LECTURE 18: MULTIPLE LINEAR REGRESSION I

NOVEMBER 4, 2025

TODAY

- Estimating the coefficients in Multiple Linear Regression (MLR)
  - It is the same OLS engine we continue to use
- Interpretation, of the regression equation
- Fitted Values and Multiple R^2

MODEL SETUP

- Response: y; predictors $x^{(1)},\ldots,x^{(p)}$
- Model: $y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \cdots + \beta_p x^{(p)} + \varepsilon$
- Goal: choose coefficients β to make predictions $\hat{y}$ close to observed y

RECAP: LEAST SQUARES ESTIMATION

- Pick β to minimize total loss $L = \Sigma_i\ \ell(y_i, \hat{y}_i)$
- Standard choice: squared-error loss $\ell = (y_i - \hat{y}_i)^2$
- Gives Ordinary Least Squares (OLS) estimates

LOSS FUNCTION: SUM OF SQUARED ERRORS (SSE)

- $SSE(\beta) = \Sigma_i (y_i - \hat{y}_i)^2$, where $\hat{y}_i = \beta_0 + \Sigma_j \beta_j x_i(j)$
- OLS picks β that minimize SSE
- Computed by lm(...) in R

FITTING WITH R

```
lm(formula = Salary ~ YearsExperience + TeamSize + EducationLevel + JobSatisfaction + Age, data = df)

Coefficients:
   (Intercept)  YearsExperience     TeamSize   EducationLevel   JobSatisfaction          Age
       44.9255           3.7244       0.3341           5.9496           -0.4402      -0.1063
```

- *Salary = 44.9255 + 3.7244*YearsExperience + 0.3341*TeamSize + 5.9496*EducationLevel – 0.4402*JobSatisfaction - 0.1063*Age*

- Use coefficients to compute predictions ŷ for any x

INTERPRETING Β $_j$

- Model: $y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \cdots + \beta_p x^{(p)} + \varepsilon$
- β_j is the **average change in y** per unit increase in $x^{(j)}$
  - with all **other predictors** held constant !
- β_0 is the **average y** when ?

INTERPRETING Β $_j$

- Model: $y = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \cdots + \beta_p x^{(p)} + \varepsilon$
- β_j is the average change in y per unit increase in $x^{(j)}$
  - with all other predictors held constant !
- β_0 is the average y when **all predictors equal zero**

EXAMPLE: YEARSEXPERIENCE COEFFICIENT

- *Salary = 44.9255 + 3.7244*YearsExperience + 0.3341*TeamSize + 5.9496*EducationLevel – 0.4402*JobSatisfaction - 0.1063*Age*

- $\hat{\beta}$_YearsExperience ≈ 3.72 ⇒ +1 year of experience ≈ + 3.72 salary units
- Interpretation holds other variables fixed: TeamSize, EducationLevel , JobSatisfaction, Age

SCALING: THE PREDICTOR VARIABLES

- Scale each variable
  - Difference from mean / standard deviation
    - $x^{(j)\prime} = \dfrac{x^{(j)} - \overline{x^{(j)}}}{s_{x^{(j)}}}$
  - aka Z-score standardization
- Centered on 0 with SD of 1

WHY SCALE THE VARIABLES (IN MLR)

- **To make the coefficients comparable**
  - After scaling, each predictor is measured in “1 standard deviation units,”
  - So we can compare which predictor has the stronger association with the response (instead of mixing years, team size, and satisfaction scores in different units)

- **To improve numerical stability**
  - When predictors are on wildly different scales (like 0–10 vs 0–100,000), the linear algebra under the hood can get unstable

- **To make the intercept easy to interpret**
  - With centered predictors, the intercept is roughly the average Salary (when each predictor is at its mean), which is often more meaningful than “Salary when YearsExperience = 0, TeamSize = 0, Age = 0,” etc.

SCALED COEFFICIENTS

```
> ftScale <- lm(Salary ~ YearsExperience + TeamSize + EducationLevel + JobSatisfaction + Age, data = tempDF)
> cat("Coefficients with variables scaled:\n")
Coefficients with variables scaled:
> print(coef(ftScale))
  (Intercept)   YearsExperience   TeamSize   EducationLevel   JobSatisfaction   Age
   92.5975000     22.1816102     1.4580938     4.7170071      -1.2303120      -0.6774317
```

- **YearsExperience (+22.18):** Holding other predictors fixed, a **+1 SD** increase in YearsExperience is associated with about **+22.18** higher Salary
- **EducationLevel (+4.72):** Holding others fixed, a **+1 SD** increase in EducationLevel corresponds to about **+4.72** higher Salary
- **JobSatisfaction (–1.23):** Holding others fixed, a **+1 SD** increase in JobSatisfaction is associated with about **–1.23** lower Salary
- **Intercept (92.60):** With all standardized predictors at 0 (i.e., each at its mean), expected Salary ≈ **92.60** (in our Salary units)
- (This aspect is not quite covered in the text)

CORRELATIONS AMONGST VARIABLES

```
cor(df[c("YearsExperience","TeamSize","EducationLevel","JobSatisfaction","Age")])

pairs(df[, c("Salary", "YearsExperience", “Age")])
```

	YearsExperience	**TeamSize**	**EducationLevel**	**JobSatisfaction**	**Age**
YearsExperience	1.00000000	0.88935042	-0.02998766	-0.29641256	0.90601988
TeamSize	0.88935042	1.00000000	-0.05440416	-0.25672396	0.79330777
EducationLevel	-0.02998766	-0.05440416	1.00000000	0.08677718	-0.06339764
JobSatisfaction	-0.29641256	-0.25672396	0.08677718	1.00000000	-0.27328225
Age	0.90601988	0.79330777	-0.06339764	-0.27328225	1.00000000

RECAP

lm(formula = Salary ~ YearsExperience + TeamSize + EducationLevel + JobSatisfaction + Age, data = df)

Residuals:

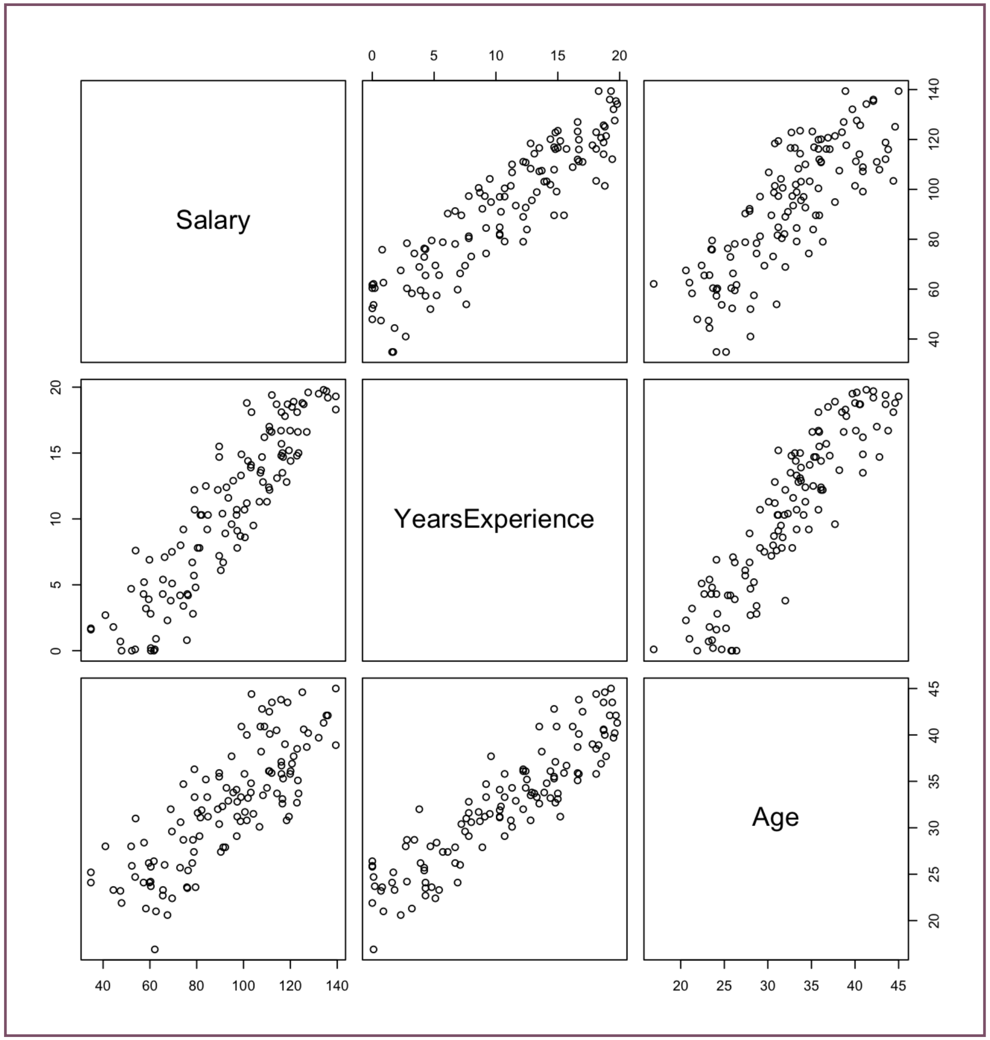
Min	1Q	Median	3Q	Max
-30.0833	-6.0215	0.5906	7.0405	21.8689

Coefficients:

	Estimate	Std. Error	t value	Pr(>ltl)
(Intercept)	44.9255	8.6548	5.191	9.23e-07 ***
YearsExperience	3.7244	0.4894	7.611	8.54e-12 ***
TeamSize	0.3341	0.4607	0.725	0.470
EducationLevel	5.9496	1.1656	5.104	1.34e-06 ***
JobSatisfaction	-0.4402	0.3441	-1.279	0.203
Age	-0.1063	0.3415	-0.311	0.756

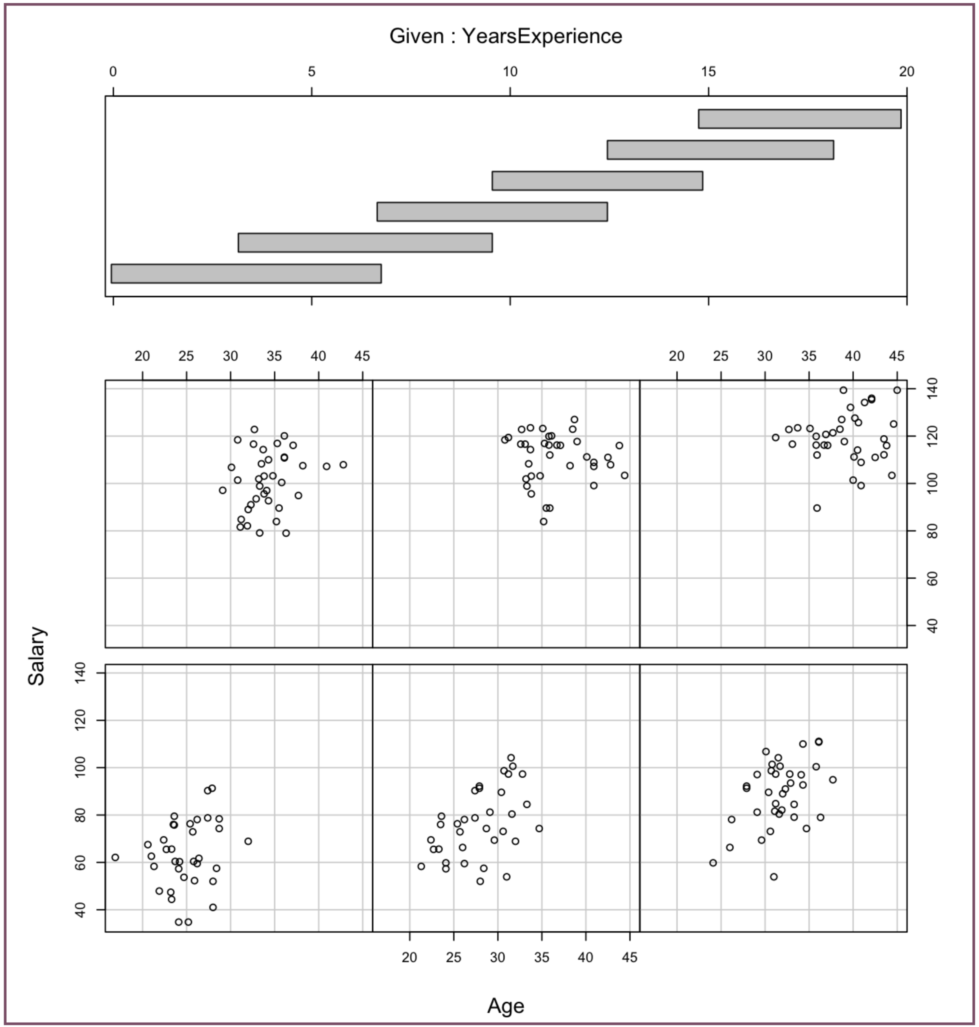
CORRELATIONS

- pairs(df[, c("Salary", "YearsExperience", "Age")])



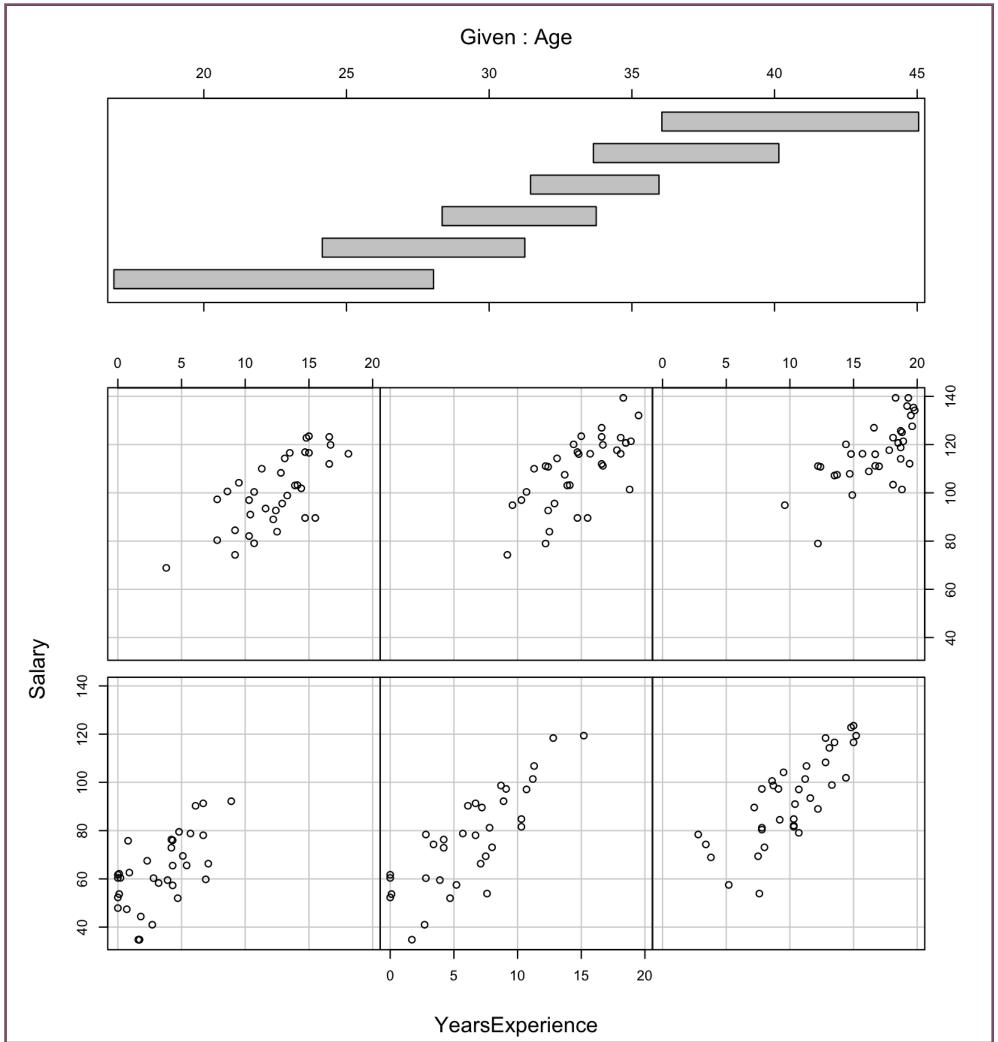
COPLOT: KEEP THE **OTHER** FIXED

- coplot(Salary ~ **Age I YearsExperience**, data=df)



COPLOT: KEEP THE **OTHER** FIXED

- coplot(Salary ~ **YearsExperience I Age**, data=df)



WHAT IS **AGE** WELL CORRELATED WITH ?

	YearsExperience	TeamSize	EducationLevel	JobSatisfaction	Age
YearsExperience	1.00000000	0.88935042	-0.02998766	-0.29641256	$\boxed{0.90601988}$
TeamSize	0.88935042	1.00000000	-0.05440416	-0.25672396	$\boxed{0.79330777}$
EducationLevel	-0.02998766	-0.05440416	1.00000000	0.08677718	-0.06339764
JobSatisfaction	-0.29641256	-0.25672396	0.08677718	1.00000000	-0.27328225
Age	0.90601988	0.79330777	-0.06339764	-0.27328225	1.00000000

MODEL **WITHOUT** THE CORRELATED VARIABLES (OF YEARS EXPERIENCE, TEAMSIZE)

ft2 <- lm(**Salary ~ EducationLevel + JobSatisfaction + Age**, data=df)

summary(ft2)

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -18.7733     8.6169  -2.179   0.0314 *  
EducationLevel    6.8262     1.6368   4.170 5.89e-05 ***
JobSatisfaction  -0.9001     0.4818  -1.868   0.0642 .  
Age               3.1812     0.2109  15.081  < 2e-16 ***
```

- Look at the significance (p-value) now !

COLLINEARITY AND SIGNIFICANCE

- When predictors move together, it’s hard to isolate individual effects
- High correlation among x’s inflates coefficient uncertainty
- Significance can disappear despite strong joint predictive power

CORRELATION ≠ CAUSATION

CORRELATION ≠ CAUSATION

- β_j reflects association after controlling for other x's in data
- Does not by itself establish causal effect
- Randomized interventions would be needed for causality claims

FITTED VALUES
MULTIPLE R^2
ADJUSTED R^2

FITTED VALUES ($\hat{Y}_i$)

- For each observation x_i, $\hat{y}_i = \hat{\beta}_0 + \Sigma_j \hat{\beta}_j x_{i(j)}$
- R stores fitted values in fitted(ft)
- Compare $\hat{y}_i$ vs y_i to assess fit

MULTIPLE R^2

- $R^2 = (\text{cor}(y, \hat{y}))^2$
- Measures fraction of variance in y explained by the model
- cor(df$Salary, fitted(ft1))^2
  - 0.8535076
- How about when variables are scaled ?
- How about for the model when we took YearsExperience and TeamSize out ?

AJUSTED R^2

```
Residual standard error: 10.03 on 113 degrees of freedom
Multiple R-squared:  0.8535,    Adjusted R-squared:  0.8457
F-statistic: 109.7 on 6 and 113 DF,  p-value: < 2.2e-16
```

- **Adjusted** R^2 is like R^2 with a fairness penalty for extra predictors
  - So it only goes up if a new variable genuinely improves the model
- Compared to R^2 : It can stay the same or go down
- Use it to compare models on the same dataset and response
  - Higher adjusted R^2 means better fit after accounting for model size

INTERPRETING R^2

- Higher R^2 suggests better in-sample fit
- But high R^2 does not guarantee causal interpretation, or out-of-sample performance !
- Use plots and validation in practice