутствие зависимости, является предел

$$\lim_{n\to\infty}\frac{[xy]}{n}=0. \qquad (8.1)$$

Частным случаем предела (8.1) является свойство независимости случайных погрешностей, представленное пределом (1.12).

8.2. Количественные характеристики линейной стохастической зависимости

Линейная стохастическая зависимость не может быть точно выражена функциональной зависимостью, например, в виде параболы, изображенной на рис.8.1 или другими строгими зависимостями - логарифмической, показательной и т.д. Вместе с тем, существуют количественные характеристики, достаточно точно описывающие тесноту связи между величинами *x* и *y*. Изучением количественных характеристик, описывающих тесноту связей между случайными величинами занимается теория корреляции.

Корреляция — статистическая взаимосвязь двух или нескольких случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми). При этом, изменения одной или нескольких из этих величин приводят к систематическому изменению другой или других величин.

Одной из характеристик оценки тесноты связи по опытным (апостериорных) данным величин *x* и *y* является корреляционный момент

$$k=\frac{[xy]}{n}-\bar{x}\bar{y}, \qquad (8.2)$$

где n - объем выборки, т.е. количество пар x,y;

$\bar{x}=\frac{[x]}{n}$ - среднее значение *x*;

$\bar{y}=\frac{[y]}{n}$ - среднее значение y.

Величина k зависит от размерности величин x и y и поэтому она не совсем удобна для оценивания тесноты связи этих величин.

Наиболее эффективным критерием оценивания тесноты связи измеренных геодезических величин является выборочный коэффициент корреляции, вычисляемый по формуле

$$r = \frac{k}{m_x m_y}, \tag{8.3}$$

где $m_x = \sqrt{\frac{[x^2]}{n} - \bar{x}^2}$, $m_y = \sqrt{\frac{[y^2]}{n} - \bar{y}^2}$ - эмпирические среднеквадратические отклонения (погрешности) величин x и y.

Свойства коэффициента корреляции:

1. Коэффициент корреляции принимает значения в интервале от -1 до +1, т.е. справедливо неравенство $-1 \le r \le +1$.

2. Когда коэффициент корреляции равен +1 или -1, между величинами x и y существует линейная функциональная зависимость вида

$$y = ax + c \text{ или } x = by + d.$$

3. Если $r = 0$, то между величинами x и y линейная зависимость отсутствует, но могут существовать более сложные зависимости.

Коэффициент корреляции, вычисленный по опытным данным в общем случае является величиной случайной. Поэтому при значениях $r < 0.5$, возникает вопрос, подтверждает ли вычисленное значение r наличие стохастической связи величин x и y или оно есть следствием каких-то случайных факторов? Другими словами, является ли r величиной значимой?

При $n > 50$ критерием значимости может служить среднеквадратическая погрешность

$$m_r = \frac{1 - r^2}{\sqrt{n}}. \tag{8.4}$$

Стохастическая связь между величинами x и y считается установленной, если

$$|r| > 3m_r. \tag{8.5}$$

При $n < 50$ критерием значимости могут служить критические значения коэффициента корреляции при $r = 0$, приведенные в табл. 8.1. Если при объеме выборки n и заданной вероятности 0.75, 0.90,…,0.995 вычисленное значение r больше приведенного в таблице, то с вероятностью p можно утверждать, что $r > 0$ и стохастическая зависимость между величинами x и y существует.

Пример 8.2. По выборке $n = 16$ вычислен коэффициент корреляции $r = 0.72$. На пересечении строки $n = 16$ и столбца $p = 0.995$ находим критическое значение, равное 0.6226. Так как 0.72 > 0.6226, с вероятностью $p = 0.995$ можем утверждать, что величины x и y имеют стохастическую зависимость. Критические значения для коэффициента корреляции, когда $\rho = 0$, $P = \left\{ r \le \text{табл. знач.} \middle| \rho = 0 \right\} = \gamma$.

Таблица 8.1

Исходные данные для оценивания зависимости случайных величин

n	Вероятность наличия зависимости между случайными величинами					
	0.75	**0.90**	**0.95**	**0.975**	**0.99**	**0.995**
3	0.7071	0.9511	0.9877	0.9969	0.9995	0.9999
4	0.5000	0.8000	0.9000	0.9500	0.9800	0.9900
5	0.4040	0.6870	0.8054	0.8783	0.9343	0.9587
6	0.3473	0.6084	0.7293	0.8114	0.8822	0.9172
7	0.3091	0.5509	0.6694	0.7545	0.8329	0.8745
8	0.2811	0.5067	0.6215	0.7067	0.7887	0.8343
9	0.2596	0.4716	0.5822	0.6664	0.7498	0.7977
10	0.2423	0.4428	0.5493	0.6319	0.7155	0.7646
11	0.2281	0.4187	0.5214	0.6021	0.6851	0.7348
12	0.2161	0.3981	0.4973	0.5760	0.6581	0.7079
13	0.2058	0.3802	0.4762	0.5529	0.6339	0.6835
14	0.1968	0.3646	0.4575	0.5324	0.6120	0.6614
15	0.1890	0.3507	0.4409	0.5140	0.5923	0.6411
16	0.1820	0.3383	0.4259	0.4973	0.5742	0.6226
17	0.1757	0.3271	0.4124	0.4822	0.5577	0.6055
18	0.1700	0.3170	0.4000	0.4683	0.5426	0.5897
19	0.1649	0.3077	0.3887	0.4555	0.5285	0.5751
20	0.1602	0.2992	0.3783	0.4438	0.5155	0.5614
21	0.1558	0.2914	0.3687	0.4329	0.5034	0.5487
22	0.1518	0.2841	0.3598	0.4227	0.4921	0.5368
23	0.1481	0.2774	0.3515	0.4132	0.4815	0.5256
24	0.1447	0.2711	0.3438	0.4044	0.4716	0.5151
25	0.1415	0.2653	0.3365	0.3961	0.4622	0.5052
30	0.1281	0.2407	0.3061	0.3610	0.4226	0.4629
35	0.1179	0.2220	0.2826	0.3338	0.3916	0.4296
40	0.1098	0.2070	0.2638	0.3120	0.3665	0.4026
45	0.1032	0.1947	0.2483	0.2940	0.3457	0.3801
50	0.0976	0.1843	0.2353	0.2787	0.3281	0.3610

Если установлено, что между величинами x и y - связь существенная, может быть составлено так называемое *уравнение регрессии* – функция, описывающая стохастическую связь:

$$y_i = \bar{y} + r\frac{m_y}{m_x}(x_i - \bar{x}). \tag{8.6}$$

8.3. Зависимые случайные погрешности в геодезии

Зависимые случайные погрешности в геодезии встречаются редко, но тем не менее иногда имеют место.

Пусть заданы два ряда зависимых случайных погрешностей $\Delta_{1,}\Delta_2,\ldots,\Delta_n$ и $\Delta'_{1,}\Delta'_2,\ldots,\Delta'_n$ полученные в условиях, характеризуемых стандартами σ и σ'. Тогда на основании (8.3), (8.4) можем записать

$$k = \lim_{n\to\infty}\frac{[\Delta\Delta']}{n}, \tag{8.7}$$

$$r = \frac{k}{\sigma\sigma'}. \tag{8.8}$$

В случае зависимых погрешностей формула основной теоремы в отличие от (4.2) принимает вид

$$\sigma_y = \sqrt{\left(\frac{\partial y}{\partial x_1}\right)^2\sigma_1^2 + \left(\frac{\partial y}{\partial x_2}\right)^2\sigma_2^2 + \ldots + \left(\frac{\partial y}{\partial x_t}\right)^2\sigma_t^2 + 2\sum r_{ij}\frac{\partial y \partial y}{\partial x_i \partial y_j}\sigma_i\sigma_j}\,. \tag{8.9}$$

Простейшим примером зависимых случайных погрешностей являются погрешности смежных углов при измерении способом круговых приемов. В самом деле, погрешность общего направления 2, если уменьшает угол β_1, то неизбежно увеличивает угол β_2 и наоборот. Данное обстоятельство порождает отрицательную стохастическую зависимость с коэффициентом корреляции $r = -0.5$.

Более сложные случаи зависимых погрешностей рассматриваются во второй части данного курса.

Таким образом, рассмотрены основные сведения случайных величин и зависимых погрешностях, которые встречаются в геодезической практике. Приведены основные виды зависимостей случайных величин и количественные характеристики линейных стохастических зависимостей.

СПИСОК ЛИТЕРАТУРЫ

1.

2. Большаков, В.Д. Практикум по теории математической обработки геодезических измерений [Текст]: учебное пособие для вузов / В. Д. Большаков, Ю.И. Маркузе. – М.: Недра, 1984, 352 с.

3. Свешников, А.А. Основы теории ошибок [Текст]: учебное пособие / А.А. Свешников. – Л., Изд-во Ленингр. Ун-та, 1972, 121 с.

3. Бурмистров, Г.А. Основы способа наименьших квадратов. - М.: Госгеолтехиздат, 1963.-392 с.

2. Гайдаев П.А., Большаков В.Д. Теория математической обработки геодезических измерений. – М.: Недра, 1969.- 400 с.

3. Кемниц Ю.В. Теория ошибок измерений. – М.: Недра, 1962. – 175 с.

5. Оуэн Д.Б. Сборник статистических таблиц. – М.: ВЦ АН СССР, 1973. – 586 с.

6. Хейфец Б.С., Данилевич Б.Б. Практикум по инженерной геодезии. – М., Недра, 1973. – 320 с.