LECTURE 19: MULTIPLE LINEAR REGRESSION II

MEASURING FIT

NOVEMBER 6, 2025

TODAY

- Residuals and Residual Sum of Squares (RSS)
- Behaviour of RSS (and R^2) when variables are added/removed
- Residual Degrees of Freedom & Residual Standard Error (RSE)
- The F-Statistic, ANOVA (analysis of variance)

RESIDUALS

- Residual for obs i: $r_i = y_i - \hat{y}_i$
- Vector form: $r = y - \hat{y}$
  - Stored in lm objects as residuals(model)
- Interpretation: (signed) **prediction error** at each data point

RESIDUALS (IN R)

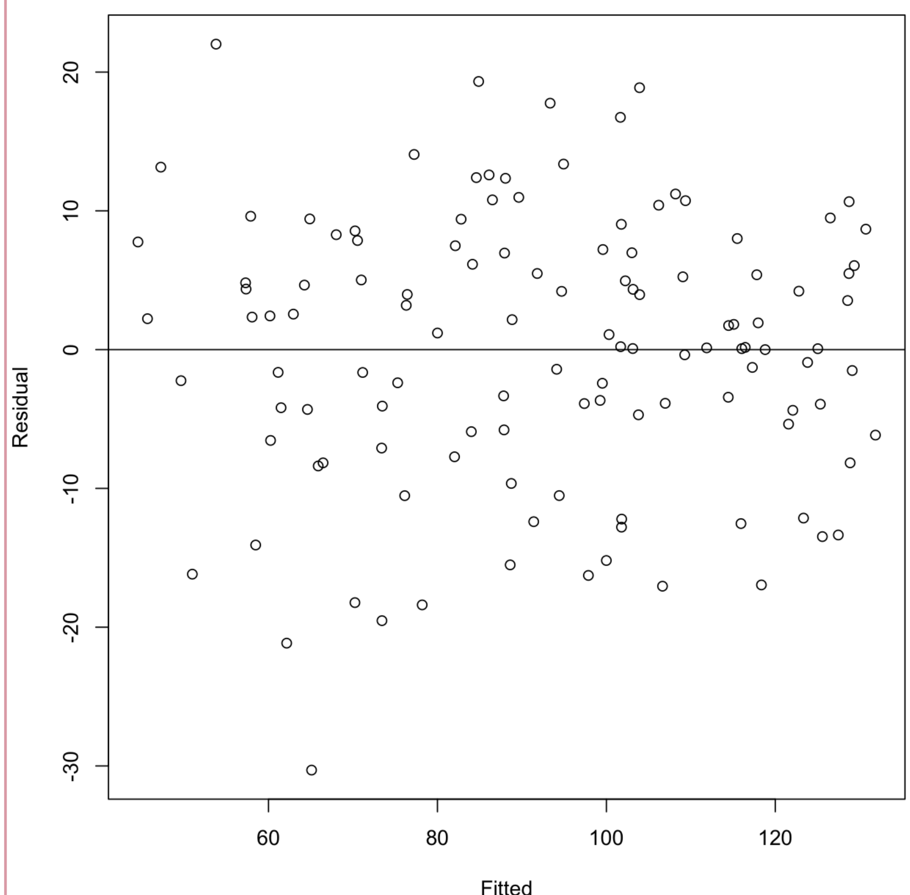
```
ft1 <- lm(Salary ~ ., data = df)
head(residuals(ft1))
plot(fitted(ft), residuals(ft), xlab='Fitted',ylab='Residual')
abline(h=0)
```

- lm returns residuals by default: residuals(ft) or ft$residuals
- First look: head(residuals(ft))
- Pair residuals with fitted values for diagnostics

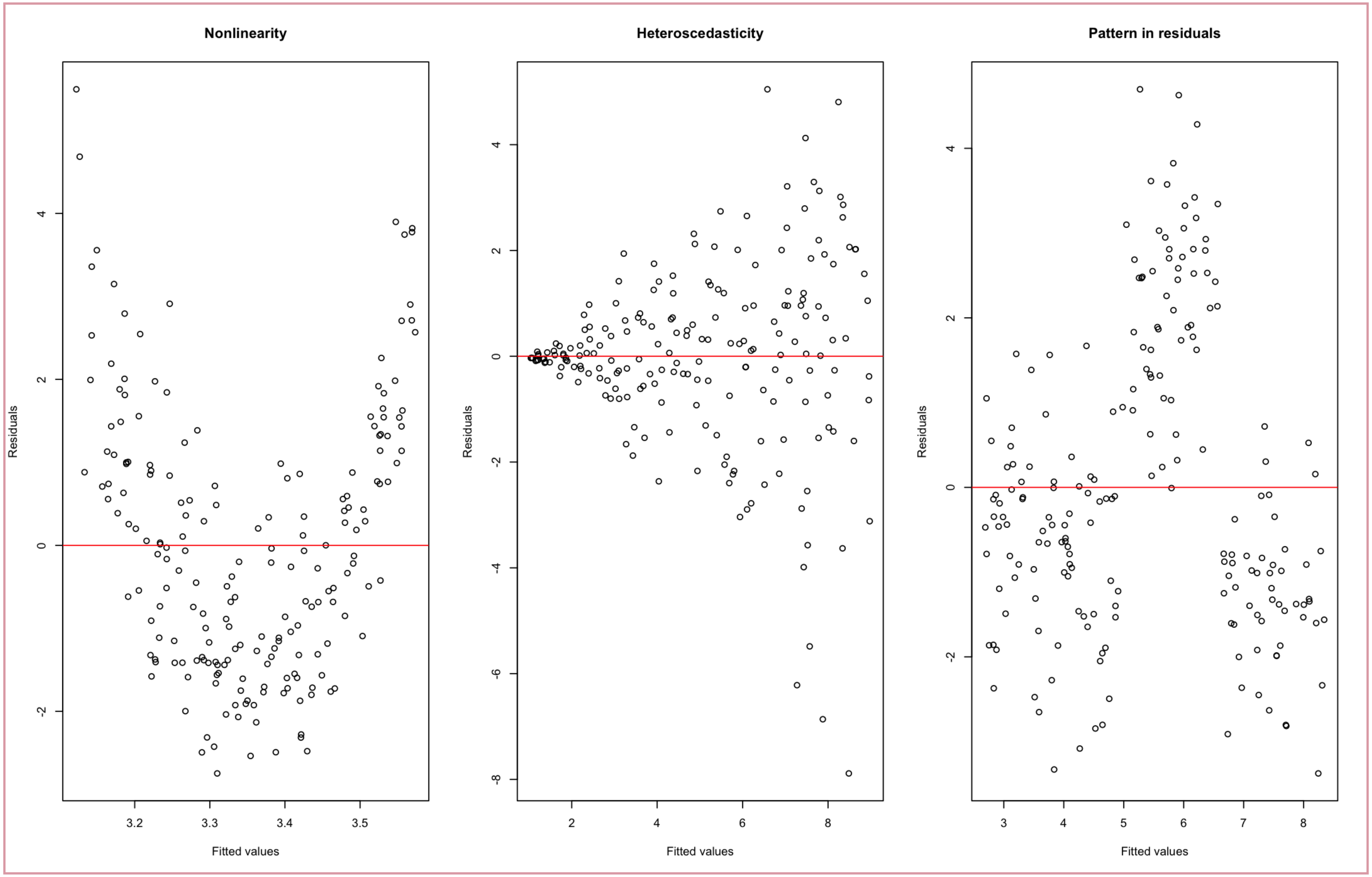
RESIDUAL PLOT (AGAINST FITTED Ŷ)

- Check for:
  - Nonlinearity
  - Heteroscedasticity
  - Patterns
- A good linear model shows residuals scattered randomly around 0
- Systematic patterns ⇒ model

  misspecification or transforms needed



RESIDUAL PLOT



RESIDUAL SUM OF SQUARES (RSS)

- The Residual Sum of Squares: RSS
- $RSS = \Sigma_i \, r_i^2 = \Sigma_i \, (y_i - \hat{y}_i)^2$
- Least squares (OLS) chooses coefficients to **minimize RSS**
- Lower RSS ⇒ better in-sample fit (for given data & predictors)

COMPUTING RSS IN R

```
> rss.ft1 = sum((ft1$residuals)^2)
> rss.ft1
[1] 11359.07
```

```
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
```

- Units of RSS: those of the response/target variable
- Residual standard error ?
  - **Residual Standard Error (RSE)**: $\mathrm{RSE} = \sqrt{\frac{\mathrm{RSS}}{\mathrm{df}}}$
- Compare RSS across models fitted to the same y
  - Model 1: Salary ~ YearsExperience + Age
  - Model 2: Salary ~ YearsExperience + Age + TeamSize + EducationLevel

RELATIONSHIP: R^2 AND RSS

```
> rss <- sum(residuals(ft1)^2)
> tss <- sum( (df$Salary - mean(df$Salary))^2 )
> R2 <- 1 - rss / tss
> R2
[1] 0.8535076
> cor(df$Salary, fitted(ft1))^2
[1] 0.8535076
```

- $R^2 = 1 - RSS/TSS$, where $TSS = \Sigma_i (y_i - \bar{y})^2$
  - Equivalently: $R^2 = cor(y, \hat{y})^2$
- Note that
  - TSS depends **only on y**
  - RSS depends on **model fit**

RSS AND VARIABLES

```
ft1 <- lm(Salary ~ ., data=df)

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



> rss.ft1
[1] 11359.07

> rss.ft2
[1] 23032.33
```

RSS AND VARIABLES

```
ft1 <- lm(Salary ~ ., data=df)

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

ft3 <- lm(Salary ~ EducationLevel + JobSatisfaction + YearsExperience,
data=df)



> rss.ft1
[1] 11359.07

> rss.ft2
[1] 23032.33


> rss.ft3
??
```

RSS AND VARIABLES

```
ft1 <- lm(Salary ~ EducationLevel + JobSatisfaction + Age, data=df)

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

ft3 <- lm(Salary ~ EducationLevel + JobSatisfaction + YearsExperience,
data=df)



> rss.ft1
[1] 11359.07

> rss.ft2
[1] 23032.33


> rss.ft3
[1] 11427.18
```

INTERPRETING R^2 VIA RESIDUALS

- Remember: $R^2 = 1 - RSS/TSS$
- Small RSS relative to TSS $\Rightarrow$ R^2 close to 1
- Large RSS (big residuals) $\Rightarrow$ R^2 closer to 0
- Use R^2 cautiously: it is in-sample and non-causal

HOW **RSS** BEHAVES WHEN VARIABLES CHANGE

- Adding a predictor can only decrease (or leave unchanged) RSS
- Removing a predictor can only increase (or leave unchanged) RSS
- Reason: OLS re-optimizes over a larger/smaller parameter space

HOW **R^2** BEHAVES WHEN VARIABLES CHANGE

- Adding predictors can only increase (or leave unchanged) R^2
- Removing predictors can only decrease (or leave unchanged) R^2
- Adjusted R^2 may decrease when adding weak predictors (penalizes p)

REMOVING VARIABLES

```
> summary(ft1)$r.squared
[1] 0.8535076

> summary(ft2)$r.squared
[1] 0.7029631

> summary(ft3)$r.squared
[1] ??
```

REMOVING VARIABLES

```
> summary(ft1)$r.squared
[1] 0.8535076

> summary(ft2)$r.squared
[1] 0.7029631

> summary(ft3)$r.squared
[1] 0.8526291
```

CAUTIONS IN ADDING PREDICTORS

- Don't just keep adding variables because R^2 always goes up; that can be misleading
  - R^2 **will** go up, when you add more variables
- Make sure the variables you include make sense in the real-world context and are not almost copies of each other
- Check how well the model predicts on new data (using a test set or cross-validation), not just on the data you used to fit it.

MODEL AND PARAMETERS (N AND P)

- We fit a multiple linear regression model:
  - $\text{Salary} = \beta_0 + \beta_1 \cdot \text{YearsExperience} + \beta_2 \cdot \text{TeamSize} + \beta_3 \cdot \text{EducationLevel} + \beta_4 \cdot \text{JobSatisfaction} + \beta_5 \cdot \text{Age} + \varepsilon$
- n = number of observations (rows in df).
- p = number of predictors (here p = 5; β_0 is the intercept, not counted in p).

RESIDUAL DEGREES OF FREEDOM

- We estimate p + 1 parameters: β_0 and the p slopes.
- Each estimated parameter uses up one degree of freedom
- Residual degrees of freedom: df = n – p – 1.
- This is the number of 'free' residuals left after fitting the model

RSS AND RESIDUAL STANDARD ERROR (RSE)

- Residual for observation i: $r_i = y_i - \hat{y}_i$.
- Residual Sum of Squares (RSS): $RSS = \Sigma_i \, r_i^2 = \Sigma_i \, (y_i - \hat{y}_i)^2$.
- Residual Standard Error (RSE) is the *typical size of a residual*

ADJUSTED R^2

- $$R^2_{\text{adj}} = 1 - \frac{\frac{\text{RSS}}{n-p-1}}{\frac{\text{TSS}}{n-1}}$$
- R^2 : fraction of total variation (in y) explained by the model
- **Adjusted** R^2 : corrects R^2 for the number of predictors used
- It compares average unexplained variance to average total variance

ADJUSTED R^2 : INTERPRETATION

- If you add a predictor that does not help much, RSS does not drop enough
- So, adjusted R^2 can stay the same or go down (penalty for extra parameters).
- For our single Salary model, adjusted R^2 is a 'fair' version of R^2 that accounts for model size.

F-STATISTIC

- We test H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$
  - no linear relationship with y (Salary)
- TSS = $\Sigma_i (y_i - \bar{y})^2$ and RSS = $\Sigma_i (y_i - \hat{y}_i)^2$.
- Regression Sum of Squares (RegSS) = TSS – RSS
- F compares 'variance explained per predictor' to 'variance left per residual degree of freedom'
- $$F = \frac{\frac{\mathrm{TSS} - \mathrm{RSS}}{p}}{\frac{\mathrm{RSS}}{n - p - 1}}$$

F-STATISTIC – INTERPRETATION

- If H_0 is true, explained variance and unexplained variance are similar, so $F \approx 1$
- A large F means the model explains much more variation than expected from random noise.
- In summary(ft), the F-statistic tests whether the whole set of predictors jointly has a meaningful linear relationship with Salary

ANOVA

```
> anova(ft1)
Analysis of Variance Table

Response: Salary
                 Df Sum Sq Mean Sq  F value    Pr(>F)    
YearsExperience   1  63433   63433 631.0284 < 2.2e-16 ***
EducationLevel    1   2521    2521  25.0834 2.042e-06 ***
Age               1     13      13   0.1263    0.7229    
HoursWeek         1      4       4   0.0373    0.8471    
TeamSize          1     49      49   0.4855    0.4874    
JobSatisfaction   1    162     162   1.6101    0.2071    
Residuals       113  11359     101                       
```

- ANOVA table: splits the total variation in Salary into parts explained by each predictor and the leftover residual variation.
- Sum Sq for a variable: how much extra variation in Salary that variable explains (given the variables listed above it)
- Mean Sq for a variable: its Sum Sq divided by its Df, i.e., the average extra variation in Salary explained per degree of freedom for that variable.
- F-value for a variable: compares “extra variation explained by this variable” to “average variation left in the residuals.”