Charlie Carpenter

BIOS 6643

Homework 1

1a.

1b.

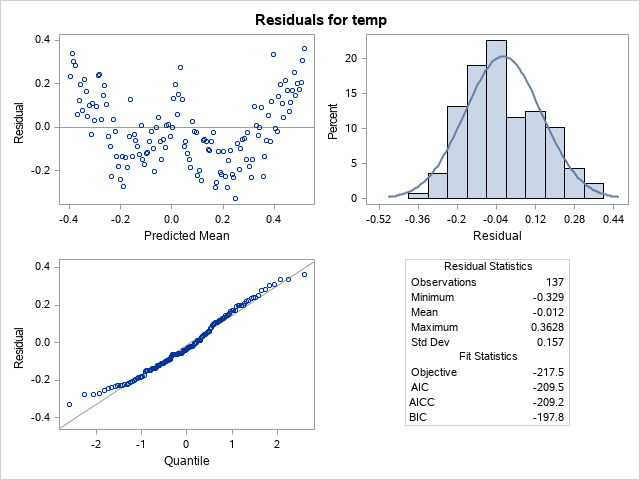
1c.

1d. is a stationary process as it has a constant expected value (0) and its correlation does not depend on .

2. Let where is the observed value for subject *i* at time *j*.

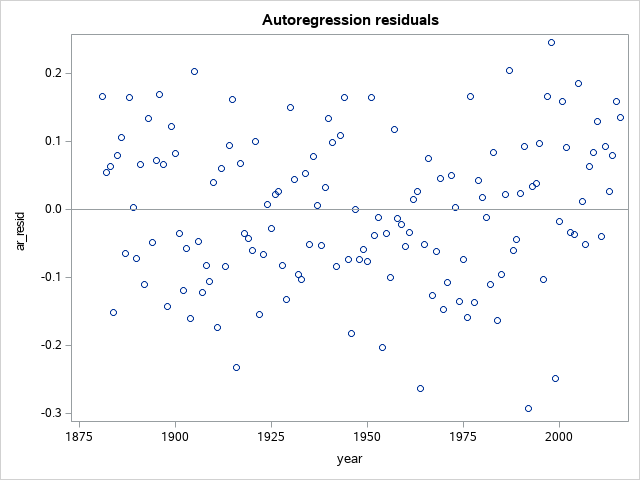
1. Change score models use as their response. This is a simple model that corresponds to a paired t-test which is an easily understandable analysis, however this is forcing to have a slope of 1.
2. Baseline as covariate models use as their response and have as a covariate along with any others that would have been in the change score model. Since is a combination of two variables it will have a larger variance than using .
3. A hybrid model is a combination of the two, using as the response while controlling for the baseline measurement . This is essentially doing the same thing as the baseline as covariate model, but will change the slope of by 1. This model can make the covariate of interest harder to interpret.
4. Longitudinal models will model each time point and allow us to obtain estimates for both while allowing for time-varying covariates, at unique variance structure for each time point, and the ability to model covariances between time points. That being said, for small data sets with 2 time points and few covariates, methods i-iii would still be appropriate and possibly easier to explain to collaborators.

3a.



The “W” pattern in the residuals vs fitted values plot (top left) indicates a quartic trend in average temp over time. Just looking at this it appears we are missing a large piece of the data structure.

3b.



3c.

This plot of residuals indicates that there is no left over structure that isn’t being accounted for as was apparent in part 3a. There is a random spread around 0. The “stall” from 1950 to 1975 is still visible as a slight dip in the residuals, but the spread of the residuals seems reasonably consistent. The mean appears to drop around this time period, not the variance.

3d.

The estimated increase in mean temperature is 0.067 ºC per decade.

| **Solution for Fixed Effects** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Effect** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| **Intercept** | -13.0092 | 1.6243 | 0 | -8.01 | . |
| **year** | 0.006710 | 0.000834 | 135 | 8.05 | <.0001 |

Appendix:

PROC Import datafile= '/folders/myfolders/sasuser.v94/BIOS 6643/Data/global\_temp\_anomalies.csv'

Out = global\_temp;

RUN;

\*PROC Print data=global\_temp;

PROC Mixed data=global\_temp method=ml plots=residualpanel;

model temp=year / solution outp=tempout;

repeated / type=ar(1) subject=intercept;

Run;

\*Proc print data=tempout;

PROC Sgplot data = tempout;

scatter x = year y = resid;

refline 0;

title 'Residuals vs. Fitted Plot';

Run;

\* Making AR(1) residuals accounting for AR(1) str;

DATA tempout;

set tempout;

\* Predicted - residual at time t-1

\* Covariance estimate from PROC Mixed;

y\_tilde = pred + 0.7335\*lag(resid);

ar\_resid = temp - y\_tilde;

Run;

\*\* Plotting residuals;

PROC sgplot data=tempout;

scatter x=year y = ar\_resid;

refline 0;

title 'Autoregression residuals';

Run;