

Лабораторная работа

Тема: Прогнозирование спортивных данных на основе уравнения линейной регрессии и линий тренда.

Задание

В таблице представлена динамика спортивных результатов, показанная спортсменом на тренировках в течение года. Требуется спрогнозировать результат спортсмена, который он покажет в следующем месяце. Выполнить прогноз на основе уравнения линейной регрессии, линейной и квадратичной линий тренда. Оценить абсолютную и относительную погрешности такого прогноза.

Ход выполнения работы

1. Перенести данные своего варианта в книгу MS Excel на лист 1
2. Изобразить данные таблицы на диаграмме (тип диаграммы - точечная).
3. Вычислить коэффициент корреляции и сделать вывод о зависимости или независимости результата спортсмена от месяца.
4. Построить уравнение линейной регрессии, и изобразить его на диаграмме.
5. Вычислить значение результата, соответствующего 13 месяцу с помощью построенного уравнения.
6. Оценить абсолютную и относительную погрешность.
7. Выполнить восполнение данных на основе линейной и полиномиальной (степень 2) линий трендов.
8. Оценить абсолютную и относительную погрешность такого восполнения (см. таблицу 1).
9. Сделать выводы по работе.

Таблица 1

Результаты, показанные спортсменом в 13 месяце

Номер варианта	Результат спортсмена
1	2,71
2	0,85
3	0,94
4	2,68
5	0,95
6	2,74
7	2,54
8	1,19
9	1,11
10	1,76
11	0,5
12	2,57
13	3,71
14	1,85
15	1,94

Номер варианта	Результат спортсмена
16	3,68
17	1,95
18	3,74
19	3,54
20	2,19
21	2,11
22	2,76
23	1,5
24	3,57
25	3,67
26	4,79
27	4,71
28	3,58
29	2,96
30	5,54

Методические указания

Для выявления зависимости между случайными величинами $X = \{x_i\}_{i=1}^n$ и $Y = \{y_i\}_{i=1}^n$ рассматривают коэффициент корреляции

$$r_{xy} = \frac{K_{xy}}{\sigma_x \cdot \sigma_y}$$

где σ_x, σ_y - среднеквадратические отклонения случайных величин X и Y соответственно, а K_{xy} - коэффициент ковариации.

$$\sigma_x = \sqrt{D_x};$$

$$\sigma_y = \sqrt{D_y}.$$

$$K_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$$

$$D_x = \frac{1}{n} \sum_{i=1}^n (\bar{X} - x_i)^2; D_y = \frac{1}{n} \sum_{i=1}^n (\bar{Y} - y_i)^2. \bar{X} = \frac{1}{n} \sum_{i=1}^n x_i; \bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

Известно, что в общем случае $-1 \leq r_{xy} \leq 1$. Причем, если он близок к нулю, то зависимости между величинами нет, если r_{xy} близок к единице, то зависимость прямо пропорциональная, а если близок к -1, то обратно пропорциональная.

Для построения уравнения линейной регрессии $y = a \cdot x + b$ требуется вычислить коэффициенты a и b по формулам:

$$a = \frac{r_{xy} \cdot \sigma_y}{\sigma_x}; b = \bar{Y} - a \cdot \bar{X}.$$

Excel имеет специальный аппарат для графического анализа моделей, в том числе построение аппроксимационных зависимостей (*линий тренда* $P(x)$) по данной таблице $\{x_i, y_i\}$, которые приближенно отражают функциональную связь $y=f(x)$.

Линии тренда обычно используются в задачах прогнозирования. Такие задачи решают с помощью методов регрессионного анализа. С помощью регрессионного анализа можно показать тенденцию изменения рядов данных, экстраполировать их (то есть продолжить линию тренда вперед или назад за пределы известных данных). Можно также построить *линию скользящего среднего*, которая сглаживает случайные флуктуации, яснее демонстрирует модель и прослеживает тенденцию изменения данных.

Линиями тренда можно дополнить ряды данных, представленные на линейчатых диаграммах, гистограммах, графиках, биржевых, точечных и пузырьковых диаграммах. Нельзя дополнить линиями тренда ряды данных на объемных диаграммах, лепестковых диаграммах, круговых и кольцевых диаграммах.

Excel позволяет выбрать один из пяти типов линии тренда – линейный, логарифмический, экспоненциальный, степенной или полиномиальный (2...6 степени) и проверить (по различным критериям), какой из типов лучше всего подходит в данной ситуации. Критерием может служить критерий R^2 (коэффициент детерминации), автоматическое вычисление которого встроено в диалоговое окно **Линия тренда**. Чем ближе коэффициент детерминации к единице, тем лучше тренд.

Для построения линии тренда нужно выделив линию графика. Выполнить **Вставка | Линия тренда**. (или, щелкнув правой клавишей мыши, вызвать контекстно-зависимое меню и выбрать **Добавить линию тренда**). Выбрать тип линии - линейную, полиномиальную (n=2): **Тип | Построение линии тренда** (рис. 1).

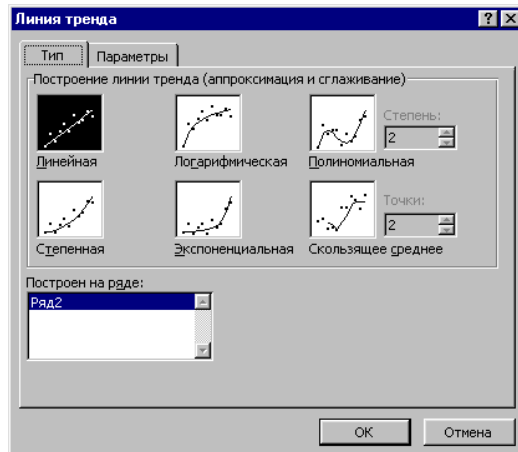


Рис. 1 - Диалоговое окно **Линия тренда** |

Для того, чтобы получить функциональную зависимость на закладке **Параметры** поставьте галочку против **Показывать уравнение на диаграмме** (рис. 2).

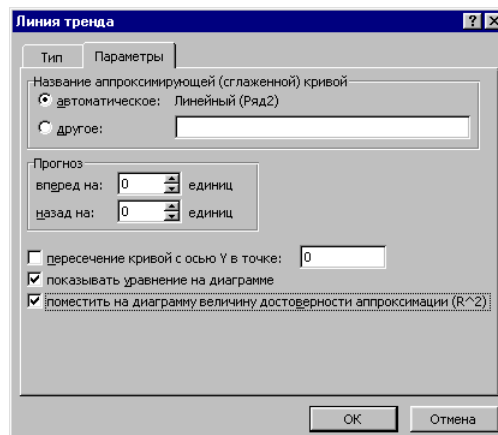


Рис. 2 – Диалоговое окно **Линия тренда** | **Параметры**