Christen May
Mason Pearce
MW 11:00am

# March 31st

## peakdemand.csv

Spent 30 minutes working on the following problem to finish up modeling variation in time (period and seasonality)

```
library(mosaic)

# we see a cyclical as well as an overall increasing effect)
plot(peakdemand$PeakDemand ~ peakdemand$period, data = peakdemand)


# temperature smile - at the extremes of temperature (parabola characteristic)
# super hot and super cold we use more energy (light and heat and ac - maybe)
plot(peakdemand$PeakDemand ~ peakdemand$DailyTemp, data = peakdemand)

# now let's look at what we think might be affecting the trend
# there is definitely a seasonal effect => something that looks like a sine wave
# maybe once we've adequately adjusted for temperature, maybe we will no longer see any seasonality
# how do we get rid of cyclical things? - dummy variables

# 80% of the time we find the trend and find the seasonal effects to get the answer
# we don't have enough time to cover the other 20%

# Trend:
# add a time index
N = nrow(peakdemand)
peakdemand$period = 1:N
head(peakdemand)


lm1 = lm(peakdemand$PeakDemand ~ peakdemand$period, data= peakdemand)
plot(resid(lm1))
coef(lm1)


# replot the trend to see how this linear model fits
plot(peakdemand$PeakDemand ~ peakdemand$period, data = peakdemand)
abline(lm1, col = 'red')

# add temperature and temperature squared (quadratic model - temp smile) and dummy variables for Sat and Sun
```

Christen May
Mason Pearce
MW 11:00am

```
lm2 = lm(peakdemand$PeakDemand ~ peakdemand$period + peakdemand$DailyTemp +
I(peakdemand$DailyTemp^2) + peakdemand$Sat + peakdemand$Sun, data = peakdemand)

# now we compare the fit
# we first plot the residual model of demand versus the period
# then we look at the residuals of the model with the temperature and Saturday and Sunday included
# you can see that lm2 doesn't cover the lines made by the lm1 plot (what's attributed to period)
plot(resid(lm1), type = 'l')
lines(resid(lm2), col = 'red')

#does month have any affect?
lm3 = lm(peakdemand$PeakDemand ~ peakdemand$period + peakdemand$DailyTemp +
I(peakdemand$DailyTemp^2) + peakdemand$Sat + peakdemand$Sun + factor(peakdemand$Month),
data = peakdemand)
plot(resid(lm3), type = 'l')
summary(lm3)
anova(lm3)

lmstep = step(lm3, direction = 'backward')

# we look at the r^2 of the summary to see which predictor is best to remove
# removing any of them worsens the model, so we're at the end
# we stop here and say that's as far as we go


# compare fits
# side note - type = 'l' = line graph rather than dot plot
plot(peakdemand$PeakDemand ~ peakdemand$period, data = peakdemand, type = 'l')
lines(fitted(lm3) ~ peakdemand$period, data=peakdemand, col='red')

# you can see that the red almost completely covers the black, so this model fits much better

# now let's talk about what is being left on the table
# our data seems to bow over time (higher in the middle than the edges) - very very slight
# and we've fit a linear model - here it's okay because it's a super small bow
# but maybe in another scenario you might want to use quadratic

# what's the best tool to predict temp tomorrow => temp today
# lagged predictors - autoregressive (predicting peak demand today on peak demand yesterday) => this
is in the notes, but not at all required info
```

# Probability

Rules: (Kolmogrorov's Axioms)

1. Probabilities must sum to one (mutually exclusive)
    a. Jump off a cliff ➔ live or die (no in between)

Christen May
Mason Pearce
MW 11:00am

2. Probabilities for disjoint events are added together
   a. Disjoint event = can't both happen
   b. Go to UT or go to OU
3. Probabilities are numbers between 1 and 0 (or percentages between 1 and 100)

More complex rules:

1. Addition rule – (union rule) –
   a. P(A u B) = P(A) + P(B) – P(A, B)
      (A or B) (A and B) – joint event (both occur)
      Probability that you're a lady or from dallas = probability that you're a lady + probability that you're from dallas – probability that you're a lady and from dallas
2. Multiplication Rule
   a. P(A,B) = P(A) * P(B|A)
      "|" = conditional upon (given)
      P(B|A) = Probablity of B given that A is true

"But what does it all mean?"

Definition:

1. Frequency Interpretation – (Vegas Interpretation) – probability is the long run limiting frequency – where it evens out after a lot of trials
   P(Black 31) = # times black 31 comes up / # spins of roulette wheel
   If you spin the wheel a million times and you count how many times you get black 31. Divide one by the other and that number is your limiting frequency and what we call our probability
2. Degree of Belief Interpretation – (Wall Street Interpretation) – (page 3 of this week's notes – you should read them! => don't have to know the math, but this validates this interpretation mathematically)
   How do you look outside and decide the probability that it will rain (this afternoon is a one-time event, so this doesn't work for frequency interpretation) => we look at behavior
   Example: Dr. Scott will pay you $100 if it rains – how much will you pay in order to be written this contract that pays out if the event comes true
      This causes you to put your money where your mouth is and that number is your personal belief of that likelihood
      How likely your phone will break in the next year => 80% (so if the phone costs 100 dollars would you be willing to pay 80 bucks for a contract – no maybe not drop that number) this helps you refine your personal belief about the likelihood of the event happening
      These contracts actually exist – prediction market

Christen May
Mason Pearce
MW 11:00am

**(you have to forget risk arbitrage (no riskless profit – you can't make money just based on discrepancies in value or anything like that)and net present value)**
This doesn't use any predictors or variables, it just uses public opinion – but if enough people are betting on it then it will actually have pretty legitimate forecasting ability

Bayes' Rule (Learning Rule) (Bayesian Updating rule)

Comes from the multiplication rule

$P(A,B) = P(A) * P(B|A)$
"|" = conditional upon (given)
$P(B|A)$ = Probablity of B given that A is true

We could just as easily flip the variables
$P(A,B) = P(A) * P(B|A) = P(B) * P(A|B)$

So we isolate one of these and solve for one variable
$P(A|B) = P(A) * P(B|A)/ P(B)$
$= P(A) (P(B|A)/P(B))$

Mathematical description of how we ought to update probabilities when we get new information $P(A|B)$ – probability that A happened given B

$P(A)$ – prior probability

$P(B|A)/P(B)$ – update

Next Time:

G = event that accused is guilty

D = accused DNA "matches" the DNA at the crime scene

$P(G) \rightarrow P(G|D)$