Poisson Regression for the Medfly Data

```
> # first fit linear regression to logA_1, to get starting values
> reg.1 = lm(logA_1 ~ south + loc_host.south + alt + dist)

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.7320340  0.1807736  15.113  < 2e-16 ***
south          -0.8103601  0.2066947  -3.921 0.000180 ***
loc_host.south  0.7214667  0.2210740   3.263 0.001593 **
alt             0.0002521  0.0005232   0.482 0.631178
dist           -0.0347156  0.0094032  -3.692 0.000395 ***
---
# reg.1 uses log_10 and glm uses ln, so adjust scales
reg.1.start = log(10) * reg.1$coeff
  (Intercept)          south loc_host.south            alt           dist
  6.2907407616  -1.8659230582   1.6612385214   0.0005804539 -0.0799357257

now fit Poisson regression
poisson.1 = glm(A ~ south + loc_host.south + alt + dist,
          family = poisson(link=log),
          start=reg.1.start)

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     7.025e+00  8.050e-03 872.732  < 2e-16 ***
south          -1.420e+00  1.771e-02 -80.207  < 2e-16 ***
loc_host.south  1.306e+00  1.865e-02  70.009  < 2e-16 ***
alt             1.514e-04  3.898e-05   3.885 0.000102 ***
dist           -5.152e-02  7.733e-04 -66.617  < 2e-16 ***
---
Number of Fisher Scoring iterations: 7
> # what if we start with log(mean) and zeros?
> mean.start = c(log(mean(A)), 0, 0, 0, 0)
 [1] 6.072863 0.000000 0.000000 0.000000 0.000000

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     7.025e+00  8.050e-03 872.734  < 2e-16 ***
south          -1.420e+00  1.771e-02 -80.208  < 2e-16 ***
loc_host.south  1.306e+00  1.865e-02  70.010  < 2e-16 ***
alt             1.514e-04  3.898e-05   3.885 0.000102 ***
dist           -5.152e-02  7.733e-04 -66.618  < 2e-16 ***
---
Number of Fisher Scoring iterations: 5
```

A comparison of linear and Poisson regression results

> summary(reg.1)

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     2.7320340  0.1807736  15.113  < 2e-16 ***
south          -0.8103601  0.2066947  -3.921 0.000180 ***
loc_host.south  0.7214667  0.2210740   3.263 0.001593 **
alt             0.0002521  0.0005232   0.482 0.631178
dist           -0.0347156  0.0094032  -3.692 0.000395 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7955 on 84 degrees of freedom
Multiple R-squared: 0.3387,   Adjusted R-squared: 0.3072
F-statistic: 10.76 on 4 and 84 DF,  p-value: 4.359e-07
```

> summary(poisson.1)

```
Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     7.025e+00  8.050e-03 872.732  < 2e-16 ***
south          -1.420e+00  1.771e-02 -80.207  < 2e-16 ***
loc_host.south  1.306e+00  1.865e-02  70.009  < 2e-16 ***
alt             1.514e-04  3.898e-05   3.885 0.000102 ***
dist           -5.152e-02  7.733e-04 -66.617  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(Dispersion parameter for poisson family taken to be 1)

```
    Null deviance: 70797  on 88  degrees of freedom
Residual deviance: 47739  on 84  degrees of freedom
AIC: 48299
```

- coefficients are much too significant
- probably need to allow for overdispersion; i.e. do not assume dispersion parameter (phi) equals 1, but estimate it. will do this later using quasi-likelihood.

```
> mean(A)
[1] 433.9213
> var(A)
[1] 464220.7
> var(A) / mean(A)
[1] 1069.827
```

this ratio should be close to 1 for a Poisson distribution