# Homework 1

Ricardo Andres Calvo Mendez

January 28, 2026

## Question 1

### a. Fit a regression model with `lpsa` as the response and `lcavol` as the predictor. Show the $R^2$ of this model.

The following `R` code was used to compute the $R^2$ value.

R code

```
library(faraway)
data(prostate, package = "faraway")
model <- lm(lpsa ~ lcavol, data = prostate)
summary(model)
print(paste("R-squared estimator: ",
↪  summary(model)$r.squared))
```

The results are summarized below:

R output

```
Call:
lm(formula = lpsa ~ lcavol, data = prostate)

Residuals:
     Min       1Q   Median       3Q      Max
-1.67625 -0.41648  0.09859  0.50709  1.89673

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.50730    0.12194   12.36   <2e-16 ***
lcavol       0.71932    0.06819   10.55   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7875 on 95 degrees of freedom
Multiple R-squared:  0.5394, Adjusted R-squared:  0.5346
F-statistic: 111.3 on 1 and 95 DF,  p-value: < 2.2e-16

[1] "The R-squared estimator: 0.53943190877902"
```

Therefore, the $R^2$ value of the model is:

$$\boxed{R^2 \approx 0.5394}$$

## b. Add `lweight`, `svi`, `lbph`, `age`, `lcp`, `pgg45` and `gleason` as predictors to the regression model. Show the $R^2$ of this model.

The following `R` code was used to compute the $R^2$ value.

R code

```r
library(faraway)
data(prostate, package = "faraway")
model <- lm(lpsa ~ lcavol + lweight + svi + lbph + age +
↪  lcp + pgg45 + gleason, data = prostate)
summary(model)
print(paste("R-squared estimator: ",
↪  summary(model)$r.squared))
```

The results are summarized below:

R output

```
Call:
lm(formula = lpsa ~ lcavol + lweight + svi + lbph + age + lcp +
    pgg45 + gleason, data = prostate)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7331 -0.3713 -0.0170  0.4141  1.6381

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.669337   1.296387   0.516  0.60693
lcavol       0.587022   0.087920   6.677 2.11e-09 ***
lweight      0.454467   0.170012   2.673  0.00896 **
svi          0.766157   0.244309   3.136  0.00233 **
lbph         0.107054   0.058449   1.832  0.07040 .
age         -0.019637   0.011173  -1.758  0.08229 .
lcp         -0.105474   0.091013  -1.159  0.24964
pgg45        0.004525   0.004421   1.024  0.30886
gleason      0.045142   0.157465   0.287  0.77503
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7084 on 88 degrees of freedom
Multiple R-squared:  0.6548, Adjusted R-squared:  0.6234
F-statistic: 20.86 on 8 and 88 DF,  p-value: < 2.2e-16

[1] "The R-squared estimator: 0.654754085299709"
```

Therefore, the $R^2$ value of the model is:

$$\boxed{R^2 \approx 0.6548}$$

## c. Compare the $R^2$ of these two models. Explain why you observe such a comparison result.

The $R^2$ estimator measures the proportion of the variance in the response variable that is explained by the linear model. By definition, $R^2$ is non-decreasing when predictors are added. In these two examples, we observe that adding more variables leads to a larger value of the $R^2$ estimator.

Moreover, the increase from $R^2 = 0.5394$ to $R^2 = 0.6548$ indicates that the additional predictors increase the amount of variance explained in the observed data.

However, since $R^2$ always increases as more predictors are added, a larger $R^2$ alone is not sufficient to conclude that a model is superior. Therefore, when comparing models with different numbers of predictors, additional criteria such as the adjusted $R^2$ or predictive performance should be considered.

## d. Use the method introduced in lecture slides to manually fit the model in $b$. Construct a matrix $X$, a response vector $y$, and then obtain the *Least Squares estimator*. Compare the manually estimated parameters with the result from the `lm` function.

This is the `R` code used to compute the *Least Squares estimator*

R code

```
library(faraway)
data(prostate, package = "faraway")
x <- model.matrix(~ lcavol + lweight + svi + lbph + age +
↪  lcp + pgg45 + gleason, data = prostate)
y <- prostate$lpsa
xtxi <- solve(t(x) %*% x)
print(xtxi %*% t(x) %*% y)
```

Here are the results:

R output

```
                  [,1]
(Intercept)  0.669336698
lcavol       0.587021826
lweight      0.454467424
svi          0.766157326
```

```
lbph         0.107054031
age         -0.019637176
lcp         -0.105474263
pgg45        0.004525231
gleason      0.045141598
```

Here, the *resulting column* presents the same values as the column *Estimate* in the previous linear model (b.)

## e. Consider the model in part b. For each parameter associated with a predictor, conduct the following hypothesis test $(\alpha = 0.05)$ using the manual method in lecture notes.

We define the following hypotheses:

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

Let $(\alpha = 0.05)$ and $t$ statistic as:

$$t = \frac{\widehat{\beta}_i - \beta_i}{\widehat{se(\widehat{\beta}_i)}}$$

This is the R code used to manually compute the hypothesis test:

R code

```
library(faraway)
data(prostate, package = "faraway")
x <- model.matrix(~ lcavol + lweight + svi + lbph + age +
↪  lcp + pgg45 + gleason, data = prostate)
n <- nrow(x)
p <- ncol(x)
alpha <- 0.05
y <- prostate$lpsa
xtxi <- solve(t(x) %*% x)
beta_hat <- xtxi %*% t(x) %*% y
y_hat <- x %*% beta_hat
sigma2_hat <- sum((y - y_hat)^2) / (n - p)
t <- beta_hat / diag(sigma2_hat * xtxi)^0.5
p_value <- 2 * pt(-abs(t), df = n - p)
significant <- p_value < alpha
print(t)
print(significant)
```

Here are the results: the first output reports the t-statistics, and the second output indicates which predictors lead to rejection of the null hypothesis.

R output

```
                  [,1]
(Intercept)  0.5163091
lcavol       6.6767493
lweight      2.6731423
svi          3.1360157
lbph         1.8315735
age         -1.7575995
lcp         -1.1588861
pgg45        1.0235350
gleason      0.2866779
             [,1]
(Intercept) FALSE
lcavol       TRUE
lweight      TRUE
svi          TRUE
lbph        FALSE
age         FALSE
lcp         FALSE
pgg45       FALSE
gleason     FALSE
```

## f. Use `lm` and summary in `R` to do the above test. Are the test statistics the same as those computed in e.)?

We reproduce the results from part b. to compare with the ones of e.

R output

```
Call:
lm(formula = lpsa ~ lcavol + lweight + svi + lbph + age + lcp +
    pgg45 + gleason, data = prostate)

Residuals:
    Min      1Q  Median      3Q     Max
-1.7331 -0.3713 -0.0170  0.4141  1.6381

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.669337   1.296387   0.516  0.60693
lcavol       0.587022   0.087920   6.677 2.11e-09 ***
lweight      0.454467   0.170012   2.673  0.00896 **
svi          0.766157   0.244309   3.136  0.00233 **
lbph         0.107054   0.058449   1.832  0.07040 .
age         -0.019637   0.011173  -1.758  0.08229 .
lcp         -0.105474   0.091013  -1.159  0.24964
pgg45        0.004525   0.004421   1.024  0.30886
gleason      0.045142   0.157465   0.287  0.77503
---
```

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

Residual standard error: 0.7084 on 88 degrees of freedom
Multiple R-squared:  0.6548, Adjusted R-squared:  0.6234
F-statistic: 20.86 on 8 and 88 DF,  p-value: < 2.2e-16

[1] "The R-squared estimator: 0.654754085299709"
```

Here we can see that the column `t value` contains the same values as the t-statistics computed in e. except that the manually computed values are reported with more decimal places. Additionally, the `summary()` function marks the level of statistical significance using asterisks (*), where a larger number of asterisks corresponds to a smaller significance level $\alpha$. A single asterisk indicates significance at the $\alpha = 0.05$ level.

We observe that the predictors *lcavol*, *lweight*, and *svi* have more than one asterisk, indicating that they are statistically significant at the chosen significance level. These predictors are exactly the ones marked as `TRUE` in the manual procedure, confirming the correctness of the manual hypothesis testing results.

## g. Compute a 95% *CI* for the parameter associated with each predictor.

The following `R` code computes the 95% confidence intervals:

R code

```r
library(faraway)
data(prostate, package = "faraway")
x <- model.matrix(~ lcavol + lweight + svi + lbph + age +
↪  lcp + pgg45 + gleason, data = prostate)
n <- nrow(x)
p <- ncol(x)
alpha <- 0.05
y <- prostate$lpsa
xtxi <- solve(t(x) %*% x)
beta_hat <- xtxi %*% t(x) %*% y
y_hat <- x %*% beta_hat
sigma2_hat <- sum((y - y_hat)^2) / (n - p)
lower_bound <- beta_hat - qt(1 - alpha / 2, df = n - p) *
↪  sqrt(diag(sigma2_hat * xtxi))
upper_bound <- beta_hat + qt(1 - alpha / 2, df = n - p) *
↪  sqrt(diag(sigma2_hat * xtxi))
print(data.frame(Lower = lower_bound, Upper = upper_bound))
```

Here are the results:

R output

```
                  Lower       Upper
(Intercept) -1.906960983 3.245634379
lcavol       0.412298699 0.761744954
lweight      0.116603435 0.792331414
svi          0.280644232 1.251670420
lbph        -0.009101499 0.223209561
age         -0.041840618 0.002566267
lcp         -0.286344443 0.075395916
pgg45       -0.004260932 0.013311395
gleason     -0.267786053 0.358069248
```

# Question 2

Let $y \in \mathbb{R}^n$ and $X \in \mathbb{R}^{n \times p}$ with $n \geq p$. Consider the linear model

$$y = X\beta + \epsilon,$$

where $\beta \in \mathbb{R}^p$ is unknown. Assume:

1. $X$ is a fixed design matrix with full column rank $p$ so $X^\top X$ is invertible.
2. $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ with $\sigma^2 > 0$ (equivalently, $\epsilon_i$ are i.i.d. $\mathcal{N}(0, \sigma^2)$).

## a. Derive the OLS estimator

Starting from $\min_\beta \|y - X\beta\|_2^2$, show that the unique minimizer is

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

To show this, start with converting this square to something easier to manipulate:

$$\hat{\beta} = \|y - X\beta\|_2^2 = (y - X\beta)^T (y - X\beta)$$

Remember that the transpose here is to match the correct dimensions for matrix multiplication.

Now as we want to minimize this, let make the derivative of this equals to zero:

$$\frac{\partial \, (y - X\beta)^T (y - X\beta)}{\partial \beta} \; = \; 0$$

Now following the rules of matrix derivative $X^T X \approx x^2$ and $X\beta \approx ax$ we solve the derivative as follows:

$$\frac{\partial \, (y - X\beta)^T (y - X\beta)}{\partial \beta} = -2(y - X\beta)X = -2X^T (y - X\beta) = 0$$

Now solving:

$$
\begin{aligned}
-2X^T(y - X\beta) &= 0 \\
-X^T(y - X\beta) &= 0 \\
-X^Ty + X^TX\beta &= 0 \\
X^TX\beta &= X^Ty \\
\beta &= (X^TX)^{-1}X^Ty
\end{aligned}
$$

The procedure to derive the least squares formula has been presented.

## b. Unbiasedness

Show that $\mathbb{E}[\hat{\beta}] = \beta$

To show this, we replace $y$ from the linear model definition into the least squares formula and solve:

$$
\begin{aligned}
\hat{\beta} &= (X^TX)^{-1}X^Ty \\
\hat{\beta} &= (X^TX)^{-1}X^T(X\beta + \epsilon) \\
\hat{\beta} &= (X^TX)^{-1}X^TX\beta + (X^TX)^{-1}X^T\epsilon \\
\hat{\beta} &= \beta + (X^TX)^{-1}X^T\epsilon
\end{aligned}
$$

Now applying expected value to both sides of the equation and remembering that $\beta$ and $X$ are constants, and $\epsilon$ has mean 0 according to our assumptions:

$$
\begin{aligned}
\mathbb{E}[\hat{\beta}] &= \mathbb{E}[\beta + (X^TX)^{-1}X^T\epsilon] \\
\mathbb{E}[\hat{\beta}] &= \mathbb{E}[\beta] + \mathbb{E}[(X^TX)^{-1}X^T\epsilon] \\
\mathbb{E}[\hat{\beta}] &= \beta + (X^TX)^{-1}X^T\mathbb{E}[\epsilon] \\
\mathbb{E}[\hat{\beta}] &= \beta + (X^TX)^{-1}X^T(0) \\
\mathbb{E}[\hat{\beta}] &= \beta
\end{aligned}
$$

The procedure to derive the expected value formula has been presented.

## c. Variance

Show that $Var(\hat{\beta}) = \sigma^2(X^TX)^{-1}$

To show this, we make again the substitution of $y$ from the linear model definition into the least squares formula:

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

Now, using the property that says that if $A$ is fixed, then $Var(A\epsilon) = A \cdot Var(\epsilon) \cdot A^T$, we compute the variance of $\hat{\beta}$ as follows:

$$
\begin{aligned}
Var(\hat{\beta}) &= Var(\beta + (X^T X)^{-1} X^T \epsilon) \\
Var(\hat{\beta}) &= Var(\beta) + Var((X^T X)^{-1} X^T \epsilon) \\
Var(\hat{\beta}) &= Var((X^T X)^{-1} X^T \epsilon) \\
Var(\hat{\beta}) &= (X^T X)^{-1} X^T \cdot Var(\epsilon) \cdot ((X^T X)^{-1} X^T)^T \\
Var(\hat{\beta}) &= (X^T X)^{-1} X^T \cdot \sigma^2 I \cdot ((X^T X)^{-1} X^T)^T \\
Var(\hat{\beta}) &= (X^T X)^{-1} X^T \cdot \sigma^2 I \cdot X(X^T X)^{-1} \\
Var(\hat{\beta}) &= \sigma^2 (X^T X)^{-1} \cdot (X^T \cdot X) \cdot (X^T X)^{-1} \\
Var(\hat{\beta}) &= \sigma^2 (X^T X)^{-1}
\end{aligned}
$$

The procedure to derive the variance formula has been presented.

## d. Distribution

Using that linear transformations of multivariate normal vectors are normal, prove that

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

To show this, we use again the substitution of linear model definition into least squares formula

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

We can see here that as $X$ is fixed (it is constant) the only random variable is $\epsilon$ which follows the Normal distribution $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ so that, $\hat{\beta}$ also follows the given Normal distribution.

now using the proof from b. where the expected value (or mean) of $\hat{\beta}$ is $\mathbb{E}[\hat{\beta}] = \beta$ and the proof from c. where the variance of $\hat{\beta}$ is $Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$. We can determine that $\hat{\beta}$ follows the Normal distribution:

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

# Question 3

You have regressed $y$ on variables $x_1, x_2, \ldots, x_p$. Your friend, Bob, has regressed $y$ on the variables $z_1, z_2, \ldots, z_p$, where

$$z_j \;=\; c_{j0} \;+\; \sum_{k=1}^{p} c_{jk} x_k$$

That is, Bob has applied a linear transformation to the predictors (but not to the response).

### a. Show that Bob's $n \times (p+1)$ design matrix $Z$ is related to yours via $Z = XC$ for some $(p+1) \times (p+1)$ matrix $C$; explain how the entries in $C$ are related to coefficients $c_{jk}$.

We define $X$ and $Z$ matrices using the variables $x_1, x_2, \ldots, x_p$ and $z_1, z_2, \ldots, z_p$ as follows:

$$X = \begin{pmatrix} 1 & x_1 & x_2 & \cdots & x_p \end{pmatrix}$$

$$Z = \begin{pmatrix} 1 & z_1 & z_2 & \cdots & z_p \end{pmatrix}$$

We define an arbitrary matrix $C$ as follows

$$C = \begin{pmatrix} 1 & c_{10} & c_{20} & \cdots & c_{p0} \\ 0 & c_{11} & c_{21} & \cdots & c_{p1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & c_{1p} & c_{2p} & \cdots & c_{pp} \end{pmatrix}$$

Using the matrix multiplication definition (row-column dot product) we define $Z = XC$ following the definition provided in the question:

$$z_j \ = \ c_{j0} \ + \ \sum_{k=1}^{p} c_{jk} x_k$$

So we can conclude that exists a matrix $C$ such that $Z = XC$

### b. Show that the predictions from your model and Bob's model are exactly equal, if $C$ is invertible.

The predictions using matrices $X$ and $Z$ can be as follows:

$$\hat{\beta} = X(X^T X)^{-1} X^T y$$

$$\hat{\gamma} = Z(Z^T Z)^{-1} Z^T y$$

We know that $C$ is invertible, so the procedure continue as follows showing that $\hat{\beta} = \hat{\gamma}$:

$$
\begin{aligned}
\hat{\beta} &= \hat{\gamma} \\
X(X^TX)^{-1}X^Ty &= Z(Z^TZ)^{-1}Z^Ty \\
X(X^TX)^{-1}X^Ty &= XC((XC)^T(XC))^{-1}(XC)^Ty \\
X(X^TX)^{-1}X^Ty &= XC(C^TX^TXC)^{-1}C^TX^Ty \\
X(X^TX)^{-1}X^Ty &= XCC^{-1}(X^TX)^{-1}(C^T)^{-1}C^TX^Ty \\
X(X^TX)^{-1}X^Ty &= X(X^TX)^{-1}X^Ty
\end{aligned}
$$

Here we can see that both predictions are the same.