Seasonal Difference Exploration

Hao Zheng (hz2770)

5/6/2022

Explore the seasonal differences

```
load("./dt_long.RData")
load("./ID_in.RData")
load("./beta.res.postmean.RData")

dt_season <-
  dt_long %>%
  drop_na() %>%
  filter(ID %in% ID_in) %>%
  distinct(ID, .keep_all = TRUE) %>%
  select(ID, Season, Month, Nature) %>%
  mutate(Month = factor(Month, levels = month.name))
```

```
season_diff <-
  merge(dt_season, beta.res.postmean, by = c("ID")) %>%
  janitor::clean_names()

# Beta0
intercept.fit <- lm(intercept ~ month + season + nature, data = season_diff)
# Beta1
wind_prev.fit <- lm(wind_prev ~ month + season + nature, data = season_diff)
# Beta2
lat_change.fit <- lm(lat_change ~ month + season + nature, data = season_diff)
# Beta3
long_change.fit <- lm(long_change ~ month + season + nature, data = season_diff)
#Beta4
wind_change.fit <- lm(wind_change ~ month + season + nature, data = season_diff)

summary(intercept.fit)
```

```
##
## Call:
## lm(formula = intercept ~ month + season + nature, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43433 -0.03822  0.00234  0.04052  0.38709
##
## Coefficients:
```

```
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.4810021  0.3902677  11.482   <2e-16 ***
## monthApril      0.0232609  0.1113880   0.209   0.8346
## monthMay        0.0259813  0.0942017   0.276   0.7828
## monthJune       0.0275693  0.0922175   0.299   0.7651
## monthJuly       0.0125400  0.0918533   0.137   0.8914
## monthAugust    -0.0198034  0.0913669  -0.217   0.8285
## monthSeptember -0.0070528  0.0912856  -0.077   0.9384
## monthOctober    0.0093435  0.0913761   0.102   0.9186
## monthNovember   0.0145692  0.0924416   0.158   0.8748
## monthDecember   0.0057977  0.0976110   0.059   0.9527
## season         -0.0003419  0.0001895  -1.804   0.0717 .
## natureET        0.0008449  0.0298315   0.028   0.9774
## natureNR        0.0008122  0.0484835   0.017   0.9866
## natureSS        0.0141564  0.0205164   0.690   0.4904
## natureTS        0.0118370  0.0166932   0.709   0.4785
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09022 on 682 degrees of freedom
## Multiple R-squared:  0.03095,    Adjusted R-squared:  0.01105
## F-statistic: 1.556 on 14 and 682 DF,  p-value: 0.0866
```

```
summary(wind_prev.fit)
```

```
##
## Call:
## lm(formula = wind_prev ~ month + season + nature, data = season_diff)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.135790 -0.019433  0.002041  0.022272  0.061680
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     1.343e+00  1.215e-01  11.058  < 2e-16 ***
## monthApril      1.479e-02  3.467e-02   0.427 0.669679
## monthMay       -1.180e-04  2.932e-02  -0.004 0.996789
## monthJune       5.393e-03  2.870e-02   0.188 0.850987
## monthJuly       1.540e-02  2.859e-02   0.539 0.590174
## monthAugust     2.332e-02  2.843e-02   0.820 0.412418
## monthSeptember  2.610e-02  2.841e-02   0.919 0.358560
## monthOctober    2.108e-02  2.844e-02   0.741 0.458718
## monthNovember   2.461e-02  2.877e-02   0.856 0.392526
## monthDecember   8.824e-03  3.038e-02   0.290 0.771531
## season         -2.252e-04  5.899e-05  -3.817 0.000147 ***
## natureET        3.733e-03  9.284e-03   0.402 0.687709
## natureNR       -1.461e-02  1.509e-02  -0.969 0.333111
## natureSS       -3.330e-03  6.385e-03  -0.522 0.602172
## natureTS       -5.998e-03  5.195e-03  -1.155 0.248693
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02808 on 682 degrees of freedom
```

```
## Multiple R-squared:  0.0659, Adjusted R-squared:  0.04672
## F-statistic: 3.437 on 14 and 682 DF,  p-value: 1.998e-05
```

```
summary(lat_change.fit)
```

```
##
## Call:
## lm(formula = lat_change ~ month + season + nature, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90321 -0.07062  0.00781  0.07691  0.95935
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     4.131e-02  6.667e-01   0.062    0.951
## monthApril      1.656e-02  1.903e-01   0.087    0.931
## monthMay        7.088e-02  1.609e-01   0.440    0.660
## monthJune      -7.088e-03  1.575e-01  -0.045    0.964
## monthJuly      -9.091e-03  1.569e-01  -0.058    0.954
## monthAugust    -5.225e-02  1.561e-01  -0.335    0.738
## monthSeptember -3.611e-02  1.559e-01  -0.232    0.817
## monthOctober   -2.862e-02  1.561e-01  -0.183    0.855
## monthNovember   2.400e-02  1.579e-01   0.152    0.879
## monthDecember  -5.431e-02  1.668e-01  -0.326    0.745
## season          3.655e-05  3.238e-04   0.113    0.910
## natureET       -7.020e-02  5.096e-02  -1.378    0.169
## natureNR        5.897e-03  8.283e-02   0.071    0.943
## natureSS       -1.352e-03  3.505e-02  -0.039    0.969
## natureTS       -1.545e-02  2.852e-02  -0.542    0.588
##
## Residual standard error: 0.1541 on 682 degrees of freedom
## Multiple R-squared:  0.02561,    Adjusted R-squared:  0.005609
## F-statistic:  1.28 on 14 and 682 DF,  p-value: 0.2137
```

```
summary(long_change.fit)
```

```
##
## Call:
## lm(formula = long_change ~ month + season + nature, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.30817 -0.03599  0.00530  0.04273  0.50782
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.8336700  0.3531651  -2.361   0.0185 *
## monthApril      0.0416468  0.1007984   0.413   0.6796
## monthMay        0.0632772  0.0852459   0.742   0.4582
## monthJune       0.0556884  0.0834504   0.667   0.5048
## monthJuly       0.0361214  0.0831208   0.435   0.6640
## monthAugust     0.0123691  0.0826807   0.150   0.8811
```

```
## monthSeptember  0.0212965  0.0826071   0.258   0.7966
## monthOctober    0.0341549  0.0826890   0.413   0.6797
## monthNovember   0.0263450  0.0836532   0.315   0.7529
## monthDecember   0.0422468  0.0883312   0.478   0.6326
## season          0.0002184  0.0001715   1.273   0.2033
## natureET       -0.0263888  0.0269955  -0.978   0.3287
## natureNR        0.0030556  0.0438742   0.070   0.9445
## natureSS        0.0126339  0.0185659   0.680   0.4964
## natureTS       -0.0231521  0.0151062  -1.533   0.1258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08164 on 682 degrees of freedom
## Multiple R-squared:  0.05042,    Adjusted R-squared:  0.03093
## F-statistic: 2.586 on 14 and 682 DF,  p-value: 0.001201
```

```
summary(wind_change.fit)
```

```
##
## Call:
## lm(formula = wind_change ~ month + season + nature, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40181 -0.04476 -0.00309  0.04544  0.35691
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2.890e-01  3.809e-01   0.759    0.448
## monthApril      3.618e-02  1.087e-01   0.333    0.739
## monthMay       -1.629e-02  9.195e-02  -0.177    0.859
## monthJune       2.377e-02  9.001e-02   0.264    0.792
## monthJuly       1.308e-02  8.965e-02   0.146    0.884
## monthAugust     3.124e-02  8.918e-02   0.350    0.726
## monthSeptember  4.448e-02  8.910e-02   0.499    0.618
## monthOctober    3.505e-02  8.919e-02   0.393    0.694
## monthNovember   2.091e-02  9.023e-02   0.232    0.817
## monthDecember   1.142e-02  9.527e-02   0.120    0.905
## season          9.048e-05  1.850e-04   0.489    0.625
## natureET       -2.092e-02  2.912e-02  -0.719    0.473
## natureNR       -2.173e-02  4.732e-02  -0.459    0.646
## natureSS       -2.385e-02  2.003e-02  -1.191    0.234
## natureTS       -1.750e-02  1.629e-02  -1.074    0.283
##
## Residual standard error: 0.08806 on 682 degrees of freedom
## Multiple R-squared:  0.02104,    Adjusted R-squared:  0.000943
## F-statistic: 1.047 on 14 and 682 DF,  p-value: 0.404
```

```
sum0 <- summary(intercept.fit)$coefficients[,c(1,4)]
sum1 <- summary(wind_prev.fit)$coefficients[,c(1,4)]
sum2 <- summary(lat_change.fit)$coefficients[,c(1,4)]
sum3 <- summary(long_change.fit)$coefficients[,c(1,4)]
sum4 <- summary(wind_change.fit)$coefficients[,c(1,4)]
```

	Beta 0		Beta 1		Beta 2		Beta 3	
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)
(Intercept)	4.4810021	0.0000000	1.3431063	0.0000000	0.0413063	0.9506172	-0.8336700	0.0185275
monthApril	0.0232609	0.8346449	0.0147943	0.6696787	0.0165579	0.9306863	0.0416468	0.6796126
monthMay	0.0259813	0.7827813	-0.0001180	0.9967888	0.0708822	0.6597505	0.0632772	0.4581672
monthJune	0.0275693	0.7650618	0.0053935	0.8509869	-0.0070875	0.9641298	0.0556884	0.5047909
monthJuly	0.0125400	0.8914489	0.0154032	0.5901741	-0.0090910	0.9538180	0.0361214	0.6640154
monthAugust	-0.0198034	0.8284715	0.0233206	0.4124181	-0.0522548	0.7378961	0.0123691	0.8811234
monthSeptember	-0.0070528	0.9384385	0.0261005	0.3585599	-0.0361073	0.8169707	0.0212965	0.7966351
monthOctober	0.0093435	0.9185853	0.0210829	0.4587183	-0.0286163	0.8546050	0.0341549	0.6796975
monthNovember	0.0145692	0.8748155	0.0246144	0.3925264	0.0239972	0.8792681	0.0263450	0.7529105
monthDecember	0.0057977	0.9526542	0.0088244	0.7715305	-0.0543131	0.7447475	0.0422468	0.6326060
season	-0.0003419	0.0717253	-0.0002252	0.0001471	0.0000365	0.9101708	0.0002184	0.2032812
natureET	0.0008449	0.9774141	0.0037334	0.6877086	-0.0702038	0.1687975	-0.0263888	0.3286540
natureNR	0.0008122	0.9866387	-0.0146142	0.3331114	0.0058967	0.9432660	0.0030556	0.9444979
natureSS	0.0141564	0.4904257	-0.0033299	0.6021721	-0.0013517	0.9692484	0.0126339	0.4964264
natureTS	0.0118370	0.4785102	-0.0059979	0.2486925	-0.0154533	0.5880814	-0.0231521	0.1258337

```
kable(cbind(sum0, sum1, sum2, sum3, sum4)) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed")) %>%
  add_header_above(c(" " = 1, "Beta 0" = 2, "Beta 1" = 2, "Beta 2" = 2, "Beta 3" = 2, "Beta 4" = 2))
```

Now based on the estimated Bayesion model from previous questions, we need to explore the seasonal difference. We can fit the 5 estimated beta values against the three predictors: $X_{i,1}$: the month of the year the ith hurricane started, $X_{i,2}$:the year of the ith hurricane and $X_{i,3}$: the nature of the ith hurricane. The beta values obtained from previous Gibbs Sampler MCMC method contains the mean value of $\beta_{0,i}$, $\beta_{1,i}$, $\beta_{2,i}$, $\beta_{3,i}$ and $\beta_{4,i}$ for each of the 697 unique hurricanes.

According to the summary, the R squared value for all the five fitted linear models are quite small, which may indicate bad fit. In addition, most coefficients for the model are not significant. However, for those significant coefficients, we could infer a potential relationship between the certain predictors and the beta values. For the intercept $\beta_{0,i}$,only the intercept for the fitted model is significant. As for $\beta_{1,i}$, the value for the wind speed at time t, the intercept and the season predictor are both quite significant. Therefore, as year increase, the coefficient for $Y_{i,t}$ may decrease a little. For $\beta_{2,i}$, $\beta_{3,i}$ and $\beta_{4,i}$ indicating the latitude, longitude and wind speed change, none of the fitted coefficients are quite significant.

In conclusion, for different months, there is no significant differences observed. But for different year, the effect the current wind speed has on the next wind speed may decrease for a small amount.

```
# Try to fit the beta model only with the variable season(which indicates the year)
# Beta0
intercept.fit.new <- lm(intercept ~ season, data = season_diff)
# Beta1
wind_prev.fit.new <- lm(wind_prev ~ season, data = season_diff)
# Beta2
lat_change.fit.new <- lm(lat_change ~ season, data = season_diff)
# Beta3
long_change.fit.new <- lm(long_change ~ season, data = season_diff)
#Beta4
wind_change.fit.new <- lm(wind_change ~ season, data = season_diff)

summary(intercept.fit.new)
```

```
##
## Call:
## lm(formula = intercept ~ season, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.43803 -0.03487  0.00150  0.04113  0.37030
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.5142875  0.3580505  12.608   <2e-16 ***
## season      -0.0003543  0.0001803  -1.965   0.0498 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09053 on 695 degrees of freedom
## Multiple R-squared:  0.005526,   Adjusted R-squared:  0.004095
## F-statistic: 3.862 on 1 and 695 DF,  p-value: 0.04979
```

```
summary(wind_prev.fit.new)
```

```
##
## Call:
## lm(formula = wind_prev ~ season, data = season_diff)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.131834 -0.019873  0.000843  0.023479  0.060412
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.3448481  0.1126207  11.941  < 2e-16 ***
## season      -0.0002178  0.0000567  -3.842 0.000133 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02848 on 695 degrees of freedom
## Multiple R-squared:  0.0208, Adjusted R-squared:  0.01939
## F-statistic: 14.76 on 1 and 695 DF,  p-value: 0.0001332
```

```
summary(lat_change.fit.new)
```

```
##
## Call:
## lm(formula = lat_change ~ season, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90041 -0.06349  0.00594  0.07630  1.07461
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.056e-01  6.117e-01  -0.173    0.863
```

```
## season       8.776e-05  3.079e-04   0.285    0.776
##
## Residual standard error: 0.1547 on 695 degrees of freedom
## Multiple R-squared:  0.0001168,  Adjusted R-squared:  -0.001322
## F-statistic: 0.08122 on 1 and 695 DF,  p-value: 0.7757
```

```
summary(long_change.fit.new)
```

```
##
## Call:
## lm(formula = long_change ~ season, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31082 -0.02940  0.00746  0.04013  0.57851
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.0267628  0.3273439  -3.137  0.00178 **
## season       0.0003188  0.0001648   1.934  0.05347 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08277 on 695 degrees of freedom
## Multiple R-squared:  0.005355,   Adjusted R-squared:  0.003924
## F-statistic: 3.742 on 1 and 695 DF,  p-value: 0.05347
```

```
summary(wind_change.fit.new)
```

```
##
## Call:
## lm(formula = wind_change ~ season, data = season_diff)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.40370 -0.04331 -0.00361  0.04853  0.36826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.051e-01  3.486e-01   0.875    0.382
## season      9.024e-05  1.755e-04   0.514    0.607
##
## Residual standard error: 0.08814 on 695 degrees of freedom
## Multiple R-squared:  0.0003802,  Adjusted R-squared:  -0.001058
## F-statistic: 0.2644 on 1 and 695 DF,  p-value: 0.6073
```

```
sum0.new <- summary(intercept.fit.new)$coefficients[,c(1,4)]
#colnames(sum0.new) <- c("Estimate0", "p.value0")
sum1.new <- summary(wind_prev.fit.new)$coefficients[,c(1,4)]
#colnames(sum1.new) <- c("Estimate1", "p.value1")
sum2.new <- summary(lat_change.fit.new)$coefficients[,c(1,4)]
#colnames(sum2.new) <- c("Estimate2", "p.value2")
```

	Beta 0		Beta 1		Beta 2		Beta 3		
	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estimate	Pr(>|t|)	Estim
(Intercept)	4.5142875	0.0000000	1.3448481	0.0000000	-0.1056332	0.8629385	-1.0267628	0.001781	0.3051
season	-0.0003543	0.0497902	-0.0002178	0.0001332	0.0000878	0.7757368	0.0003188	0.053474	0.0000

```r
sum3.new <- summary(long_change.fit.new)$coefficients[,c(1,4)]
#colnames(sum3.new) <- c("Estimate3", "p.value3")
sum4.new <- summary(wind_change.fit.new)$coefficients[,c(1,4)]
#colnames(sum4.new) <- c("Estimate4", "p.value4")

kable(cbind(sum0.new, sum1.new, sum2.new, sum3.new, sum4.new)) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed")) %>%
  add_header_above(c(" " = 1, "Beta 0" = 2, "Beta 1" = 2, "Beta 2" = 2, "Beta 3" = 2, "Beta 4" = 2))
```

Now we also fit linear models for the season variables (corresponding to the year) to seek for potential evidence of the statement :"the wind speed has been increasing over years". In order to analyze this question, need to inspect on the second model, which corresponds to the wind speed and the year. As we can see from the table, the coefficient for year is quit significant with a negative estimates quite close to zero. Therefore, we can infer that as the year increase, the wind speed may decrease a little, which is contrary to the statement. However, it's quite match with the results shown in the figures in the initial EDA session, which indicates the mean wind speed tends to decrease over years.