Семинар 5.

Задача 1. Пусть $\hat{\sigma}^2$ — стандартная оценка дисперсии σ^2 и матрица $\hat{\sigma}^2\left(X'X\right)^{-1}$ — несмещенная оценка ковариационной матрицы $\operatorname{Var}\left(\widehat{\beta}_{\text{OLS}}\right)$. Тогда имеем

$$t=\frac{\widehat{\beta}_1+\widehat{\beta}_2-(\beta_1+\beta_2)}{\sqrt{\widehat{Var}(\widehat{\beta}_1)+\widehat{Var}(\widehat{\beta}_2)+2\widehat{Cov}(\widehat{\beta}_1,\widehat{\beta}_2)}}\sim t(n-k),$$

и

$$\Bigg(\widehat{\beta}_1+\widehat{\beta}_2-t_c\sqrt{\widehat{Var}(\widehat{\beta}_1)+\widehat{Var}(\hat{\beta}_2)+2\widehat{Cov}(\widehat{\beta}_1,\widehat{\beta}_2)};$$

$$\widehat{\beta}_1+\widehat{\beta}_2+t_c\sqrt{2\widehat{Var}(\widehat{\beta}_1)+\widehat{Var}(\widehat{\beta}_2)+2\widehat{Cov}(\widehat{\beta}_1,\widehat{\beta}_2)}$$

является 95%-ным доверительным интервалом для суммы $\beta_1+\beta_2$, где $t_c=t_{0.025}(n-k)$. Его ширина равна, $2t_c\sqrt{\widehat{Var}(\widehat{\beta}_1)+\widehat{Var}(\widehat{\beta}_2)+2\widehat{Cov}(\widehat{\beta}_1,\widehat{\beta}_2)}$, в то время как длины 95%-ных доверительных интервалов для β_1 и β_2 равны $2t_c\sqrt{\widehat{Var}(\widehat{\beta}_1)}$ и $2t_c\sqrt{\widehat{Var}(\widehat{\beta}_2)}$ соответственно. Нетрудно понять, что если $\widehat{\beta}_1,\widehat{\beta}_2$ достаточно сильно отрицательно коррелированы (коэффициент корреляции близок к -1), то длина доверительного интервала для $\beta_1+\beta_2$ будет меньше, чем длина каждого из доверительных интервалов для β_1 и β_2.

Задача 2. Рассмотрим модель $Y_i=\beta_1+\beta_2X_{i2}+\beta_3X_{i3}+\beta_4X_{i4}+\varepsilon_i$. При оценке модели по 24 наблюдениям оказалось, что $RSS=15$, $\sum(Y_i-\bar{Y}-X_{i3}+\bar{X}_3)^2=20$. На уровне значимости 1% протестируйте гипотезу

$$H_0:\begin{cases}\beta_2+\beta_3+\beta_4=1\\ \beta_2=0\\ \beta_3=1\\ \beta_4=0\end{cases}.$$

Решение: Заметим, что в основной гипотезе есть зависимые ограничения, оставим только независимые:

$$H_0:\begin{cases}\beta_2=0\\ \beta_3=1\\ \beta_4=0\end{cases}$$

Ограниченная модель имеет вид:

$$Y_i=\beta_1+X_{i3}+\varepsilon_i$$

Переносим ε_i в левую часть, и получим оценку коэффициента β_1:

$$\hat{\beta}_1 = \bar{Y} - \bar{X}_3$$

Теперь можно найти RSS_R:

$$RSS_R = \sum_{i=1}^{24}(Y_i - \hat{\hat{Y}}_i)^2 = \sum_{i=1}^{24}(Y_i - \bar{Y} + \bar{X}_3 - X_{i3})^2 = 20$$

Осталось найти значение F-статистики, которая при верной H_0 имеет распределение $F_{3,20}$:

$$F_{obs} = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k_{UR})} = \frac{(20-15)/3}{15/(24-4)} = 20/9$$

Так как $F_{obs} < F_{3,20;0.99} = 4.94$, оснований отвергать нулевую гипотезу нет.