Explanation for R output

1 Model inference

For model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, ..., n$$

We CAN DO

- estimation of: β_0, β_1 and σ^2
- confidence intervals for: β_0 and β_1,
- testing (t-statistics, F-statistics, ANOVA, P-values),
  - t-statistics for β_0, β_1
  - F-statistics for β_1, but it is equivalent to testing regression model
  - ANOVA for analysis of the regression model
  - all can be tested based on their corresponding P-values
- prediction of: EY, confidence interval for EY, prediction confidence interval for Y.
- relevant concepts: LSE, S.E. P-value, t-value, F-value, R^2, MSE, SSE, SSR, MSR, SST, b_0, b_1, $\hat{\sigma}$, degree of freedom, ...

2 Using R

To estimate the model

$$\text{outputname} = \text{lm}(Y \sim X)$$

To show the estimation

$$\text{summary(outputname)}$$

To predict new $X_a, X_b, ..., X_m$

$$\text{new} = \text{data.frame}(X = c(X_a, X_b, ..., X_m))$$

$$\text{predictname} = \text{predict(outputname, new, interval="confidence", level=0.95)}$$

To show the ANOVA table

anova(outputname)

To plot data

plot(X, Y)

To add the fitted line

lines(X, outputname$fitted)

To add the upper limit and lower limit and form a confidence band

3 How to read the output of R

4 Example

For the heights of husband and wife data, **(data)**

- **(code)**

  xy = read.table('HeightsHusbandWife.dat')

  X = xy$V1 # Husband's height

  Y = xy$V2 # Wife's height

  reg = lm(Y∼X)

  summary(reg)

  anova(reg)

  plot(X, Y, xlim=c(120, 220), ylim=c(110, 200))

  new = data.frame(X=c(120:220))

  pred = predict(reg, new, interval="confidence", level=0.90)

  lines(new$X, pred[,1])

  lines(new$X, pred[,2], col="red")

  lines(new$X, pred[,3], col="red")

- output (how to read it)

  Regression Estimation summary

  Coefficients:

	Estimate	Std. Error	t value	$Pr(>\|t\|)$	
(Intercept)	41.93015	10.66162	3.933	0.000161	***
X	0.69965	0.06106	11.458	$< 2e-16$	***

  —

  Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

  Residual standard error: 5.928 on 94 degrees of freedom
  Multiple R-squared: 0.5828, Adjusted R-squared: 0.5783
  F-statistic: 131.3 on 1 and 94 DF, p-value: $< 2.2e-16$

  Analysis of Variance Table

  Response: Y

	Df	Sum Sq	Mean Sq	F value	$Pr(>F)$	
X	1	4613.7	4613.7	131.29	$< 2.2e-16$	***
Residuals	94	3303.3	35.1			

  —

  Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1