

Ridge regression

Treating multicollinearity

STAT 615 (Baron)

```
DATA C;
INFILE homes firstobs=3;
INPUT price footage land rooms baths;

PROC REG OUTEST=RIDGE OUTVIF RIDGE=0 to 3.0 by 0.05;
MODEL price = footage land rooms baths;

DATA slopes; SET ridge(where=(_type_='RIDGE'));
DATA vif; SET ridge(where=(_type_='RIDGEVIF'));

PROC PLOT data=slopes;
PLOT footage*_RIDGE_='F' land *_RIDGE_='L'
rooms *_RIDGE_='R' baths*_RIDGE_='B' / overlay;

PROC PLOT data=vif;
PLOT footage*_RIDGE_='F' land *_RIDGE_='L'
rooms *_RIDGE_='R' baths*_RIDGE_='B' / overlay;

proc print data=slopes;
proc print data=vif;

RUN;
```

type: 'parms' = $\hat{b}^{(0)}$
'ridge' = $\hat{b}^{(2)}$

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	89.18575	20.17285	4.42	<.0001
footage	1	0.00017763	0.00014595	1.22	0.2255
land	1	-0.02337	0.01261	-1.85	0.0659
rooms	1	-0.93825	2.23118	-0.42	0.6747
baths	1	0.37931	4.28200	0.09	0.9295

Plot of footage*_RIDGE_ Symbol used is 'F'.
Plot of land*_RIDGE_ Symbol used is 'L'.
Plot of rooms*_RIDGE_ Symbol used is 'R'.
Plot of baths*_RIDGE_ Symbol used is 'B'.

