Решения задач 2018-2019

(благодаря любезности Б.Б.Демешева)

1. а) $\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2+\lambda\sum_{i=1}^{k}|\hat{\beta}_i|$

   б) Вес второго слагаемого в целевой функции будет выше, и коэффициенты будут зануляться.

   в) Оценки коэффициентов будут стремиться к тем, которые были бы получены при обычном МНК.

2. а) $t_{obs}=(0.375-1)/0.085\approx-7.35, t_{0.975;197}=\pm1.97$. Основная гипотеза отвергается.

   б) Необходимо проверить гипотезу о том, что $\beta_1=\beta_2$. Иначе говоря, что $\beta_1-\beta_2=0$. Найдём стандартную ошибку:

   $\mathrm{Var}(\hat{\beta}_1-\hat{\beta}_2)=0.125^2+0.085^2-2\cdot(-0.0096)=0.04025$, следовательно, $se(\hat{\beta}_1-\hat{\beta}_2)\approx0.205$

   Тогда $t_{obs}=(0.6029-0.375)/0.205\approx1.11, t_{0.975;197}=\pm1.97$. Основная гипотеза не отвергается.

   в) $t_{obs}=-0.964/0.707\approx-1.3635, t_{crit}=\pm1.97$. Основная гипотеза не отвергается.

   г) $F_{0.95;3,194}=2.65$

   $$F=\frac{(0.954-0.9)/3}{(1-0.954)/194}=75.913$$

   Основная гипотеза отвергается.

3. Выпишем ограниченную и неограниченную модели:

   $$\begin{aligned}&R: Y_i=\beta_0+\beta_1X_i+\beta_2Z_i+\varepsilon_i\\&UR: Y_i=\beta_0+\Delta_0d_i+(\beta_1+\Delta_1d_i)X_i+(\beta_2+\Delta_2d_i)Z_i+\varepsilon_i,\end{aligned}$$

   где d_i дамми-перменная, обозначающая городских жителей:

   $$d_i=\begin{cases}1, & \text{для жителей города}\\0, & \text{для сельских жителей}\end{cases}$$

   Проверять будем следующую гипотезу:

   $$\begin{cases}H_0:\Delta_0=\Delta_1=\Delta_2=0\\H_a:\Delta_0^2+\Delta_1^2+\Delta_2^2>0\end{cases}$$

Посчитаем значение статистики:

$$F_{obs} = \frac{(900-500)/3}{500/(500-6)} \approx 131.73$$

Критическое значение: $F_{0.95;3,494} = 2.62$. Основная гипотеза отвергается, то есть зависимость нельзя считать единой.

4. а) Собственные числа матрицы $\tilde{X}^T\tilde{X}$: $\lambda_1 = 1.85$, $\lambda_2 = 1$, $\lambda_3 = 0.15$. Параметр обусловленности: $\sqrt{1.85/0.15}$.

   б) Выясним, сколько главных компонент нужно найти:

$$\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} = \frac{1.85}{3} = 0.61(6) < 0.7$$
$$\frac{\lambda_1+\lambda_2}{\lambda_1+\lambda_2+\lambda_3} = \frac{2.85}{3} = 0.95 > 0.7$$

   Необходимо найти две главные компоненты. Например, подойдут следующие:

$$z_1 = \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \qquad z_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

5. а)
   - Линейная модель. При увеличении общей площади на 1 м^2 продажи в расчёте на квадратный метр снизятся на 15.25 гульденов.
   - Полулогарифмическая модель. При увеличении общей площади на 1 м^2 продажи в расчёте на квадратный метр снизятся на 0.24%.
   - Модель в логарифмах. При увеличении общей площади на 1% продажи в расчёте на квадратный метр снизятся на 0.31%.

   б) Можно воспользоваться PE-тестом. Чтобы выбрать между линейной и полулогарифмическ моделями, на первом шаге находим оценки $\widehat{\ln Sales}$ и $\widehat{Sales}$. На втором — оцениваем вспомогательные регрессии:

$$\begin{aligned} \ln Sales_i &= \gamma_0 + \gamma_1 Size_i + \gamma_2 Nfull_i + \gamma_3 Ntemp_i + \gamma_4 Owner_i \\ &\quad + \theta_1(\widehat{Sales}_i - \exp\left(\widehat{\ln Sales_i}\right)) + \varepsilon_{1i} \\ Sales_i &= \delta_0 + \delta_1 Size_i + \delta_2 Nfull_i + \delta_3 Ntemp_i + \delta_4 Owner_i \\ &\quad + \theta_2(\ln\widehat{Sales}_i - \widehat{\ln Sales_i}) + \varepsilon_{2i} \end{aligned}$$

   Если коэффициент θ_1 незначим, то выбирается полулогарифимическая модель. Если коэффициент θ_2 незначим, то выбирается линейная модель. Если оба коэффициента значимы или незначимы одновременно, то выбор сделать невозможно.
   Для выбора между линейной и линейной в логарифмах моделями процедура аналогична.