Семинар 2.
Решение.

1. Рассмотрим нормальную классическую линейную модель множественной регрессии $y_i = \beta_1 + \beta_2 x_{i2} + ... + \beta_k x_{ik} + \varepsilon_i$ с неслучайными регрессорами. Дополнительно известно, что на самом деле $\beta_2 = ... = \beta_k = 0$.

   (a) Найдите $\mathbb{E}(R^2)$.

   (b) Найдите $\mathbb{E}(R^2_{adj.})$.

   (c) Покажите, что $nR^2 \sim \chi^2(k-1)$.

   Решение:

(a) Модель без ограничений:

$$Y_i = \beta_1 + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i.$$

Модель с ограничениями (истинная модель!):

$$Y_i = \beta_1 + \varepsilon_i.$$

Тогда F-статистика при справедливости H_0 имеет следующий вид:

$$F = \frac{R^2(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k).$$

Выразим R^2:

$$R^2(n-k) = F(1-R^2)(n-k).$$

Утверждение №1: Если $X \sim F(k_1, k_2)$, то $Y = \frac{\frac{k_1}{k_2}X}{1+\frac{k_1}{k_2}X} \sim Beta\left(\frac{k_1}{2}, \frac{k_2}{2}\right)$.

Используя утверждение №1, получаем:

$$R^2 = \frac{(k-1)F}{(n-k)+(k-1)F} = \frac{\frac{k-1}{n-k}F}{1+\frac{k-1}{n-k}F} \sim Beta\left(\frac{k-1}{2}, \frac{n-k}{2}\right).$$

Тогда чтобы посчитать математическое ожидание R^2, надо вспомнить, чему равно математическое ожидание для $Beta\left(\frac{k-1}{2}, \frac{n-k}{2}\right)$:

$$E(R^2) = \frac{\frac{k-1}{2}}{\frac{k-1}{2} + \frac{n-k}{2}} = \frac{k-1}{n-1}.$$

Что нам даёт полученный результат? Математическое ожидание коэффициента детерминации линейно по k. То есть даже при включении в модель лишних факторов

R^2 все равно продолжает линейно расти! Однако мы знаем, что истинная модель — это регрессия на константу, и матожидание ее коэффициента детерминации должно быть равно нулю. В следующем пункте покажем, что использование скорректированного коэффициента детерминации позволяет преодолеть такое нежелательное свойство коэффициента детерминации как линейный рост по параметру k (число регрессоров).

(b) Скорректированный коэффициент детерминации имеет вид:

$$R^2_{adj.} = 1 - \left(1 - R^2\right) \frac{n-1}{n-k}.$$

Рассчитаем математическое ожидание:

$$E(R^2_{adj.}) = E\left(1 - \left(1 - R^2\right) \frac{n-1}{n-k}\right) = 1 - \frac{n-1}{n-k} + \frac{n-1}{n-k} E(R^2) =$$

$$= 1 - \frac{n-1}{n-k} + \frac{n-1}{n-k} \frac{k-1}{n-1} = 0.$$

Скорректированный R^2 помог решить проблему линейного роста по k!

(c) Заметим, что выражение для R^2 из пункта (a) можно переписать в следующем виде:

$$R^2 = \frac{\frac{k-1}{n-k} F}{1 + \frac{k-1}{n-k} F} = \frac{\chi^2_{(k-1)}}{\chi^2_{(n-1)}}.$$

Тогда рассмотрим $\lim_{n\to\infty} nR^2 = \lim_{n\to\infty} n \frac{\chi^2_{(k-1)}}{\chi^2_{(n-1)}} = \lim_{n\to\infty} \frac{\chi^2_{(k-1)}}{\frac{\chi^2_{(n-1)}}{n-1} \frac{n-1}{n}}$.
Заметим, что $\lim_{n\to\infty} \frac{n-1}{n} = 1$, а $\lim_{n\to\infty} \frac{\chi^2_{(n-1)}}{n-1} = 1$.
Следовательно, $\lim_{n\to\infty} nR^2 = \chi^2_{(k-1)}$.