

**Контролно по приложна статистика,  
“Приложна математика”, III курс, май 2017г.**

Време за работа: 45 минути.

Оценяване: Всяка задача дава по 1 точка. 3 точки – среден (3), 6 точки – отличен (6).

**Задача 1.** Интерпретирайте формулата  $E(Y - \hat{Y})^2 = [f(X) - \hat{f}(X)]^2 + Var(\epsilon)$ , където  $Y = f(X) + \epsilon$  е регресионен модел.

**Задача 2.** Какво наричаме overfitting (прекомерно нагаждане) към даден статистически метод?

**Задача 3.** Опишете метода с предварително избиране (forward selection) на най-добро подмножество от предиктори в линейна регресия.

**Задача 4.** Интерпретирайте резултатите от оценката на линеен модел в R:

	Coefficient	Std. error	t-statistic	p-value
Intercept	531.00	46.32	11.464	< 0.0001
ethnicity[Asian]	-18.69	65.02	-0.287	0.7740
ethnicity[Caucasian]	-12.50	56.68	-0.221	0.8260

**Задача 5.** Определете коя от следните формули отговаря на  $AIC$ ,  $BIC$ ,  $R^2$ :

$$1 - \frac{RSS}{TSS} \qquad \frac{1}{n\hat{\sigma}^2} (RSS + 2d\hat{\sigma}^2) \qquad \frac{1}{n} (RSS + \log(n)d\hat{\sigma}^2)$$

**Задача 6.** Формулата  $\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$  определя модел на ...