MAS 372 Supplemental Questions

Sec 2.2

Supplemental Question: (Modified SOA)

You are given the results on the right from a regression model.
(a) Calculate the mean squared error (MSE).
(b) Calculate the MSE for an estimator $\hat{g}$ that predicts $\hat{g}(x_i) = 5$ for all x_i.

Obs.	y_i	$\hat{f}(x_i)$
1	2	4
2	5	3
3	6	9
4	8	3
5	4	6

Sec 3.1, 3.6.1, 3.6.2

Supplemental Question:

In order to get a better feel for how linear regression works, do this question without using R.

An ordinary least squares model with one variable (Concentration) and an intercept was fit to the observed data on the right in order to estimate the response.

Obs.	Concen.	Response
1	2	7
2	4	9
3	5	10
4	6	15
5	8	14

(a) Calculate $\hat{\beta}_0$ and $\hat{\beta}_1$.
(b) Calculate the residual standard error.
(c) Find a 95% confidence interval for β_1.
(d) Test the hypothesis $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$.

Sec 3.2, 3.6.2, 3.6.3

Supplemental Question:

The chart on the right shows the TSS and RSS for five different models, all fit using the same 30 observations and ordinary least squares regression. You test $H_0 : \beta_3 = \beta_4 = 0$ against the alternative that at least one of $\beta_3, \beta_4 \neq 0$.

Variables	TSS	RSS
Intercept Only	19851	3456
X_1	19851	3003
X_1, X_2	19851	2781
X_1, X_2, X_3	19851	2121
X_1, X_2, X_3, X_4	19851	2104

(a) Calculate your test statistic and estimate your p-value.
(b) Which model will have the largest R^2 value. Do you think this will be the best model?
(c) Which model has the smallest RSE?

Sec 3.3.1, 3.3.2, 3.6.4-7

Supplemental Question:

(a) In your model from exercise 3.10(e), include an interaction term between your two predictors. Code this two different ways, once using the $*$ symbol and once using the : symbol. Do you think that including the interaction term improves the model?

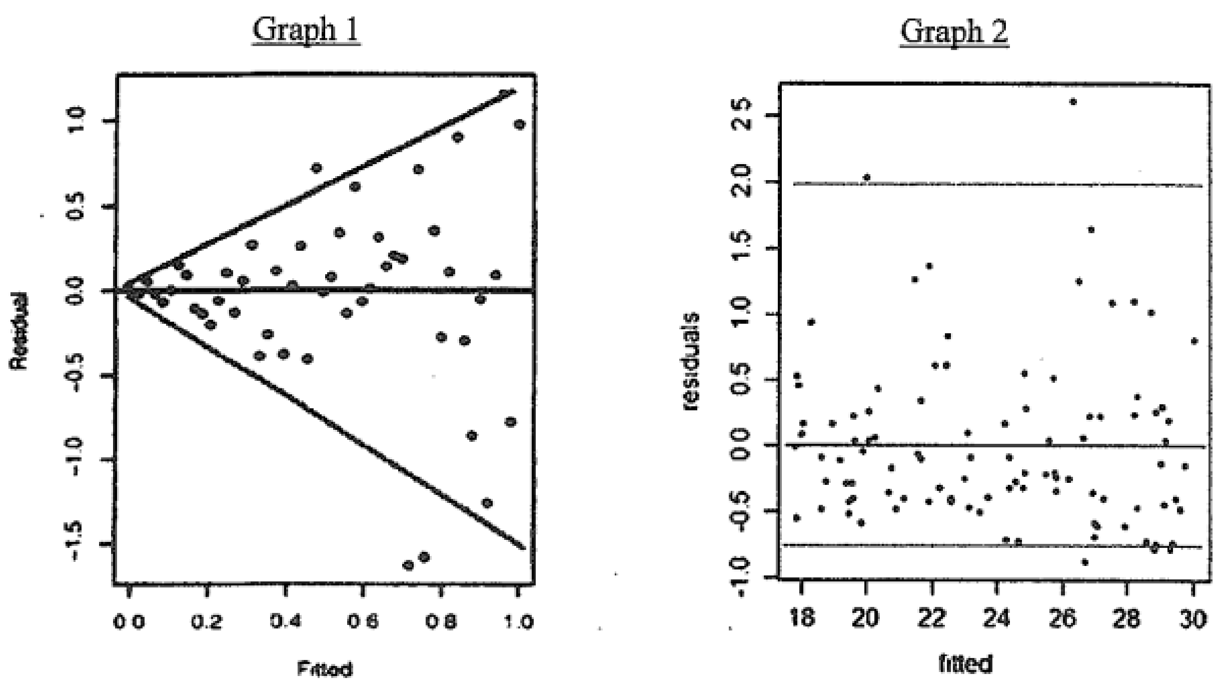
(b) In your model from exercise 3.10(e), include a $(\text{price})^2$ term. Code this two different ways, once using the $\wedge 2$ symbol and once using the poly command. Do you think that including the quadratic term improves the model?

(c) In part (b), your answers should be slightly different for the two different ways that you coded your answer in R. Look up the poly command in R's help to figure out why. Then adjust your code using the poly command to produce the same results as when the $\wedge 2$ symbol is used.

Sec 3.3.3, 3.6.2, 3.6.3

Supplemental Question 1:

You are given the following two graphs comparing the fitted values to the residuals of two different linear models (image from CAS):



Determine if the following statements are true or false. Briefly justify your answers.

(a) Graph 1 indicates the data is heteroskedastic.

(b) Graph 2 indicates the data is non-normal

Supplemental Question 2:

You are given the following information related to the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$.
For the model $x_1 = a_0 + a_1 x_2 + a_2 x_3$ the Regression SS is 3.73 and the Error SS is 3.04.
For the model $x_2 = b_0 + b_1 x_1 + b_2 x_3$ the Regression SS is 24.74 and the Error SS is 226.34.
For the model $x_3 = c_0 + c_1 x_1 + c_2 x_2$ the Regression SS is 214,258 and the Error SS is 85,883.
The Error SS is another term for the RSS. The Regression SS is the difference between the TSS and the RSS, so can be thought of as the portion of the TSS explained by the regression.
Calculate the variance inflation factor for the variable which exhibits the greatest collinearity in the original model.

Sec 4.1-4.3, 4.6.1, 4.6.2

Supplemental Question 1:

For exercise 4.6, give an interpretation in words for the coefficients β_0, β_1, and β_2 in the context of this problem.

Supplemental Question 2:

X	2	3	4	6	10
Y	0	1	0	1	1

The chart on the right shows pairs of values of a predictor X and a response Y, where Y is binary.
You are fitting a logistic regression model to the data using maximum likelihood to find the coefficients. Write down your likelihood function.

Sec 4.6.2, Ag 3.1, 3.2

Notes about Problems 3.7 and 3.8:

The crabs data set is on the course canvas page. Note that the variable y has already been created for you in the data set.
In previous assignments we used the glm function in R to perform logistic regression, but by changing the family command we can use glm to fit various different generalized linear models. Look up 'family' in R's help to see the various options available for the assumed distribution of the response variable. For each distribution, the default link function is listed next to the family name. In the description of link lower down on the help page there is a listing of the different links that can be used for each family.
For example, we have seen how to use a binomial response variable with the default logit link to carry out logistic regression. If we instead wanted to use the probit link we would need to use family = binomial(link="probit"). For problem 3.7(b) we want to treat Y as binomial but use a linear model, so an identity link. The Gaussian model with the identity link is the ordinary least squares regression that studied in the last chapter.

Sec Ag 3.3, Sups

Supplemental Question:

You are predicting the number of years newly planted trees will survive based on the density of deer per square mile in the area of the trees. The trees are observed just once a year, so the response variable is discrete. You create a Poisson loglinear model and get coefficients $\beta_0 = 2$ and $\beta_1 = -0.25$.

(a) For a tree in an area with 5 deer per square mile, what does the model predict as the average number of years a new tree tree will survive? What does the model predict for the variance of the number of years a new tree will survive?

(b) Interpret the parameter estimates in the context of this problem.

Sec Ag 3.4, 3.5, Sup

Supplemental Question 1:

A set of n observations, $y_1, y_2, \ldots, y_n$, are assumed to be exponentially distributed, $f(y_i) = \exp[-y_i/\theta_i]/\theta_i$ for $i = 1, 2, \ldots, n$. So note that θ_i is the expected value for y_i. A GLM is fit to the data with the following model specification: $\ln(\theta_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{2i}^2 + \beta_4 x_{3i}$. The parameter estimates $\hat{\beta}_k$ are 2.99, -0.27, -0.67, -0.16, and 0.91 respectively. For the second observation, the predictor values x_{j2} are 0, 0.6, and 1, the observed response is 137.6, and the leverage is 0.091. Calculate the standardized residual for the second observation.

Supplemental Question 2:

You are given $y_1, y_2, \ldots, y_n$, observations from independent and Poisson distributed random variables with respective means μ_i for $i = 1, 2, \ldots, n$. We want to predict μ_i using a predictor with value x_i for $i = 1, 2, \ldots, n$. A Poisson GLM was fitted to the data with a log-link function and systematic component $\beta_0 + \beta_1 x_i$.

(a) Let $\hat{\mu}_i$ be the prediction for the mean μ_i. Express $\hat{\mu}_i$ in terms of $\hat{\beta}_0$, $\hat{\beta}_1$, and x_i.

(b) For the saturated model, what will be the prediction for the mean μ_i?

(c) Show that for the unsaturated model the maximum value of the log-likelihood function is $\sum_{i=1}^{n}[-\hat{\mu}_i + y_i \ln(\hat{\mu}_i) - \ln(y_i!)]$.

(d) Show that the deviance for the unsaturated model is $2\sum_{i=1}^{n}[y_i \ln(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)]$.

Teamwork Formal Presentation and Submission Problems 1

3.13: For problem 13(h), please let the noise variance to be 0.05; for 13(i), please let the noise variance to be 0.5. After finish problem 13 (a-j), in your analysis report, please give a paragraph to compare the model performance on the three different data sets. What conclusion you can obtain?

Supplemental Question 1

Consider the Boston data set from the chapter. Your goal to create a linear model to predict the per capita crime rate using the other variables in the data set. This question is much more open ended than many of the other problems we have done - you need to come up

with a good model and justify your process, but there are multiple different models that are appropriate. As part of your work you should do the following (not necessarily in this order):
(i) Decide what variables to include in your model.
(ii) Consider at least one transformation of a predictor (the transformed variable may or may not be included in your final model).
(iii) Look at diagnostic plots for your model and identify any potential issues.
(iv) Use a transformation of the response data to address a potential issue.

On canvas, submit your R code used to explore different possible models as well as the final selected model.

Then write a paragraph describing to a non-technical audience the elements of your model (think of your audience as a criminal justice major interested in crime rates but with no experience with regression).

Lastly, write a paragraph describing to a technical audience describing why you chose the model that you did (think of your audience as another member of the class).

Sec 6.1, 6.5

Supplemental Question (Modified CAS):
Three statisticians were given a dataset and asked to build a model to predict claim frequency using any of 5 independent predictors $\{1, 2, 3, 4, 5\}$ as well as an intercept $\{I\}$. When evaluating the models they all use R-squared to compare models with the same number of parameters and AIC to compare models with different numbers of parameters. The results of all possible models are shown below.

Model	# of Non Intercept Parameters	Parameters	R^2	AIC
1	0	I	0	1.9
2	1	I, 1	0.56	1.4
3	1	I, 2	0.57	1.2
4	1	I, 3	0.55	1.6
5	1	I, 4	0.52	1.7
6	1	I, 5	0.51	1.8
7	2	I, 1, 2	0.61	1.0
8	2	I, 1, 3	0.64	0.5
9	2	I, 1, 4	0.63	0.8
10	2	I, 1, 5	0.69	0.0
11	2	I, 2, 3	0.61	1.0
12	2	I, 2, 4	0.62	0.9
13	2	I, 2, 5	0.68	0.2
14	2	I, 3, 4	0.66	0.4
15	2	I, 3, 5	0.64	0.5
16	2	I, 4, 5	0.60	1.1

Model	# of Non Intercept Parameters	Parameters	R^2	AIC
17	3	I, 1, 2, 3	0.73	1.3
18	3	I, 1, 2, 4	0.71	1.5
19	3	I, 1, 2, 5	0.72	1.4
20	3	I, 1, 3, 4	0.75	1.0
21	3	I, 1, 3, 5	0.76	0.8
22	3	I, 1, 4, 5	0.79	0.2
23	3	I, 2, 3, 4	0.78	0.6
24	3	I, 2, 3, 5	0.74	1.2
25	3	I, 2, 4, 5	0.75	1.1
26	3	I, 3, 4, 5	0.73	1.3
27	4	I, 1, 2, 3, 4	0.88	1.6
28	4	I, 1, 2, 3, 5	0.80	2.1
29	4	I, 1, 2, 4, 5	0.87	1.8
30	4	I, 1, 3, 4, 5	0.83	2.0
31	4	I, 2, 3, 4, 5	0.85	1.9
32	5	I, 1, 2, 3, 4, 5	0.90	3.5

(a) One statistican uses best subset selection. Which model do they choose?
(b) The second statistician uses forward stepwise selection. Which model do they choose?
(c) The third statistician uses backwards stepwise selection. Which model do they choose?

Sec 6.3, 6.7

Supplemental Question 1:
To better understand PCA and PCR, we will explore both in this problem by doing some hand computation and estimation in a small example. You have two predictors, X_1 and X_2, and a response variable Y. You have 5 observations: $(0, 3, 4)$, $(0, 4, 5)$, $(3, 0, 31)$, $(3, 3, 33)$, and $(4, 0, 40)$, each in the format (X_1, X_2, Y). The two predictors have the same SD so you do not need to standardize the variables.

(a) Plot the values of the two predictor variables for the observations.
(b) Find the means of the two predictors.
(c) Using the results of (a) and (b), estimate the first principal component. Briefly explain how you made your estimate.
(d) Estimate the second principal component.
(e) Based on your answers to (c) and (d), estimate the component scores for $(3, 0, 31)$.
(f) In a couple of sentences, describe how we would carry out PCR using one component in this case.

Supplemental Question 2:

Using the same data as supplement question 1, we will now explore PLS a bit.
(a) Estimate the correlations between each of the predictors and the response.
(b) Based on your answer to (a), estimate the first component Z_1. You can just use weights equal to the correlations - you do not need to worry about scaling.
(c) Find the value of Z_1 for each of the observations based on your component in part (b).
(d) In a couple of sentences, describe how we would carry out PLS using one component in this case.

Supplemental Question 3 - (Note: Do this after the R problem 6.9):
Look at the the summaries for the PCR and PLS fits in your results for exercise 6.9. When there are two components, which method explains more of the variance in the predictor variables X? Which method explains more of the variance in the response variable Apps? Briefly explain why both these results make sense.

Sec 6.4, 6.6

Supplemental Question 1:
In this problem we will explore some of the issues that occur when there are a large number of predictors compared to observations. We will have 100 predictors and 200 observations. All of our predictors and the response will be randomly generated, so there is not any actual relationship between the variables.

(a) Create a matrix with 100 columns (representing our 100 predictor variables) and 200 rows (representing the 200 observations) with each entry a random number from a standard normal distribution. Label the columns as 1 through 100.
Hint: See what the command
variables = matrix(rnorm(5*10),nrow=5, dimnames=list(NULL, paste(1:10)))
does and then adjust appropriately.
(b) Create a vector y of 200 random numbers from a standard normal distribution to represent the response variables. Combine y and your predictors from part (a) into one data frame.
(c) Fit an ordinary linear regression model for y in terms of all of the predictors.
(d) How many predictors were statistically significant at the 5% level in your model from part (c)? Why would you expect a result like this?
(e) Fit an ordinary linear regression model for y in terms of three of the predictors that we most significant. Do you have evidence that at least one of the coefficients is non-zero?

Supplemental Question 2:
(a) Building on supplement 1, now lets consider the case where we have 195 predictors and 200 observations. Repeat parts (a-c) from supplement 1 for this case. What is the R^2 value

in this case? How would you explain this R^2 value to a classmate.
(b) Now repeat parts (a-c) from supplement 1 for the case of 300 predictors and 200 observations. What happens?

Sec 7.1-7.4, 7.8.1, 7.8.2

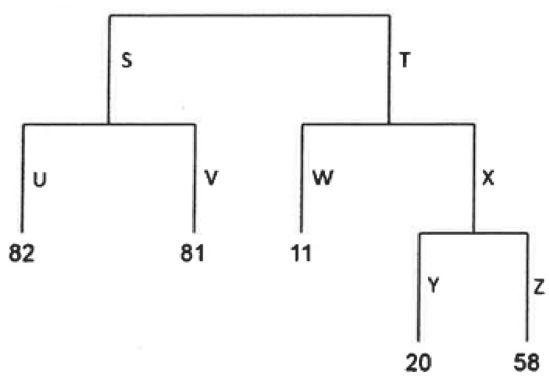
Supplemental Question:
For a given data set you want to perform a piecewise polynomial regression with knots at $X = \{10, 20, 30, 40\}$. For each of the following, determine the number of degrees of freedom in the model. Justify your work with a computation or brief explanation for each.
(a) Piecewise Linear Regression (b) Piecewise Cubic Regression.
(c) Linear Spline (d) Cubic Spline (e) Natural Cubic Spline

Sec 7.5-7.7, 7.8.2, 7.8.3

Supplemental Question :
(a) In local regression, if you made a graph of training MSE versus span what would you expect the graph to show? Briefly justify your answer.
(b) Repeat part (a) for test MSE.

Sec 8.1, 8.3.1, 8.3.2

Supplemental Question (Modified CAS):



You are given the unpruned decision tree on the left. The values at each terminal node are the residual sums of squares at that node. If the tree was pruned at S the RSS at S would be 251. If the tree was pruned at T the RSS at T would be 209. If the tree was pruned at X the RSS at X would be 86. The RSS for the model with only one node is 486. You use the cost complexity pruning algorithm with the tuning parameter, α, equal to 9 in order to prune the tree. What tree results from applying the algorithm?

Teamwork Formal Presentation and Submission Problem 2

Notes about Exercise 6.11
Use a validation set approach to determine which of your models works best. This is in addition to any CV used to fit a particular model. For best subset selection you can use a statistic to fit the model rather than CV.
For your R code, submit your work for fitting your various models and for computing the test MSE for each.

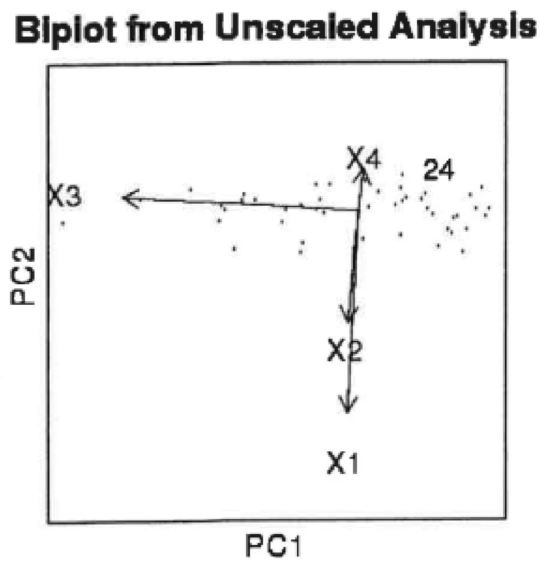
For your written work, write a two paragraphs, each describing why you selected the model that you did. The first should be aimed at a non-technical audience (think of your audience as a criminal justice major interested in crime rates but with no experience with regression). The second should be aimed at a technical audience (think of your audience as another member of the class).
Then give a brief answer to the question in part (c).
You do NOT need to give a separate written answer to the question in part (a).

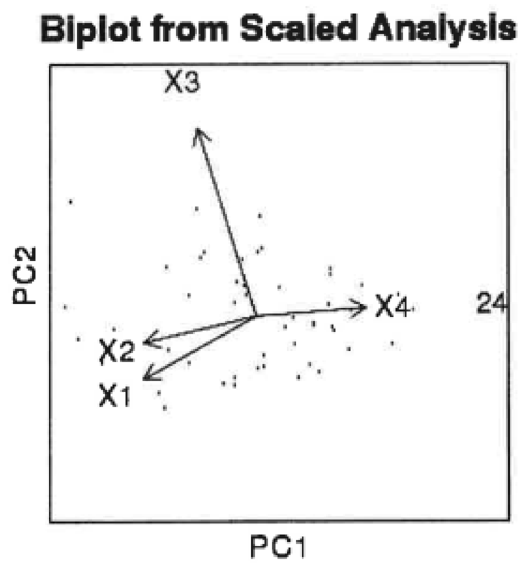
<u>Supplemental Question 1</u>
Complete problem 8.10 from the text.
For part (e), first use linear regression with a subset of predictors chosen using one of the methods from chapter 6 (use an appropriate statistic to determine the size of the subset to select, NOT CV). Then, using this subset of predictors, fit some sort of model from chapter 7. Find the test MSE for your two models.
After completing part (g), decide which one of the models you would select to predict Salary.
For your R code, submit your work for fitting your various models and for computing the test MSE for each.
For your written work, write a two paragraphs, each describing why you selected the model that you did. The first should be aimed at a non-technical audience (think of your audience as a baseball fan with no background in college-level mathematics or statistics) and should include an appropriate description of how your model works. The second should be aimed at a technical audience (think of your audience as another member of the class).

Sec 12.4, 12.5

Supplemental Question 1 (CAS):

You perform two separate principal component analyses on the same four variables in a particular data set: X1, X2, X3, and X4. The first analysis centers but does not scale the variables, and the second analysis centers and scales the variables. The biplots of the first two principal components produced from these analyses are shown below. The location of Observation 24 is labeled on the plots as well.





Decide if each of the following statements is demonstrated in the biplots or not. Briefly explain your answers.

(a) $X1$ is more highly correlated with $X2$ than with $X3$.

(b) $X3$ has the highest variance of these four variables.

(c) Observation 24 has a relatively large, positive value for $X4$.

Supplemental Question 2:

Decide if each of the following statements is true or false. Briefly justify your answers.

(a) The proportion of variance explained by an additional principal component increases as more principal components are added.

(b) The cumulative proportion of variance explained increases as more principal components are added.

(c) Using all possible principal components provides the best understanding of the data.

(d) A scree plot provides a method for determining the number of principal components to use.

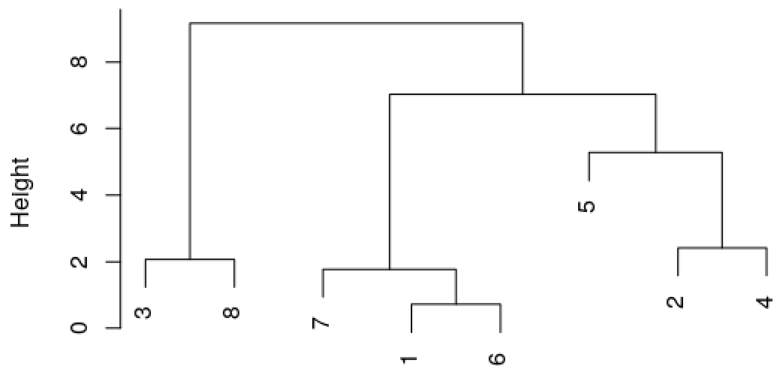
Notes about Question 12.8:

Formula (12.8) assumes that the data is centered, so I recommend using centered and scaled data throughout the problem. The scale() function can do this quite quickly. To save a bit of work in part (b), instead of getting the principal component loadings from the output of prcomp(), instead get the principal component scores from the output of prcomp(). You then will need to recognize what in formula (12.8) corresponds to the principal component scores.

Sec 12.4, 12.5 Part2

Supplemental Question:

Consider the dendrogram shown on the right. Give the sequence of sets of clusters that can be formed by cutting this dendrogram.



Sec TS: 1.1, 1.2, 1.3

Supplemental Question 1:

The data below are the monthly number of copies of a textbook sold over the past two years.

Month	J	F	M	A	M	J	J	A	S	O	N	D
Year 1	52	18	16	14	16	16	18	50	20	16	18	20
Year 2	60	22	17	19	21	23	25	60	28	22	24	28

You compute a centered, year-long moving average of the number of copies sold. Since a year has 12 months, which is even, your moving average will weight the current month and the five months before and after the current month equally, and then give half weight to the month six months before the current month and the month six months after the current month.

(a) Compute the centered, year-long moving average of the number of copies sold for December of year 1 and for January of year 2.

(b) Give an example of a question that the company might be interested in for which the centered, year-long moving average would be useful data. Then give an example of a question that the company might be interested in for which the centered, year-long moving average would not be useful.

Supplemental Question 2:

I would recommend completing Problem 1.3 before doing this problem.

(a) For each of the three times series x_t in Problem 1.3 parts (a,b,c), compute the mean

functions $\mu_x(t)$. For the series in part (a) you can assume $x_1 = w_1$ and $x_2 = w_2$.
(b) For each of the three moving average time series v_t in Problem 1.3 parts (a,b,c), compute the mean functions $\mu_v(t)$.

Supplemental Question 3:
In the reading we looked at smoothing a data series using a moving average. Another possibility would be **exponential smoothing**. Exponential smoothing uses a weighted average of all of the past values of a series. Let w be a weight between 0 and 1. Then for a time series x_t the exponential smoothed estimate of the time series value at time t is

$$\frac{x_t + wx_{t-1} + w^2x_{t-2} + w^3x_{t-2} + \cdots}{1/(1-w)}$$

Note that since $w \in (0,1)$ the sum of the coefficients $1 + w + w^2 + w^3 + \cdots = 1/(1-w)$. Thus since we also divide the entire expression by $1/(1-w)$ the sum of the coefficients is one and we are taking a weighted average of all of the previous time series values. Also since $w \in (0,1)$ the higher powers of w are smaller, so the average gives more weight to the more recent time series values.

However, this formula requires an infinite number of past values of the time series, which we almost never have! Therefore, in practice we cut off the series in the numerator at some initial time zero and define the exponential smoothed estimate of the time series to be

$$\widehat{s}_t = \frac{x_t + wx_{t-1} + w^2x_{t-2} + w^3x_{t-2} + \cdots + w^tx_0}{1/(1-w)}$$

This is no longer a true moving average, but is close enough to work well in practice.

(a) [Does not use R] Show that for a time series x_t and a corresponding (truncated) exponential smoothed estimate $\widehat{s}_t$, the following recursive relationship holds:

$$\widehat{s}_t = (1-w)x_t + w\widehat{s}_{t-1}$$

(b) [Does not use R] In a couple of sentences, explain how the value of w affects the exponential smoothed estimate. How is the estimate different if w is close to 1 as opposed to if w is close to 0?
(c) Consider the time series x_t in Exercise 1.3 part (c). Use the result from part (a) above along with the filter function to compute the exponential smoothed estimate of x_t. Compute the estimate for a few different values of w and plot your results. Do your results agree with what you would expect based on your answer to part (b)?

Sec TS: 1.3, 1.4

Supplemental Question:
(a) Compute the autocovariance function $\gamma_x(s,t)$ for the time series in Problem 1.3(b).
(b) Compute the autocovariance function $\gamma_x(s,t)$ for the time series in Problem 1.3(c).

Sec TS: 3.1

Supplemental Question 1:
Consider the AR(2) model $x_t = \frac{3}{4}x_{t-1} - \frac{1}{8}x_{t-2} + w_t$.
(a) Identify the autoregressive operator $\phi(B)$ for this model.
(b) Use the idea of matching coefficients to find the first three coefficients in the linear process representation of this times series, $x_t = \psi(B)w_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$.
(c) Find a recursive formula for ψ_j in terms of ψ_{j-1} and ψ_{j-2} for $j > 2$.

Supplemental Question 2:
Consider the ARMA(p,q) model $x_t = \dfrac{-5}{6}x_{t-1} - \dfrac{1}{6}x_{t-2} + \dfrac{8}{12}w_{t-1} + \dfrac{1}{12}w_{t-2} + w_t$.
(a) Write this model in the form $\phi(B)x_t = \theta(B)w_t$ where $\phi(B)$ is the AR operator and and $\theta(B)$ is the MA operator.
(b) Using your answer from part (a), identify parameter redundancy and then write the model as an ARMA(1,1) model in the form $x_t = \cdots$.

Supplemental Question 3:
(a) Simulate two different AR(1) series, both of length 100, one with $\phi = 0.9$ and one with $\phi = 0.4$. Use the ar function in R to fit an auto-regressive model to your simulated series. Do you get reasonable results?
(b) Simulate two different MA(1) series, both of length 100, one with $\theta = 0.9$ and one with $\theta = -0.9$. Use the ar function in R to fit an auto-regressive model to your simulated series. What happens?
(c) For the $\phi = 0.9$ model from part (a) and the $\theta = 0.9$ model from part (b), simulate 100 different times series, each of length 100, and fit an auto-regressive model to each. For each fitted model, record the estimate of the first coefficient. Then compute the mean and sd of the first coefficient estimates for each of the two models. What do you notice?

Sec TS: 3.1, 3.3

Supplemental Question:
Consider the ARMA(1,2) model $x_t = \frac{-1}{3}x_{t-1} - \frac{1}{4}w_{t-2} + w_t$.
(a) State why this model is causal. Then use coefficient matching to find the first four terms in the linear process representation of this times series, $x_t = \psi(B)w_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$. Check your answer using R.
(b) State why this model is invertible. Then use coefficient matching to find the first four

terms in the invertible representation of this times series, $w_t = \pi(B)x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}$. Check your answer using R.

Sec 3.4

Supplemental Question 1:

Consider a stationary AR(1) model $x_t = \beta_0 + \beta_1 x_{t-1} + w_t$.

(a) Choose non-zero values for β_0 and β_1. Let the variance of w_t be 1. Then simulate 100 values of for the AR(1) model with your parameters. For this problem, do not use the arima.sim function. Instead, simulate your values recursively.

(b) Use regression of x_t on x_{t-1} to estimate the values of β_0 and β_1 for your simulated data.

(c) Plot the ACF of the residuals. What do you observe?

(d) From the output of your regression, estimate the value of $V(w_t)$ for your model.

(e) When we previously studied regression we showed that the variance of the response variable equaled the variance of the errors. This is not the case for regression on a lagged variable. Find the relationship between the variance of x_t and $V(w_t)$ in terms of the β_i for the model in this problem.

(f) Use the arima function in R to fit an AR(1) model to the simulated data. The estimate of β_1 should be close to the true value but the estimate of the intercept won't be close to β_0. Use the documentation for the arima function to explain why this difference occurred.

Supplemental Question 2: (Adapted from Cowpertwait)

(a) Simulate 100 values of a time series $x_t = 70 + 2t - 3t^2 + z_t$ where z_t is the AR(1) process $z_t = 0.5z_{t-1} + w_t$ and w_t is white noise with SD 25.

(b) Use least squares to fit a quadratic model to the series x_t. Give the coefficient estimates and the SE's of the coefficient estimates.

(c) Plot the ACF of the residuals. What do you observe?

(d) The method of generalized least squares can be used to better deal with the autocorrelation in the residual series. This method maximizes likelihood while taking into account the autocorrelation in the data. In R, the gls function in the nlme library can be used to carry out generalized least squares. Note that you need to specify the correlation structure that you are assuming, in our case this is corAR1(0.5). Use generalized least squares to fit a quadratic model to the series x_t. Give the coefficient estimates and the SE's of the coefficient estimates.

(e) How do your values from (b) and (d) compare? Give a conceptual justification for why one of the sets of SE estimates is larger than the other.

(f) Plot the ACF of the residuals for the gls regression. What do you observe?

Supplemental Question 3:

Given an AR(p) model and values for a time series up to the present we will often want to predict what the future values of the time series will be. This is most often done recursively using the chain rule of forecasting. For an AR(p) model $x_t = \beta_0 + \sum_{i=1}^{p} \beta_i x_{t-i} + w_t$, assume that we know the values of x_t for $t \leq T$. We want forecasts $\hat{x}_{T+k}$ for $k > 0$. For convenience, let $\hat{x}_t = x_t$ for $t \leq T$. Then define the k-step ahead forecast $\hat{x}_{T+k} = \beta_0 + \sum_{i=1}^{p} \beta_i \hat{x}_{T+k-i}$.

(a) Consider the AR(2) process $x_t = 0.5x_{t-1} + 0.2x_{t-2} + w_t$. You know that $x_1 = 100$, $x_2 = 300$, and $x_3 = 200$. Forecast x_4, x_5, and x_6.

(b) What will happen to your forecast for x_t as t gets very large? Briefly explain why this makes sense.

(c) Consider an AR(1) process $x_t = \beta_0 + \beta_1 x_{t-1} + w_t$. Assuming that β_0 and β_1 are known, what is the variance of the one step ahead forecast error $x_{T+1} - \hat{x}_{T+1}$?

(d) Use induction to show that the variance of the k-step ahead forecast error $x_{T+k} - \hat{x}_{T+k}$ is $\sigma_w^2(1 + \beta_1^2 + \cdots + \beta_1^{2(k-1)})$.

(e) Now assume that you have 100 observations that you know are from an AR(1) process but for which you know nothing about the model parameters. You first estimate the model parameters and then make a k-step ahead forecast. You want to create a prediction interval fo the true x_{T+k} value. Ignore the error due to having to estimate the parameters. You assume that the true value follows a t-distribution with mean the forecast and SD given by part (d). How many degrees of freedom should you use?

Supplemental Question 4:

We now use R to implement some of the estimation and forecasting described in the previous exercises. Consider the recruitment data set rec from Example 1.5.

(a) Create an ACF and PACF for the sample data. Based on these graphs, briefly explain why an ARMA(p,0) model might be appropriate and what value of p should be used.

(b) In R we can use the arima function to fit an ARMA model (and later more general ARIMA models) to a data set. The syntax is similar to the arima.sim function that we have used in the past. Use this function to fit the ARMA(p,0) model you chose from part (a) as well as ARMA(p-1,0), ARMA(p+1,0), ARMA(0,2), and ARMA(p,2) models. Does the model you picked have the smallest AIC.

(c) We can apply the predict function to an arima object in R to get k-step ahead predictions and the corresponding forecast SEs. See ?predict.Arima for more details. Use R to forecast the next two years of values for the recruitment series using your AR(p) model.

(d) Plot the ACF of the residuals of your model. What do you notice?

(e) Make a plot like the one on page 110 showing the observed data for this time series along with the forecast values and the forecast interval. Hint: See the code on page 111. Be sure to understand what the code is doing!

(f) Repeat parts (c,d,e) for the ARMA(0,2) model.

Teamwork Formal Presentation and Submission Problem 3

Supplemental Question 1:

Consider the global temperature data set from Example 1.2 on page 3.

(a) Fit an ARMA(2,1) model to the data.

(b) Use your fitted model to predict the temperature values for the next 100 years.

(c) Plot the past data and the predictions on the same graph in different colors. Then comment on how well you think your model performed.

Supplemental Question 2:

Consider the time series $x_t = \frac{5}{2}x_{t-1} - 2x_{t-2} + \frac{1}{2}x_{t-3} + w_t$ where w_t is white noise with $\sigma_w^2 = 1$.

(a) Is x_t causal? Show your work.

(b) Use seed(3) to simulate 1000 random values of x_t.

(c) Fit an AR(3) model to all of the 1000 values of x_t and then to just the first 100 values of x_t. Your intercept should be non-zero and you may get a warning message. Why does this occur?

(d) Let $y_t = \nabla x_t$. Calculate the 999 simulated values of y_t corresponding to your simulated values of x_t.

(e) Find a formula for y_t in terms of y_{t-1} and y_{t-2} and the w_i.

(f) Use R to fit an appropriate AR(p) model to y_t. You likely will still get a warning message. Why?

(g) Let $z_t = \nabla^2 x_t = \nabla y_t$. Calculate the 998 simulated values of z_t corresponding to your simulated values of x_t.
(h) Find a formula for z_t in terms of z_{t-1} and the w_i.
(i) Use R to fit an appropriate AR(p) model to z_t.
(j) Find an approximate 95% confidence interval for the parameter(s) in your model. How does your confidence interval compare to the parameter value(s) you expected?

Supplemental Question 3:
Consider the data set ARMA available on Canvas. This data shows the historical daily values of air quality in a classroom (higher values are better and an average value is 0).
Fit an appropriate ARMA(p,q) model to this data.

For your R code, submit your work for how you determined which model to use and how you fit the model that you picked.
For your written work, write a two paragraphs.
The first paragraph should describe how you picked the model you chose and should be aimed at a technical audience (think of your audience as another member of the class).
The second paragraph should describe how your model of the time series works and should be aimed at a non-technical audience (think of your audience as a fellow student interested in how air quality varies and who has no background in college-level mathematics or statistics).

Sec TS: 3.9, 5.3, Sup

Supplemental Question:
(a) Write out the difference equation form for an $\text{ARIMA}(1,0,1) \times (0,2,0)_4$ model.
(b) Identify the following as an $\text{ARIMA}(p,d,q) \times (P,D,Q)_s$ model with d and D both non-zero: $x_t = x_{t-1} + x_{t-10} - x_{t-11} + w_t + 0.4w_{t-1} - 0.4w_{t-10} - 0.16w_{t-11}$.