

1. a) The columns of X are linearly independent, so $\text{rank}(X) = p$. This in turn dictates that $\text{rank}(QR) = p$. Since the rank of the product of two matrices can be no greater than the smallest rank of the factors, we must conclude that $\text{rank}(Q) = \text{rank}(R) = p$, since they both can't have greater rank than p . Since R has full rank, it must be invertible.
- b) Multiplying both sides of the linear system through by Q , we get that

$$\begin{aligned} R\tilde{\beta} &= Q^T Y \\ QR\tilde{\beta} &= QQ^T Y \\ X\tilde{\beta} &= Y \end{aligned}$$

So we may deduce that Y is in the span of X . Thus, for $H = X(X^T X)^{-1} X^T$, $HY = Y$, since $HZ = Z$ for all vectors Z in the span of X . Starting with this, we get that

$$\begin{aligned} HY &= Y \\ X(X^T X)^{-1} X^T Y &= Y \\ X\hat{\beta} &= Y \end{aligned}$$

Since, $\tilde{\beta}$ is unique, we may conclude that $\tilde{\beta} = \hat{\beta}$.

- c) The "normal equation" to calculate $\hat{\beta}$ ran into a singularity at $\epsilon = 1e - 5$, while the QR method of calculation still worked. The R code used for this calculation is attached in the back.
- d) The QR method and the linear model returned the following results for $\epsilon = 1e - 5$:

Coefficient	QR Method	Linear Model
Intercept	2.862e+1	2.862e+1
pop15	-4.620e-1	-4.620e-1
pop75	-1.151e+4	-1.151e+4
dpi	-3.356e-4	-3.356e-4
ddpi	4.088e-1	4.088e-1
pop75 + $u\epsilon$	1.151e+4	1.151e+4

It is not surprising that they returned the same estimates, since we were told that the linear model in R is using the QR method of calculation. The one surprising result is that the two pop75 variables had such big coefficients, but this is due to them being highly collinear, which is why their coefficients added together are almost 0.

2. AY and BY are jointly normally distributed, since they are both linear combinations of Y , which is drawn from a multivariate normal distribution. We can also show that their covariance is 0:

$$\begin{aligned} \text{Cov}[AY, BY] &= E[AY(BY)^T] \\ &= E[AYY^T B^T] \\ &= E[\|Y\|^2 AB] \\ &= E[0] \\ &= 0 \end{aligned}$$

Since AY and BY are jointly normally distributed, this is sufficient to prove their independence. Now, since YAY and YBY are merely linear combinations of the variables in AY and BY , which are pairwise independent, YAY and YBY must also be linearly independent.

3. Since A is symmetric, we can eigendecompose it, so that $A = PDP^T$, where D is a diagonal matrix of A 's eigenvalues, and P is an orthogonal matrix of A 's eigenvectors. It is the case that, where L is the set of all eigenvalues of A :

$$p = \text{tr}(A) = \text{tr}(D) = \sum_{\lambda \in L} \lambda$$

Since all of A 's eigenvalues are either 0 or 1, this means that p is the number of non-zero eigenvalues of A . Also,

$$\begin{aligned}
e^T A e &= \text{tr}(e^T A e) \\
&= \text{tr}(e^T P D P^T e) \\
&= \text{tr}(e^T D P^T P e) \\
&= \text{tr}(e^T D e) \\
&= e^T D e \\
&= \sum_{i=1}^n e_i^2 I_{\{j: d_{jj}=1\}}(i)
\end{aligned}$$

The last line is based on the fact that the eigen values of A , which are also the diagonal of D , are all either 0 or 1. Clearly, we will have p terms in our summand, since that is how many eigenvalues of 1 there are. Since each e_i is independent of all other e_j , this means $e^T A e$ is the sum of p independent squared normals, or in other words

$$e^T A e \sim \chi_p^2$$

4.

$$\begin{aligned}
\text{Cov}(\hat{e}, \hat{Y}) &= E \left[\hat{e} \hat{Y}^T \right] - E[\hat{e}] E[\hat{Y}]^T \\
&= E[0_{n \times n}] - 0_{n \times 1} Y^T && \text{Since } \hat{e} \text{ and } \hat{Y} \text{ are orthogonal} \\
&= 0_{n \times n}
\end{aligned}$$

5. 3.1 a) After running the model and getting the coefficient for age and its standard error, the confidence intervals are:

$$90\% \text{ CI} = \beta_{age} \pm SE \cdot t_{rdf}^{.05} = -.0196 \pm .0112 \cdot 1.6624 = (-.0382, -.0012)$$

$$95\% \text{ CI} = \beta_{age} \pm SE \cdot t_{rdf}^{.025} = -.0196 \pm .0112 \cdot 1.9873 = (-.0418, .0026)$$

Since the 90% CI doesn't include 0, but the 95% CI does, we can conclude the pvalue for β_{age} is between .05 and .1.

- c) The simulation, the code for which is in my code supplement, run on 1000 permutations of lpsa, had 8.8% of the results with a higher t value for β_{age} .
 - d) Running the F test with the null hypothesis that a model just with lcaivol, lweight, and svi is the real model, one would expect to see at least as large an improvement from adding four variables as we saw by adding lbph, age, lcp, and gleason 21.67% of the time. Thus, we can not reject the null hypothesis, and prefer the simpler model.
- 3.2 a) Running a linear model predicting taste in terms of the other three variables, only H2S and Lactic are statistically significant.
- b) Transforming Acetic and H2S back from the log scale, H2S ceases to be statistically significant, Acetic remains insignificant, and Lactic remains significant.

```

library('faraway')

#####
# Problem 1
#####

# Load the data and create the matrices
data('savings')
attach(savings)
n<-nrow(savings)
x<-as.matrix(data.frame(rep(1,n),savings[-1]))
y<-as.matrix(savings[1])

### Part c)
# Loop over epsilons and try methods until failure
eps <- .1
norm.diff <- NULL
continue <- TRUE
while (continue) {
  eps <- eps *.1
  # Add pop75 + noise
  x.i<-cbind(x,as.matrix(x[,3]+ eps*rnorm(n)))

  #See if the QR method still works
  QR.success <-tryCatch(
    {
      q<-qr.Q(qr(x.i))
      r<-qr.R(qr(x.i))
      beta.qr <- solve(r, t(q) %*% y)
      TRUE
    },
    error=function(e) return(FALSE),
    warning=function(e) return(FALSE)
  )

  # See if the normal method still works
  Normal.success <-tryCatch(
    {
      beta.x.i <- solve(t(x.i)%*%x.i) %*% t(x.i) %*% y
      TRUE
    },
    error=function(e) return(FALSE),
    warning=function(e) return(FALSE)
  )

  # Keep going and calculate norm of difference if both succeed
  continue <- QR.success & Normal.success
  if (continue) norm.diff <- c(norm.diff,norm(beta.qr-beta.x.i))
}
if (!QR.success) print(paste('The QR method failed at an epsilon of ',toString(eps)))

```

```

if (!Normal.success) print(paste('The normal equation failed at an epsilon of ',toString
#### Part d)
model<-lm(y~x.i)
summary(model)
beta.qr

#####
# Problem 5
#####

data(prostate)

#### 3.1.a)
m<-lm(lpsa~lcavol + lweight + age +lbph + svi + lcp + gleason + pgg45, data=prostate)
coe<-summary(m)$coef
ci.9 <- c(coe[4,1] + qt(.05,88)*coe[4,2],coe[4,1] - qt(.05,88)*coe[4,2])
ci.95 <- c(coe[4,1] + qt(.025,88)*coe[4,2],coe[4,1] - qt(.025,88)*coe[4,2])

#### 3.1.c)
nb<-1000
n<-nrow(prostate)
x<-as.matrix(prostate[1:8])
y<-as.matrix(prostate[9])

# Thus function pulls the t score from a regression predicting the y permuted
est.t.age <- function(n) summary(lm(y[sample(n,rep=FALSE)]~x))$coef[4,3]

# Which we now run on a vector of the draw size
t.age <- sapply(rep(n,nb),est.t.age)
pval<- mean(abs(t.age)>abs(coe[4,3]))

#### 3.1.d)
#only lcavol, lweight, and svi had a pvalue of less than .05
rm<-lm(lpsa~lcavol + lweight + svi, data=prostate)
anova(m,rm)

#### 3.2.a)
data(cheddar)
summary(lm(taste ~ Acetic + H2S + Lactic, data=cheddar))

#### 3.2.b)
summary(lm(cheddar$taste ~ exp(cheddar$Acetic) + exp(cheddar$H2S) + cheddar$Lactic))

```