

# 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()
colnames(season_diff)[2] <- "year"

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

summary(intercept.fit)

##
## Call:
## lm(formula = intercept ~ month + year + 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
## year          -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 + year + 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
## year          -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 + year + 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
year	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 + year + 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
## year              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 + year + 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
## year            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)]
```

	Beta 0		Beta 1		Beta 2		Beta 3		Beta 4	
	Est	pval	Est	pval	Est	pval	Est	pval	Est	pval
(Intercept)	4.481	0.000	1.343	0.000	0.041	0.951	-0.834	0.019	0.289	0.448
monthApril	0.023	0.835	0.015	0.670	0.017	0.931	0.042	0.680	0.036	0.739
monthMay	0.026	0.783	0.000	0.997	0.071	0.660	0.063	0.458	-0.016	0.859
monthJune	0.028	0.765	0.005	0.851	-0.007	0.964	0.056	0.505	0.024	0.792
monthJuly	0.013	0.891	0.015	0.590	-0.009	0.954	0.036	0.664	0.013	0.884
monthAugust	-0.020	0.828	0.023	0.412	-0.052	0.738	0.012	0.881	0.031	0.726
monthSeptember	-0.007	0.938	0.026	0.359	-0.036	0.817	0.021	0.797	0.044	0.618
monthOctober	0.009	0.919	0.021	0.459	-0.029	0.855	0.034	0.680	0.035	0.694
monthNovember	0.015	0.875	0.025	0.393	0.024	0.879	0.026	0.753	0.021	0.817
monthDecember	0.006	0.953	0.009	0.772	-0.054	0.745	0.042	0.633	0.011	0.905
year	0.000	0.072	0.000	0.000	0.000	0.910	0.000	0.203	0.000	0.625
natureET	0.001	0.977	0.004	0.688	-0.070	0.169	-0.026	0.329	-0.021	0.473
natureNR	0.001	0.987	-0.015	0.333	0.006	0.943	0.003	0.944	-0.022	0.646
natureSS	0.014	0.490	-0.003	0.602	-0.001	0.969	0.013	0.496	-0.024	0.234
natureTS	0.012	0.479	-0.006	0.249	-0.015	0.588	-0.023	0.126	-0.017	0.283

```

sum4 <- summary(wind_change.fit)$coefficients[,c(1,4)]

p5_sum <- cbind(sum0, sum1, sum2, sum3, sum4)
cbind(sum0, sum1, sum2, sum3, sum4) %>%
  knitr::kable(digits = 3, col.names = c("Est", "pval", "Est", "pval", "Est", "pval", "Est", "pval", "Est", "pval"), "E
  kable_styling(bootstrap_options = c("condensed")) %>%
  add_header_above(c(" " = 1, "Beta 0" = 2, "Beta 1" = 2, "Beta 2" = 2, "Beta 3" = 2, "Beta 4" = 2))

save(p5_sum, file = "./tables/p5_sum.RData")

```

Now based on the estimated Bayesian model from previous questions, we need to explore the seasonal difference. We can fit 5 models using 5 estimated beta values against the three predictors:  $X_{i,1}$ : the month of the year the  $i$ th hurricane started,  $X_{i,2}$ : the year of the  $i$ th hurricane and  $X_{i,3}$ : the nature of the  $i$ th 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, which is of the size  $697 * 5$ .

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 with a p-value larger than 0.05. However, for those significant coefficients, we could infer a potential relationship between the certain predictors and the beta coefficients respectively. We should consult the previous Bayesian model:

$$Y_i(t+6) = \beta_{0,i} + \beta_{1,i}Y_i(t) + \beta_{2,i}\Delta_{i,1}(t) + \beta_{3,i}\Delta_{i,2}(t) + \beta_{4,i}\Delta_{i,3}(t) + \epsilon_i(t)$$

to interpret the change of the influence on  $Y_{i,t+6}$  as the value of the predictor changes.

For the fitted coefficients of  $\beta_0$  to  $\beta_4$ , the intercept cannot show information about seasonal difference since they indicate when holding all the predictors zero, the value for the corresponding  $\beta$ . We can only observe that the year is quite significant in the model for  $\beta_0$ ,  $\beta_1$  with both negative estimates close to zero. Therefore, as the year increase, the coefficient of the intercept and  $Y_{i,t}$  may decrease a little, which means for the Bayesian model, the wind speed when holding all the variables zero and the effect the previous wind speed has will decrease over years. Apart from seasonal difference, some other predictors are quite significant, such as natureET for  $\beta_2$ , natureTS for  $\beta_3$ .

	Beta 0		Beta 1		Beta 2		Beta 3		Beta 4	
	Est	pval	Est	pval	Est	pval	Est	pval	Est	pval
(Intercept)	3.837	0.000	0.894	0.000	0.161	0.000	-0.350	0.000	0.442	0.000
seasonSummer	-0.031	0.205	0.015	0.044	-0.098	0.017	-0.047	0.034	0.036	0.120
seasonAutumn	-0.024	0.325	0.021	0.005	-0.091	0.025	-0.043	0.046	0.049	0.035
seasonWinter	-0.019	0.654	0.003	0.794	-0.098	0.164	-0.009	0.802	0.015	0.711

```
# Try to fit the beta model only with the four seasons
season_diff <- as.data.frame(season_diff) %>%
  mutate(month = recode(month, April = "Spring"),
         month = recode(month, May = "Spring"),
         month = recode(month, June = "Summer"),
         month = recode(month, July = "Summer"),
         month = recode(month, August = "Summer"),
         month = recode(month, September = "Autumn"),
         month = recode(month, October = "Autumn"),
         month = recode(month, November = "Autumn"),
         month = recode(month, December = "Winter"),
         month = recode(month, January = "Winter"),
         month = factor(month, levels = c("Spring", "Summer", "Autumn", "Winter")))
colnames(season_diff)[3] <- "season"

# Beta0
intercept.fit.2 <- lm(intercept ~ season, data = season_diff)
# Beta1
wind_prev.fit.2 <- lm(wind_prev ~ season, data = season_diff)
# Beta2
lat_change.fit.2 <- lm(lat_change ~ season, data = season_diff)
# Beta3
long_change.fit.2 <- lm(long_change ~ season, data = season_diff)
# Beta4
wind_change.fit.2 <- lm(wind_change ~ season, data = season_diff)

sum0_2 <- summary(intercept.fit.2)$coefficients[,c(1,4)]
sum1_2 <- summary(wind_prev.fit.2)$coefficients[,c(1,4)]
sum2_2 <- summary(lat_change.fit.2)$coefficients[,c(1,4)]
sum3_2 <- summary(long_change.fit.2)$coefficients[,c(1,4)]
sum4_2 <- summary(wind_change.fit.2)$coefficients[,c(1,4)]

p5_sum2 <- cbind(sum0_2, sum1_2, sum2_2, sum3_2, sum4_2)
cbind(sum0_2, sum1_2, sum2_2, sum3_2, sum4_2) %>%
  knitr::kable(digits = 3, col.names = c("Est", "pval", "Est", "pval", "Est", "pval", "Est", "pval", "Est", "pval", "Est", "pval"),
  kable_styling(bootstrap_options = c("condensed")) %>%
  add_header_above(c(" " = 1, "Beta 0" = 2, "Beta 1" = 2, "Beta 2" = 2, "Beta 3" = 2, "Beta 4" = 2))

save(p5_sum2, file = "./tables/p5_sum2.RData")
```

We also try to represent the months as four seasons and fit a model for  $\beta$  against them. Each model has three dummy variables corresponding to the three seasons except Spring. The latter three rows of estimate shows how the value of  $\beta$  differentiate between Spring and the other three seasons respectively. If with a rather small p-value, we can conclude the existence of seasonal difference. Therefore, by constructing model in this way, we find that  $\beta_1$  and  $\beta_4$  will increase a little as season changes from Spring to Summer, then

to Autumn, which means a season difference of the effect  $Y_{i,t}$  and  $\Delta_{i,3}(t)$  has on the wind speed. For  $\beta_2$ ,  $\beta_3$ , Summer and Autumn may lead to a slightly smaller effect of  $\Delta_{i,1}(t)$ ,  $\Delta_{i,2}(t)$  have on the wind speed compared to Spring.

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

summary(intercept.fit.new)
```

```
##
## Call:
## lm(formula = intercept ~ year, 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 ***
## year        -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 ~ year, 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 ***
## year        -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 ~ year, 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
## year         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 ~ year, 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 **
## year         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 ~ year, data = season_diff)
##
## Residuals:
```



	Beta 0		Beta 1		Beta 2		Beta 3		Beta 4	
	Est	pval	Est	pval	Est	pval	Est	pval	Est	pval
(Intercept)	4.514	0.00	1.345	0	-0.106	0.863	-1.027	0.002	0.305	0.382
year	0.000	0.05	0.000	0	0.000	0.776	0.000	0.053	0.000	0.607

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

p5_sum_new <- cbind(sum0.new, sum1.new, sum2.new, sum3.new, sum4.new)

cbind(sum0.new, sum1.new, sum2.new, sum3.new, sum4.new) %>%
  knitr::kable(digits = 3, col.names = c("Est", "pval", "Est", "pval", "Est", "pval", "Est", "pval", "Est", "pval"), "E",
  kable_styling(bootstrap_options = c("condensed")) %>%
  add_header_above(c(" " = 1, "Beta 0" = 2, "Beta 1" = 2, "Beta 2" = 2, "Beta 3" = 2, "Beta 4" = 2))

save(p5_sum_new, file = "./tables/p5_sum_new.RData")
```

Now fit linear models for  $\beta$  against 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 model which corresponds to the wind speed and the year. For  $\beta_2$  model, the estimate of year is significant, although it’s really close to zero. Therefore, we can infer that as the year increases, the impact past wind speed has on the current wind speed may decrease a little, which cannot provide support for 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.

In conclusion, for different months, there is no significant differences observed. Over years, the effect the wind speed 6 months ago has on the current wind speed may decrease a little. And there is no evidence to support the statement in task 5.