

Homework 2.

Theory question

Question 1. Derive the normal equations and the $\hat{\beta}$ from the multivariate version of the multiple regression likelihood

$$\frac{\partial \ell(\beta, \sigma^2)}{\partial \beta} = \frac{\partial}{\partial \beta} \left[-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta) \right]$$

Since $-\frac{n}{2} \log(2\pi\sigma^2)$ is constant with respect to β :

$$\frac{\partial \ell}{\partial \beta} = 0 - \frac{1}{2\sigma^2} \frac{\partial}{\partial \beta} [(Y - X\beta)'(Y - X\beta)]$$

Let $Q = (Y - X\beta)'(Y - X\beta) = Y'Y - 2\beta'X'Y + \beta'X'X\beta$. Then:

$$\frac{\partial Q}{\partial \beta} = -2X'Y + 2X'X\beta$$

Set the derivative to zero:

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

$$X'X\beta = X'Y$$

$$\hat{\beta}_{MLE} = (X'X)^{-1}X'Y$$

Question 2. Demonstrate that $\hat{\beta}$ is an unbiased estimator of β

$$\begin{aligned}Y &= X\beta + \epsilon \\ \hat{\beta} &= (X'X)^{-1}X'Y \\ \hat{\beta} &= (X'X)^{-1}X'(X\beta + \epsilon) \\ \hat{\beta} &= (X'X)^{-1}X'(X\beta) + (X'X)^{-1}X'(\epsilon) \\ \hat{\beta} &= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'(\epsilon) \\ \hat{\beta} &= \beta + (X'X)^{-1}X'(\epsilon) \\ E(\hat{\beta}) &= \beta + E((X'X)^{-1}X'(\epsilon)) \\ E(\hat{\beta}) &= \beta + (X'X)^{-1}X' E(\epsilon) \\ E(\epsilon) &= 0 \quad (\text{by assumption})\end{aligned}$$

$$\boxed{E(\hat{\beta}) = \beta}$$

Question 3. Demonstrate that $\hat{\beta}$ follows a multivariate normal distribution

$$\begin{aligned}\text{Cov}(\hat{\beta}) &= \text{Cov}((X'X)^{-1}X'\epsilon) \\ &= (X'X)^{-1}X' \cdot \text{Cov}(\epsilon) \cdot [(X'X)^{-1}X']' \\ &= (X'X)^{-1}X' \cdot (\sigma^2 I_n) \cdot X(X'X)^{-1} \\ &= (\sigma^2 I_n)(X'X)^{-1}X'X(X'X)^{-1} \\ &= (\sigma^2 I_n)(X'X)^{-1}\end{aligned}$$

$$E(\hat{\beta}) = \beta \quad \text{Cov}(\hat{\beta}) = (\sigma^2 I_n)(X'X)^{-1}$$

A **linear transformation** of a multivariate normal is also a **multivariate normal**, and

$$\hat{\beta} = A + C(\epsilon)$$

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

Therefore

Since $\epsilon \sim N(0, \sigma^2 I_n)$ and $\hat{\beta}$ is a linear transformation of ϵ , it follows that:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$$

US Judge ratings

Regression Analysis for USJudgeRatings

This will be split into 5 step, 1. Preliminary data exploration to understand relationship between target variable RTEN and PREP and plotting the 2 variables against each other. 2. Simple Linear Relationship: Building a simple linear regression between RTEN and PREP and Plotting residuals against trial preparation variable to see if the relationship is linear 3. Variable selection: Running an F test to see if additional variables are needed 4. Multiple Linear Relationship: using an algorithmic approach to see if other variables are needed 5. Building final model and interpretation

STEP 1. DATA PREPARATION

```
# 1. Basic structure & missingness
```

```
cat("Missing values:", sum(is.na(USJudgeRatings)), "\n")
```

```
## Missing values: 0
```

```
cat("\nNo missing values -- ok to proceed with whole dataset")
```

```
##
```

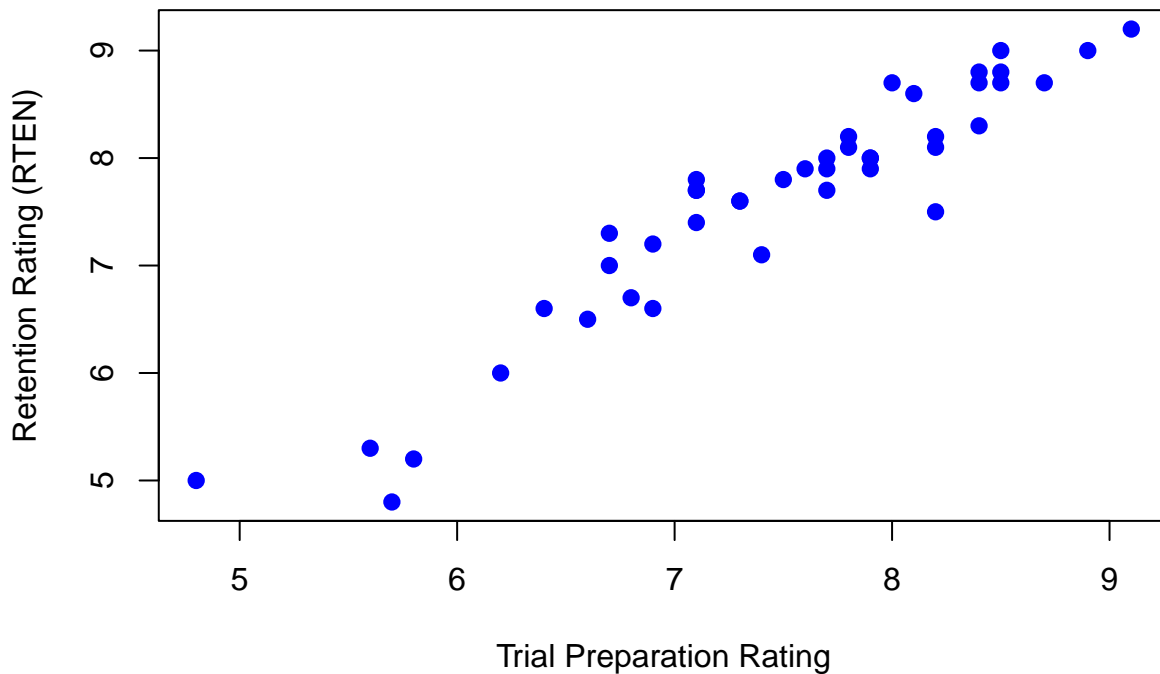
```
## No missing values -- ok to proceed with whole dataset
```

Zooming into the trial preparation vs retention, we see that there is a linear relationship between the 2 variables - this justifies our selection for a simple linear regression between RTEN and PREP

```
# Focused plot for Trial Preparation vs Retention Rating
```

```
plot(USJudgeRatings$PREP, USJudgeRatings$RTEN,  
     xlab = "Trial Preparation Rating",  
     ylab = "Retention Rating (RTEN)",  
     main = "Primary Relationship: Trial Preparation vs Retention",  
     pch = 16, col = "blue", cex = 1.2)
```

Primary Relationship: Trial Preparation vs Retention



STEP 2. SIMPLE LINEAR MODEL OF RTEN AGAINST PREP

We see that a 1 unit increase in the preparedness of the judge increases their retention by 1.09 points. This model also has a high adjusted

$$R^2$$

of 0.90, indicating that over 90% of the variation in judge retention can be explained by the judge preparation variable alone.

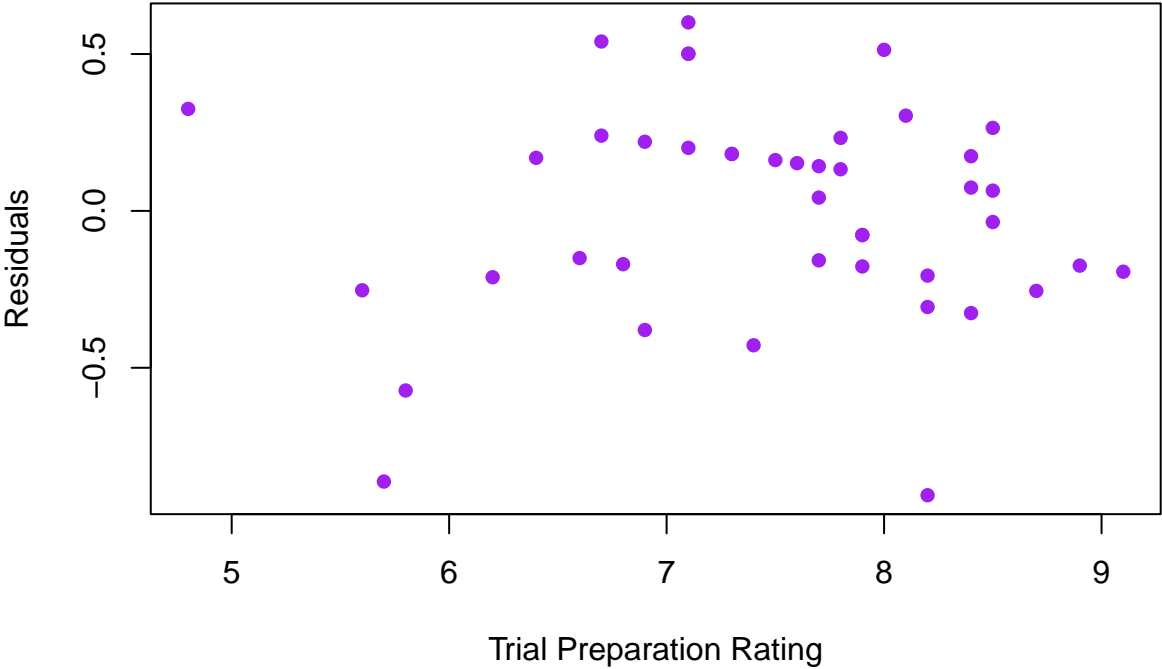
```
model_prep <- lm(RTEN ~ PREP, data = USJudgeRatings)
summary(model_prep)
```

```
##
## Call:
## lm(formula = RTEN ~ PREP, data = USJudgeRatings)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90625 -0.20008  0.06453  0.21065  0.60091
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.59257    0.42265  -1.402   0.168
## PREP         1.09742    0.05615  19.543 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.347 on 41 degrees of freedom
## Multiple R-squared:  0.9031, Adjusted R-squared:  0.9007
## F-statistic: 381.9 on 1 and 41 DF,  p-value: < 2.2e-16
```

We see that the plot between the residuals and the trial preparation variable resemble a random scatter, suggesting that there is in fact a linear relationship between trial preparation and judge retention.

```
plot(USJudgeRatings$PREP, residuals(model_prep),
     xlab = "Trial Preparation Rating",
     ylab = "Residuals",
     main = "Residuals vs Trial Preparation",
     pch = 16, col = "purple")
```

Residuals vs Trial Preparation



STEP 3. CHECKING IF OTHER COVARIATES ARE NEEDED

Running an F test against all the other variables to see if the other variables are needed.

For the F test, the null hypothesis states that all coefficients for the additional judicial rating variables (besides PREP) are equal to zero, meaning they provide no additional explanatory power for predicting retention ratings beyond what is already captured by trial preparation alone.

We see that the P value of the F test is near zero, and statistically significant at the 0.1% level, which presents strong evidence against rejecting the null; this indicates that other variables are needed in the model.

```
# Complex model: PREP + all other variables
model_complex <- lm(RTEN ~ ., data = USJudgeRatings) # . means all other variable
# F-test to compare nested models
f_test_result <- anova(model_prep, model_complex, test='F')
print(f_test_result)
```

```
## Analysis of Variance Table
##
## Model 1: RTEN ~ PREP
## Model 2: RTEN ~ CONT + INTG + DMNR + DILG + CFMG + DECI + PREP + FAMI +
##      ORAL + WRIT + PHYS
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      41 4.9354
## 2      31 0.4274 10      4.508 32.696 1.052e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

STEP 4. MODEL SELECTION AND VARIABLE SELECTION

Using the leaps package, we see that the model with both a high R^2 and a more parsimonious model, with only 5 covariates.

```
library(leaps)

##### Adjusted R^2 for Judicial Ratings #####
best_adjR2 <- leaps(x = USJudgeRatings[, -which(names(USJudgeRatings) == "RTEN")],
  y = USJudgeRatings$RTEN,
  method = 'adjr2')

# Create readable output
models <- best_adjR2$which
adjR2 <- cbind(models, round(best_adjR2$adjr2, digits = 5))
colnames(adjR2) <- c(names(USJudgeRatings)[names(USJudgeRatings) != "RTEN"], 'adjR2')

# Display models sorted by adjusted R^2 (highest first)
top_10_adjR2 <- adjR2[order(best_adjR2$adjr2, decreasing = TRUE), ][1:10,]
print(top_10_adjR2)
```

```
##   CONT INTG DMNR DILG CFMG DECI PREP FAMI ORAL WRIT PHYS   adjR2
## 7    0    1    1    0    1    1    0    1    1    0    1 0.98972
## 7    0    1    1    0    1    1    0    0    1    1    1 0.98969
## 5    0    1    1    0    0    1    0    0    1    0    1 0.98967
## 6    0    1    1    0    1    1    0    0    1    0    1 0.98962
## 6    0    1    1    0    0    1    0    1    1    0    1 0.98961
## 6    0    1    1    0    0    1    0    0    1    1    1 0.98960
## 6    0    1    0    0    1    1    0    1    1    0    1 0.98952
## 7    1    1    1    0    1    1    0    0    1    0    1 0.98951
## 8    0    1    1    1    1    1    0    1    1    0    1 0.98950
## 8    1    1    1    0    1    1    0    1    1    0    1 0.98948
```

Based on this R^2 selection criteria, let's build $RTEN \sim INTG, DMNR, DECI, ORAL, PHYS$

```
model_r2 = lm(RTEN ~ INTG + DMNR + DECI + ORAL + PHYS, data=USJudgeRatings)
summary(model_r2)
```

```
##
## Call:
## lm(formula = RTEN ~ INTG + DMNR + DECI + ORAL + PHYS, data = USJudgeRatings)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
	-0.240656	-0.069026	-0.009474	0.068961	0.246402

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.20433	0.43611	-5.055	1.19e-05 ***
INTG	0.37785	0.10559	3.579	0.000986 ***
DMNR	0.15199	0.06354	2.392	0.021957 *
DECI	0.16672	0.07702	2.165	0.036928 *
ORAL	0.29169	0.10191	2.862	0.006887 **
PHYS	0.28292	0.04678	6.048	5.40e-07 ***

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
##
## Residual standard error: 0.1119 on 37 degrees of freedom
## Multiple R-squared: 0.9909, Adjusted R-squared: 0.9897
## F-statistic: 806.1 on 5 and 37 DF, p-value: < 2.2e-16
```

Lets also check the BIC – we see as expected, the model with the lowest BIC is the model with 5 covariates- it is also of note that the preparation variable is not included.

```
# Get the top 10 models by adjusted R²
top_10_adjR2 <- adjR2[order(best_adjR2$adjR2, decreasing = TRUE),][1:10,]

# Calculate BIC for each of the top 10 models
n <- nrow(USJudgeRatings)
bic_values <- numeric(10)

for(i in 1:10) {
  # Get which variables are in this model
  model_vars <- names(which(top_10_adjR2[i, 1:11] == 1)) # First 11 columns are variables

  # Build the formula
  if(length(model_vars) > 0) {
    formula <- as.formula(paste("RTEN ~", paste(model_vars, collapse = " + ")))
    model <- lm(formula, data = USJudgeRatings)
    bic_values[i] <- BIC(model)
  } else {
    # Intercept-only model
    model <- lm(RTEN ~ 1, data = USJudgeRatings)
    bic_values[i] <- BIC(model)
  }
}

# Create final results table - convert to data frame first
top_10_with_bic <- data.frame(top_10_adjR2, BIC = bic_values)

# Sort by BIC (lower is better)
top_10_with_bic[order(top_10_with_bic$BIC), ]
```

##	CONT	INTG	DMNR	DILG	CFMG	DECI	PREP	FAMI	ORAL	WRIT	PHYS	adjR2	BIC
## X5	0	1	1	0	0	1	0	0	1	0	1	0.98967	-46.47553
## X6	0	1	1	0	1	1	0	0	1	0	1	0.98962	-43.65629
## X6.1	0	1	1	0	0	1	0	1	1	0	1	0.98961	-43.60802
## X6.2	0	1	1	0	0	1	0	0	1	1	1	0.98960	-43.59449
## X6.3	0	1	0	0	1	1	0	1	1	0	1	0.98952	-43.23883
## X7	0	1	1	0	1	1	0	1	1	0	1	0.98972	-41.53361
## X7.1	0	1	1	0	1	1	0	0	1	1	1	0.98969	-41.39501
## X7.2	1	1	1	0	1	1	0	0	1	0	1	0.98951	-40.65669
## X8	0	1	1	1	1	1	0	1	1	0	1	0.98950	-38.11787
## X8.1	1	1	1	0	1	1	0	1	1	0	1	0.98948	-38.04348

Step 5. building the final model and interpretation

```
summary(model_r2)
```

```
##
## Call:
## lm(formula = RTEN ~ INTG + DMNR + DECI + ORAL + PHYS, data = USJudgeRatings)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.240656 -0.069026 -0.009474  0.068961  0.246402
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.20433    0.43611  -5.055 1.19e-05 ***
## INTG         0.37785    0.10559   3.579 0.000986 ***
## DMNR         0.15199    0.06354   2.392 0.021957 *
## DECI         0.16672    0.07702   2.165 0.036928 *
## ORAL         0.29169    0.10191   2.862 0.006887 **
## PHYS         0.28292    0.04678   6.048 5.40e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1119 on 37 degrees of freedom
## Multiple R-squared:  0.9909, Adjusted R-squared:  0.9897
## F-statistic: 806.1 on 5 and 37 DF,  p-value: < 2.2e-16
```

Overall Model Significance F-statistic = 806.1 with p-value < 2.2e-16 Adjusted R-squared: 0.9897

Extremely significant - the model as a whole explains retention ratings

Interpretation of Coefficients:

Judicial Integrity (INTG) - For each 1-point increase in integrity rating, retention rating increases by 0.378 points, holding other judicial qualities constant

Demeanor (DMNR) - Each 1-point increase in demeanor rating corresponds to a 0.152 point increase in retention rating, adjusting for other factors

Decisiveness (DECI)- More decisive judges receive 0.167 point higher retention ratings per 1-point decisiveness increase

Oral Ruling Quality (ORAL) - Superior oral rulings boost retention ratings by 0.292 points per 1-point quality increase

Physical Ability (PHYS) - Physical ability has the strongest impact: 0.283 point retention increase per 1-point ability increase

Model Performance: $R^2 = 0.9909$: 99.09% of retention rating variance explained

Adjusted $R^2 = 0.9897$: Exceptional fit even after penalizing for 5 predictors

Key Insight: Physical ability and judicial integrity emerge as the most powerful predictors, while your original focus on trial preparation (PREP) was correctly excluded. Lawyers appear to value overall judicial competence and presence more than specific trial skills.