1 Часть 1. Нарушение предпосылок теоремы Гаусса-Маркова

В каждой части необходимо решить одну любую задачу по желанию.

1. Мартовский Заяц и Безумный Шляпник почти всё время пьют чай. Известно, что количество выпитого за день чая (в чашках) зависит от количества пирожных (в штуках) и печенья (в штуках). Алиса, гостившая у героев в течение 25 дней, заметила, что если оценить зависимость выпитого чая от закуски для Мартовского Зайца и Шляпника, то получится регрессия с $RSS = 11.5$:

   $$\widehat{Tea}_i = 6 + 0.5Biscuit_i + 1.5Cake_i$$

   Чтобы понять, удачную ли модель она построила, Алиса оценила ещё одну регрессию с $RSS = 9.5$:

   $$\widehat{\widehat{Tea}}_i = 12.7 + 0.65Biscuit_i - 0.8Cake_i - 0.59\widehat{Tea}_i^2 + 0.03\widehat{Tea}_i^3$$

   Помогите Алисе понять, верную ли спецификацию модели она выбрала

   (a) Проведите подходящий тест

   (b) Сформулируйте основную и альтернативную гипотезы

   (c) Алиса решила проверить первоначальную короткую модель на наличие гетероскедастичности с помощью теста Уайта. Выпишите уравнение регрессии, которое она должна оценить.

2. Пионеры, Крокодил Гена и Чебурашка собирали металлолом несколько дней подряд. В распоряжение иностранной шпионки, гражданки Шапокляк, попали ежедневные данные по количеству собранного металлолома: вектор g — для Крокодила Гены, вектор h — для Чебурашки и вектор x — для Пионеров. Гена и Чебурашка собирали вместе, поэтому выборочная корреляция $\mathrm{sCorr}(g, h) = -0.9$. Гена и Чебурашка собирали независимо от Пионеров, поэтому выборочные корреляции $\mathrm{sCorr}(g, x) = 0$, $\mathrm{sCorr}(h, x) = 0$. Если регрессоры g, h и x центрировать и нормировать, то получится матрица $\tilde{X}$.

   (a) Найдите параметр обусловленности матрицы $(\tilde{X}'\tilde{X})$

   (b) Вычислите одну или две главные компоненты (выразите их через вектор-столбцы матрицы $\tilde{X}$), объясняющие не менее 70% общей выборочной дисперсии регрессоров

   (c) Шпионка Шапокляк пытается смоделировать ежедневный выпуск танков, y. Выразите коэффициенты регрессии $y = \beta_1 + \beta_2 g + \beta_3 h + \beta_4 x + \varepsilon$ через коэффициенты регрессии на главные компоненты, объясняющие не менее 70% общей выборочной дисперсии.

3. Обследовав выборку из 27 домохозяйств, исследователь оценил уравнение регрессии:

   $$\frac{\widehat{Exp}_i}{Size_i} = 926 + 235\frac{1}{Size_i} + 0.3\frac{Income_i}{Size_i}$$

   где Exp_i — месячные затраты i-го домохозяйства на питание в рублях, $Income_i$ — месячный доход домохозяйства (также в рублях), $Size_i$ — число членов домохозяйства. Известен коэффициент детерминации, $R^2 = 0.3$.

   (a) Каково, согласно оценённой модели, ожидаемое различие в затратах на питание между двумя домохозяйствами с одинаковым доходом, первое из которых больше второго на одного человека?

(b) Известно, что нормировка переменных модели на размер семьи $Size_i$ была проведена с целью устранения гетероскедастичности в модели $Exp_i = \beta_1 + \beta_2 Size_i + \beta_3 Income_i + \varepsilon_i$. Какое предположение сделал исследователь о виде гетероскедастичности?

(c) Для проверки правильности выбранной спецификации было оценено ещё одно уравнение:

$$\widehat{\frac{Exp_i}{Size_i}} = 513 + 1499\frac{1}{Size_i} + 0.5\frac{Income_i}{Size_i} - 0.001\left(\widehat{\frac{Exp_i}{Size_i}}\right)^2$$

Известно, что $R^2 = 0.4$. Даёт ли эта проверка основание считать модель исследователя неверно специфицированной? Используйте уровень значимости 1%

2 Часть 2. Временные ряды

В каждой части необходимо решить одну любую задачу по желанию.

1. На графике представлены данные по уровню озера Гурóн в футах в 1875-1972 годах:

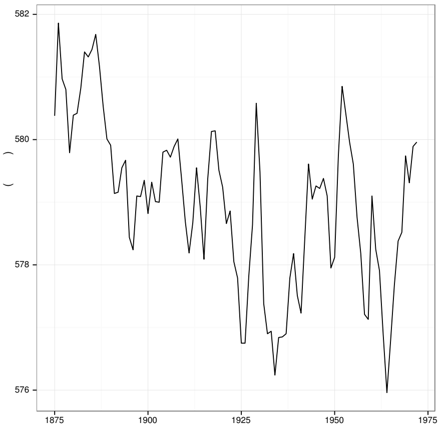
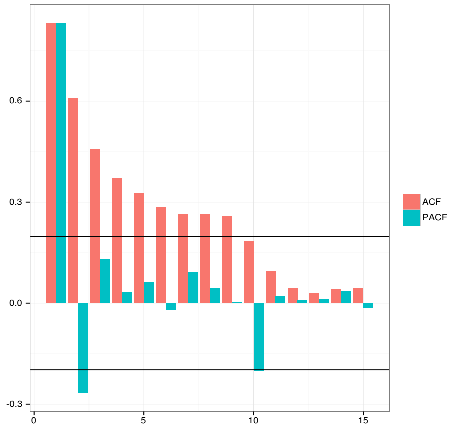


График автокорреляционной и частной автокорреляционной функций:



(a) Судя по графикам, какие модели класса ARMA или ARIMA имеет смысл оценить?

(b) По результатам оценки некоей модели ARMA с двумя параметрами, исследователь посчитал оценки автокорреляционной функции для остатков модели. Известно, что для остатков модели первые три выборочные автокорреляции равны соответственно 0.0047, -0.0129 и -0.063. С помощью подходящей статистики проверьте гипотезу о том, что первые три корреляции ошибок модели равны нулю.

Ответы:

(a) [2], $AR(2)$, т.к. две первые частные корреляции значимо отличаются от нуля, а гипотезы о том, что каждая последующая равна нулю не отвергаются.

(b) [8], Можно использовать одну из двух статистик

$$\text{Ljung-Box} = n(n+2)\sum_{k=1}^{3}\frac{\hat{\rho}_k^2}{n-k} = 0.4289$$

$$\text{Box-Pierce} = n\sum_{k=1}^{3}\hat{\rho}_k^2 = 0.4076$$

Критическое значение хи-квадрат распределения с 3-мя степенями свободы для $\alpha = 0.05$ равно $\chi^2_{3,crit} = 7.8147$. Вывод: гипотеза H_0 об отсутствии корреляции ошибок модели не отвергается. Разбалловка: [4] — формула статистики, [1] — подсчёт Q_{obs}, [1] — знание закона распределения, [1] — подсчёт Q_{crit}, [1] — вывод

2. Винни-Пух пытается выявить закономерность в количестве придумываемых им каждый день ворчалок. Винни-Пух решил разобраться, является ли оно стационарным процессом, для этого он собрал данные за 100 дней и оценил регрессию

$$\Delta\hat{y}_t = \underset{(0.5)}{4.5} - \underset{(0.1)}{0.4}y_{t-1} + \underset{(0.5)}{0.7}\Delta y_{t-1}$$

Из-за опилок в голове Винни-Пух забыл, какой тест ему нужно провести, то ли Доктора Ватсона, то ли Дикого Фуллера.

(a) [2] Аккуратно сформулируйте основную и альтернативную гипотезы

(b) [4] Проведите подходящий тест на уровне значимости 5%

(c) [2] Сделайте вывод о стационарности ряда

(d) [2] Почему Сова не советовала Винни-Пуху пользоваться широко применяемым в Лесу t-тестом?

Ответы:

(a) H_0: ряд содержит единичный корень, $\beta = 0$; H_0: ряд не содержит единичного корня, $\beta < 0$

(b) $ADF = -0.4/0.1 = -4$, $ADF_{crit} = -2.89$, H_0 отвергается

(c) Ряд стационарен

(d) При верной H_0 ряд не стационарен, и t-статистика имеет не t-распределение, а распределение Дики-Фуллера.

3. Продавец мороженного оценил динамическую модель объёмов продаж:

$$\ln\hat{Q}_t = 26.7 + 0.2\ln\hat{Q}_{t-1} - 0.6\ln P_t$$

Здесь Q_t — число проданных в день t вафельных стаканчиков, а P_t — цена одного стаканчика в рублях. Продавец также рассчитал остатки $\hat{e}_t$.

(a) Чему, согласно полученным оценкам, равна долгосрочная эластичность объёма продаж по цене?

(b) Предположим, что продавец решил проверить наличие автокорреляции первого порядка с помощью теста Бройша-Годфри. Выпишите уравнение регрессии, которое он должен оценить.

3 Часть 3. Методы оценивания

В каждой части необходимо решить одну любую задачу по желанию.

1. Данные описываются моделью $y_t = \beta + \varepsilon_t$, где ε_t — стационарный AR(1) процесс, $\varepsilon_t = u_t + \rho\varepsilon_{t-1}$, $u \sim N(0;\sigma^2 \cdot I)$. Имеются наблюдения $y' = (1, 2, 0, 0, 1)$.
   (a) Выпишите функцию правдоподобия для оценивания параметров β, σ^2, ρ
   (b) Найдите оценки неизвестных параметров
2. Винни-Пух знает, что мёд бывает правильный, $honey_i = 1$, и неправильный, $honey_i = 0$. Пчёлы также бывают правильные, $bee_i = 1$, и неправильные, $bee_i = 0$. По 100 своим попыткам добыть мёд Винни-Пух составил таблицу сопряженности:

	$honey_i = 1$	$honey_i = 0$
$bee_i = 1$	12	36
$bee_i = 0$	32	20

   Используя метод максимального правдоподобия Винни-Пух хочет оценить логит-модель для прогнозирования правильности мёда с помощью правильности пчёл:

$$\ln\left(\frac{\mathbb{P}(honey_i = 1)}{\mathbb{P}(honey_i = 0)}\right) = \beta_1 + \beta_2 bee_i$$

   (a) Выпишите функцию правдоподобия для оценки параметров β_1 и β_2
   (b) Оцените неизвестные параметры
   (c) С помощью теста отношения правдоподобия проверьте гипотезу о том, правильность пчёл не связана с правильностью мёда на уровне значимости 5%.
   (d) Держась в небе за воздушный шарик, Винни-Пух неожиданно понял, что перед ним неправильные пчёлы. Помогите ему оценить вероятность того, что они делают неправильный мёд.
3. Пусть p — неизвестная вероятность выпадения орла при бросании монеты. Из 100 испытаний 42 раза выпал «Орел» и 58 — «Решка». Протестируйте на 5%-ом уровне значимости гипотезу о том, что монетка — «правильная» с помощью:
   (a) теста Вальда
   (b) теста множителей Лагранжа
   (c) теста отношения правдоподобия

   Функция правдоподобия, $l(p) = 42 \ln p + (100 - 42) \ln(1 - p)$.
   Численно, $l(\hat{p}) = l(0.42) = -68.0292$, $l(p_0) = l(0.5) = -69.3147$
   Критическая точка χ^2 распределения, 3.8415
   Наблюдаемое значение LR статистики, $LR = 2.571$