1. Регрессионная модель имеет вид $y_i = \beta_1 + \beta_x x_i + \beta_z z_i + \beta_w w_i + u_i$. Исследователь Феофан оценил эту модель по 20 наблюдениям и оказалось, что $R^2 = 0.8$. Феофан хочет проверить гипотезу H_0 о том, что $\beta_x = \beta_z$ и одновременно $\beta_z + \beta_w = 0$. Предпосылки теоремы Гаусса-Маркова на ошибки u_i выполнены, кроме того, u_i нормально распределены.

   а) Какую вспомогательную регрессию достаточно оценить Феофану для проверки H_0?

   б) Во вспомогательной регрессии оказалось, что $R^2 = 0.7$. Отвергается ли H_0 на 5%-ом уровне значимости?

   в) На сколько процентов изменилась несмещённая оценка дисперсии случайной ошибки при переходе ко вспомогательной регрессии?

   Решение:

   а) Подставляем ограничения и получаем

   $$y_i = \beta_1 + \beta_x(x_i + z_i - w_i) + u_i$$

   б) Применяем F-тест:

   $$F_{obs} = \frac{(R^2_{UR} - R^2_R)/n_{rest}}{(1 - R^2_{UR})/(n - k_{UR})} = \frac{(0.8 - 0.7)/2}{(1 - 0.8)/16} = 4$$

   По таблицам находим критическое значение:

   $$F_{cr,2,16} = 3.63$$

   Вывод: H_0 отвергается.

   в) На сколько процентов изменилась несмещённая оценка дисперсии случайной ошибки при переходе ко вспомогательной регрессии?

   Вспоминаем, что $\hat{\sigma}^2 = RSS/(n-k) = (1-R^2)\cdot TSS/(n-k)$. Величина TSS не изменяется. Меняется только $a = (1-R^2)/(n-k)$. В исходной регрессии $a_{old} = 0.2/16$, во вспомогательной $a_{new} = 0.3/18$. Процентное изменение равно $(a_{new} - a_{old})/a_{old} \approx 33\%$.

2. Исследователь Феофан изучает регрессию со 100 наблюдениями и 10 оцениваемыми коэффициентами. Предпосылки теоремы Гаусса-Маркова на ошибки u_i выполнены, кроме того, u_i нормально распределены.

   Феофан хочет оценить неизвестную дисперсию $\sigma^2 = \mathrm{Var}(u_i)$ по формуле $\hat{\sigma}^2 = c \cdot RSS$ так, чтобы величина среднеквадратичной ошибки была минимальной. Какое значение c получит Феофан?

   Подсказка: Феофан смутно помнит, что дисперсия χ^2-распределения с d степенями свободы равна $2d$.

   Решение:

   Заметим, что $RSS/\sigma^2 \sim \chi^2_{n-k}$, поэтому:

   $$MSE = \mathrm{Var}(\hat{\sigma}^2) + bias^2(\hat{\sigma}^2) = \sigma^4 \cdot (c^2 2(n-k) + (c(n-k)-1)^2).$$

   Минимизируя по c, получаем $c = 1/(n-k+2) = 1/92$.

3. На работе Феофан построил парную регрессию по трём наблюдениям и посчитал прогнозы $\hat{y}_i$. Придя домой он отчасти вспомнил результаты:

y_i	$\hat{y}_i$
0	1
6	?
6	?

Поднапрягшись, Феофан вспомнил, что третий прогноз был больше второго. Помогите Феофану восстановить пропущенные значения.

Решение:

На две неизвестных a и b нужно два уравнения. Эти два уравнения — ортогональность вектора остатков плоскости регрессоров. А именно:

$$\begin{cases} \sum_i (y_i - \hat{y}_i) = 0 \\ \sum_i (y_i - \hat{y}_i)\hat{y}_i = 0 \end{cases}$$

В нашем случае

$$\begin{cases} -1 + (6-a) + (6-b) = 0 \\ -1 + (6-a)a + (6-b)b = 0 \end{cases}$$

Решаем квадратное уравнение и получаем два решения: $a = 4$ и $a = 7$. Итого: $a = 4$, $b = 7$.