

Ф.И.О.: \_\_\_\_\_

1. (a) ☐ (b) ☐ (c) ☐ (d) ☐

2. (a) ☐ (b) ☐ (c) ☐ (d) ☐

3. (a) ☐ (b) ☐ (c) ☐ (d) ☐

4. (a) ☐ (b) ☐ (c) ☐

5. (a) ☐ (b) ☐ (c) ☐ (d) ☐

6. (a) ☐ (b) ☐ (c) ☐ (d) ☐

7. (a) ☐ (b) ☐ (c) ☐ (d) ☐

8. (a) ☐ (b) ☐ (c) ☐ (d) ☐

9. 

--	--	--	--	--	--

 . 

--	--	--

1. Что такое доверительная вероятность (p-value)? Отметьте справедливые утверждения.
  - (a)  $p$  показывает вероятность того, что можно отвергнуть  $H_0$
  - (b) Низкое значение  $p$  показывает, что значение тестовой статистики экстремально при условии справедливости  $H_0$
  - (c) Высокое значение  $p$  показывает, что значение тестовой статистики вполне обычно, если  $H_0$  была бы справедлива
  - (d)  $p$  оценивает вероятность получить такое значение тестовой статистики, если  $H_0$  верна
2. Отметьте условия применимости линейной регрессии
  - (a) Линейная связь
  - (b) Независимость значений  $y$  друг от друга
  - (c) Коэффициент детерминации  $R^2 > 0.79$
  - (d) Нормальное распределение остатков
3. Какими из этих свойств обладают точки, исключение которых из расчетов может сильно повлиять на ход регрессии?
  - (a) Положение непосредственно на линии регрессии
  - (b) Большая величина остатка
  - (c) Очень большие и очень маленькие значения  $x$
  - (d) Сильное отклонение от ожидаемого значения
4. Посмотрите на графики остатков (рис. 1) и отметьте истинные утверждения

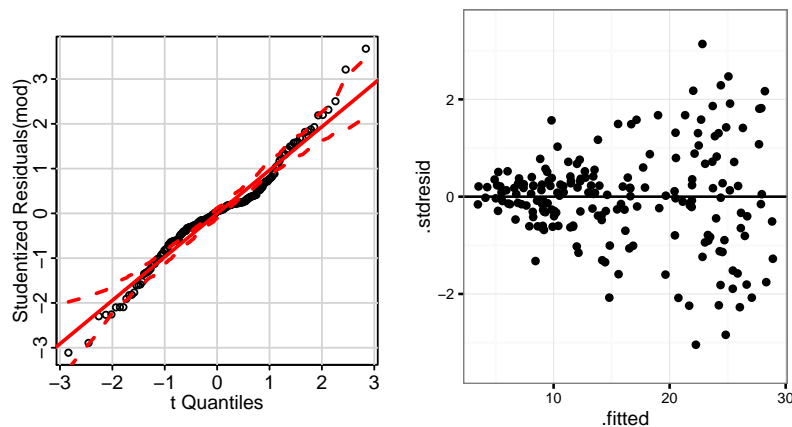


Рис. 1: Диагностические графики остатков линейной регрессии.

- (a) Нарушено условие гомогенности дисперсий
  - (b) Нарушено условие нормального распределения ошибок
  - (c) Все условия соблюдены, нарушения крайне незначительны
5. Стандартизация (z-score scaling) шкал всех предикторов перед выполнением регрессионного анализа:

- (a) не позволяет правомерно сравнить силу эффектов разнородных предикторов
  - (b) позволяет обойтись без проверки состоятельности модели
  - (c) центрует значения каждого предиктора  $X_i$  вокруг нуля вместо средней
  - (d) позволяет оценить ошибку предсказания модели
6. Во множественной регрессии, описываемой моделью  $Y = 0.01 + 1.4X_1 - 4.3X_2 + 0.8X_3$ , интерсепт это:
- (a) ожидаемое среднее значение зависимой переменной, когда все предикторы ( $X_1, X_2, X_3$ ) равны нулю
  - (b) один из параметров модели
  - (c) один из предикторов модели
  - (d) самый слабый предиктор
7. Поправка adjusted  $R^2$ :
- (a) всегда уменьшается с увеличением количества предикторов
  - (b) применима только для моделей с предварительно стандартизированными предикторами
  - (c) позволяет определить наиболее значимый предиктор множественной модели
  - (d) позволяет правомерно сравнивать модели с разным количеством предикторов
8. Во встроенном датасете `stackloss` пусть переменная `stack.loss` будет зависимой, а переменные `Air.Flow`, `Water.Temp`, `Acid.Conc.` - предикторами. Проведите регрессионный анализ и определите значение коэффициента `Water.Temp`
- (a) 1.2953
  - (b) -0.1521
  - (c) -39.9197
  - (d) NA
9. Для встроенного датасета `stackloss` рассчитайте (предскажите) значение зависимой переменной `stack.loss` при условии, что предиктор `Air.Flow` принимает значение 60, а остальные независимые переменные принимают свои средние значения. Запишите результат, округленный до третьего знака.