

Лабораторная работа

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

Задание

В таблице представлена динамика спортивных результатов, показанная спортсменом на тренировках в течение года. Данные за июль месяц оказались потеряны. Требуется восполнить данные на основе уравнения линейной регрессии, линейной и квадратичной линий тренда и оценить абсолютную и относительную погрешности таких восполнений.

Ход выполнения работы

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

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).

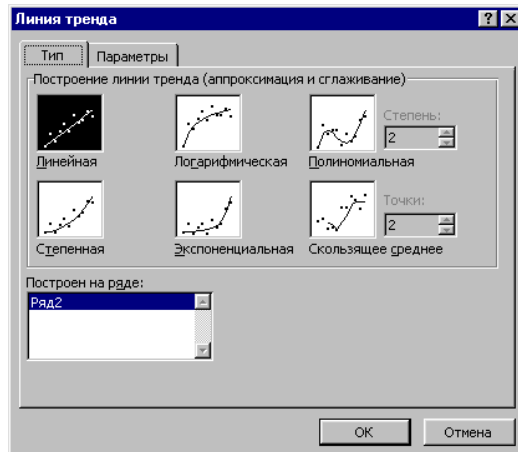


Рис. 1 - Диалоговое окно **Линия тренда** |

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

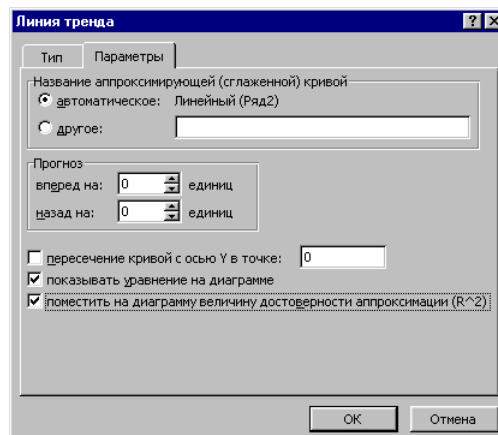


Рис. 2 – Диалоговое окно **Линия тренда** | **Параметры**