1. **VPLIV STANDARDIZACIJE PODATKOV NA REZULTATE**:

   Za različne metode standardizacije podatkov podane pod točko a) b) in c) boste izračunali parametre modela / enačbe po metodi LSE in PCA.

   Za LSE metodo uporabite model s prostim členom: y = a1 x1 + … + r

   S PCA metodo izračunamo model tako, da vse meritve vključno z izhodom zložimo v matriko podatkov X (X = [A_LOW T_H2O C_ACID I_EFF]). Naredimo svd ([P,D] = svd(F)). Kjer je F = X'X/(N-1) Izbrati morate pravilni lastni vektor P1 = P(:,i), ki nam podaja enačbo modela. S tem dobite enačbo v obliki (V = center podatkov):
   P1' * (x-V) = 0.
   To enačbo pretvorite v obliko y = a1 x1+…+r.
   PRIMER:
   [p1 p2] *([x1 y]' – [v1 v2]') = 0
   p1 x1 + p2 y – p1 v1 –p2 v2 = 0
   y = -p1/p2 x1 + p1/p2 v1 + p2/p2 v2

   Za primerjavo parametrov in dobljenih modelov morate modele dobljene pod točko a) in b) pretvoriti nazaj v dejanski prostor, saj v primeru normiranih spremenljivk dobite tudi normirane parametre.
   PRIMER:
   y_n = (y-m_y)/n_y x_n = (x –m_x) / n_x

   y_n = a_n * x_n + r_n
   (y-m_y)/n_y = a_n * (x –m_x) / n_x + r_n
   …
   y = (a_n * n_y / n_x) * x - ((a_n * n_y / n_x) *m_x - r_n * n_y – m_y)

   Ko to naredite imate za vsako metodo in vsako normiranje svoje parameter. Parametre lahko predstavite v razpredelnici.

   Naredite primerjavo med PCA in LSE modeli glede na standardizacijo. Poglejte katera metoda je manj občutljiva na standardizacijo. Poglejte katera daje bolj podobne rezultate.

   Za vsak model izračunajte napako (v originalnem prostoru): e = y – ksi*theta. Izračunajte povprečno vrednost napake in standardno deviacijo napake. Iz teh dveh vrednosti lahko ocenimo PRISTRANSKOST in KONSISTENCO modela. Dober model ni pristranski in je konsistenten. Manjša kot je standardna deviacija napake bolj je model konsistenten. Bližje kot je povprečna vrednost napake ničli, manj pristranski je model. Za primerjavo naših modelov med seboj je dovolj da primerjamo standardne deviacije napake in povprečno vrednost napake.

   Če želimo na splošno oceniti kako dober je model pa uporabimo običajno NRMSE napako:

Sqrt(Sum((yocene-y)^2) /n)/std(y)

V poročilo vključite komentar katera metoda je manj občutljiva na standardizacijo podatkov in kateri model je po vašem mnenju boljši. Tisti bolj pridni boste vključili še kakšno teoretično ozadje zakaj je ena od metod bolj občutljiva na standardizacijo podatkov.

2. **KOLINEARNOST**:

Dodamo meritev, ki je odvisna od že podane meritve. Standardiziramo podatke.

Izračunamo parametre modela z LSE metodo (ni potrebno uporabiti prostega člena).

Izračunamo parametre s PCR. Sestavimo matriko X iz vhodnih podatkov. S pomočjo PCA odstranite glavno komponento z zanemarljivim vplivom. Originalne podatke transformirate v prostor glavnih komponent T = X*Ps.
Izračunate parameter z LSE metodo THETA = (T'T)^-1 T' Y
Pretvorimo parameter iz prostora glavnih komponent THETA = Ps THETA.

Tu se ustavite in pregledate dobljene parametre. Kaj boste opazili.
a)PCR in LSE parametri so čisto drugačni, iz napake težko rečemo kateri so boljši.
b)Da sta parametra pri dodani spremenljivki in tisti, od katere je odvisna zelo podobna.
c)Da je vsota teh dveh parametrov približno enaka parametru, ki nastopa pri originalnem setu podatkov (parameter za temperaturo).
d)če naredimo razpredelnico glavnih komponent in originalnih spremenljivk boste opazili, da imata dodana odvisna spremenljivka in originalna spremenljivka, isti vplivni koeficijent na glavne komponente.

Za izbrane glavne komponente (lastne vektorje) naredimo naslednjo razpredelnico:

SPREMENLJIVKA	P_1	P_2
X1	P11	P21
X2	P12	P22
X3	P13	P23
…	..	

P11 je prvi element prvega lastnega vektorja. Za lažjo predstavo lahko v tabelo vpišemo procente, ki jih dobimo kot abs(P_1)./sum(abs(P_1)). Ti nam povejo koliko prvotne spremenljivke X1 je vsebovane v glavni komponenti.