Cheatsheet: Statistical Assumptions

This cheatsheet provides statistical codes that can be used in SAS to test 3 assumptions of linear regression: model fit, residual distribution and residual variance. It also provides statistical codes to examine 2 important factors to examine in statistical models: collinearity and influence/outliers (however, these are not assumptions of linear regression).

SAS codes for Assumptions

To view contents of the SAS practice dataset "Class"

```
proc contents data=sashelp.class;
run;
```

Assumption 1) Model Fit – make a scatterplot

```
proc sgplot data=sashelp.class;
scatter x=height y=weight;
run;
```

Assumption 2) Residual Distribution (2 step process)

Step 1: Run a linear regression and output residual and predicted terms in a new dataset

```
proc reg data=sashelp.class;
model height=weight age;
output out=resid residual=r predicted=fit;
run;
quit;
```

Step 2: Plot the output of the new dataset

```
goptions reset=all;
proc univariate data=resid normal;
var r;
qqplot r / normal(mu=est sigma=est);
run;
```

Assumption 3) Residual Variance – run a linear regression and create a predicted plot

```
proc reg data=sashelp.class;
model height=weight / spec;
plot r.*p.;
run;
quit;
```

SAS codes for Collinearity and Outliers

Collinearity – examine VIF or tolerance

```
proc reg data=sashelp.class;
model height = weight age / vif tol;
run;
quit;
```

Outliers – make a scatterplot and visually inspect the points or run proc univariate to obtain 5 highest and 5 lowest extreme observations

Scatterplot – this is a more complex code to make a scatterplot. Using the code from "Assumption 1) Model Fit" will also work.

```
proc gplot data=sashelp.class;
plot height*weight=1 / vaxis=axis1;
run;
quit;
```

Proc univariate to obtain output of extreme observations

```
proc univariate data=sashelp.class;
var height weight;
run;
quit;
```