

**A marine biologist was interested in the relationship between different coastline sites and the number of Hermit crabs inhabiting a site. The biologist counted Hermit crabs on 25 transects randomly located in each of six different coastline sites. The biologist would like to know if there are any significant differences in abundance of hermit of crabs among the 6 sites. The counts are provided below.**

site 1	site 2	site 3	site 4	site 5	site 6
0	415	0	0	0	0
0	466	0	0	1	0
22	6	4	0	1	0
3	14	13	4	2	2
17	12	5	2	2	3
0	0	1	2	1	0
0	3	1	5	2	0
7	1	4	4	29	4
11	16	4	2	2	0
11	55	36	1	2	5
73	142	407	0	0	4
33	10	0	12	13	22
0	2	0	1	0	0
65	145	18	30	19	64
13	6	4	0	1	4
44	4	14	3	3	4
20	5	0	28	26	43
27	124	24	2	30	3
48	24	52	21	5	16
104	204	314	8	4	19
233	0	245	82	94	95
81	0	107	12	1	6
22	56	5	10	9	22
9	0	6	2	3	0
2	8	2	0	0	0

```
proc means data=one mean std;
  var count;
  class site;
run;
proc glm data=one;
  class site;
  model count=site;
  output out=two residual=ehat predicted=yhat;
run;
proc univariate plot data=two;
  var ehat;
run;
proc plot data=two;
  plot ehat*yhat;
run;
data three; set one;
  y=log(count+1);
run;
proc means data=three mean std;
  var y;
  class site;
run;
proc glm data=three;
  class site;
  model y=site;
  output out=four residual=ehat predicted=yhat;
run;
proc univariate plot data=four;
  var ehat;
run;
proc plot data=four;
  plot ehat*yhat;
run;
```

Analysis Variable : count

site	N Obs	Mean	Std Dev
1	25	33.8000000	50.3851830
2	25	68.7200000	125.3536730
3	25	50.6400000	107.4379201
4	25	9.2400000	17.3860097
5	25	10.0000000	19.8410349
6	25	12.6400000	23.0106497

The GLM Procedure

Dependent Variable: count

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	76695.0400	15339.0080	2.97	0.0140
Error	144	744493.1200	5170.0911		
Corrected Total	149	821188.1600			

R-Square	Coeff Var	Root MSE	count Mean
0.093395	233.1496	71.90335	30.84000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
site	5	76695.04000	15339.00800	2.97	0.0140

Source	DF	Type III SS	Mean Square	F Value	Pr > F
site	5	76695.04000	15339.00800	2.97	0.0140





## The MEANS Procedure

Analysis Variable : y

site	N Obs	Mean	Std Dev
1	25	2.5699382	1.6372401
2	25	2.6693789	1.9314434
3	25	2.2198009	1.8612040
4	25	1.4924988	1.2511145
5	25	1.5101943	1.2460912
6	25	1.5467793	1.4696738

## The GLM Procedure

Dependent Variable: y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	38.1018490	7.6203698	3.02	0.0128
Error	144	363.6739366	2.5255134		
Corrected Total	149	401.7757856			

R-Square	Coeff Var	Root MSE	y Mean
0.094834	79.40248	1.589186	2.001432

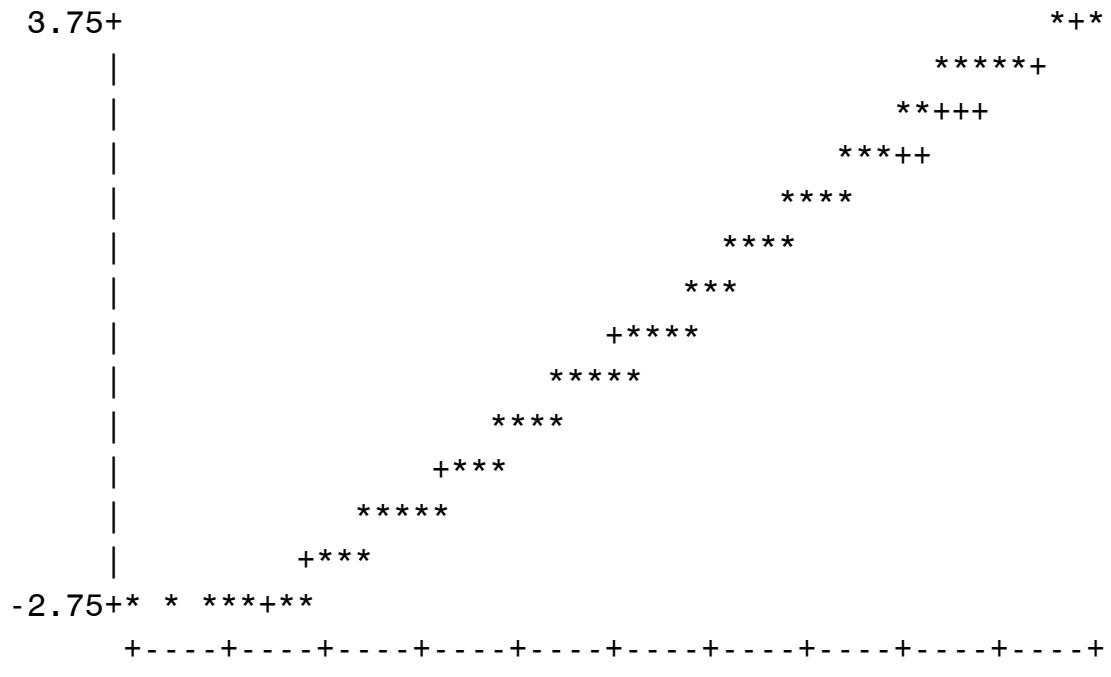
Source	DF	Type I SS	Mean Square	F Value	Pr > F
site	5	38.10184900	7.62036980	3.02	0.0128

Source	DF	Type III SS	Mean Square	F Value	Pr > F
site	5	38.10184900	7.62036980	3.02	0.0128

Stem Leaf	#	Boxplot
3 558	3	
3 0034	4	
2 56799	5	
2 12233	5	
1 5666678889999	13	
1 001112334444	12	+-----+
0 5556677889	10	
0 01111111112233344	16	+
-0 44444444444444433322111111	24	*-----*
-0 9888888877666655	16	
-1 3211	4	
-1 6555555555555555555555	23	+-----+
-2 222220	6	
-2 777766666	9	

-----+-----+-----+-----+-----

Normal Probability Plot



Plot of ehat\*yhat. Legend: A = 1 obs, B = 2 obs, etc.

