Question 1

A biologist performed a regression on how much a planet diameter (x) affects it's mass (y), Based on 30 samples. His concllusion was $\hat{y} = 10 + 0.1x$. were x is in trillions of tons (10^{12} kilogram). Which regression line would he get if he would use units of ten trillions of ton? (10^{13} kilogram)?
Guidance - compute the new ss_x, ss_{xy} as a function of the old ones. See what happens to new $b_1 = ss_{xy}/ss_x$ and $b_0 = \bar{y} - b_1\bar{x}$

<u>שאלה 1</u>

נסמן את המשתנים (חלקם הוגדרו בשאלה):
Y – המשתנה המוסבר: קוטר הכוכב.
X – המשתנה המסביר במודל המקורי "הנכון": המשקל בעשרות טריליוני טון.
Z – המשתנה המסביר במודל "השגוי": המשקל בטריליוני טון

במדגם שאסף המהנדס, כל תצפית מורכבת מנתון אחד של y_i ונתון אחד של z_i. משוואת הרגרסיה שהוא בנה אומדת את **y כפונקציה של z**, בעוד שהמנהל מעוניין במודל שמסביר את y כפונקציה של x.
כדי לחשב את האומדים במודל שנבנה, חושבו SS_{zy} ו-SS_z.
כדי לחשב את האומדים במודל הרצוי, נדרשים SS_{xy} ו-SS_x.
נשים לב לקשרים הבאים בין המשתנים:
לכל תצפית i, מתקיים $x_i = \frac{1}{10}z_i$. לכן $\bar{x} = \frac{\sum_i x_i}{n} = \frac{\sum_i \frac{1}{10}z_i}{n} = \frac{1}{10}\bar{z}$.
אם כך:

$$SS_z = \sum_i^n (z_i - \bar{z})^2 = \sum_i^n (10x_i - 10\bar{x})^2 = 100\sum_i^n (x_i - \bar{x})^2 = 100 \cdot SS_x$$

$$SS_{zy} = \sum_i^n (z_i - \bar{z})(y_i - \bar{y}) = \sum_i^n (10x_i - 10\bar{x})(y_i - \bar{y}) = 10\sum_i^n (x_i - \bar{x})(y_i - \bar{y}) = 10 \cdot SS_{xy}$$

נסמן:

- b_1^x, b_0^x – האומדים במודל שבו המשתנה המסביר הוא x (אותם מחפשים)
- b_1^z, b_0^z – האומדים במודל שבו המשתנה המסביר הוא z (הם נתונים בשאלה)

$$b_1^x = \frac{SS_{xy}}{SS_x} = \frac{\frac{1}{10}SS_{zy}}{\frac{1}{100}SS_z} = \frac{100}{10}\frac{SS_{zy}}{SS_z} = 10b_1^z = 10 \cdot 0.1 = 1$$

$$b_0^x = \bar{y} - b_1^x\bar{x} = \bar{y} - 10b_1^z \cdot \frac{1}{10}\bar{z} = \bar{y} - b_1^z\bar{z} = b_0^z = 10$$

כלומר האומד לשיפוע גדל פי 10, והאומד לחותך לא משתנה.
לכן קו הניבוי החדש הינו $\hat{y} = 10 + x$.

Question 2

Given that $\bar{x} = 432.2, \sum_{i=1}^{10} x_i^2 = 2,048,810, \sum_{i=1}^{10} y_i^2 = 103,195, \hat{y} = 5.821 + 0.195x$

Calculate:

1. $\bar{y}$
2. R^2
3. is there a positive linear connection between x and y? Guidance, use hypothesis testing.

<u>פתרון</u>

א. קו הריבועים הפחותים עובר בנקודת הממוצעים $(\bar{x},\bar{y})$ ולכן:

$$\bar{y} = b_0 + b_1 \cdot \bar{x} = 5.821 + 0.195 \cdot 432.2 = 90.1$$

ב. $r = \dfrac{SS_{xy}}{\sqrt{SS_x SS_y}}$

$$SS_x = \sum_{i=1}^{10} x_i^2 - n\bar{x}^2 = 2,408,810 - 10 \cdot 432.2^2 = 540,841.6$$

$$SSR = b_1^2 SS_x = 0.195^2 \cdot 540,841.6 = 20,565.5$$

$$SST = SS_y = \sum_{i=1}^{10} y_i^2 - n\bar{y}^2 = 103,195 - 10 \cdot 90.1^2 = 22,014.9$$

$$\Rightarrow R^2 = \frac{SSR}{SST} = \frac{20,565.5}{22,014.9} = 0.934$$

ג. קשר לינארי חיובי משמעו ש- $\beta_1 > 0$.

מערכת ההשערות: $H_0 : \beta_1 = 0, \quad H_1 : \beta_1 > 0$.

סטטיסטי המבחן: $T_{b_1} = \dfrac{b_1}{s/\sqrt{SS_x}} = \dfrac{0.195}{s/\sqrt{540,841.6}}$

$$s = \sqrt{\frac{SS_y - b_1^2 SS_x}{n-2}} = \sqrt{\frac{22,014.9 - 0.195^2 \cdot 540,841.6}{8}} = 13.46$$

$$\Rightarrow T_{b_1} = 10.65, \quad > \quad t_{1-\alpha}^{n-2} = t_{0.95}^{8} = 1.86$$

ולכן השערת האפס נדחית בר״מ 0.05, כלומר קיים קשר ליניארי חיובי.