Тех.Завдання

Власники футболу постійно конкурують за хороших гравців. Чим більше перемог, тим більша ймовірність того, що команда забезпечить хороші прибутки від бізнесу для власників. Ресурси, які кожна з 32 команд має в Національній футбольній лізі (НФЛ), відрізняються, але система драфтів розроблена таким чином, щоб нейтралізувати переваги, які можуть мати багатші команди.

**Об’єм вибірки:**32

**Назви змінних:**

* Team: назва футбольної команди;
* Wins: кількість виграних матчів;
* Avg\_Age: середній вік гравців;
* Active: рівень активності;
* Salary: зарплатня гравців.

Необхідно:

* Побудувати матричну діаграму розсіювання даних і ящики з вусами для змінних Wins, Avg\_Age,Active, Salary. Чи є потенційні регресори мультиколінеарними?
* Побудувати лінійну модель регресії вигляду

Виписати модель з оціненими коефіцієнтами і коефіцієнт детермінації. Що більше схоже на правду: що зарплатня командам призначається лінійно в залежності від активності чи що залежність між Active і Salary стохастична, випадкова?

* За допомогою підгонки в пакеті R (команда nls()) побудувати модель вигляду

Зробити повний аналіз моделі, виписати модель з оціненими коефіцієнтами. Побудувати графік predicted versus residuals

* Двома методами, за допомогою методу лінеаризації моделі і за допомогою підгонки в пакеті R (команда nls()) побудувати модель вигляду

Зробити повний аналіз моделі, виписати моделі з оціненими коефіцієнтами, показати, куди в формулу моделі входить похибка.

* Вказати найкращу з усіх моделей.

**Дані**

Team Wins Avg\_Age Active Salary

Arizona\_Cardinals 7 27.4 154384100.00 154.3841

Atlanta\_Falcons 11 26.45 157666798.00 157.666798

Baltimore\_Ravens 7 26.6 134670306.00 134.670306

Buffalo\_Bills 7 26.7 139720093.00 139.720093

Carolina\_Panthers 6 26.75 154994278.00 154.994278

Chicago\_Bears 3 26.28 144061149.00 144.061149

Cincinnati\_Bengals 6 25.52 139855574.00 139.855574

Cleveland\_Browns 1 24.25 113317279.00 113.317279

Dallas\_Cowboys 13 26.04 137061990.00 137.06199

Denver\_Broncos 9 26.11 149908645.00 149.908645

Detroit\_Lions 9 25.77 141941975.00 141.941975

Green\_Bay\_Packers 10 25.85 150044453.00 150.044453

Houston\_Texans 9 25.6 133467065.00 133.467065

Indianapolis\_Colts 8 25.62 135822350.00 135.82235

Jacksonville\_Jaguars 3 25.64 153411580.00 153.41158

Kansas\_City\_Chiefs 12 25.87 141076528.00 141.076528

Los\_Angeles\_Rams 4 25.23 146914285.00 146.914285

Miami\_Dolphins 10 26.62 130477781.00 130.477781

Minnesota\_Vikings 8 26.06 136268561.00 136.268561

New\_England\_Patriots 14 26.42 142473404.00 142.473404

New\_Orleans\_Saints 7 26.55 127273255.00 127.273255

New\_York\_Giants 11 25.68 155321043.00 155.321043

New\_York\_Jets 5 25.58 121122908.00 121.122908

Oakland\_Raiders 7 26.09 158310806.00 158.310806

Philadelphia\_Eagles 7 26.36 145518303.00 145.518303

Pittsburgh\_Steelers 11 26.06 152459177.00 152.459177

San\_Diego\_Chargers 5 25.66 138405987.00 138.405987

San\_Francisco\_49ers 2 25.77 109640818.00 109.640818

Seattle\_Seahawks 10 25.75 153648618.00 153.648618

Tampa\_Bay\_Buccaneers 9 26.23 141511896.00 141.511896

Tennessee\_Titans 9 26.53 150611974.00 150.611974

Washington\_Redskins 8 25.94 155812636.00 155.812636

**1)Будуємо матричну діаграму через пари векторів і одразу бачимо мультиколінеарний регресор (Active & Salary). Повна лінійна залежність =1 і такий самий графік.**

**ggpairs(data[, c("Wins", "Avg\_Age", "Active", "Salary")])**



**Побудова ящиків з вусами**









**2) Побудова моделі лінійної регресії**



**Маємо результат:**

**Call:**

**lm(formula = Salary ~ Active, data = data)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-5.153e-14 -4.440e-15 2.100e-15 7.636e-15 1.419e-14**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 0.000e+00 2.544e-14 0.000e+00 1**

**Active 1.000e-06 1.784e-22 5.606e+15 <2e-16 \*\*\***

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 1.226e-14 on 30 degrees of freedom**

**Multiple R-squared: 1, Adjusted R-squared: 1**

**F-statistic: 3.143e+31 on 1 and 30 DF, p-value: < 2.2e-16**

**Звідси беремо коефіцієнти a, b, R^2**

**a=0 (перехоплення)**

**b=1 (коефіцієнт при змінній Active)**

**Коефіцієнт детермінації R2=1**

**Коефіцієнт детермінації R^2 рівний 1 , що вказує на дуже високу частку поясненої варіації зарплати (Salary) за рахунок рівня активності (Active). Це свідчить про те, що модель ідеально підходить для прогнозування зарплати на основі активності.Враховуючи дуже високий коефіцієнт детермінації та статистично значущі коефіцієнти, можна зробити висновок, що зарплата команд в основному призначається лінійно в залежності від активності гравців. Таким чином, залежність між Active і Salary є лінійною ніж стохастичною або випадковою.**

**3)Підгонка за допомогою NLS**

**4)Модель за допомогою лінеаризації та підгонки**

**Лінеаризуємо за допомогою логарифмів**



**Маємо**

**Call:**

**lm(formula = log(Active) ~ log(Wins) + log(Avg\_Age), data = data)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-0.15916 -0.05100 0.01047 0.06228 0.15062**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 16.49566 2.41367 6.834 1.66e-07 \*\*\***

**log(Wins) 0.07261 0.02897 2.506 0.0181 \***

**log(Avg\_Age) 0.65428 0.74901 0.874 0.3896**

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

**Residual standard error: 0.07896 on 29 degrees of freedom**

**Multiple R-squared: 0.2976, Adjusted R-squared: 0.2491**

**F-statistic: 6.143 on 2 and 29 DF, p-value: 0.005964**

**Звідси коеф. a,b,R^2 - 16.49566, 0.07261, 0.2976**

**Підгонка за допомогою NLS**

**Розрахунок похибки**

**5)Обрання кращої моделі**

