Midterm - Stat 4840/5084

Name ______________________________

Thursday, March 6

Write directly on this exam. You may "show work" by handing in an R script, .Rmd file, or knit Markdown document.

You may use R, the internet, and any reference material. You are not allowed to communicate with anyone - no email, messaging, internet forums, AI, etc. If you happen to find exact copies of the exam questions on a "homework help" website, please bring that to the instructor's attention.

Problem 1 (10 points)

Consider a random walk model $y_{t+1} = y_t + \epsilon_t$ where ϵ_t is white noise with variance 5. Suppose $y_0 = 8$.

a. What is $E[y_3]$?
b. What is $\text{Var}(y_3)$?

Solution

a. 8
b. 15

Problem 2 (10 points)

Consider a constant mean model $y_t = \mu + \epsilon_t$ where $\mu = 10$ and ϵ_t is white noise with variance 5. Suppose $y_0 = 8$.

a. What is $E[y_3]$?
b. What is $\text{Var}(y_3)$?

Solution

a. 10
b. 5

Problem 3 (10 points)

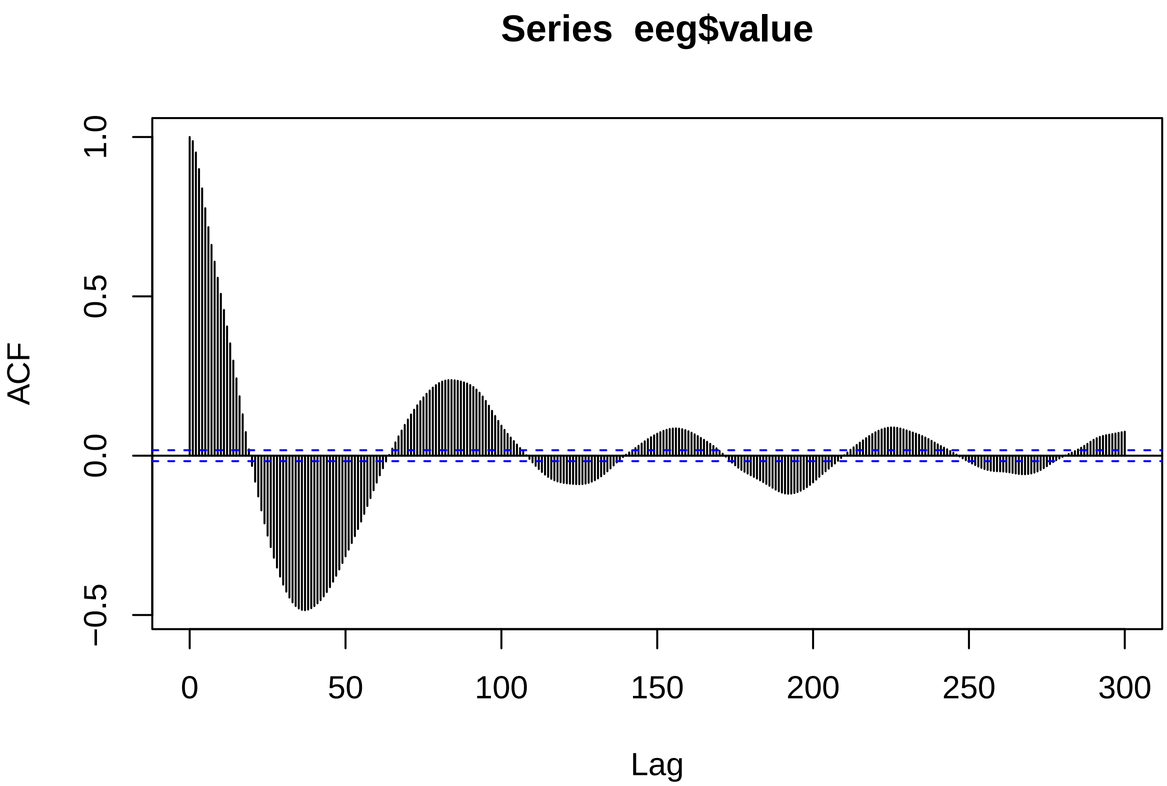
An electroencephalogram (EEG) is a record of electrical activity in the brain. The data `eeg.csv` is an EEG of a patient having a seizure, recorded at 256 values per second.

Load the data from our course data page and make an autocorrelogram with at least 300 lags.

What can you say about this EEG series by looking at its autocorrelation?

Solution

```
# data is from Cryer and Chan, see help on eeg from the TSA library.
eeg <- read.csv("https://turtlegraphics.org/timeseries/data/eeg.csv")
acf(eeg$value, lag.max = 300)
```



The series appears to be approximately periodic with period around 75. That's 1 cycle per 75/256 seconds or about 3.4Hz.

Problem 4 (10 points)

The `prices` data in the `fpp3` package has the annual price of copper ($/lb) for most of the 20th century.

Filter the data to use only recent prices from the year 1920 onward and make a time plot of the series.

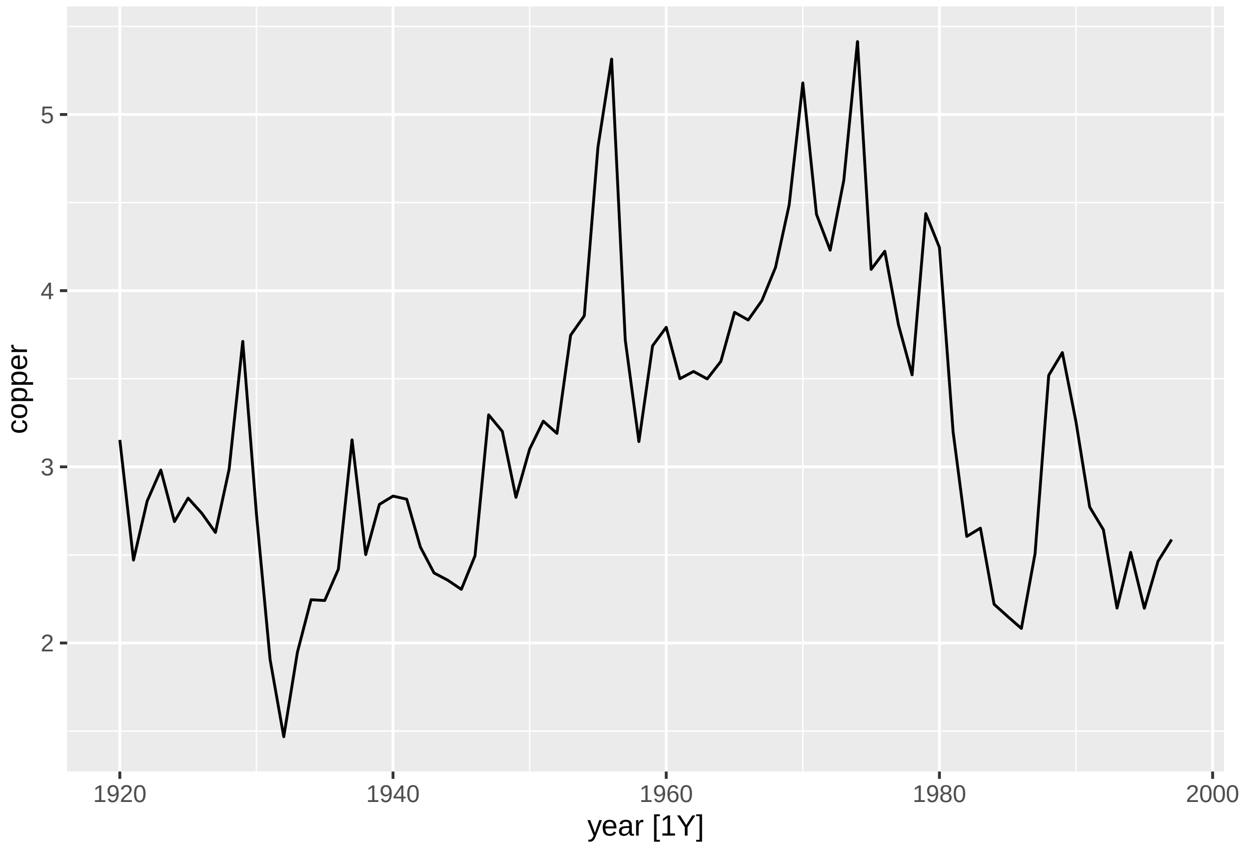
Just looking at your plot (no need for precise answers):

a. When was copper cheapest and about how much did it cost per pound?
b. When was copper expensive and about how much did it cost per pound?

Solution

Cheap around 1932, low price of $1.50 a pound. Most expensive in 1955, 1970, 1975 when it was over $5 a pound.

```
copper_prices <- prices |> filter(year >= 1920) |> select(year, copper)
copper_prices |> autoplot(copper)
```



Problem 5 (10 points)

Continue with copper pricing from 1920-1997. Fit a naive/random walk model $y_{t+1} = y_t + \epsilon_t$ to the price series. The white noise term ϵ_t has variance σ^2.

What is the model's estimate of σ^2?

Solution

```
copper_prices |> model(MEAN(copper)) |> report()
```

```
## Series: copper
## Model: MEAN
##
## Mean: 3.1917
## sigma^2: 0.717
```

Problem 6 (10 points)

Continue with your naive model for copper pricing, and investigate the residuals.

a. Do the residuals appear unbiased?
b. Are the residuals homoscedastic?
c. Run a Ljung-Box test with 15 lags to test if the residuals are autocorrelated. What p-value do you get?
d. Overall, would you say these residuals look like white noise?

Solution

a. Yes (in fact the mean is -0.007)

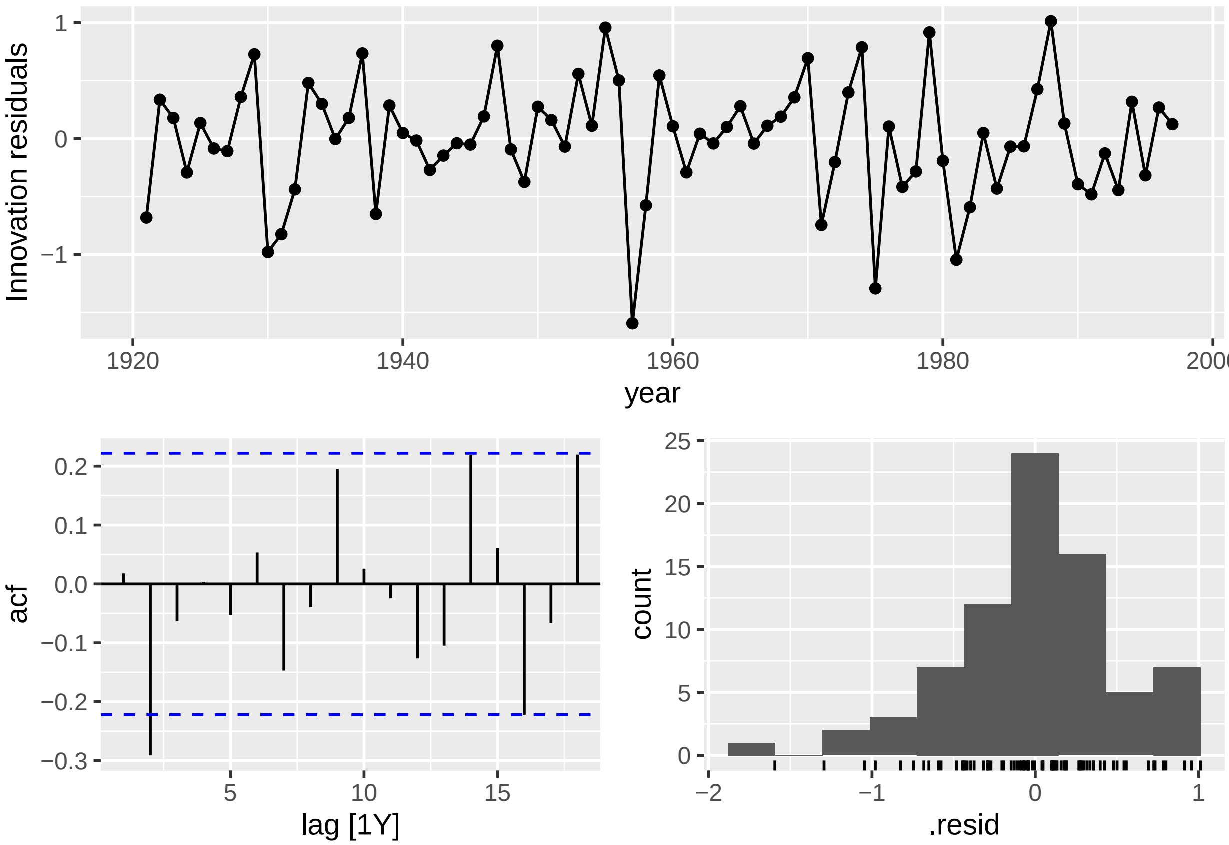
b. Yes, variance seems constant over time.
c. $p = 0.14$, no significant autocorrelation.
d. The only concern is some autocorrelation at lag 1. The model seems satisfactory but could possibly be improved to account for this.

```
copper_prices |> model(NAIVE(copper)) |> gg_tsresiduals()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_line()`).

## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).

## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_bin()`).
```




```
copper_prices |> model(NAIVE(copper)) |> augment() |> features(.resid, ljung_box, lag=15)
```

```
## # A tibble: 1 x 3
##   .model        lb_stat lb_pvalue
##   <chr>           <dbl>     <dbl>
## 1 NAIVE(copper)    20.8     0.145
```

Problem 7 (10 points)

Using your naive model for copper pricing, forecast the price of copper in 2025. Give the mean forecast and the 95% confidence interval.

Solution Estimate is 2.587, 95% CI is [-2.58, 7.75].

```r
copper_prices |> model(NAIVE(copper)) |> forecast(h=30) |> filter(year == 2025) |> hilo()
```

```
## # A tsibble: 1 x 6 [1Y]
## # Key:       .model [1]
##   .model  year      copper
##   <chr>  <dbl>      <dist>
## 1 NAIVE~  2025 N(2.6, 6.9)
## # i 3 more variables: .mean <dbl>, `80%` <hilo>, `95%` <hilo>
```

Problem 8 (10 points)

Your forecast interval in Problem 7 included negative copper prices, which are impossible. What could you do to ensure your forecast intervals never go negative?

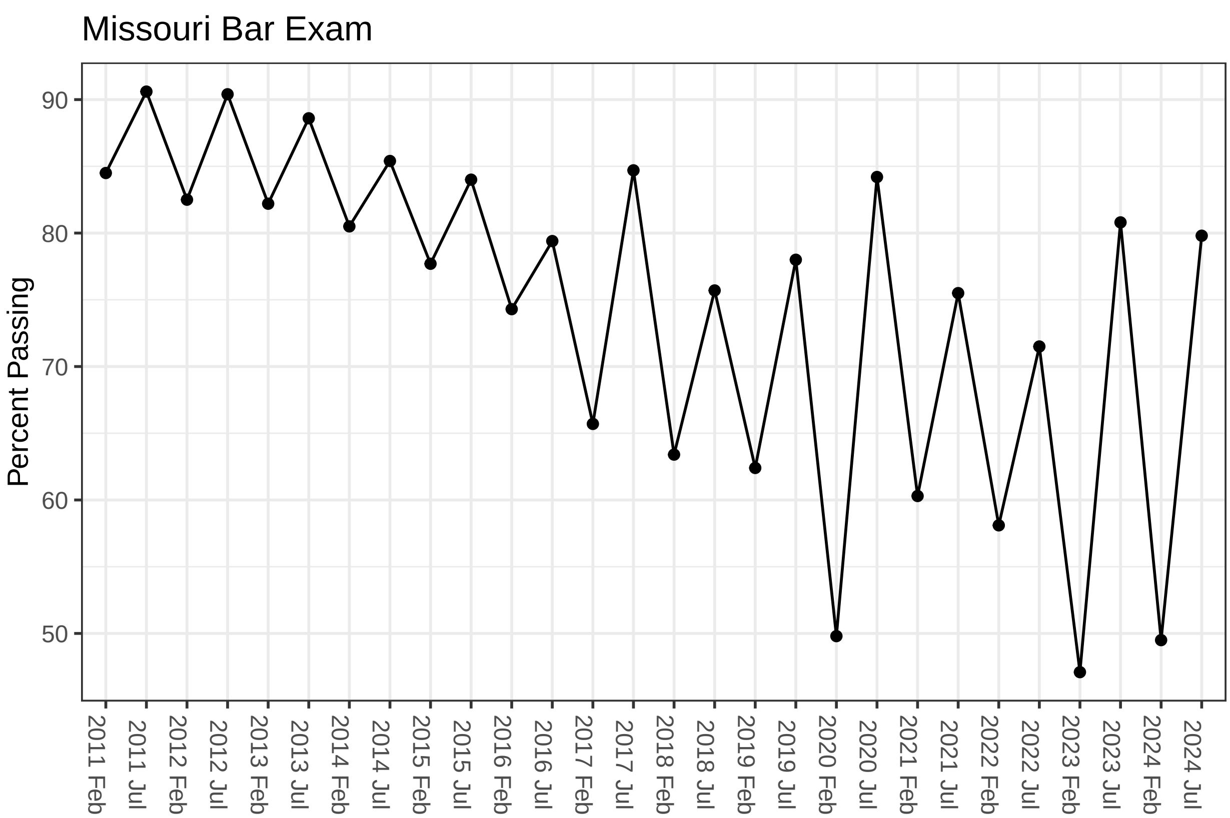
Solution Apply a log transformation to the price series when fitting the model.

Problem 9 (10 points)

To become a lawyer in the state of Missouri, you need to pass the bar exam, which is offered every February and July.

The data `MO-bar-pass-rates.csv` on our course data page contains the percentage of people that passed the bar exam since 2011.

Make the PctPass into a base R time series with frequency 2, and perform a classical `decompose()` into trend, seasonal, and random components.

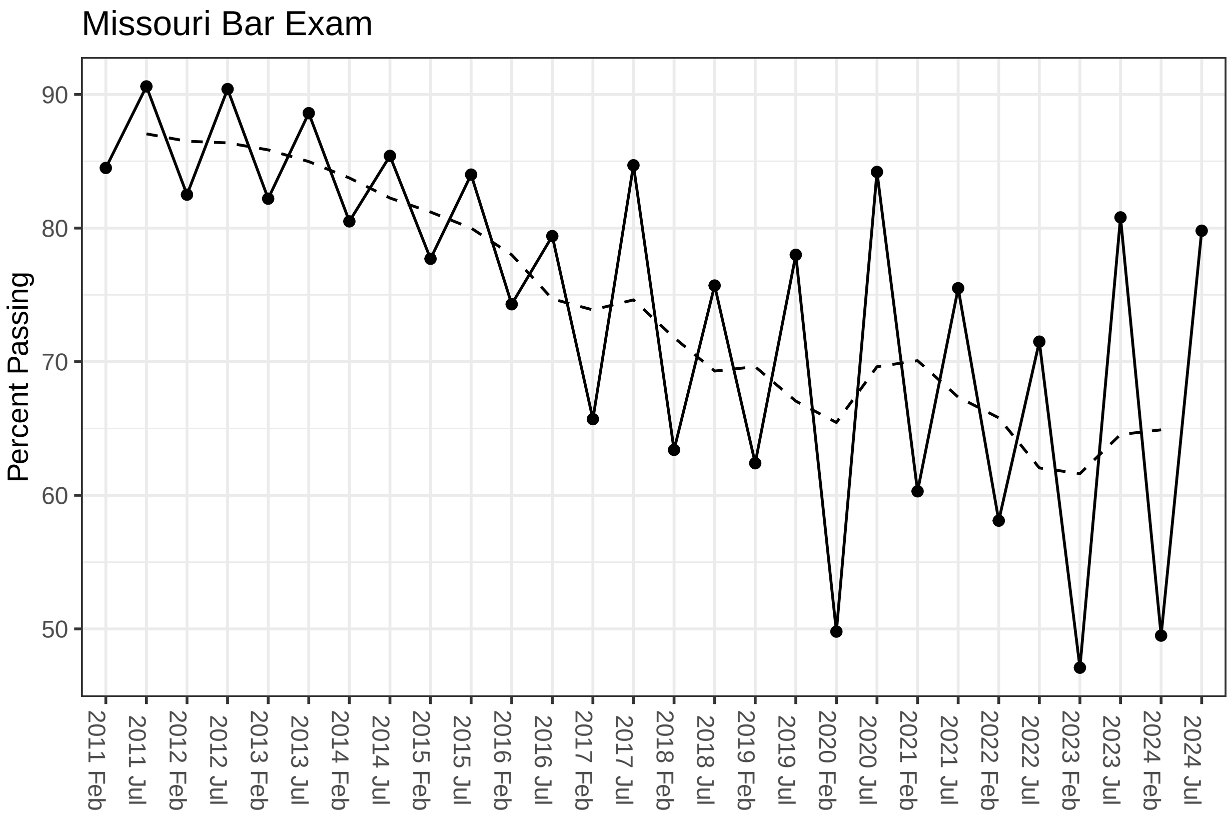


a. Sketch the trend on the graph above.

b. What is the seasonal difference between February and July pass rates?

Solution

```r
mobar |> mutate(trend = stats::filter(PctPass, c(1/4,1/2,1/4))) |>
  ggplot(aes(x=Date, y=PctPass)) + geom_point() + geom_line(aes(group=1)) +
  geom_line(aes(group=1, y=trend), linetype='dashed') +
  labs(x="", y="Percent Passing",title="Missouri Bar Exam") +
  theme_bw() + theme(axis.text.x = element_text(angle = -90))
```




```r
decompose(ts(mobar$PctPass, frequency = 2))$seasonal
```

```
## Time Series:
## Start = c(1, 1)
## End = c(14, 2)
## Frequency = 2
##  [1] -7.840385  7.840385 -7.840385  7.840385 -7.840385  7.840385 -7.840385
##  [8]  7.840385 -7.840385  7.840385 -7.840385  7.840385 -7.840385  7.840385
## [15] -7.840385  7.840385 -7.840385  7.840385 -7.840385  7.840385 -7.840385
## [22]  7.840385 -7.840385  7.840385 -7.840385  7.840385 -7.840385  7.840385
```

Problem 10 (10 points)

a. Find the 3-MA of the series $1, 1, -2, 1, 1, -2, 1, 1, -2, \ldots$

b. Give an example of a non-zero series with 5-MA equal to zero.

Solution

a. The 3-MA is entirely zeros.
b. There are many ways, but following part (a) you could do $1, 1, 1, 1, -4, 1, 1, 1, 1, -4, 1, 1, 1, 1, -4, \ldots$