Ασκήσεις Στατιστικής (Γραμμική Παλινδρόμηση)

1	Use the statistical model $$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$ to show that $\epsilon_i \sim \text{NID}(0, \sigma^2)$ implies each of the following: (a) $\mathcal{E}(Y_i) = \beta_0 + \beta_1 X_i$, (b) $\sigma^2(Y_i) = \sigma^2$, and (c) $\text{Cov}(Y_i, Y_{i'}) = 0,\ i \neq i'$. For Parts (b) and (c), use the following definitions of variance and covariance. $$\begin{aligned} \sigma^2(Y_i) &= \mathcal{E}\{[Y_i - \mathcal{E}(Y_i)]^2\} \\ \text{Cov}(Y_i,\ Y_{i'}) &= \mathcal{E}\{[Y_i - \mathcal{E}(Y_i)][Y_{i'} - \mathcal{E}(Y_{i'})]\}. \end{aligned}$$
2	Obtain the normal equations and the least squares estimates for the model $$Y_i = \mu + \beta_1 x_i + \epsilon_i,$$ where $x_i = (X_i - \overline{X})$. Compare the results to equation 1.6. (The model expressed in this form is referred to as the "centered" model; the independent variable has been shifted to have mean zero.) Εδώ ζητάει να λύσετε το πρόβλημα του προσδιορισμού των συντελεστών όπως και στο πρόβλημα που λύσαμε στην τάξη αλλά αφού πρώρα αλλαχτούν τα X. Παρατηρείστε ότι ο σταθερός συντελεστής (intercept) βγαίνει πολύ εύκολα σε αυτή την περίπτωση και είναι το μ, ο δειγματικός μέσος των Y.
3	If $\hat{\beta}_0$ and $\hat{\beta}_1$ are the least-squares estimates for the intercept and slope in a simple linear regression model, show that the least-squares equation $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ always goes through the point $(\overline{x}, \overline{y})$. [*Hint:* Substitute $\overline{x}$ for x in the least-squares equation and use the fact that $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$.]
4	Suppose that we have postulated the model $$Y_i = \beta_1 x_i + \varepsilon_i \qquad i = 1, 2, \ldots, n,$$ where the ε_i's are independent and identically distributed random variables with $E(\varepsilon_i) = 0$. Then $\hat{y}_i = \hat{\beta}_1 x_i$ is the predicted value of y when $x = x_i$ and $\text{SSE} = \sum_{i=1}^{n} [y_i - \hat{\beta}_1 x_i]^2$. Find the least-squares estimator of β_1. (Notice that the equation $y = \beta x$ describes a straight line passing through the origin. The model just described often is called the *no-intercept* model.)

<table>
<tr><td></td><td>Στό πρόβλημα αυτό πρέπει ουσιαστικά να επαναλάβετε τη διαδικασία που μας οδήγησε στον υπολογισμό των συντελεστών a και b, αλλά τώρα η ελαχιστοποίηση του αθροίσματος των τετραγώνων των αποστάσεων των σημείων από την ευθεία γίνεται μόνο για τις ευθείες που περνάνε από το 0.</td></tr>
<tr><td>5</td><td>

a Derive the following identity:

$$\begin{aligned} \text{SSE} &= \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\ &= \sum_{i=1}^{n}(y_i - \overline{y})^2 - \hat{\beta}_1 \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = S_{yy} - \hat{\beta}_1 S_{xy}. \end{aligned}$$

Notice that this provides an easier computational method of finding SSE.

b Use the computational formula for SSE derived in part (a) to prove that SSE $\leq$ S_{yy}. [*Hint:* $\hat{\beta}_1 = S_{xy}/S_{xx}$.]

Εδώ SSE είναι αυτό που στις σημειώσεις σας ονομάζεται RSS.</td></tr>
<tr><td>6</td><td>

Με το συμβολισμό των σημειώσεων δείξτε ότι το άθροισμα των υπολοίπων

$$\sum_{i=1}^{N} \widehat{Y}_i - Y_i$$

είναι πάντα 0.</td></tr>
</table>