Zadanie 1
Rozważamy model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

Estymatory $\beta_0, \beta_1, \beta_2, \sigma$ wynoszą $b_0 = 1, b_1 = 4, b_2 = 3, s = 3$

Dokonamy predykcji dla wartości $X_1 = 2, X_2 = 6$:
$\hat{Y} = 1 + 4 \cdot 2 + 3 \cdot 6 = 27$

W drugim kroku będziemy estymować wartość wariancji błędu predykcji Y_h, przy założeniu, że standardowe odchylenie estymatora wartości oczekiwanej Y_h wynosi 2, tj. $\hat{\mu_h} = 2$, dla X_1, X_2 takich jak w poprzednim punkcie. Otrzymujemy:
$s^2(pred) = s^2 + s^2(\hat{\mu_h}) = 9 + 4 = 13$.

Załóżmy, że nasz model został wytrenowany na 20 obserwacjach, a standardowe odchylenie b_1, $s(b_1)$, wynosi 1. Skonstruuję 95% przedział ufności dla β_1.
Jest on dany wzorem: $b_1 \pm t_c s(b_1)$, gdzie t_c jest kwantylem rzędu 0.95 z rozkładu studenta z 17 stopniami swobody. Korzystając z funkcji qt w R, otrzymujemy, że jest on w przybliżeniu równy 2.1. Otrzymujemy przedział ufności postaci $(1.9, 6.1)$.

Zadanie 2

Analizujemy dane korzystając z modelu $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$.
Mamy dane sumy typu I i II:

	Typ I	**Typ II**
X_1	300	30
X_2	40	25
X_3	20	?

Oraz zakładamy, że $SST = 760, n = 24$.

W pierwszym kroku wyznaczymy sumę typu II dla X_3. Z definicji, jest ona równa $SSM(X_3|X_1, X_2)$, co jest równe sumie typu I dla X_3, a więc wynosi ona 20.

W drugim kroku zbadamy istotność regresora X_1. Porównamy modele:
H_0 : dane pochodzą z modelu $Y = \beta_0 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$.
H_1 : dane pochodzą z modelu $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$.
Statystyka testowa F wynosi $\dfrac{(SSE(R) - SSE(F))/(dfE(R) - dfE(F))}{MSE(F)} = \dfrac{SSM(X_1|X_2, X_3)}{MSE(F)}$
Z tabeli odczytujemy, że $SSM(X_1|X_2, X_3) = 30$.
Wiemy, że $MSE(F) = SSE(F)/dfE(F)$. $dfE(F)$ wynosi $24 - 4 = 20$, natomiast $SSE(F) = SST - SSM = 760 - 300 - 40 - 20 = 400$.
Otrzymujemy $F = \dfrac{30}{400/20} = 1.5$
Przy użyciu R, obliczę F^*, czyli kwantyl rzędu 0.95 z rozkładu Fischera-Snedecora o 1, 20 stopniach swobody. Wynosi on 4.35, a więc jest znacznie większy od statystyki testowej F, co oznacza, że nie możemy odrzucić hipotezy $H_0 : \beta_1 = 0$.

W trzecim podpunkcie, przetestuję hipotezę $\beta_2 = 0, \beta_3 = 0$. Analogicznie jak w poprzednim kroku, porównam modele:
H_0 : dane pochodzą z modelu $Y = \beta_0 + \beta_1 X_1 + \epsilon$.
H_1 : dane pochodzą z modelu $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$.

$SSE(R) - SSE(F) = SSM(F) - SSM(R) = SSM(X_1, X_2, X_3) - SSM(X_1) = 60.$
Otrzymujemy $F = \frac{60/2}{20} = 1.5$. Wartość F^* (z rozkładu FIschera-Snedecora o 2, 20 stopniach swobody), wynosi 3.49, więc ponownie jest ona wyższa of F, przez co nie możemy odrzucić hipotezy $\beta_2 = \beta_3 = 0$

W kolejnym kroku przetestuję hipotezę $\beta_1 = \beta_2 = \beta_3 = 0$, czyli porównam modele:
H_0 : dane pochodzą z modelu $Y = \beta_0 + \epsilon$.
H_1 : dane pochodzą z modelu $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$.
Statystyka testowa F wynosi $\frac{360/3}{20} = 6$
Kwantyl rzędu 0.95 z rozkładu Fishera-Snedecora o 3, 20 stopniach swobody wyniósł $F^* = 3.09$, co jest mniejsze od F. Możemy więc odrzucić hipotezę, że $\beta_1 = \beta_2 = \beta_3 = 0$.

W przedostatnim kroku, odrzucę zmienne X_2, X_3 i zbadam hipotezę:
$H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$.
Statystyka testowa wynosi $F = \frac{MSM}{MSE} = \frac{dfE \cdot SSM}{dfM \cdot SSE} = \frac{22 \cdot SSM}{1 \cdot SSE} = \frac{22 \cdot 30}{760 - 30} = 0.9$
Kwantyl rzędu 0.95 z rozkładu Fishera-Snedecora o 1, 22 stopniach swobody wynosi 4.3, co jest większe od F, więc nie możemy odrzucić hipotezy $H_0 : \beta_1 = 0$

W ostatnim podpunkcie, policzymy korelację próbkową pomiędzy Y a X_1. Skorzystam z faktu, że współczynnik determinacji R^2 jest kwadratem korelacji próbkowej.
$$R^2 = \frac{SSM(X_1|X_2, X_3)}{SSM(X_1|X_2, X_3) + SSE(F)} = \frac{30}{430} \approx 0.069$$

Otrzymujemy, że korelacja próbkowa jest w przybliżeniu równa ± 0.26