Regressão Linear múltipla no SAS

Rui J. B. Bessa

Economia de variáveis

Escolha do modelo tem de ter em conta o número de vaiáveis incluídas vs. o R^2.
R^2 tende assimptoticamente para 1 com o aumento do número de variáveis incluídas no modelo, sejam as variáveis adicionadas significativas ou não.
A entrada de variáveis cujo F ratio seja menor que 1 piora o R^2adj, se for igual a 1 é indiferente e se for maior que 1 melhora.

Modelo básico

R^2 e R^2adj, a análise de variância e os parâmetros estimados e seu erro padrão:

```
proc reg data= name;
model Y = X1 X2 X3 … Xn;
run;
```

Modelo básico com os paramentos estandardizados

```
proc reg data= name;
model Y = X1 X2 X3 … Xn /stb;
run;
```

Dá os betas (parâmetros estandardizados) em que o intercept é 0 e todos os coeficientes estão na mesma escala. Logo quanto maior o parâmetro (em módulo) maior é o seu impacto na previsão dos Y.

Escolha de modelos

Proc Reg

```
proc reg data= name;
model Y = X1 X2 X3 … Xn /selection=forward;
run;

proc reg data= name;
model Y = X1 X2 X3 … Xn /selection=backward;
run;
```

Pode-se definir qual o “significance level for entry” de modo a impor limites para a entrada de variáveis:

```
proc reg data= name;
model Y = X1 X2 X3 … Xn /selection=forward sle=0.05;
run;
```

Neste caso só entram variáveis com P até 0.05. O default do SAS, i.e. se não definirmos nenhum “sle” é 0.50, daí ser muito permissivo na admissão de variáveis.

```
proc reg data= name;
model Y = X1 X2 X3 … Xn /selection=stepwise;
run;
```

A vantagem de usar este tipo de seleção em comparação com a “forward/backward” é que este vai incluindo as variáveis e retirando-as conforme for necessário, não se limita a incluir se a variável preencher os requisitos.

```
proc reg data= name;
model Y = X1 X2 X3 … Xn /selection=maxr;
run;
```

A seleção “maxr” escolhe as variáveis que têm maior impacto no R^2 e, como na seleção “stepwise”, adicione ou retira variáveis consoante o efeito que estas tenham no R^2.

Proc Rsquare

```
proc rsquare adjrsq cp mse sse ;
model Y = X1 X2 X3 … Xn ;
run;
```

O proc rsquare fornece uma lista ordenada do maior para o menor valor de R^2 dos vários modelos, pode ser adicionado ao comando inicial:

adjrsq – fornece o R^2 ajustado

cp – fornece o valor de “Mallow’s cp selection”

mse – fornece o “mean squared error”

sse – fornece o “sum of squared error”

A qualquer um dos procs referidos pode ser acrescentado um comando para restringir o número de variáveis que o programa irá incluir, sendo eles start= e stop=

Exemplo:

```
proc reg data= name;
model Y = X1 X2 X3 … Xn /selection=maxr start=3 stop=7;
run;
```

ou

```
proc rsquare adjrsq cp mse sse ;
model Y = X1 X2 X3 … Xn /start=3 stop=7;
run;
```

Neste caso o programa iria apenas fornecer resultados dos modelos de três a sete variáveis. Este tipo de comando pode ser usado em simultâneo ou em separado, útil quando não se quer ultrapassar um determinado número de variáveis a serem incluídas no modelo.

Pode ser utilizado no caso do “proc rsquare” o comando best=

Análise de modelos

Após a sua seleção, outros parâmetros devem ser avaliados para além do R^2. Isto pode ser feito usando o proc reg:

```
proc reg data= name;
model Y = X1 X2 X3 … Xn /vif collinoint;
run;
```

Parâmetros a ter em conta:

R^2 e R^2 adj	Quanto mais próximo de 1 melhor. Ter em conta a sua relação com o número de variáveis.
Root MSE (mean squared error)	Deve ser baixo. (subjetivo pois depende da unidade em que os dados são apresentados)
Coeficient of variation (CV=Root MSE/Dependent mean x100)	Deve ser baixo
VIF (“Variance Inflation Factors”)	Deve ser baixo, o mais próximo possível de 1, nunca ultrapassar 5
Condition Index	Acima de 10 pode implicar que que algumas dependências possam estar a afetar “regression estimates” (15-30 mau; >30 muito mau ;>100 catastrófico)

Prediction intervals (Intervalos de previsão):

Podemos obter uma file de output com os valores de y previstos pelo modelo e os intervalos de predição. É importante para avaliar como funciona o modelo na realidade e qual a janela de previsão.

Comando para aparecer os resultados no sas:

```
proc reg data = name;
model Y = X1 X2 X3 … Xn /clb clm cli;
run;
```

“clb” requests the 95% confidence intervals for the parameter (β) estimates

“clm” requests the 95% confidence interval for mean prediction

“cli” requests the 95% confidence interval for individual prediction

Comandos para ver o gráfico com os intervalos de confiança:

Primeiro criar um output do proc citado anteriormente com o seguinte comando

```
proc reg data=name ; model Y = X1 X2 X3 … Xn /clb clm cli;
output out=name2 p=name3;
run;
```

Depois usamos os dados criados com o comando anterior,

```
proc reg data=name2;
model Y=name3;
run;
```

Com este comando consegue-se ver ambos as variáveis com o respetivos intervalos de confiança em gráfico.

No caso do sas 9.2 penso que para conseguir ver os gráficos seja necessário usar o comando "ods graphics on"

```
proc reg data = name;
model Y = X1 X2 X3 … Xn /stb;
output out=name2 p=pred l95m=lowerpm u95m=upperpm l95=lowerp u95=upperp;
run;
```

É gerada uma file, depois é necessário ir ao menu>Files>export data> etc...

Neste exemplo o que é pedido são:

- Os valores estimados (p=pred),
- Os limites dos intervalos de previsão mínimos e máximos dos valores estimados para 95% de probabilidade (`l95m=lowerpm` e `u95m=upperpm`) que tomam em consideração apenas a incerteza na incerteza na estimativa dos parametros que definem a linha do modelo.
- Os limites dos intervalos de previsão mínimos e máximos dos valores estimados para 95% de probabilidade (`l95=lowerp` e `u95=upperp`) que tomam em consideração a dispersão à volta da linha do modelo, e a incerteza na estimativa dos parametros que definem a linha do modelo.

Ainda se pode pedir os - os limites de confiança mínimos e máximos dos valores estimados para 95% de probabilidade que tomam apenas em consideração a dispersão à volta da linha do modelo, e assume o modelo como 100% certo. Mas não sei as instruções para pedir isto ao SAS.

Proc útil

Proc para eliminar o output do "results viewer" pois no sas 9.3 não é possível usar "clear all"

```
ods html close; /* close previous */
ods html; /* open new */
```