Zadanie 1:
Drużyna informatyków testuje nowy program rozciągania pleców w przerwach od programowania zespołowego, by ograniczyć liczbę dolegliwości. Poniższa tabelka pokazuje codzienną liczbę minut rozciągania się przez uczestników i powiązaną z nimi liczbę dolegliwości przez cały semestr.

Czas rozciągania w minutach $= x$	0	30	10	15	5	25	35	40
Dolegliwości $= y$	4	1	2	2	3	1	0	1

(a) Znajdź prostą regresji, pokazującą zależność ilości dolegliwości od czasu rozciągania.

(b) O ile spada ilość dolegliwości z każdą minutą rozciągania?

(c) Ile minut rozciągania jest potrzebne, by zawodnik uniknął dolegliwości?

Rozwiązanie: Aby znaleźć prostą regresji posłużymy się wzorem z wykładu dla współczynników przy jednej zmiennej objaśniającej:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}.$$

Po podstawieniu naszych danych otrzymujemy:

$$\bar{x} = \frac{0 + 30 + 10 + 15 + 5 + 25 + 35 + 40}{8} = 20,$$

$$\bar{y} = \frac{4 + 1 + 2 + 2 + 3 + 1 + 0 + 1}{8} = \frac{14}{8} = 1\frac{3}{4},$$

$$\hat{\beta}_1 = \frac{-45 + (-7.5) + (-2.5) + (-1.25) + (-18.75) + (-3.75) + (-26.25) + (-15)}{(-20)^2 + 10^2 + (-10)^2 + (-5)^2 + (-15)^2 + 5^2 + 15^2 + 20^2},$$

$$\hat{\beta}_1 = \frac{-120}{1500} = -0.08,$$

$$\hat{\beta}_0 = 1.75 + 0.08 \cdot 20 = 3.35.$$

Więc prosta regresji jest określona wzorem $\hat{y} = 3.35 - 0.08 \cdot x$.

Co za tym idzie według naszego estymatora:

(a) $\hat{y} = 3.35 - 0.08 \cdot x$

(b) Co minutę rozciągania ilość dolegliwości spada o 0.08.

(c) By zawodnik uniknął dolegliwości są potrzebne 41.875 minuty rozciągania.

Rozwiązanie:

(a) Znajdź prostą regresji, pokazującą zależność ilości dolegliwości od czasu rozciągania.

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{160}{8} = 20$$

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{14}{8} = 1.75$$

$x_i - \overline{x}$	-20	10	-10	-5	-15	5	15	20
$y_i - \overline{y}$	2.25	-0.75	0.25	0.25	1.25	-0.75	-1.75	-0.75
$(x_i - \overline{x})(y_i - \overline{y})$	-45	-7.5	-2.5	-1.25	-18.75	-3.75	-26.25	-15
$(x_i - \overline{x})^2$	400	100	100	25	225	25	225	400

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{-120}{1500} = -0.08$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 1.75 + 0.08 \cdot 20 = 3.35$$

Czyli mamy:

$$\hat{y} = -0.08x + 3.35$$

(b) O ile spada ilość dolegliwości z każdą minutą rozciągania?

Ilość dolegliwości spada o 0.08 z każdą minutą rozciągania.

(c) Ile minut rozciągania jest potrzebne, by zawodnik uniknął dolegliwości?

$$-0.08x + 3.35 = 0$$

$$x = 41.875$$

Potrzebnych jest 41.875 minut. **Rozwiązanie:** a) Wiem z wykładu 6., że w przypadku $Y = [y_1, \ldots y_n]^T$, $X = [x_1, \ldots x_n]^T$, Y to zmienna objaśniana (u nas dolegliwości), a X to zmienna objaśniająca (czas rozciągania).

Regresja liniowa przedstawia się jako:

$$Y = \widehat{\beta_1}X + \widehat{\beta_0}$$

gdzie to współczynniki regresji wyznaczane z:

$$\widehat{\beta_1} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

kolejnie

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$$

x_i	y_i	y_i - $\bar{y}$	x_i -$\bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
0	4	2.25	-20	400	-45.0
30	1	-0.75	19	100	-14.25
10	2	0.25	-10	100	-2.5
15	2	0.25	-5	25	-1.25
5	3	1.25	-15	225	-18.75
25	1	-0.75	5	25	-3.75
35	0	-1.75	15	225	-26.25
40	1	-0.75	20	400	-15.0
suma				1475	-126.75

W naszym przypadku $\bar{x} = 20$, $\bar{y} = 1.75$ gdzie $\bar{x}$ średnia arytmetyczna ze zmiennych x_i a $\bar{y}$ średnia arytmetyczna ze zmiennych y_i

Z obliczeń sumując kolumny ostatnią i przedostatnią i dzieląc, tak jak we wzorze dostajemy, że

$$\bar{\beta_1} = \frac{-120}{1500} = -0.08$$

,

$$\beta_0 = \hat{y} - \beta_1 \hat{x} = 1.75 - 20 \times (-0.08) = 3.35$$

Prosta regresji można opisać równaniem: Y = -0.08X + 3.35

b) Wówczas ilość dolegliwości z każdą minutą rozciągania spada średnio o $-\beta_1$, czyli 0.08.

c) Obliczam, punkt przecięcia się prostej

$$Y = -0.08X + 3.35$$

z osią O_x i otrzymuję:

$$-0.08x_0 + 3.35 = 0$$

Co daję mi:

$$x_0 = 41.875$$

Do uniknięcia dolegliwości, zawodnik powinien ćwiczyć 42 minuty.

Zadanie 2:

Pokaż, że jeżeli $n \geq p$ oraz $rank(X) = n$ to estymator β metodą najmniejszych kwadratów

$$\hat{\beta} = arg\ \underset{\beta}{min} ||Y - X\beta||^2$$

jest postaci

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Dodatkowo jeżeli $\mathcal{E} \sim N(0, \sigma^2 Id)$ to pokaż, że $\hat{\beta}$ jest estymatorem największej wiarygodności oraz znajdź estymator największej wiarygodności σ^2

Rozwiązanie: Chcąc policzyć estymator β metodą najmniejszych kwadratów chcemy policzyć gradient wyrażenia $||Y - X\beta||^2$ oraz znaleźć takie β dla, którego wynosi on 0.

$$f(\beta) = ||Y - X\beta||^2 = (Y - X\beta)^T (Y - X\beta)$$

Określmy funkcję:

$$g : \mathbb{R}^n \to \mathbb{R}\ ;\ g(x) = x^T x$$
$$h : \mathbb{R}^p \to \mathbb{R}^n\ ;\ h(z) = Y - Xz$$

Zauważmy że $f(\beta) = g(h(\beta))$. Stąd wykorzystując regułę łańcuchową możemy łatwo policzyć pochodną $\frac{\partial f(\beta)}{\partial \beta}$

Zauważmy, że:

$$\frac{\partial g(x)}{\partial x} = 2x^T$$

$$\frac{\partial h(z)}{\partial z} = -X$$

Stąd:

$$\frac{\partial f(\beta)}{\partial \beta} = \frac{\partial g(h(\beta))}{\partial h(\beta)} \frac{\partial h(\beta))}{\partial \beta} = -2X^T(Y - X\beta)$$

Przyrównując do 0 gradient otrzymujemy:

$$(Y - X\beta)^T(-X) = 0$$
$$((Y - X\beta)^T X)^T = 0$$
$$X^T(Y - X\beta) = 0$$
$$X^T X\beta = X^T Y$$

Z założeń zadania wiemy, że $X^T X$ jest odwracalna stąd:

$$\beta = (X^T X)^{-1} X^T Y$$

Wiemy, że $\mathcal{E} \sim N(0, \sigma^2 Id)$ stąd $Y \sim N(\beta X, \sigma^2 Id)$, więc:

$$p(y_i|x_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \beta x_i)^2}{2\sigma^2}}$$

Chcemy zmaksymalizować funkcję wiarygodności $L(Y, X, \beta, \sigma^2)$

$$L(Y, X, \beta, \sigma^2) = \prod_i^n p(y_i|x_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \beta x_i)^2}{2\sigma^2}}$$

Standardowo szukamy maksimum funkcji $l = logL$

$$l(Y, X, \beta, \sigma^2) = -\frac{n}{2} log 2\pi - n \log \sigma - \frac{1}{2\sigma^2} \sum_i^n (y_i - \beta x_i)^2$$

Zauważmy, że maksymalizując funkcję wiarygodności w zależności od β chcemy zminimalizować człon $\frac{1}{\sigma^2} \sum_i^n (y_i - \beta x_i)^2$, który jest równy $\frac{1}{\sigma^2} ||Y - X\beta||^2$. Co jest tożsame z minimalizacją, którą dokonaliśmy za pomocą metody najmniejszych kwadratów. Więc $\hat{\beta}$ jest estymatorem największej wiarygodności.

$$\frac{\partial l(Y, X, \beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^3} + \frac{1}{2\sigma^4} \sum_i^n (y_i - \beta x_i)^2$$

Przyrównując do zera otrzymujemy estymator największej wiarygodności $\hat{\sigma}^2$:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i^n (y_i - \beta x_i)^2 = \frac{1}{n} ||Y - X\beta||^2$$

Rozwiązanie: Przyjmijmy, że L ma postać

$$L = \|Y - X\beta\|^2$$
$$L = \|Y - X\beta\|^2 = (Y - X\beta)^T(Y - X\beta) = Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta$$

Chcemy zminimalizować różnicę pomiędzy naszą estymacją $(X\beta)$ a wartościąrzeczywistą (Y), więc liczymy pochodną i przyrównujemy ją do zera:

$$\frac{\partial L(X,\beta)}{\partial\beta} = \frac{\partial(Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta)}{\partial\beta} = -2X^TY + 2X^TX\beta = 0$$

$$X^TX\beta = X^TY$$

$$\beta = (X^TX)^{-1}X^TY$$

Policzmy teraz β korzystając z estymatora największej wiarygodności.

$$\hat{\beta} = argmax_{\beta}\mathbb{P}(Y_i|X, \beta, \sigma^2 Id)$$

Z założeń regresji liniowej mamy, że

$$Y|X \sim N(X\beta, \sigma^2 Id)$$

Chcemy zatem znaleźć maksimum tego:

$$\mathcal{L} = \prod_{i=1}^{n}\mathbb{P}(Y_i|X_i) = \frac{1}{\sigma\sqrt{2\pi}}exp(-\frac{1}{2\sigma^2}(Y - X\beta)^2)$$

Po zlogarytmowaniu obustronnie:

$$\ell = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)$$

Wyraz wolny jest niezależny od β, mamy zatem następujący problem optymalizacyjny:

$$\hat{\beta} = argmax_{\beta} - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)$$

Wyraz $\frac{1}{2\sigma^2}$ nie zmienia nam wyniku, możemy zatem go usunąć i przekształcić nasze równanie otrzymując:

$$\hat{\beta} = argmin_{\beta}(Y - X\beta)^T(Y - X\beta).$$

Czyli dokładnie to co mieliśmy w przypadku metody najmniejszych kwadratów.

Znajdźmy teraz estymator dla parametru σ^2:

$$\hat{\sigma^2} = argmax_{\sigma^2} - \frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)^T(Y - X\beta)$$

Różniczując ze względu na σ^2 i przyrównując do 0 otrzymujemy:

$$0 = -\frac{n}{2\sigma^2} - \frac{1}{2\sigma^4}(Y - X\beta)^T(Y - X\beta)$$

$$0 = -n - \frac{1}{\sigma^2}(Y - X\beta)^T(Y - X\beta)$$

$$\sigma = \frac{1}{n}(Y - X\beta)^T(Y - X\beta)$$

Zadanie 3:
Oblicz wartość oczekiwaną i wariancję $\hat{\beta}$. Jaki ma rozkład $\hat{\beta}$, jeśli $\epsilon \sim N(0, \sigma^2)$? **Rozwiązanie:**
Z zadania poprzedniego mamy, że $\hat{\beta} = (X^TX)^{-1}X^TY$ oraz, że $Y = X\beta + \epsilon$ Policzmy najpierw wartość oczekiwaną $\hat{\beta}$:

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}((X^TX)^{-1}X^TY) = (X^TX)^{-1}X^T\mathbb{E}(Y) = (X^TX)^{-1}X^T\mathbb{E}(X\beta + \epsilon) =$$

Skorzystaliśmy z faktu, że X i β ustalone, dodatkowo wiemy, że $\mathbb{E}(\epsilon) = 0$, stąd:

$$= (X^TX)^{-1}X^T\mathbb{E}(X\beta) + (X^TX)^{-1}X^T\mathbb{E}(\epsilon) = (X^TX)^{-1}X^TX\beta = \beta.$$

Policzymy teraz wariancję $\hat{\beta}$ ponieważ $\mathbb{E}(\hat{\beta}) = \beta$:

$$VAR(\hat{\beta}) = \mathbb{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T$$

Zastanówmy się jak wygląda $(\hat{\beta} - \beta)$:

$$(\hat{\beta} - \beta) = (X^TX)^{-1}X^TY - (X^TX)^{-1}X^T(Y - \epsilon) = (X^TX)^{-1}X^T(Y - Y + \epsilon) = (X^TX)^{-1}X^T\epsilon$$

Stąd otrzymujemy:

$$VAR(\hat{\beta}) = \mathbb{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T = \mathbb{E}((X^TX)^{-1}X^T\epsilon \cdot ((X^TX)^{-1}X^T\epsilon)^T) = \mathbb{E}((X^TX)^{-1}X^T\epsilon \cdot \epsilon^TX(X^TX)^{-1}) =$$

(bo $(AB)^T = B^TA^T$) dodatkowo $(X^TX)^{-1}X^T$ to stała którą wyciągamy z wartości oczekiwanej:

$$= (X^TX)^{-1}X^T(\mathbb{E}\epsilon \cdot \epsilon^T)X(X^TX)^{-1} =$$

z założenia $VAR(\epsilon) = \sigma^2$ oraz $\mathbb{E}(\epsilon) = 0$ stąd $\mathbb{E}(\epsilon)^2 = VAR(\epsilon) + (\mathbb{E}(\epsilon))^2 = \sigma^2$, wiedząc dodatkowo, że $(X^TX)^{-1}X^TX = Id$ otrzymujemy:

$$VAR(\hat{\beta}) = (X^TX)^{-1}X^T\sigma^2X(X^TX)^{-1} = (X^TX)^{-1}X^TX\sigma^2(X^TX)^{-1} = \sigma^2(X^TX)^{-1}$$

Jeżeli założymy dodatkowo, że $\epsilon \sim N(0, \sigma^2 Id)$ łatwo zauważyć, że $\hat{\beta} \sim N(\beta, \sigma^2(X^TX)^{-1})$, ponieważ:

$$\hat{\beta} = (X^TX)^{-1}X^TY = (X^TX)^{-1}X^T(X\beta + \epsilon) = (X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\epsilon$$

gdzie ϵ jest zmienną gaussowską natomiast $\hat{\beta}$ jest tą zmienną pomnożoną razy stałą i z dodaną inną stałą - też jest gaussowska z średnią i wariancją jak wyżej.

Rozwiązanie: Przypomnijmy, że $\hat{\beta}$ dane jest wzorem $\hat{\beta} = (X^T X)^{-1} X^T y$, a o y zakładamy, że spełnia $y = X\beta + \epsilon$ (z ϵ jak w treści zadania). Bezpośrednio z definicji obliczamy

$$\mathbb{E}[\hat{\beta}] = \mathbb{E}[(X^T X)^{-1} X^T y] = (X^T X)^{-1} X^T \cdot \mathbb{E}[X\beta + \epsilon] = (X^T X)^{-1} X^T X\beta = \beta.$$

Zauważmy teraz, że zachodzi

$$\hat{\beta} - \beta = (X^T X)^{-1} X^T y - \beta = (X^T X)^{-1}(X^T y - X^T X\beta) = (X^T X)^{-1} X^T (y - \mathbb{E}[y]).$$

Transponując obie strony otrzymujemy

$$(\hat{\beta} - \beta)^T = (y - \mathbb{E}[y])^T X (X^T X)^{-1}.$$

Korzystając z powyższych, dostajemy

$$\mathbb{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] = (X^T X)^{-1} X^T \operatorname{Var}[y] X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}.$$

Na koniec zauważmy, że $\hat{\beta}$ możemy zapisać jako

$$\hat{\beta} = \beta + (X^T X)^{-1} X^T \epsilon.$$

Ponieważ $\epsilon \sim \mathrm{Norm}(0, \sigma^2 Id)$, to $(X^T X)^{-1} X^T \epsilon \sim \mathrm{Norm}(0, \sigma^2 (X^T X)^{-1})$, ze wzoru na liniową transformację rozkładu normalnego (można też powiedzieć ze jest to pewien rozkład normalny i skorzystać z faktu iż z poprzedniego podpunktu znamy wariancję). Ostatecznie $\hat{\beta} \sim \mathrm{Norm}(\beta, \sigma^2 (X^T X)^{-1})$.

Zadanie 4:
Pokaż, że predykcja $\hat{Y}$ = X $\hat{\beta}$ jest rzutem prostopadłym Y na przestrzeń liniową rozpiętą przez kolumny macierzy X.

Rozwiązanie:
X - macierz zmiennych objaśniających
$\hat{Y}$ - macierz predykcji
$\hat{\beta}$ - macierz parametrów
H - macierz daszkowa

Z wykładu wiem, że $\hat{\beta}$ jest dobrana tak, aby zminimalizować błąd RSS (resztowa suma kwadratów) i ma postać:

$$\hat{\beta} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

Zatem predykcja $\hat{Y}$ może być przedstawiona w postaci:

$$\hat{Y} = X \cdot \hat{\beta} = X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

Macierz $X \cdot (X^T \cdot X)^{-1} \cdot X^T$ to macierz daszkowa H.
Pokażę, że macierz H jest macierzą idempotentną, czyli spełnia warunek $H^2 = H$:

$$H^2 = H \cdot H = [X \cdot (X^T \cdot X)^{-1} \cdot X^T] \cdot [X \cdot (X^T \cdot X)^{-1} \cdot X^T] = X \cdot [(X^T \cdot X)^{-1} \cdot (X^T \cdot X)] \cdot (X^T \cdot X)^{-1} \cdot X^T =$$

$$= X \cdot (X^T \cdot X)^{-1} \cdot X^T$$

Zatem macierz H jest idempotentną macierzą rzutu w przestrzeni liniowej rozpiętej przez kolumny macierzy X.

Pokażę, że macierz H jest symetryczna, czyli spełnia warunek $H^T = H$:

$$H^T = ([X \cdot (X^T \cdot X)^{-1}] \cdot X^T)^T = ((X)^T)^T \cdot [X \cdot (X^T \cdot X)^{-1}]^T = X \cdot ((X^T \cdot X)^{-1})^T \cdot X^T =$$
$$= X \cdot ((X^T \cdot X)^T)^{-1} \cdot X^T = X \cdot ((X^T \cdot (X^T)^T)^{-1} \cdot X^T = X \cdot (X^T \cdot X)^{-1} \cdot X^T$$

więc H jest symetryczna.

Fakt: P jest operatorem rzutu ortogonalnego wtedy i tylko wtedy, gdy idempotentny i symetryczny.

Zatem macierz H jest macierzą rzutu ortogonalnego w przestrzeni liniowej rozpiętej przez kolumny macierzy X, więc predykcja $\hat{Y} = H \cdot Y$ jest rzutem ortogonalnym (prostopadłym) Y na przestrzeń liniową rozpiętą przez kolumny macierzy X.

Rozwiązanie: Z wykładu wiemy że $\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^Ty$.

Aby pokazć, że predykcja jest rzutem prosotpadłym wystarczy pokzać, że macierz

$$H := X(X^TX)^{-1}X^T$$

jest macierzą rzutu orotgonalnego. Aby to pokzać wystarczy udowodnić że macierz H jest macierzą idempotentną czyli $H^2 = H$ oraz, że jest macierzą symetryczną tzn. $H^T = H$.

$HX = X$, ponieważ $X(X^TX)^{-1}X^T \cdot X = X(X^TX)^{-1}(X^TX) = XI = X$

Sprawdźmy czy H jest macierzą idempotentną.

$$H^2 = X(X^TX)^{-1}X^T \cdot X(X^TX)^{-1}X^T =$$
$$X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T.$$

Zaważmy, że $(X^TX)^{-1}(X^TX) = I$, gdzie I jest macierzą identyczności.
Zatem ostatecznie dostajemy

$$X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T =$$
$$XI(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H.$$

Pozotaje pokazać że H jest symetryczna.

$$H^T = (X(X^TX)^{-1}X^T)^T$$

Mając na uwadze, że dla dowolnych macierzy ABC mamy $(ABC)^T = C^TB^TA^T$, $(A^T)^T = A$ oraz $(A^{-1})^T = (A^T)^{-1}$dostajemy

$$(X(X^TX)^{-1}X^T)^T = X^T((XX^T)^{-1})^TX =$$
$$= X^T(X^TX)^{-1}X = H.$$

Pokazaliśmy, że H jest macierzą rzutu ortogonalnego, na przestrzeń liniową rozpietą przez kolumny macierzy X.

Rozwiązanie: Przez H ozaczmy macierz rzutu ortogonalnego na p-liniową podprzestrzeń $\mathbb{R}(X)$

generowaaną przez kolumny macierzy X (obraz przekształcenia liniowego X). Wówczas, $HX = X$ oraz:

$$\hat{Y} = X(X^TX)^{-1}X^TY = HY$$

Istotnie, $\hat{Y}_i = x_i^T\hat{\beta}$ jest współrzędną Y-ową punktu odpowidającego wektorowi x_i i leżącemu na wykresie funkcji regresji. Zatem $\hat{Y} = X\hat{\beta} = (\hat{Y}_1, \ldots, \hat{Y}_n)^T$.
Sprawdzimy, czy przedstawiona macierz H jest symetryczna ($H^T = H$) i idenpotentna($H^2 = H$).
Czy $H^T = H$?

$$\begin{aligned} H^T &= (X(X^TX)^{-1}X^T)^T \\ &= X((X^TX)^{-1})^TX^T \\ &= X((X^TX)^T)^{-1}X^T \\ &= X(X^TX)^{-1}X^T \\ &= H \end{aligned}$$

Czy $H^2 = H$?

$$\begin{aligned} H^2 &= (X(X^TX)^{-1}X^T)(X(X^TX)^{-1}X^T) \\ &= X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T \\ &= X((X^TX)^T)^{-1}X^T \\ &= H \end{aligned}$$

Zadanie 5:
Niech $n > p$ oraz $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. Rozpatrzmy rozkład QR macierzy X, czyli

$$X = QR,$$

gdzie $Q \in \mathbb{R}^{n\times p}$ ortogonalna, a $R \in \mathbb{R}^{p\times p}$ górnotrójkątna. Korzystając z rozkładu QR, pokaż, że

$$\|Y - X\hat{\beta}\|^2 = \|\tilde{Q}^TY\|^2,$$

gdzie $\tilde{Q}$ to dopełnienie ortogonalne Q do bazy ortogonalnej w $\mathbb{R}^n$. Stąd wynika, że $\|Y - X\hat{\beta}\|^2/\sigma^2$ ma rozkład $\chi^2(n-p)$ oraz jest niezależny od $\hat{\beta}$. Podaj nieobciążony estymator σ^2.

Rozwiązanie: Macierz Q jest ortogonalna, więc $Q^TQ = I_p$ (ale nie $QQ^T = I_n$, bo Q nie jest kwadratowa). Weźmy dopełnienie ortogonalne $\tilde{Q}$ macierzy Q i niech $[Q\tilde{Q}]$ oznacza macierz $n \times n$ taką, że na pierwszych p kolumnach są kolumny z Q, a na pozostałych $n-p$ kolumnach są kolumny z $\tilde{Q}$. Wtedy $\tilde{Q}^T\tilde{Q} = I_{n-p}$ oraz macierz $[Q\tilde{Q}]$ jest ortogonalna (z definicji dopełnienia), tzn. $I_n = [Q\tilde{Q}][Q\tilde{Q}]^T = QQ^T + \tilde{Q}\tilde{Q}^T$. Ponadto macierz R jest odwracalna, bo na slajdach zakładaliśmy, że X jest maksymalnego rzędu. Mamy więc

$$\begin{aligned} &\|Y - X\hat{\beta}\|^2 = \|Y - X(X^TX)^{-1}X^TY\|^2 = \|(I_n - QR(R^TQ^TQR)^{-1}R^TQ^T)Y\|^2 = \\ &= \|(I_n - QR(R^TR)^{-1}R^TQ^T)Y\|^2 = \|(I_n - QRR^{-1}R^{-T}R^TQ^T)Y\|^2 = \|(I_n - QQ^T)Y\|^2 = \\ &= \|\tilde{Q}\tilde{Q}^TY\|^2 = Y^T\tilde{Q}\tilde{Q}^T\tilde{Q}\tilde{Q}^TY = Y^T\tilde{Q}\tilde{Q}^TY = \|\tilde{Q}^TY\|^2. \end{aligned}$$

Nieobciążonym estymatorem σ^2 jest $\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{n-p}$, bo

$$\mathbb{E}\,\hat{\sigma}^2 = \frac{\sigma^2}{n-p}\,\mathbb{E}\,\frac{\|Y - X\hat{\beta}\|^2}{\sigma^2} = \frac{\sigma^2}{n-p}(n-p) = \sigma^2.$$

Skorzystaliśmy tu ze wskazówki i z tego, że dla zmiennej o rozkładzie $\chi^2(\nu)$ jej wartość oczekiwana wynosi ν.