SDS 387 HW 3

Maurice Diesendruck, Sukyung Park, Bowei Yan, Michael Zhang

March 1, 2015

The general linear model includes covariates for each proxy, and each proxy's interaction with latitude and latitude squared. Both the general model and the proposed X design matrix have rank 22, meaning that only 22 of the columns are linearly independent. The results of the linear model show that proxies are not significantly different, and that latitude (and latitude squared) are not significant factors informing temperature change. Other choices of models may prove that proxies are significantly different; but here the model pivots on proxy 1, which has high variance, making differences between proxy 1 and others insignificant.

Model Description

Our general linear model matrix is $Y = X\beta + \epsilon$. In this problem:

$$\beta = [p_1, \ldots, p_8, p_1 \cdot \text{lat}, \ldots, p_8 \cdot \text{lat}, p_1 \cdot \text{lat}^2, \ldots, p_8 \cdot \text{lat}^2]$$

Where X is a design matrix, and ϵ is our error term where $E[\epsilon] = 0$ and $cov(\epsilon) = \sigma^2$ where σ^2 is the `sdev` variable squared. The rank of our design matrix is 22. The proxy of 6 is linearly dependent and thus the rank is $24 - 2 = 22$.

The rank of X is also 22, and the columns of this design matrix spans the same space as designed in 1. This function creates one intercept term and the differences of each proxy, 2 through 8, against the first proxy along with the interaction of the difference in proxy effects with latitude and latitude squared.

Results and Discussion

Our model is run with the following command:

```
> climate_glm = glm(deltaT ~ factor(proxy)*poly(latitude,2),
+                   data=climate, weights = 1/(sdev)^2)
```

For the coefficients on only the proxy factors, the estimate gives us the estimated difference in proxy 2 through 8 against proxy 1 in the effect on the response variable. For the latitude and latitude squared coefficients, the estimate gives us the change in the response variable when the latitude and latitude squared values increases by one.

The interaction between the proxies and the latitude variables gives us the estimated difference in the effect latitude (and latitude squared) has on the response variable between proxy 2 through eight against proxy 1.

The model we fit only shows the difference in effects in the response variable between proxy 1 and the rest of the proxies. The model given in this problem does not show any systematic differences. The model output will show there are no significant difference in proxies against proxy 1.

Similarly, we see that latitude does not contribute a significant effect in the response variable. The terrestrial estimate variable is linearly dependent on the proxy variable, but the model does not include this variable anyway so we cannot say anything about its effect on the response variable. For reference, see the "Confidence Intervals" section.

Confidence Intervals

	2.5 %	97.5 %
(Intercept)	-7.338271	1.751642
factor(proxy)2	-4.107586	5.000160
factor(proxy)3	-4.083327	5.530283
factor(proxy)4	-7.887664	2.036816
factor(proxy)5	-5.013540	4.406146
factor(proxy)6	-8.727285	2.445894
factor(proxy)7	-7.482627	2.132327
factor(proxy)8	-8.288608	1.784665
poly(latitude, 2)1	-23.768271	17.749910
poly(latitude, 2)2	-59.601630	44.870835
factor(proxy)2:poly(latitude, 2)1	-17.358567	24.457074
factor(proxy)3:poly(latitude, 2)1	-16.559587	29.686907
factor(proxy)4:poly(latitude, 2)1	-18.724330	42.464487
factor(proxy)5:poly(latitude, 2)1	-22.143519	23.925868
factor(proxy)6:poly(latitude, 2)1	NA	NA
factor(proxy)7:poly(latitude, 2)1	-30.653892	22.500140
factor(proxy)8:poly(latitude, 2)1	-43.979623	26.564115
factor(proxy)2:poly(latitude, 2)2	-49.244687	55.411004
factor(proxy)3:poly(latitude, 2)2	-50.845066	56.583022
factor(proxy)4:poly(latitude, 2)2	-42.378602	80.929201
factor(proxy)5:poly(latitude, 2)2	-36.664993	69.025360
factor(proxy)6:poly(latitude, 2)2	NA	NA
factor(proxy)7:poly(latitude, 2)2	-50.930772	57.554583
factor(proxy)8:poly(latitude, 2)2	-40.251131	77.573605