

# Исследование гидроэкологических данных с использованием среды программирования R

Павлов Александр Сергеевич

Кафедра Теории Вероятностей и Математической Статистики

13 мая 2014 г.



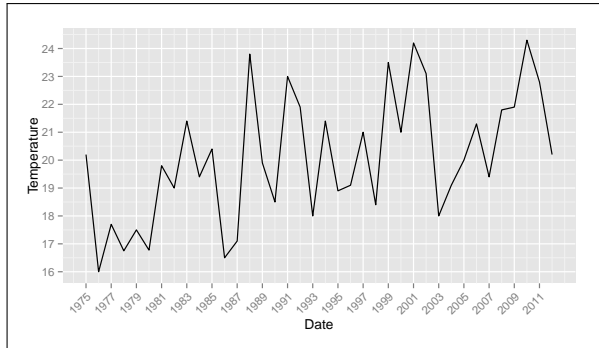
# Постановка задачи

- ▶ Исследование и статистическая обработка гидроэкологических данных в пакете R
- ▶ Вычисление и анализ описательных статистик
- ▶ Проверка исходных данных на нормальность
- ▶ Идентификация законов распределения
- ▶ Корреляционный анализ
- ▶ Регрессионный анализ
- ▶ Вариограммный анализ
- ▶ Кригинг



# Исходные данные

Исходные данные получены от учебно-научного центра «Нарочанская биологическая станция им. Г.Г.Винберга» . На рисунке представлена выборка наблюдений за температурой воды в июле месяце в период с 1975 по 2012 годы.

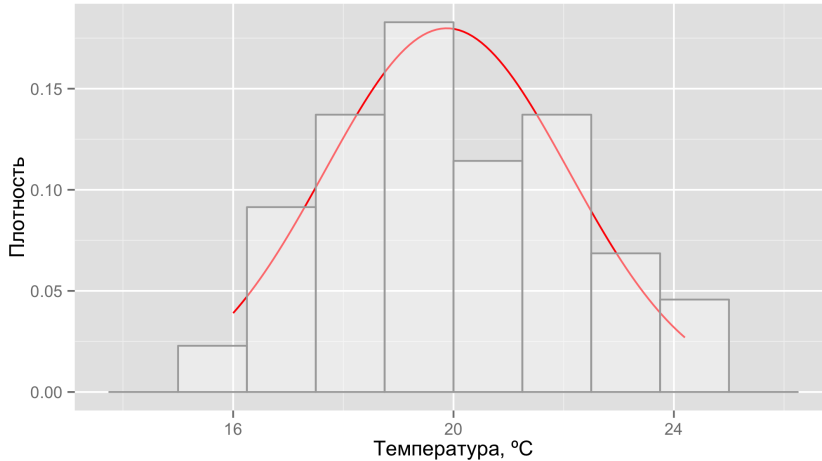


# Основные описательные статистики

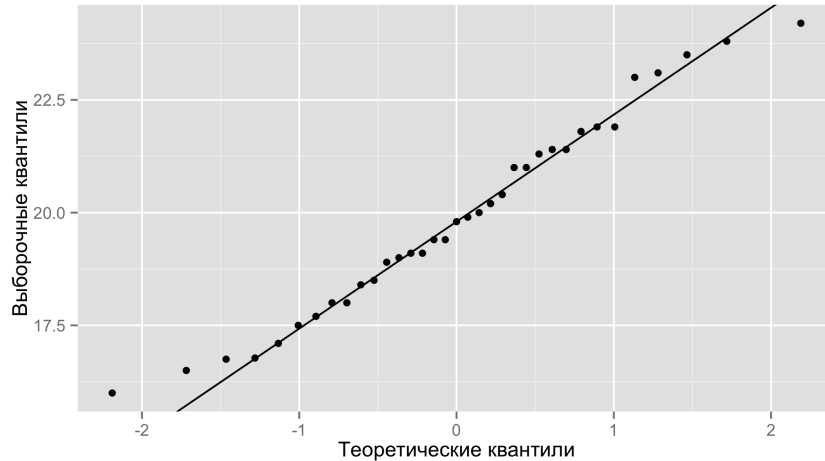
	Значение
Среднее	20.08
Медиана	19.95
Минимум	16.00
Максимум	24.30
Размах	8.30
Квартильный размах	3.28
Дисперсия	5.24
Стандартное отклонение	2.29
Коэффициент вариации	26.10
Стандартная ошибка	0.37
Асимметрия	0.14
Ошибка асимметрии	0.38
Эксцесс	-0.85
Ошибка эксцесса	0.75



# Гистограмма наблюдаемых температур



# График квантилей



# Статистические критерии в R

```
> shapiro.test():
```

Shapiro-Wilk normality test

```
data: Temperature  
W = 0.9727, p-value = 0.4706
```

```
> pearson.test():
```

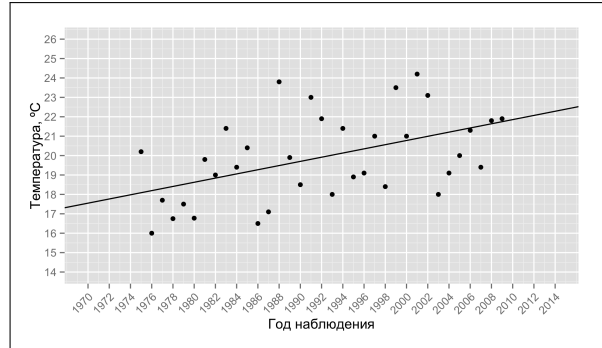
Pearson chi-square normality test

```
data: Temperature  
P = 1.7895, p-value = 0.938
```



# Диаграмма рассеяния

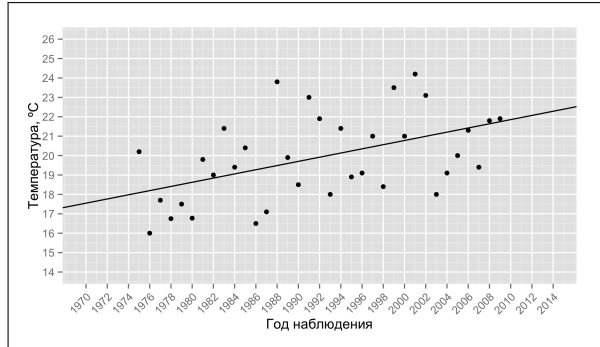
- ▶ Выборочный коэффициент корреляции:  $r_{xt} = 0.52$





# Диаграмма рассеяния

- ▶ Выборочный коэффициент корреляции:  $r_{xt} = 0.52$
- ▶ При уровне значимости  $\alpha = 0.05$  выборочный коэффициент является значимым



## Проверка критерия значимости средствами пакета R

```
> cor.test(method="pearson")
```

Pearson's product-moment correlation

data: Temperature and time

t = 3.6801, df = 36, p-value = 0.0007579

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.2439316 0.7218701

sample estimates:

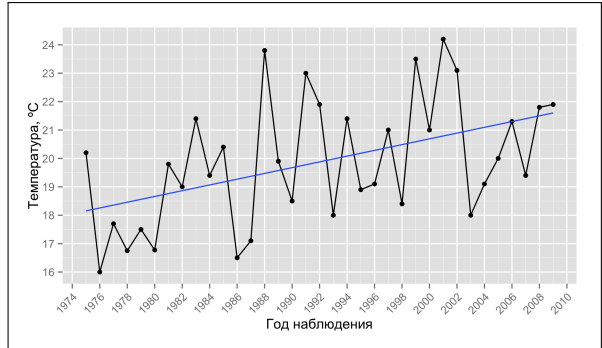
cor

0.5228432



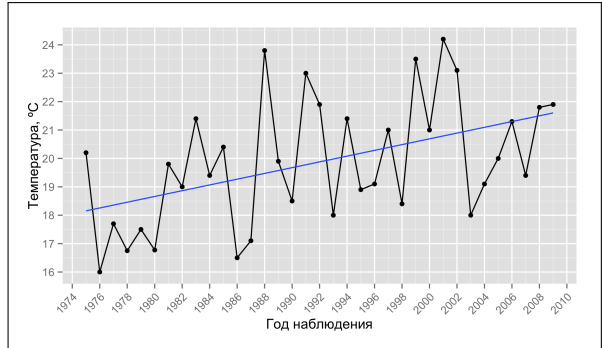
# Временной ряд

- Модель временного ряда:  
 $X = T + E$



# Временной ряд

- ▶ Модель временного ряда:  
 $X = T + E$
- ▶ Уравнение тренда:  
 $x = 0.108t + 17.98$



# Оценка модели

- ▶ С помощью критерия Стьюдента доказана значимость коэффициентов модели регрессии



# Оценка модели

- ▶ С помощью критерия Стьюдента доказана значимость коэффициентов модели регрессии
- ▶ F-критерий Фишера при уровне значимости  $\alpha = 0.05$  показал адекватность модели

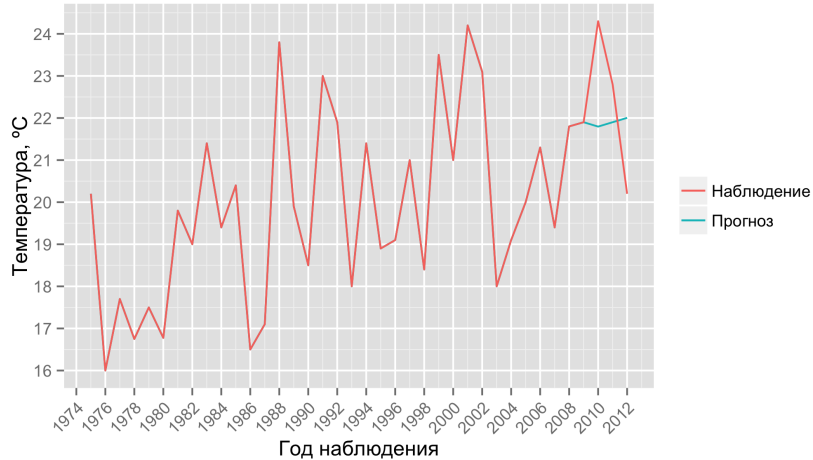


# Оценка модели

- ▶ С помощью критерия Стьюдента доказана значимость коэффициентов модели регрессии
- ▶ F-критерий Фишера при уровне значимости  $\alpha = 0.05$  показал адекватность модели
- ▶ Точность модели невысока, поскольку коэффициент детерминации  $\eta_{x(t)}^2 = 0.275$

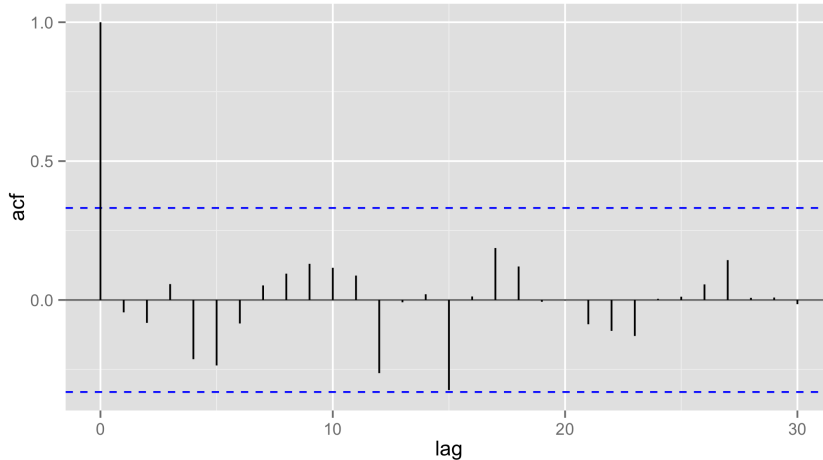


# Прогноз

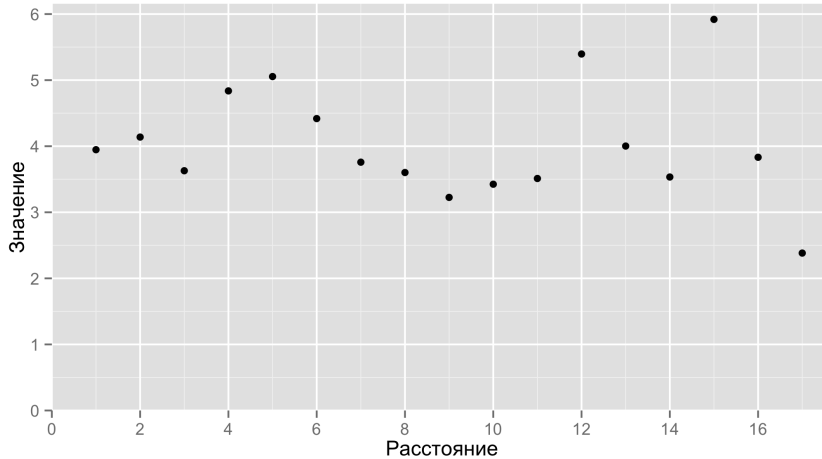




# Автокорреляционная функция



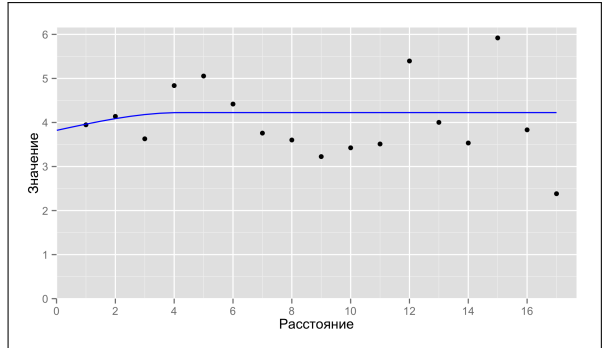
# График экспериментальной вариограммы



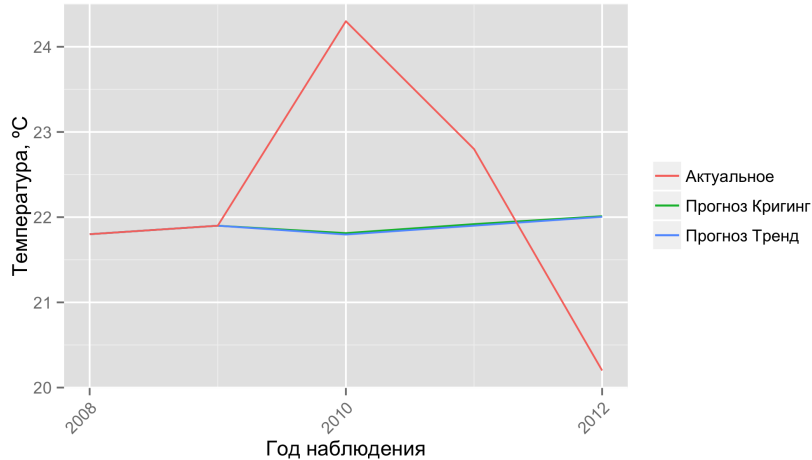
# График экспериментальной вариограммы с подобранной моделью

Параметры модели:

- ▶ Модель: Сферическая с эффектом самородков
- ▶ Эффект самородков: 3.82
- ▶ Порог: 4.22
- ▶ Ранг: 4.19



# Прогнозирование методом ординарного кригинга



Спасибо за внимание

