Changing the Reference Level of a Factor in a GLM Model Does Not   
Change the Overall Model Results

John Wallace

The second recommendation of the GFSSC (https://docs.google.com/document/d/1p891XH3e12wDIZsUdKpWcBtpXqshs9a5/edit):

*Other recommendations of the GFSSC on how to improve the method included:*

* *…*
* *Explore the sensitivity of the abundance index for vermillion rockfish to the reference level against which all the data used in the model are scaled.  The logistic model makes an arbitrary choice for the reference level, and depending on the software, the reference levels for the fixed effects are probably associated with the factor levels from the data point that happens to occur first (or last) in the data set.  If these reference levels are not supported by appreciable observations from all years in the series, then the resulting annual index values may not be reliable.*

Is addressed by the explanation (with R code) of Jim Greene (12th Mar 2022) here:

<https://www.researchgate.net/post/Does_changing_the_reference_level_in_an_R_model_cause_any_difference_in_results>

Greene shows the reported contrasts in a GLM summary can change when a different factor reference level is selected but that the results of the overall model does not change. That concept is expanded on that below with the Hook & Link Survey data.

What is unique for a factor in a GLM are the estimated differences of the pairwise comparisons, regardless whether one level is set to zero or a different contrast treatment is used. That is why a 2 level factor only has a sign change when the other level is set to zero. However, the summary output of a GLM >2 level factor is only a subset of all pairwise factors. Say a factor has four levels (3 degrees of freedom). In R, the combination of 4 items taking 2 at a time is:

combn(4, 2)

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1 1 1 2 2 3

[2,] 2 3 4 3 4 4

If the first level is set to zero then the GLM summary for the factor with four levels will only show the estimated differences for 3 (equal to the degrees of freedom) of the pairwise comparisons: (1, 2), (1, 3), and (1, 4) . However, the other three differences are not given. In R, the function stats::relevel() re-orders the levels of a factor so that the reference level specified is first (and hence set to zero with the default treatment contrasts). Two additional relevel() calls, one with a 2nd level reference and the other with a 3rd reference is one way to see all six pairwise differences.

Vermilion, thru 2019, has been caught on 118 of the 121 sites. Site is treated as a factor in the GLM and has 117 degrees of freedom. The number of different pairwise combinations is 6,903:

dim(combn(118, 2))

[1] 2 6903

The 117 pairwise differences shown in the model summary are only a fraction of the total (127/6903= 0.0169).

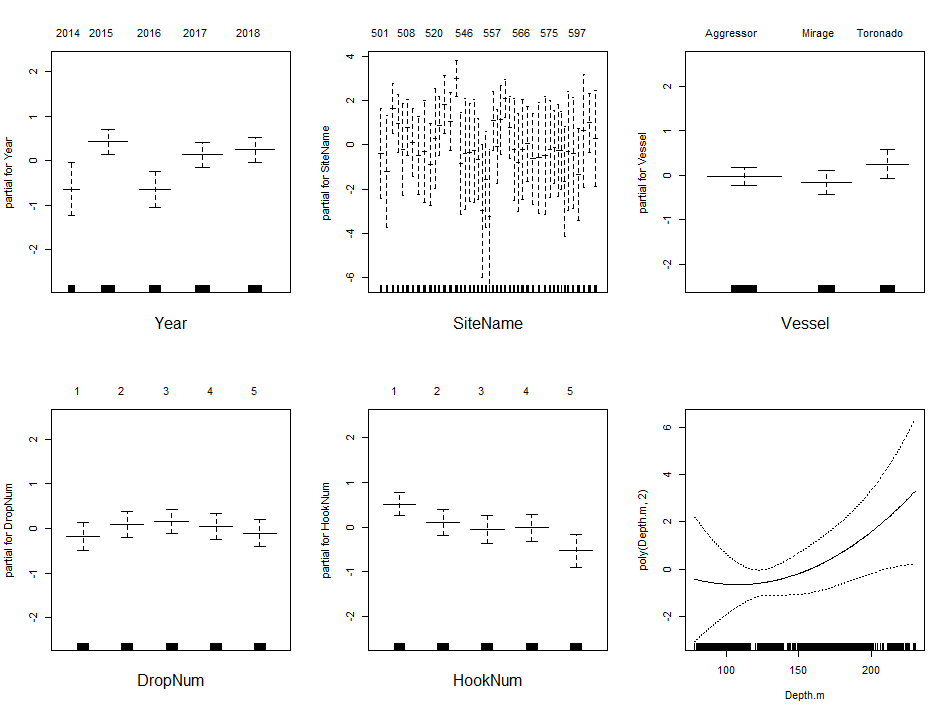
Using the Cowcod data inside the CCA, a GLM model with default Site ‘501’ set to zero gives the following figure, summary GLM output, and yearly index:

dim(combn(46, 2)) # 46 sites inside the CCA have caught Cowcod (out of 79 sites as of 2019).

[1] 2 1035

dev.new(); par(mfrow = c(2, 3))

plot.Gam(Glm <- glm(NumCow ~ Year + SiteName + Vessel + DropNum + HookNum + poly(Depth.m, 2), data = CowCod.2018.CC.In, family = binomial), scale = 5, se = TRUE)



summary(Glm)

Call:

glm(formula = NumCow ~ Year + SiteName + Vessel + DropNum + HookNum +

poly(Depth.m, 2), family = binomial, data = CowCod.2018.CC.In)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.6527 -0.1734 -0.1089 -0.0777 3.6709

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -5.634174 1.122711 -5.018 5.21e-07 \*\*\*

Year2015 1.056229 0.343729 3.073 0.00212 \*\*

Year2016 -0.009766 0.394643 -0.025 0.98026

Year2017 0.763814 0.359699 2.123 0.03371 \*

Year2018 0.874550 0.361456 2.420 0.01554 \*

SiteName502 -0.814525 1.858448 -0.438 0.66118

SiteName503 2.036290 1.211274 1.681 0.09274 .

SiteName506 1.349011 1.232385 1.095 0.27368

SiteName507 0.172081 1.435932 0.120 0.90461

SiteName508 1.169394 1.163709 1.005 0.31495

SiteName509 0.493452 1.269564 0.389 0.69751

SiteName512 -0.085415 1.411946 -0.060 0.95176

SiteName514 0.092868 1.495193 0.062 0.95047

SiteName518 -0.499450 1.530247 -0.326 0.74413

SiteName520 0.679355 1.464380 0.464 0.64270

SiteName525 1.249222 1.254567 0.996 0.31938

SiteName526 2.218295 1.300433 1.706 0.08804 .

SiteName527 1.447481 1.145896 1.263 0.20652

SiteName531 3.386144 1.068468 3.169 0.00153 \*\*

SiteName543 -0.450674 1.604677 -0.281 0.77883

SiteName546 -0.020164 1.570024 -0.013 0.98975

SiteName547 0.026433 1.461238 0.018 0.98557

SiteName548 0.112273 1.499170 0.075 0.94030

SiteName549 -0.255500 1.465097 -0.174 0.86156

SiteName551 -2.586888 2.044829 -1.265 0.20584

SiteName553 -1.198341 1.652456 -0.725 0.46834

SiteName554 -2.871422 2.033039 -1.412 0.15784

SiteName557 1.473574 1.273923 1.157 0.24739

SiteName558 0.315725 1.449405 0.218 0.82756

SiteName559 1.521580 1.426693 1.067 0.28619

SiteName560 2.488208 1.051524 2.366 0.01797 \*

SiteName561 1.180996 1.174898 1.005 0.31481

SiteName563 0.165243 1.491343 0.111 0.91177

SiteName564 -0.416478 1.546238 -0.269 0.78766

SiteName566 0.196739 1.473313 0.134 0.89377

SiteName568 0.434835 1.484229 0.293 0.76955

SiteName569 -0.251057 1.436903 -0.175 0.86130

SiteName573 -0.204481 1.566236 -0.131 0.89613

SiteName574 -0.104149 1.631159 -0.064 0.94909

SiteName575 0.198947 1.453351 0.137 0.89112

SiteName581 0.275513 1.398039 0.197 0.84377

SiteName582 0.139130 1.431287 0.097 0.92256

SiteName584 0.116750 1.502684 0.078 0.93807

SiteName586 -1.253449 1.750082 -0.716 0.47385

SiteName590 0.095607 1.636829 0.058 0.95342

SiteName592 0.012994 1.568782 0.008 0.99339

SiteName597 -0.946427 1.607126 -0.589 0.55593

SiteName598 1.016238 1.562395 0.650 0.51541

SiteName609 1.380375 1.149784 1.201 0.22993

SiteName611 0.675488 1.422831 0.475 0.63497

VesselMirage -0.140488 0.212154 -0.662 0.50784

VesselToronado 0.282219 0.232909 1.212 0.22562

DropNum2 0.264774 0.243441 1.088 0.27676

DropNum3 0.335791 0.239661 1.401 0.16118

DropNum4 0.224934 0.246469 0.913 0.36144

DropNum5 0.077354 0.254926 0.303 0.76156

HookNum2 -0.408147 0.211154 -1.933 0.05324 .

HookNum3 -0.565099 0.219472 -2.575 0.01003 \*

HookNum4 -0.529912 0.217654 -2.435 0.01491 \*

HookNum5 -1.045472 0.255090 -4.098 4.16e-05 \*\*\*

poly(Depth.m, 2)1 93.620079 67.926761 1.378 0.16813

poly(Depth.m, 2)2 42.514555 36.643643 1.160 0.24596

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2000.6 on 13313 degrees of freedom

Residual deviance: 1734.1 on 13252 degrees of freedom

AIC: 1858.1

Number of Fisher Scoring iterations: 8

Index <- rbind(cbind(Estimate = 0, 'Std. Error' = 0), summary(Glm)$coeff[2:5,1:2])

row.names(Index)[1] <- "Year2014"

Index

Estimate Std. Error

Year2014 0.000000000 0.0000000

Year2015 1.056229189 0.3437288

Year2016 -0.009765786 0.3946432

Year2017 0.763813537 0.3596992

Year2018 0.874550269 0.3614564

anova(Glm, test ='Cp')

Analysis of Deviance Table

Model: binomial, link: logit

Response: NumCow

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Cp

NULL 13313 2000.6 2002.6

Year 4 13.870 13309 1986.7 1996.7

SiteName2 45 222.647 13264 1764.0 1864.1

Vessel 2 2.722 13262 1761.3 1865.3

DropNum 4 3.179 13258 1758.2 1870.2

HookNum 4 19.565 13254 1738.6 1858.6

poly(Depth.m, 2) 2 4.476 13252 1734.1 1858.1

Looking at the number of Cowcod by site, ‘531’ has the most Cowcod over all years. Each year had at least one Cowcod at this site.

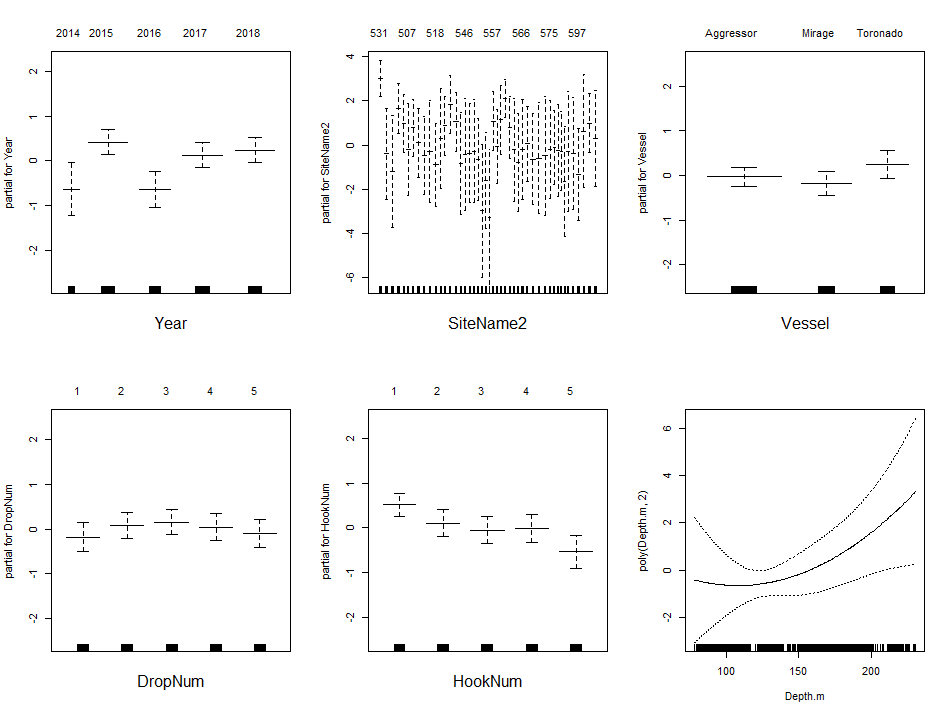
Site ‘531’ is changed to the first level reference site using relevel():

# \*\*\* When using plot.Gam(), relevel() cannot be inside the formula call \*\*\*

dev.new(); par(mfrow = c(2, 3))

CowCod.2018.CC.In$SiteName2 <- relevel(SiteName, '531', '1')

plot.Gam(Glm2 <- glm(NumCow ~ Year + SiteName2 + Vessel + DropNum + HookNum + poly(Depth.m, 2), data = CowCod.2018.CC.In, family = binomial), scale = 5, se = TRUE)



summary(Glm2)

Call:

glm(formula = NumCow ~ Year + SiteName2 + Vessel + DropNum +

HookNum + poly(Depth.m, 2), family = binomial, data = CowCod.2018.CC.In)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.6527 -0.1734 -0.1089 -0.0777 3.6709

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -2.248030 0.537575 -4.182 2.89e-05 \*\*\*

Year2015 1.056229 0.343729 3.073 0.002120 \*\*

Year2016 -0.009766 0.394643 -0.025 0.980258

Year2017 0.763814 0.359699 2.123 0.033714 \*

Year2018 0.874550 0.361456 2.420 0.015541 \*

SiteName2501 -3.386144 1.068468 -3.169 0.001529 \*\*

SiteName2502 -4.200669 1.476923 -2.844 0.004452 \*\*

SiteName2503 -1.349854 0.474461 -2.845 0.004441 \*\*

SiteName2506 -2.037133 0.555720 -3.666 0.000247 \*\*\*

SiteName2507 -3.214063 1.028032 -3.126 0.001769 \*\*

SiteName2508 -2.216750 0.653624 -3.391 0.000695 \*\*\*

SiteName2509 -2.892692 0.742032 -3.898 9.69e-05 \*\*\*

SiteName2512 -3.471559 0.840874 -4.129 3.65e-05 \*\*\*

SiteName2514 -3.293276 1.287006 -2.559 0.010501 \*

SiteName2518 -3.885594 0.991926 -3.917 8.96e-05 \*\*\*

SiteName2520 -2.706789 1.226175 -2.208 0.027279 \*

SiteName2525 -2.136922 0.582438 -3.669 0.000244 \*\*\*

SiteName2526 -1.167849 0.604267 -1.933 0.053277 .

SiteName2527 -1.938663 0.769830 -2.518 0.011792 \*

SiteName2543 -3.836818 1.125865 -3.408 0.000655 \*\*\*

SiteName2546 -3.406308 1.415816 -2.406 0.016133 \*

SiteName2547 -3.359712 1.213213 -2.769 0.005618 \*\*

SiteName2548 -3.273871 1.293681 -2.531 0.011385 \*

SiteName2549 -3.641644 0.910577 -3.999 6.35e-05 \*\*\*

SiteName2551 -5.973032 1.711711 -3.490 0.000484 \*\*\*

SiteName2553 -4.584485 1.172196 -3.911 9.19e-05 \*\*\*

SiteName2554 -6.257566 1.683728 -3.716 0.000202 \*\*\*

SiteName2557 -1.912570 0.586824 -3.259 0.001117 \*\*

SiteName2558 -3.070419 0.864131 -3.553 0.000381 \*\*\*

SiteName2559 -1.864564 0.827379 -2.254 0.024223 \*

SiteName2560 -0.897936 0.458058 -1.960 0.049960 \*

SiteName2561 -2.205148 0.792357 -2.783 0.005385 \*\*

SiteName2563 -3.220901 1.278174 -2.520 0.011738 \*

SiteName2564 -3.802622 1.072149 -3.547 0.000390 \*\*\*

SiteName2566 -3.189405 1.243223 -2.565 0.010305 \*

SiteName2568 -2.951309 0.915121 -3.225 0.001260 \*\*

SiteName2569 -3.637201 1.030794 -3.529 0.000418 \*\*\*

SiteName2573 -3.590625 1.408715 -2.549 0.010807 \*

SiteName2574 -3.490294 1.511444 -2.309 0.020930 \*

SiteName2575 -3.187197 1.199238 -2.658 0.007868 \*\*

SiteName2581 -3.110631 0.798406 -3.896 9.78e-05 \*\*\*

SiteName2582 -3.247014 1.034696 -3.138 0.001700 \*\*

SiteName2584 -3.269394 0.949490 -3.443 0.000575 \*\*\*

SiteName2586 -4.639593 1.318601 -3.519 0.000434 \*\*\*

SiteName2590 -3.290537 1.515412 -2.171 0.029902 \*

SiteName2592 -3.373150 1.411609 -2.390 0.016868 \*

SiteName2597 -4.332572 1.108170 -3.910 9.24e-05 \*\*\*

SiteName2598 -2.369906 1.481759 -1.599 0.109735

SiteName2609 -2.005769 0.778629 -2.576 0.009994 \*\*

SiteName2611 -2.710656 1.261599 -2.149 0.031667 \*

VesselMirage -0.140488 0.212154 -0.662 0.507843

VesselToronado 0.282219 0.232909 1.212 0.225622

DropNum2 0.264774 0.243441 1.088 0.276758

DropNum3 0.335791 0.239661 1.401 0.161182

DropNum4 0.224934 0.246469 0.913 0.361440

DropNum5 0.077354 0.254926 0.303 0.761556

HookNum2 -0.408147 0.211154 -1.933 0.053245 .

HookNum3 -0.565099 0.219472 -2.575 0.010029 \*

HookNum4 -0.529912 0.217654 -2.435 0.014906 \*

HookNum5 -1.045472 0.255090 -4.098 4.16e-05 \*\*\*

poly(Depth.m, 2)1 93.620079 67.926761 1.378 0.168126

poly(Depth.m, 2)2 42.514555 36.643643 1.160 0.245961

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2000.6 on 13313 degrees of freedom

Residual deviance: 1734.1 on 13252 degrees of freedom

AIC: 1858.1

Number of Fisher Scoring iterations: 8

Index <- rbind(cbind(Estimate = 0, 'Std. Error' = 0), summary(Glm2)$coeff[2:5,1:2])

row.names(Index)[1] <- "Year2014"

Index

Estimate Std. Error

Year2014 0.000000000 0.0000000

Year2015 1.056229189 0.3437288

Year2016 -0.009765786 0.3946432

Year2017 0.763813537 0.3596992

Year2018 0.874550269 0.3614564

anova(Glm2, test = 'Cp')

Analysis of Deviance Table

Model: binomial, link: logit

Response: NumCow

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Cp

NULL 13313 2000.6 2002.6

Year 4 13.870 13309 1986.7 1996.7

SiteName 45 222.647 13264 1764.0 1864.1

Vessel 2 2.722 13262 1761.3 1865.3

DropNum 4 3.179 13258 1758.2 1870.2

HookNum 4 19.565 13254 1738.6 1858.6

poly(Depth.m, 2) 2 4.476 13252 1734.1 1858.1

Note the when setting the site with the most Cowcod as the reference, which also happens to be extreme compared to the norm, many more significant pairwise differences (out of the possible 1,035) are shown in the GLM summary compared with the summary when the default site of ‘501’ is set to zero. Also in the figures, observe that plot.Gam() does calculate a value for the levels which have been set to zero and Site ‘531’ can be seen to be moved to the first position in the second figure.

Even though the second model’s summary shows more significant pairwise differences, observe that the AIC, analysis of deviance table, yearly index estimates, and the standard error of those estimates are exactly the same for both models.

Lastly, note that for the binomial model, where the dispersion is assumed to be 1, Mallows' *Cp* statistic “which is closely related to AIC (and multiple of it if the dispersion is known).” Is equal to the AIC [see the R help for anova.glm()].