

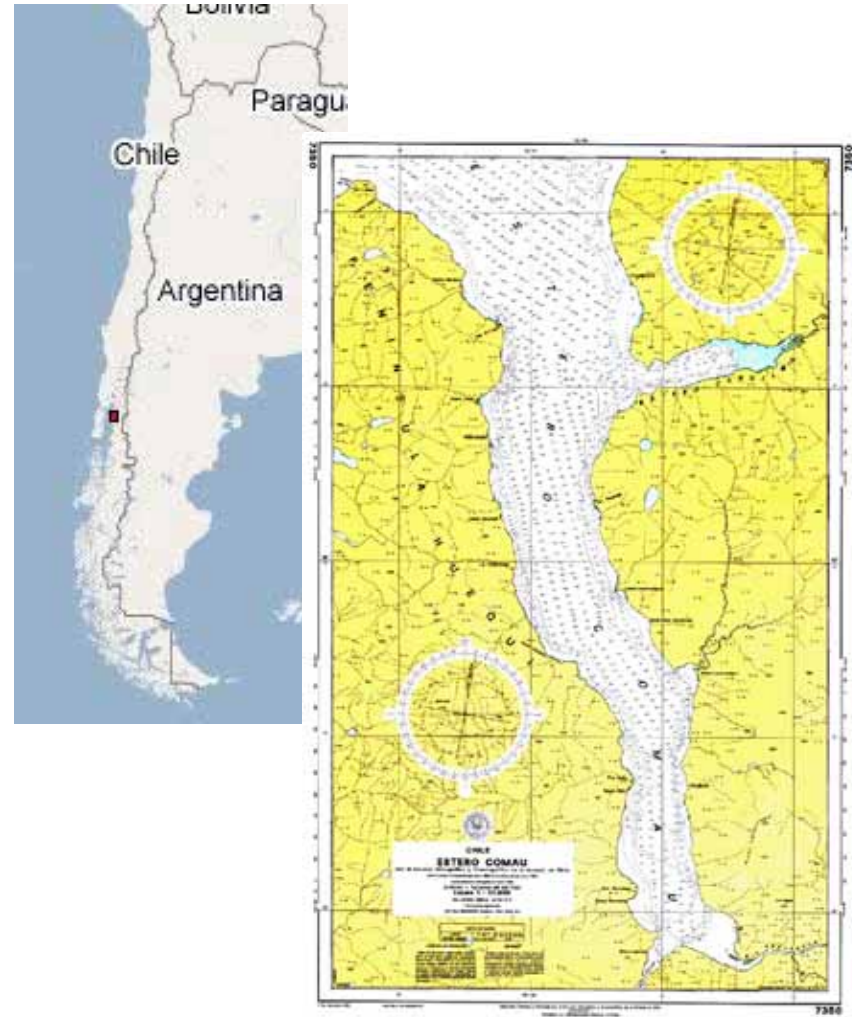
# Results on some crab larvae in some fjord

By: Liz Atwood



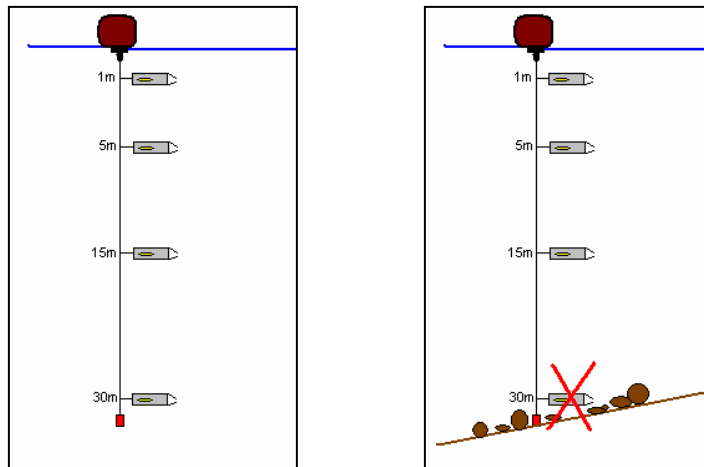
# A bit of background

- Samples from the Comau Fjord in Chilean Patagonia
- Site characterized by
  - High annual rainfall (6-7m)
  - Strong persistent pycnocline
  - Large tides (7m)
  - Rapid change in depth (0-400m depth in less than  $\frac{1}{4}$  km)
- How does this persistent pycnocline affect life history characteristics of species that normally have diel vertical migration?



# Sampling design

- Nightly catches were made every 3-4 days over 4 months (Sept-Dec 2006)
- Samples were taken by light trap →
- Each night, samples were taken at different depths (1, 5, 15, and 30m)

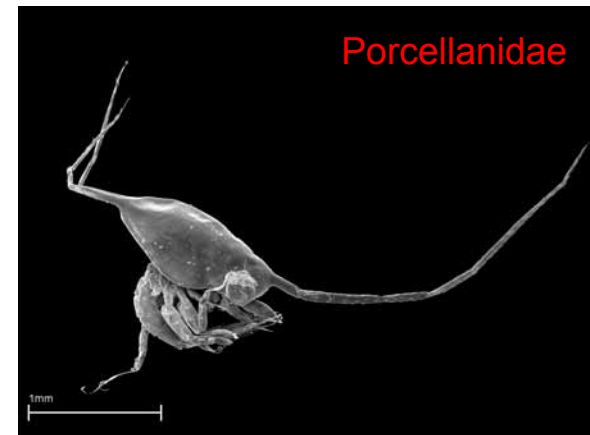


- Animals were filtered out by screen, and later counted

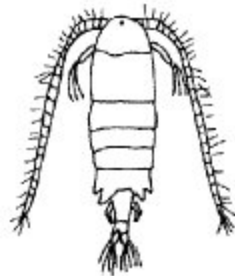
# Animals being sampled

## Biological variables

- Larval crabs (zoea and megalopa) – continuous



- Other animals recorded – presence/absence



(discrete)

[http://www.glf.dfo-mpo.gc.ca/os/bysea-enmer/images/img\\_copepod.jpg](http://www.glf.dfo-mpo.gc.ca/os/bysea-enmer/images/img_copepod.jpg)  
<http://museumvictoria.com.au/crust/images/crusties/isopod1/j31513.jpg>  
[http://animaldiversity.ummz.umich.edu/site/resources/Grzimek\\_inverts/Cumacea/Diastylis\\_rathkei.jpg/badge.jpg](http://animaldiversity.ummz.umich.edu/site/resources/Grzimek_inverts/Cumacea/Diastylis_rathkei.jpg/badge.jpg)

# Environmental variables

- Temperature and salinity recorded for every sample
  - Continuous temperature data (deg. C) from surface to 35m
  - Salinity (ppt) measured from traps (samples from 1,5,15,30m)
  - **Both are continuous data**
- These data used to obtain density estimates

# Environmental variables

(Something complicated & somewhat relevant)

One of the many formulas available for calculating seawater density ( $\rho$ )

$$\rho = C(p) + \beta(p)S - \alpha(T, p)T - \gamma(T, p)(35 - S)T$$

where units are 'km' for  $p$  (these data are in m), 'psu' for  $S$ , '°C' for  $T$ , and

$$C(p) = 999.83 + 5.053p - .048p^2$$

$$\beta(p) = .808 - .0085p$$

$$\alpha(T, p) = .0708(1 + .351p + .068(1 - .0683p)T)$$

$$\gamma(T, p) = .003(1 - .059p - .012(1 - .064p)T)$$

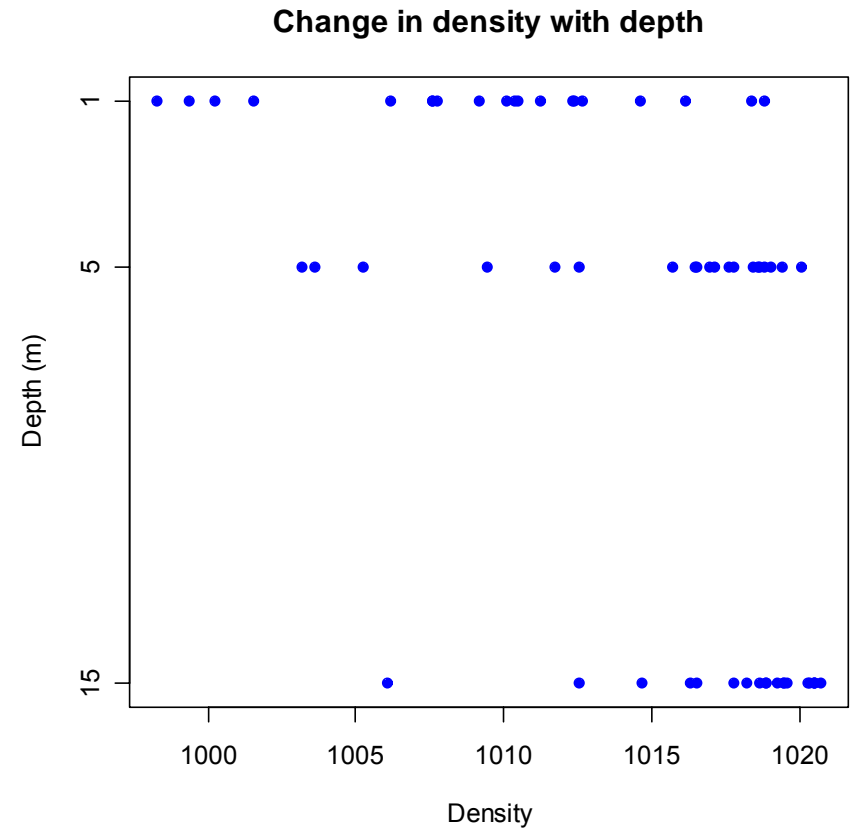
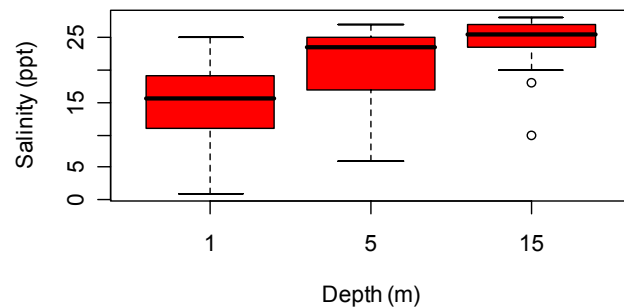
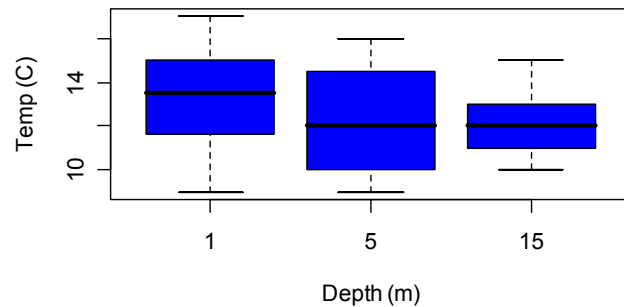
For  $30 \leq S \leq 40$ ,  $-2 \leq T \leq 30$ ,  $p \leq 6$  km: good to .16 kg/m<sup>3</sup>

For  $0 \leq S \leq 40$ , good to .3 kg/m<sup>3</sup>

**Take home message: density is a function of temperature and salinity**

# Environmental variables

- Temperature and salinity recorded for every sample, used to obtain density estimates (something ugly)
- Easy to see that density is less variable as depth increases



# Environmental variables

- And for the most part they are less variable

– Density

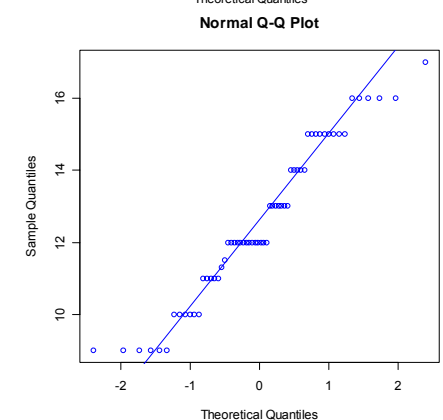
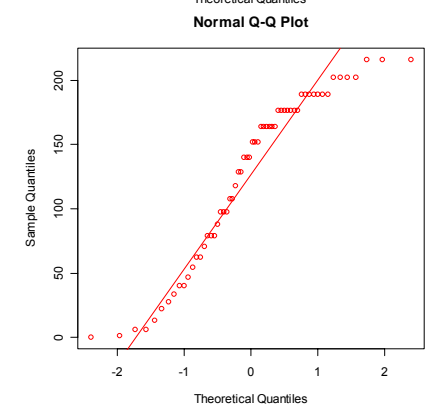
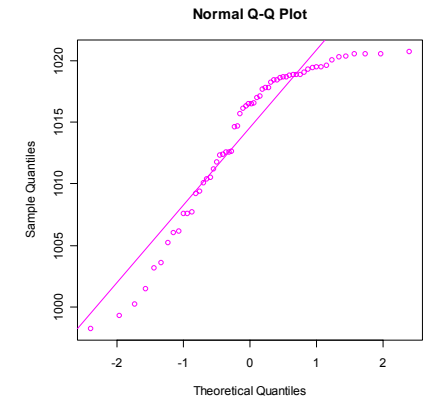
Vari	Sum Sq	Mean Sq	F val	Pr(>F)
Df				
depth	777.42	388.71	15.243	4.979e-06
2				***
Res	1453.52	25.50		
57				

– Salinity  
(transformed)

Vari	Sum Sq	Mean Sq	F val	Pr(>F)
Df				
depth	94060	47030	16.647	2.024e-06
2				***
Res	161036	2825		
57				

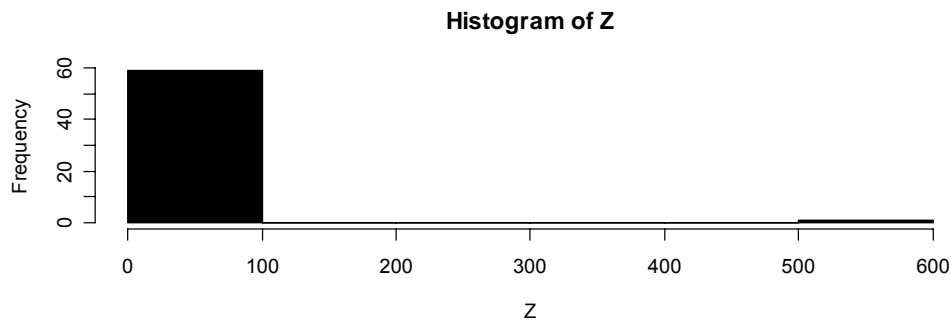
– Temperature

Vari	Sum Sq	Mean Sq	F val	Pr(>F)
Df				
depth	13.406	6.703	1.440	0.2455
2				
Res	265.403	574.66		
57				



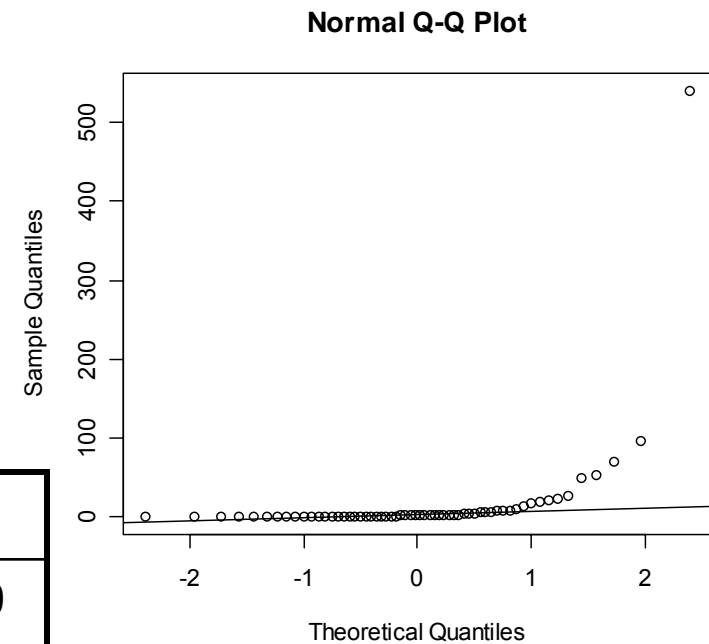
# Non-normal catch data

- Catch data of zoeas highly skewed



– Variances very different

1m	5m	15m
252.1974	553.9237	14346.3789 (yikes!)



- Used a Box-Cox transformation

# Normalized catch data

- Result of box-cox transform

- Lambda max = **-0.3888385**

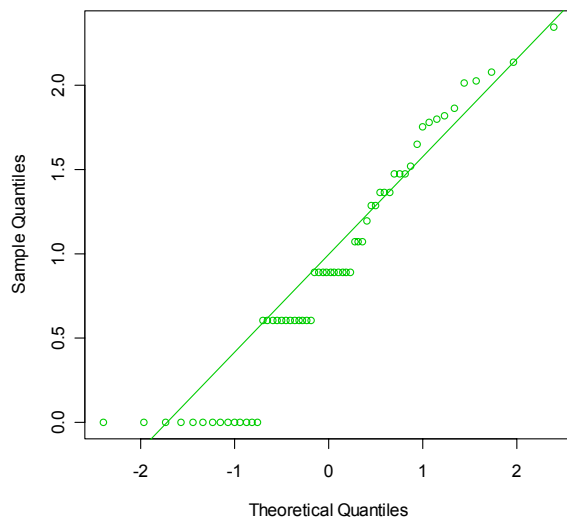
$$Z_{trans} = \frac{(Z_{old} + 1)^\lambda}{\lambda}$$

- Normalized number of zoea

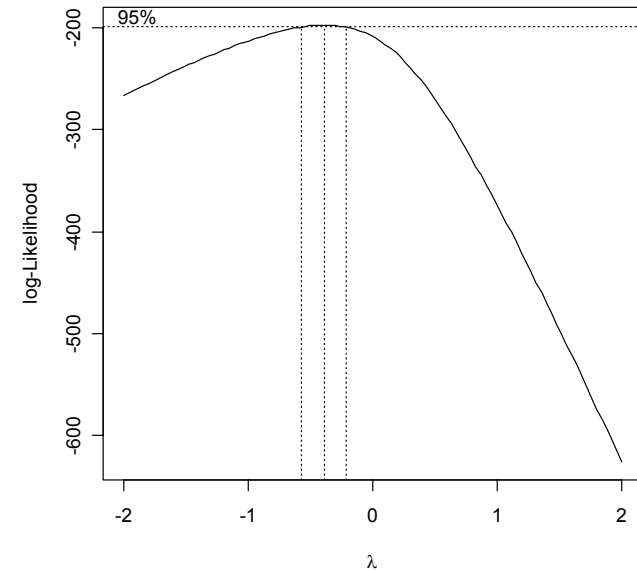
- Variances much smaller

**(0.37, 0.53, 0.49)**

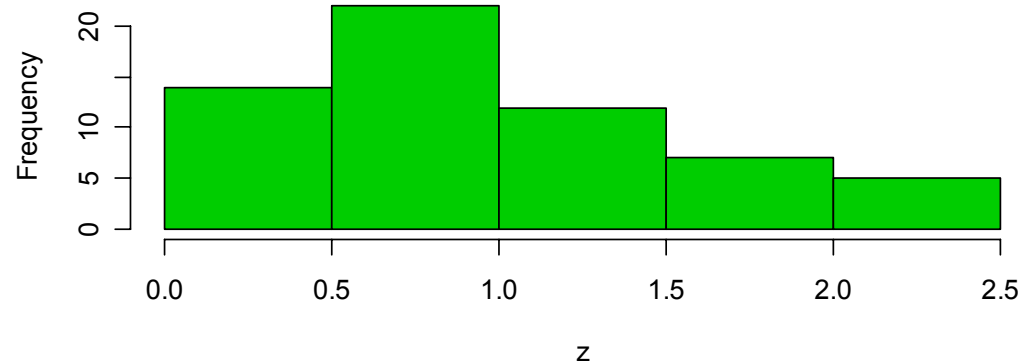
Normal Q-Q Plot



Box-Cox transform curve



Histogram of z



# Trying to find important variables

- Looking at variables together is a pretty long list
  - Abiotic: temp, sal, density, depth
  - Biological: Isopods, Nauplii, Cumacea, Cladocera, Copepoda, Amphipoda, Euphausids, Ostracoda
  - 13 variables total
- Beginning with just the environmental variables (temp, salinity, depth, density)
- Investigate best model by backwards elimination

# Backwards elimination

Linear model 1:

$$z \sim T*S*depth*D+T*S*depth+T*S*D+T*depth*D+S*depth*D+T*S+T*depth+S*depth+T*D+S*D+depth*D+ T+S+depth+D)$$

Remove interaction terms,  
testing each removal via  
ANOVA

Final model:

$$z \sim T*D+T+S+D$$

anova (lm.fin) →

Vari.	Df	Sum Sq	Mean Sq	F value	Pr(>F)
T	1	1.8749	1.8749	8.3155	0.00560 **
D	1	3.9750	3.9750	17.6302	9.891e-05 ***
S	1	1.5305	1.5305	6.7882	0.01178 *
T:D	1	7.3101	7.3101	32.4219	4.994e-07 ***
Res	55	12.4007	0.2255		

Result looks good, but this seems like a haphazard way to go about model selection (plus it is a lot of boring coding!)

# Step AIC

- Nifty stepAIC function already in R

$$AIC = 2p - 2 \log \left( \frac{f(x|\theta_2)}{f(x|\theta_1)} \right)$$

or in the case that the errors are normally distributed

$$AIC = 2p + n \left[ \log \left( \frac{2\pi \sum_{i=1}^n (Y_i - f(X_i))^2}{n} \right) + 1 \right]$$

- Whole data set is too big, crashes computer(s)
- So first look at environmental variables
  - (temp, sal, depth, density)

```
zenv.lm <- lm(z ~ .^4, data = zenv)
```

```
zenv.step <- stepAIC(zenv.lm)
```

– The model with the lowest AIC (-75.66747) is

```
z ~ T + S + D + depth + T:S + T:depth + S:D + S:depth +  
D:depth + T:S:depth + S:D:depth
```

Still too complicated!!

# Step AIC

## (environmental variables)

- Start with model  

$$z \sim T + S + D + \text{depth} + T:S + T:\text{depth} + S:D + S:\text{depth} + D:\text{depth} + T:S:\text{depth} + S:D:\text{depth}$$
- Similar to the backwards elimination, remove variables and test for significant reduction in variability
- Widdle away to get down to  

$$z \sim T + S + T:S$$
- Model performs well in explaining variability of zoea abundance (and I feel better about not including density)

Vari.	Df	Sum Sq	Mean Sq	F value	Pr(>F)
T	1	1.8749	1.8749	6.8604	0.011318 *
S	1	3.8859	3.8859	14.2191	0.000394 ***
T:S	1	1 6.0262	6.0262	22.0508	1.761e-05 ***
Res	56	15.3042	0.2733		

# Step AIC

- Same procedure with the bio variables
- Model with lowest AIC (-58.582) is

$z \sim I + N + Cu + Cl + Co + A + E + O + I:Cl + N:Cu$   
 $+ N:Cl + N:Co + N:E + N:O + Cl:Co + Cl:A + Co:O$   
 $+ A:O + N:Cl:Co + N:Co:O$

- And the simpler model after widdling  
 $z \sim Cu + Cl + Cl:A$

Vari.	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Cu	1	2.4064	2.4064	6.8074	0.01167 *
Cl	1	1.4697	1.4697	4.1576	0.04627 *
A	1	0.7061	0.7061	1.9975	0.16319
Cl:A	1	3.0668	3.0668	8.6756	0.00472 **
Res	55	19.4423	0.3535		

Key (or reminder what things are)

I - Isopoda



N - Nauplii



Cu - Cumacea



Cl - Cladocera



Co - Copepoda



A - Amphipoda



E - Euphausiids



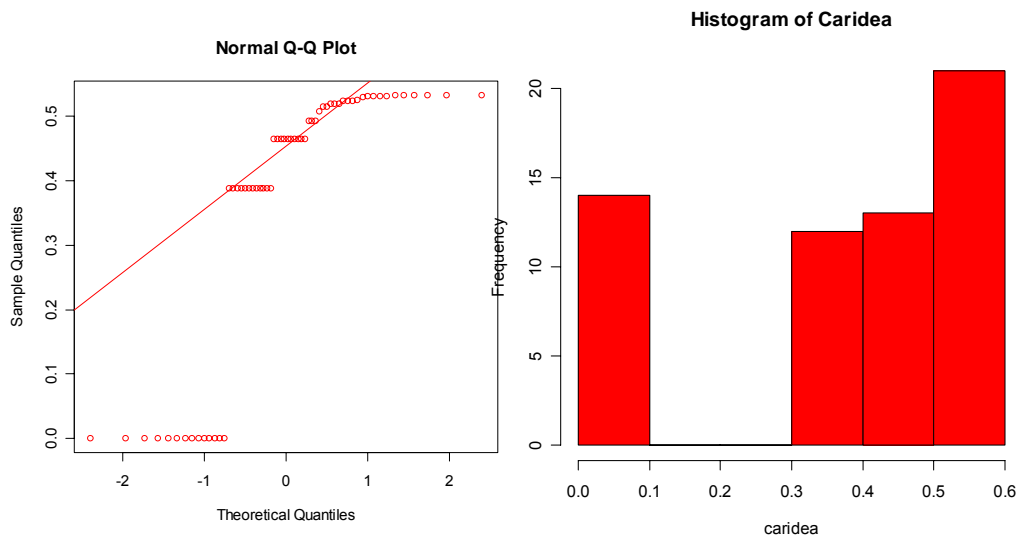
O - Ostracoda



# Step AIC

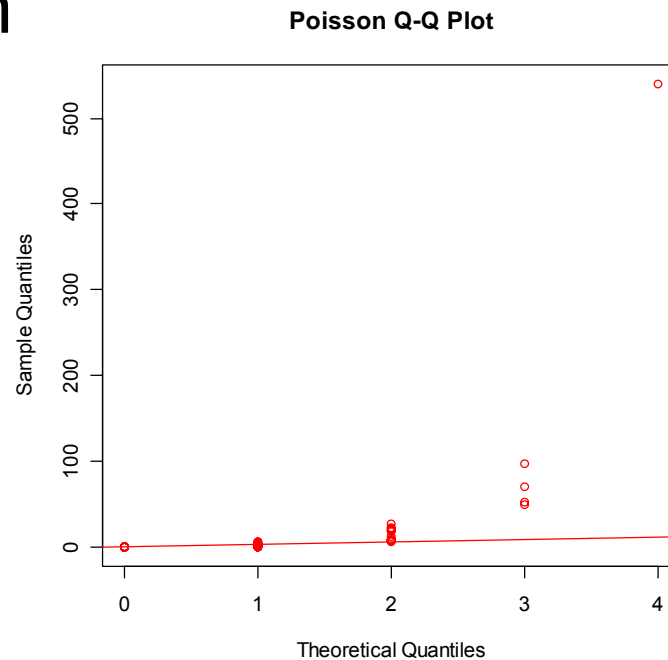
- Looking into the family groups, data on families are highly variable
- Even with abundant families (eg Caridea), Box-Cox transformation not working

Family group	Total caught
Sergestidae	344
Porcellanidae	1
Caridea	158
Cancridae/ Xanthidae/ Majidae	159
Galatheidae	5
Paguridae	36
Grapsidae	77
Hymenosomidae	42
Atelecyclidae/ Belliidae	166



# This is not normal data

- Catch data is characteristically distributed  
Poisson



mmm... not quite what I  
was looking for

- Need to check for mysterious things like over-dispersion
- Whole 'nother can of fish (or crabs)

Questions?

