

# Practical Aspects of Predictive Models

*Nick Hanewinckel, FSA, CERA, AVP and Actuary, HannoverRe*

## Contents

<b>Build a Simple Model</b>	<b>1</b>
But is this safe? . . . . .	6
Evaluate . . . . .	7
<b>So... What's wrong with my model?</b>	<b>9</b>
Gender! . . . . .	9
<b>Graphs with Count CI</b>	<b>14</b>

## Build a Simple Model

We will use ILEC data.

```
pth <- 'H:/Mortality Research/ILEC_2009-15 Data 20180601.txt'
dat <- fread(pth,stringsAsFactors = TRUE,nrows=1000000,check.names = TRUE)
setnames(dat,'Number.of.Deaths','Deaths')
setnames(dat,'Policies.Exposed','Exposure')
dat <- dat[Exposure>0]
#Possibly remove below
summary(dat)
```

```
## Observation.Year Common.Company.Indicator.57 Preferred.Indicator
## Min. :2009 Min. :0.0000 Min. :0.00000
## 1st Qu.:2009 1st Qu.:1.0000 1st Qu.:0.00000
## Median :2009 Median :1.0000 Median :0.00000
## Mean :2009 Mean :0.9977 Mean :0.01167
## 3rd Qu.:2009 3rd Qu.:1.0000 3rd Qu.:0.00000
## Max. :2009 Max. :1.0000 Max. :1.00000
##
## Gender Smoker.Status Insurance.Plan Issue.Age
## Female:558717 NonSmoker:586145 Other: 4358 Min. : 0.00
## Male :434304 Smoker :287898 Perm :280579 1st Qu.:22.00
## Unknown :118978 Term :359686 Median :36.00
## UL :159799 Mean :35.92
## ULSG : 69787 3rd Qu.:50.00
## VL : 95119 Max. :99.00
## VLSG : 23693
##
## Duration Attained.Age Age.Basis Face.Amount.Band
## Min. : 1.00 Min. : 0.00 Min. :0.0000 50000-99999 :195882
## 1st Qu.: 7.00 1st Qu.: 35.00 1st Qu.:0.0000 100000-249999:190080
## Median :13.00 Median : 50.00 Median :0.0000 25000-49999 :164091
## Mean : 14.75 Mean : 49.67 Mean :0.4624 10000-24999 :113045
## 3rd Qu.:20.00 3rd Qu.: 64.00 3rd Qu.:1.0000 250000-499999:112372
## Max. :104.00 Max. :117.00 Max. :1.0000 1-9999 : 75381
## (Other) :142170
##
## Issue.Year Number.of.Preferred.Classes Preferred.Class
```

##	Min.	:1906	Min.	:2.0	Min.	:1.0			
##	1st Qu.:	1989	1st Qu.:	2.0	1st Qu.:	1.0			
##	Median	:1996	Median	:3.0	Median	:2.0			
##	Mean	:1995	Mean	:2.9	Mean	:1.9			
##	3rd Qu.:	2002	3rd Qu.:	4.0	3rd Qu.:	2.0			
##	Max.	:2009	Max.	:4.0	Max.	:4.0			
##		NA's	:981434		NA's	:981434			
##	SOA.Anticipated.Level.Term.Period			SOA.Guaranteed.Level.Term.Period					
##	N/A (Not Term):633335			N/A (Not Term) :633335					
##	Not Level Term: 28938			Unknown : 90625					
##	Unknown :330748			10 yr guaranteed: 73134					
##				20 yr guaranteed: 56041					
##				15 yr guaranteed: 47609					
##				5 yr guaranteed : 38725					
##				(Other) : 53552					
##	SOA.Post.level.term.indicator			Select_Ultimate_Indicator					
##	N/A (Not Term) :633335			Select :788586					
##	Not Level Term : 28938			Ultimate:204435					
##	Post Level Term : 73789								
##	Unknown Level Term Period: 90625								
##	Within Level Term :166334								
##									
##									
##	Deaths		Death.Claim.Amount		Exposure				
##	Min.	: 0.0000	Min.	: 0	Min.	: 0.003			
##	1st Qu.:	0.0000	1st Qu.:	0	1st Qu.:	0.830			
##	Median	: 0.0000	Median	: 0	Median	: 2.504			
##	Mean	: 0.1279	Mean	: 5016	Mean	: 17.644			
##	3rd Qu.:	0.0000	3rd Qu.:	0	3rd Qu.:	10.514			
##	Max.	:61.0000	Max.	:10000000	Max.	:4149.075			
##									
##	Amount.Exposed		Expected.Death.QX7580E.by.Amount						
##	Min.	: 0	Min.	: 0.0					
##	1st Qu.:	52479	1st Qu.:	136.0					
##	Median	: 244932	Median	: 876.3					
##	Mean	: 1907163	Mean	: 10771.2					
##	3rd Qu.:	1012432	3rd Qu.:	5340.9					
##	Max.	:518463138	Max.	:2388270.2					
##									
##	Expected.Death.QX2001VBT.by.Amount			Expected.Death.QX2008VBT.by.Amount					
##	Min.	: 0.0	Min.	: 0.0					
##	1st Qu.:	114.3	1st Qu.:	87.6					
##	Median	: 772.7	Median	: 579.4					
##	Mean	: 7981.9	Mean	: 5814.1					
##	3rd Qu.:	4572.6	3rd Qu.:	3351.6					
##	Max.	:1885685.4	Max.	:1822913.9					
##									
##	Expected.Death.QX2008VBTLU.by.Amount			Expected.Death.QX2015VBT.by.Amount					
##	Min.	: 0.0	Min.	: 0.0					
##	1st Qu.:	122.6	1st Qu.:	77.7					
##	Median	: 824.6	Median	: 513.5					
##	Mean	: 8055.8	Mean	: 5056.6					
##	3rd Qu.:	4780.7	3rd Qu.:	2925.0					
##	Max.	:1822913.9	Max.	:1719741.9					

```
##
## Expected.Death.QX7580E.by.Policy Expected.Death.QX2001VBT.by.Policy
## Min. : 0.00000 Min. : 0.00000
## 1st Qu.: 0.00190 1st Qu.: 0.00155
## Median : 0.01047 Median : 0.00927
## Mean : 0.18611 Mean : 0.15333
## 3rd Qu.: 0.05923 3rd Qu.: 0.05035
## Max. :181.12821 Max. :171.59811
##
## Expected.Death.QX2008VBT.by.Policy Expected.Death.QX2008VBTLU.by.Policy
## Min. : 0.00000 Min. : 0.00000
## 1st Qu.: 0.00116 1st Qu.: 0.00171
## Median : 0.00683 Median : 0.00980
## Mean : 0.13077 Mean : 0.15922
## 3rd Qu.: 0.03745 3rd Qu.: 0.05197
## Max. :162.38112 Max. :167.28668
##
## Expected.Death.QX2015VBT.by.Policy
## Min. : 0.00000
## 1st Qu.: 0.00103
## Median : 0.00598
## Mean : 0.11821
## 3rd Qu.: 0.03298
## Max. :151.79925
##
```

```
set.seed(12345)
```

```
ind <- createDataPartition(dat$Attained.Age,p=.3,list=FALSE)
```

```
dat[,Set:='Training']
```

```
dat[ind,Set:='Testing']
```

```
mod1 <- glm(formula=Deaths~(bs(Attained.Age,knots = c(25,62)) + Duration + Smoker.Status)^2 + Face.Amount,
             offset=log(Exposure),family = poisson,data=dat[Set=='Training'])
```

```
summary(mod1)
```

```
##
## Call:
## glm(formula = Deaths ~ (bs(Attained.Age, knots = c(25, 62)) +
##   Duration + Smoker.Status)^2 + Face.Amount.Band, family = poisson,
##   data = dat[Set == "Training"], offset = log(Exposure))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -13.2071  -0.2360  -0.1033  -0.0424   5.6168
##
## Coefficients:
##
##              Estimate
## (Intercept)      -7.709904
## bs(Attained.Age, knots = c(25, 62))1    0.409686
## bs(Attained.Age, knots = c(25, 62))2   -0.827869
## bs(Attained.Age, knots = c(25, 62))3    2.148896
## bs(Attained.Age, knots = c(25, 62))4    6.757042
## bs(Attained.Age, knots = c(25, 62))5    7.989304
## Duration        -0.785075
## Smoker.StatusSmoker    0.892088
```

## Smoker.StatusUnknown	0.358675
## Face.Amount.Band10000-24999	-0.031124
## Face.Amount.Band100000-249999	-0.237866
## Face.Amount.Band1000000-2499999	-0.390293
## Face.Amount.Band10000000+	-9.665525
## Face.Amount.Band25000-49999	-0.050614
## Face.Amount.Band250000-499999	-0.376833
## Face.Amount.Band2500000-4999999	-0.717341
## Face.Amount.Band50000-99999	-0.101755
## Face.Amount.Band500000-999999	-0.342443
## Face.Amount.Band5000000-9999999	-0.220637
## bs(Attained.Age, knots = c(25, 62))1:Duration	0.788608
## bs(Attained.Age, knots = c(25, 62))2:Duration	0.834532
## bs(Attained.Age, knots = c(25, 62))3:Duration	0.791366
## bs(Attained.Age, knots = c(25, 62))4:Duration	0.805620
## bs(Attained.Age, knots = c(25, 62))5:Duration	0.788419
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker	-1.055104
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker	-0.234136
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker	1.049364
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker	-0.769430
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker	-2.598913
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown	-1.040301
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown	0.548431
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown	0.047581
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown	0.433941
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown	-0.797005
## Duration:Smoker.StatusSmoker	-0.013401
## Duration:Smoker.StatusUnknown	-0.022267
##	Std. Error
## (Intercept)	0.460859
## bs(Attained.Age, knots = c(25, 62))1	0.538904
## bs(Attained.Age, knots = c(25, 62))2	0.468368
## bs(Attained.Age, knots = c(25, 62))3	0.473473
## bs(Attained.Age, knots = c(25, 62))4	0.471461
## bs(Attained.Age, knots = c(25, 62))5	0.555402
## Duration	0.091605
## Smoker.StatusSmoker	0.829477
## Smoker.StatusUnknown	0.480386
## Face.Amount.Band10000-24999	0.010366
## Face.Amount.Band100000-249999	0.016388
## Face.Amount.Band1000000-2499999	0.058173
## Face.Amount.Band10000000+	40.965572
## Face.Amount.Band25000-49999	0.013123
## Face.Amount.Band250000-499999	0.029312
## Face.Amount.Band2500000-4999999	0.223906
## Face.Amount.Band50000-99999	0.014260
## Face.Amount.Band500000-999999	0.040170
## Face.Amount.Band5000000-9999999	0.316460
## bs(Attained.Age, knots = c(25, 62))1:Duration	0.091643
## bs(Attained.Age, knots = c(25, 62))2:Duration	0.092387
## bs(Attained.Age, knots = c(25, 62))3:Duration	0.091175
## bs(Attained.Age, knots = c(25, 62))4:Duration	0.092051
## bs(Attained.Age, knots = c(25, 62))5:Duration	0.091232
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker	0.967052

```

## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker      0.814014
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker      0.863029
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker      0.824737
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker      1.038420
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown     0.594504
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown     0.474571
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown     0.523757
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown     0.483209
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown     0.649477
## Duration:Smoker.StatusSmoker                                0.001802
## Duration:Smoker.StatusUnknown                                0.001101
##
## z value Pr(>|z|)
## (Intercept)                                                -16.729 < 2e-16
## bs(Attained.Age, knots = c(25, 62))1                      0.760 0.447123
## bs(Attained.Age, knots = c(25, 62))2                     -1.768 0.077135
## bs(Attained.Age, knots = c(25, 62))3                      4.539 5.66e-06
## bs(Attained.Age, knots = c(25, 62))4                     14.332 < 2e-16
## bs(Attained.Age, knots = c(25, 62))5                     14.385 < 2e-16
## Duration                                                  -8.570 < 2e-16
## Smoker.StatusSmoker                                       1.075 0.282159
## Smoker.StatusUnknown                                       0.747 0.455282
## Face.Amount.Band10000-24999                               -3.003 0.002677
## Face.Amount.Band100000-249999                             -14.515 < 2e-16
## Face.Amount.Band1000000-2499999                           -6.709 1.96e-11
## Face.Amount.Band10000000+                                  -0.236 0.813477
## Face.Amount.Band25000-49999                               -3.857 0.000115
## Face.Amount.Band250000-499999                             -12.856 < 2e-16
## Face.Amount.Band2500000-4999999                           -3.204 0.001356
## Face.Amount.Band50000-99999                               -7.135 9.65e-13
## Face.Amount.Band500000-999999                             -8.525 < 2e-16
## Face.Amount.Band5000000-9999999                           -0.697 0.485676
## bs(Attained.Age, knots = c(25, 62))1:Duration            8.605 < 2e-16
## bs(Attained.Age, knots = c(25, 62))2:Duration            9.033 < 2e-16
## bs(Attained.Age, knots = c(25, 62))3:Duration            8.680 < 2e-16
## bs(Attained.Age, knots = c(25, 62))4:Duration            8.752 < 2e-16
## bs(Attained.Age, knots = c(25, 62))5:Duration            8.642 < 2e-16
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker -1.091 0.275250
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker -0.288 0.773629
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker  1.216 0.224020
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker -0.933 0.350851
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker -2.503 0.012323
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown -1.750 0.080142
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown  1.156 0.247830
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown  0.091 0.927615
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown  0.898 0.369165
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown -1.227 0.219767
## Duration:Smoker.StatusSmoker                             -7.436 1.04e-13
## Duration:Smoker.StatusUnknown                             -20.225 < 2e-16
##
## (Intercept) ***
## bs(Attained.Age, knots = c(25, 62))1
## bs(Attained.Age, knots = c(25, 62))2
## bs(Attained.Age, knots = c(25, 62))3 ***
## bs(Attained.Age, knots = c(25, 62))4 ***

```

```

## bs(Attained.Age, knots = c(25, 62))5          ***
## Duration                                     ***
## Smoker.StatusSmoker
## Smoker.StatusUnknown
## Face.Amount.Band10000-24999                  **
## Face.Amount.Band100000-249999                 ***
## Face.Amount.Band1000000-2499999              ***
## Face.Amount.Band10000000+
## Face.Amount.Band25000-49999                  ***
## Face.Amount.Band250000-499999                 ***
## Face.Amount.Band2500000-4999999              **
## Face.Amount.Band50000-99999                  ***
## Face.Amount.Band500000-999999                 ***
## Face.Amount.Band5000000-9999999
## bs(Attained.Age, knots = c(25, 62))1:Duration ***
## bs(Attained.Age, knots = c(25, 62))2:Duration ***
## bs(Attained.Age, knots = c(25, 62))3:Duration ***
## bs(Attained.Age, knots = c(25, 62))4:Duration ***
## bs(Attained.Age, knots = c(25, 62))5:Duration ***
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker *
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown .
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown
## Duration:Smoker.StatusSmoker                  ***
## Duration:Smoker.StatusUnknown                 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 376471  on 695111  degrees of freedom
## Residual deviance: 141303  on 695076  degrees of freedom
## AIC: 237179
##
## Number of Fisher Scoring iterations: 14
dat[,ExpectedDeaths:=predict(mod1,newdata=dat,type='response')]

```

But is this safe?

```

dat[,.(minAttAge=min(Attained.Age),maxAttAge=max(Attained.Age)),by=Set]

##           Set minAttAge maxAttAge
## 1: Training           0         117
## 2: Testing            0         116

```

This could have been very *unsafe* if our Testing set was outside our Training set range. Without spline specifications or a saved spline object, different range. Don't leave safety to chance!

## Evaluate

```
dat[,.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),by=.(Set)]
```

```
##           Set Exposure AtoEcount
## 1: Training 12223979 1.0000000
## 2: Testing  5296742 0.9958213
```

```
dat[,.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),by=.(Set,Gender)]
```

```
##           Set Gender Exposure AtoEcount
## 1: Training Female  7059382 0.9353740
## 2: Testing Female  3058705 0.9373719
## 3: Training  Male  5164596 1.1828763
## 4: Testing  Male  2238037 1.1546157
```

```
dat[,.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),by=.(Set,Smoker.Status)]
```

```
##           Set Smoker.Status Exposure AtoEcount
## 1: Training      NonSmoker 7727521.7 1.0000000
## 2: Testing      NonSmoker 3317189.2 0.9978522
## 3: Training       Smoker 1230646.2 1.0000000
## 4: Testing       Smoker  529103.3 0.9859085
## 5: Training     Unknown 3265810.8 1.0000000
## 6: Testing     Unknown 1450449.9 0.9975501
```

```
dat[Attained.Age %between% c(55,75),.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths))]
```

```
##           Set Attained.Age Exposure AtoEcount
## 1: Training           55 255020.18 1.0721165
## 2: Testing           55 109920.32 0.9989118
## 3: Training           56 245303.27 0.9773148
## 4: Testing           56 107593.83 0.9863232
## 5: Training           57 240546.97 0.9565151
## 6: Testing           57  99771.87 1.0876255
## 7: Training           58 223251.52 0.9780054
## 8: Testing           58 100453.01 1.0640886
## 9: Testing           59  92027.86 0.9415547
## 10: Training          59 220556.96 0.9906852
## 11: Testing           60  94487.06 0.9671019
## 12: Training           60 214252.16 0.9885888
## 13: Training           61 221700.08 0.9484066
## 14: Testing           61  95334.26 1.0112960
## 15: Testing           62  86624.04 1.0123946
## 16: Training           62 199585.80 1.0282370
## 17: Training           63 163853.79 1.0258405
## 18: Testing           63  68385.06 1.0503520
## 19: Testing           64  67007.47 1.0754881
## 20: Training           64 156172.64 1.0065864
## 21: Testing           65  66712.85 1.0194761
## 22: Training           65 156428.86 1.0297971
## 23: Testing           66  65446.42 1.0085190
## 24: Training           66 150424.29 1.0333590
## 25: Testing           67  56513.60 1.0411604
## 26: Training           67 130692.79 0.9909812
## 27: Training           68 120226.50 1.0251569
```

```
## 28: Testing      68  49999.15 0.9671737
## 29: Training     69 112353.83 1.0089841
## 30: Testing      69  47131.22 1.0465654
## 31: Training     70 102214.57 1.0989784
## 32: Testing      70  49001.44 0.9786845
## 33: Training     71 102298.81 0.9746647
## 34: Testing      71  41827.19 1.0493836
## 35: Training     72  93754.59 0.9655554
## 36: Testing      72  43430.38 0.9320485
## 37: Training     73  92141.64 0.9956065
## 38: Testing      73  41661.49 0.9824580
## 39: Training     74  90117.15 1.0095543
## 40: Testing      74  37307.40 1.0266543
## 41: Training     75  85351.35 0.9856738
## 42: Testing      75  34822.62 1.0369340
##           Set Attained.Age Exposure AtoEcount
```

## Even Better

kableExtra gives pretty table formats!

```
kable(
  dcast(dat[,.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),by=.(Set,Gender)],Gender)
```

Gender	AtoEcount_Testing	AtoEcount_Training	Exposure_Testing	Exposure_Training
Female	0.937	0.935	3058705	7059382
Male	1.155	1.183	2238037	5164596

```
kable(
  dcast(dat[,.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),by=.(Set,Smoker.Status),Smoker.Status)
```

Smoker.Status	AtoEcount_Testing	AtoEcount_Training	Exposure_Testing	Exposure_Training
NonSmoker	0.998	1	3317189.2	7727522
Smoker	0.986	1	529103.3	1230646
Unknown	0.998	1	1450449.9	3265811

```
kable(
  dcast(dat[Attained.Age %between% c(55,75)],.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),Attained.Age)
```



Attained.Age	AtoEcount_Testing	AtoEcount_Training	Exposure_Testing	Exposure_Training
55	0.999	1.072	109920.32	255020.18
56	0.986	0.977	107593.83	245303.27
57	1.088	0.957	99771.87	240546.97
58	1.064	0.978	100453.01	223251.52
59	0.942	0.991	92027.87	220556.96
60	0.967	0.989	94487.06	214252.16
61	1.011	0.948	95334.26	221700.08
62	1.012	1.028	86624.04	199585.80
63	1.050	1.026	68385.06	163853.79
64	1.075	1.007	67007.47	156172.64
65	1.019	1.030	66712.85	156428.86
66	1.009	1.033	65446.42	150424.29
67	1.041	0.991	56513.60	130692.79
68	0.967	1.025	49999.15	120226.50
69	1.047	1.009	47131.22	112353.83
70	0.979	1.099	49001.44	102214.57
71	1.049	0.975	41827.19	102298.81
72	0.932	0.966	43430.38	93754.59
73	0.982	0.996	41661.49	92141.64
74	1.027	1.010	37307.40	90117.15
75	1.037	0.986	34822.62	85351.35

## So... What's wrong with my model?

### Gender!

Gender was never given as a variable, but is obviously crucial. Don't get so lost in the technical details that you forget your actuarial expertise!

```
mod2 <- glm(formula=Deaths~(bs(Attained.Age,knots = c(25,62)) + Duration + Smoker.Status + Gender)^2 + 1,
             offset=log(Exposure),family = poisson,data=dat[Set=='Training'])
summary(mod2)
```

```
##
## Call:
## glm(formula = Deaths ~ (bs(Attained.Age, knots = c(25, 62)) +
##   Duration + Smoker.Status + Gender)^2 + Face.Amount.Band,
##   family = poisson, data = dat[Set == "Training"], offset = log(Exposure))
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -8.8545  -0.2317  -0.0999  -0.0405   5.8488
##
## Coefficients:
##                                     Estimate
## (Intercept)                        -7.438271
## bs(Attained.Age, knots = c(25, 62))1 -0.889553
## bs(Attained.Age, knots = c(25, 62))2 -1.191162
## bs(Attained.Age, knots = c(25, 62))3  1.527021
## bs(Attained.Age, knots = c(25, 62))4  6.214860
## bs(Attained.Age, knots = c(25, 62))5  8.516730
## Duration                          -0.830185
```

## Smoker.StatusSmoker	0.922567
## Smoker.StatusUnknown	0.183878
## GenderMale	-0.147184
## Face.Amount.Band10000-24999	-0.055012
## Face.Amount.Band100000-249999	-0.371124
## Face.Amount.Band1000000-2499999	-0.607737
## Face.Amount.Band10000000+	-9.826274
## Face.Amount.Band25000-49999	-0.121562
## Face.Amount.Band250000-499999	-0.541470
## Face.Amount.Band2500000-4999999	-0.901552
## Face.Amount.Band50000-99999	-0.200319
## Face.Amount.Band500000-999999	-0.538820
## Face.Amount.Band5000000-9999999	-0.412472
## bs(Attained.Age, knots = c(25, 62))1:Duration	0.820143
## bs(Attained.Age, knots = c(25, 62))2:Duration	0.887968
## bs(Attained.Age, knots = c(25, 62))3:Duration	0.838416
## bs(Attained.Age, knots = c(25, 62))4:Duration	0.855735
## bs(Attained.Age, knots = c(25, 62))5:Duration	0.843568
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker	-1.030222
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker	-0.299266
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker	1.097092
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker	-0.742927
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker	-2.796522
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown	0.072093
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown	0.752044
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown	0.592052
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown	0.875573
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown	-1.816552
## bs(Attained.Age, knots = c(25, 62))1:GenderMale	2.155471
## bs(Attained.Age, knots = c(25, 62))2:GenderMale	0.676932
## bs(Attained.Age, knots = c(25, 62))3:GenderMale	0.848032
## bs(Attained.Age, knots = c(25, 62))4:GenderMale	1.053431
## bs(Attained.Age, knots = c(25, 62))5:GenderMale	-1.517048
## Duration:Smoker.StatusSmoker	-0.014106
## Duration:Smoker.StatusUnknown	-0.025903
## Duration:GenderMale	-0.012787
## Smoker.StatusSmoker:GenderMale	-0.035678
## Smoker.StatusUnknown:GenderMale	-0.090312
##	Std. Error
## (Intercept)	0.534781
## bs(Attained.Age, knots = c(25, 62))1	0.641226
## bs(Attained.Age, knots = c(25, 62))2	0.536099
## bs(Attained.Age, knots = c(25, 62))3	0.559718
## bs(Attained.Age, knots = c(25, 62))4	0.541992
## bs(Attained.Age, knots = c(25, 62))5	0.666535
## Duration	0.093010
## Smoker.StatusSmoker	0.821699
## Smoker.StatusUnknown	0.510662
## GenderMale	0.464207
## Face.Amount.Band10000-24999	0.010478
## Face.Amount.Band100000-249999	0.016844
## Face.Amount.Band1000000-2499999	0.058513
## Face.Amount.Band10000000+	40.925215
## Face.Amount.Band25000-49999	0.013429

## Face.Amount.Band250000-499999	0.029664
## Face.Amount.Band2500000-4999999	0.223979
## Face.Amount.Band50000-99999	0.014644
## Face.Amount.Band500000-999999	0.040534
## Face.Amount.Band5000000-9999999	0.316501
## bs(Attained.Age, knots = c(25, 62))1:Duration	0.092984
## bs(Attained.Age, knots = c(25, 62))2:Duration	0.093834
## bs(Attained.Age, knots = c(25, 62))3:Duration	0.092572
## bs(Attained.Age, knots = c(25, 62))4:Duration	0.093482
## bs(Attained.Age, knots = c(25, 62))5:Duration	0.092624
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker	0.960052
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker	0.806293
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker	0.855137
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker	0.817046
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker	1.029565
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown	0.636730
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown	0.502387
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown	0.561384
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown	0.514331
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown	0.709111
## bs(Attained.Age, knots = c(25, 62))1:GenderMale	0.569735
## bs(Attained.Age, knots = c(25, 62))2:GenderMale	0.455570
## bs(Attained.Age, knots = c(25, 62))3:GenderMale	0.499572
## bs(Attained.Age, knots = c(25, 62))4:GenderMale	0.468373
## bs(Attained.Age, knots = c(25, 62))5:GenderMale	0.679314
## Duration:Smoker.StatusSmoker	0.001799
## Duration:Smoker.StatusUnknown	0.001353
## Duration:GenderMale	0.001271
## Smoker.StatusSmoker:GenderMale	0.021046
## Smoker.StatusUnknown:GenderMale	0.069530
##	z value Pr(> z )
## (Intercept)	-13.909 < 2e-16
## bs(Attained.Age, knots = c(25, 62))1	-1.387 0.165360
## bs(Attained.Age, knots = c(25, 62))2	-2.222 0.026290
## bs(Attained.Age, knots = c(25, 62))3	2.728 0.006368
## bs(Attained.Age, knots = c(25, 62))4	11.467 < 2e-16
## bs(Attained.Age, knots = c(25, 62))5	12.778 < 2e-16
## Duration	-8.926 < 2e-16
## Smoker.StatusSmoker	1.123 0.261541
## Smoker.StatusUnknown	0.360 0.718790
## GenderMale	-0.317 0.751195
## Face.Amount.Band10000-24999	-5.250 1.52e-07
## Face.Amount.Band100000-249999	-22.033 < 2e-16
## Face.Amount.Band1000000-2499999	-10.386 < 2e-16
## Face.Amount.Band10000000+	-0.240 0.810250
## Face.Amount.Band25000-49999	-9.052 < 2e-16
## Face.Amount.Band250000-499999	-18.253 < 2e-16
## Face.Amount.Band2500000-4999999	-4.025 5.69e-05
## Face.Amount.Band50000-99999	-13.679 < 2e-16
## Face.Amount.Band500000-999999	-13.293 < 2e-16
## Face.Amount.Band5000000-9999999	-1.303 0.192498
## bs(Attained.Age, knots = c(25, 62))1:Duration	8.820 < 2e-16
## bs(Attained.Age, knots = c(25, 62))2:Duration	9.463 < 2e-16
## bs(Attained.Age, knots = c(25, 62))3:Duration	9.057 < 2e-16

```

## bs(Attained.Age, knots = c(25, 62))4:Duration          9.154 < 2e-16
## bs(Attained.Age, knots = c(25, 62))5:Duration          9.107 < 2e-16
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker -1.073 0.283231
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker -0.371 0.710516
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker  1.283 0.199512
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker -0.909 0.363200
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker -2.716 0.006603
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown 0.113 0.909853
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown 1.497 0.134409
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown 1.055 0.291595
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown 1.702 0.088689
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown -2.562 0.010415
## bs(Attained.Age, knots = c(25, 62))1:GenderMale         3.783 0.000155
## bs(Attained.Age, knots = c(25, 62))2:GenderMale         1.486 0.137305
## bs(Attained.Age, knots = c(25, 62))3:GenderMale         1.698 0.089599
## bs(Attained.Age, knots = c(25, 62))4:GenderMale         2.249 0.024504
## bs(Attained.Age, knots = c(25, 62))5:GenderMale        -2.233 0.025535
## Duration:Smoker.StatusSmoker                          -7.843 4.41e-15
## Duration:Smoker.StatusUnknown                       -19.143 < 2e-16
## Duration:GenderMale                                -10.057 < 2e-16
## Smoker.StatusSmoker:GenderMale                       -1.695 0.090033
## Smoker.StatusUnknown:GenderMale                     -1.299 0.193980
##
## (Intercept)                                           ***
## bs(Attained.Age, knots = c(25, 62))1
## bs(Attained.Age, knots = c(25, 62))2                *
## bs(Attained.Age, knots = c(25, 62))3                **
## bs(Attained.Age, knots = c(25, 62))4                ***
## bs(Attained.Age, knots = c(25, 62))5                ***
## Duration                                              ***
## Smoker.StatusSmoker
## Smoker.StatusUnknown
## GenderMale
## Face.Amount.Band10000-24999                          ***
## Face.Amount.Band100000-249999                        ***
## Face.Amount.Band1000000-2499999                      ***
## Face.Amount.Band10000000+
## Face.Amount.Band25000-49999                          ***
## Face.Amount.Band250000-499999                        ***
## Face.Amount.Band2500000-4999999                      ***
## Face.Amount.Band50000-99999                          ***
## Face.Amount.Band500000-999999                        ***
## Face.Amount.Band5000000-9999999
## bs(Attained.Age, knots = c(25, 62))1:Duration        ***
## bs(Attained.Age, knots = c(25, 62))2:Duration        ***
## bs(Attained.Age, knots = c(25, 62))3:Duration        ***
## bs(Attained.Age, knots = c(25, 62))4:Duration        ***
## bs(Attained.Age, knots = c(25, 62))5:Duration        ***
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusSmoker
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusSmoker **
## bs(Attained.Age, knots = c(25, 62))1:Smoker.StatusUnknown

```

```

## bs(Attained.Age, knots = c(25, 62))2:Smoker.StatusUnknown
## bs(Attained.Age, knots = c(25, 62))3:Smoker.StatusUnknown
## bs(Attained.Age, knots = c(25, 62))4:Smoker.StatusUnknown .
## bs(Attained.Age, knots = c(25, 62))5:Smoker.StatusUnknown *
## bs(Attained.Age, knots = c(25, 62))1:GenderMale ***
## bs(Attained.Age, knots = c(25, 62))2:GenderMale
## bs(Attained.Age, knots = c(25, 62))3:GenderMale .
## bs(Attained.Age, knots = c(25, 62))4:GenderMale *
## bs(Attained.Age, knots = c(25, 62))5:GenderMale *
## Duration:Smoker.StatusSmoker ***
## Duration:Smoker.StatusUnknown ***
## Duration:GenderMale ***
## Smoker.StatusSmoker:GenderMale .
## Smoker.StatusUnknown:GenderMale
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 376471 on 695111 degrees of freedom
## Residual deviance: 139053 on 695067 degrees of freedom
## AIC: 234946
##
## Number of Fisher Scoring iterations: 14
dat[,ExpectedDeaths:=predict(mod2,newdata=dat,type='response')]
dcast(dat[,.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),by=.(Set,Gender)],Gender

## Gender AtoEcount_Testing AtoEcount_Training Exposure_Testing
## 1: Female 1.002642 1 3058705
## 2: Male 0.986408 1 2238037
## Exposure_Training
## 1: 7059382
## 2: 5164596
dcast(dat[,.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths)),by=.(Set,Smoker.Status)]

## Smoker.Status AtoEcount_Testing AtoEcount_Training Exposure_Testing
## 1: NonSmoker 0.9943633 1 3317189.2
## 2: Smoker 0.9858622 1 529103.3
## 3: Unknown 1.0040091 1 1450449.9
## Exposure_Training
## 1: 7727522
## 2: 1230646
## 3: 3265811
dcast(dat[Attained.Age %between% c(55,75),.(Exposure=sum(Exposure),AtoEcount=sum(Deaths)/sum(ExpectedDeaths))

## Attained.Age AtoEcount_Testing AtoEcount_Training Exposure_Testing
## 1: 55 1.0132702 1.0762337 109920.32
## 2: 56 0.9954793 0.9826095 107593.83
## 3: 57 1.0998418 0.9610064 99771.87
## 4: 58 1.0595273 0.9898453 100453.01
## 5: 59 0.9526151 0.9941510 92027.86
## 6: 60 0.9742206 0.9935847 94487.06
## 7: 61 1.0067549 0.9563971 95334.26

```

```
## 8:      62      1.0130036      1.0306148      86624.04
## 9:      63      1.0534794      1.0293042      68385.06
## 10:     64      1.0819661      1.0073799      67007.47
## 11:     65      1.0200304      1.0322421      66712.85
## 12:     66      1.0123864      1.0321181      65446.42
## 13:     67      1.0327804      0.9907282      56513.60
## 14:     68      0.9639919      1.0220320      49999.15
## 15:     69      1.0393270      1.0067638      47131.22
## 16:     70      0.9671339      1.0956089      49001.44
## 17:     71      1.0351315      0.9714626      41827.19
## 18:     72      0.9161917      0.9645330      43430.38
## 19:     73      0.9754527      0.9897771      41661.49
## 20:     74      1.0093033      1.0087121      37307.40
## 21:     75      1.0331251      0.9787935      34822.62
##      Attained.Age AtoEcount_Testing AtoEcount_Training Exposure_Testing
##      Exposure_Training
## 1:      255020.18
## 2:      245303.27
## 3:      240546.97
## 4:      223251.52
## 5:      220556.96
## 6:      214252.16
## 7:      221700.08
## 8:      199585.80
## 9:      163853.79
## 10:     156172.64
## 11:     156428.86
## 12:     150424.29
## 13:     130692.79
## 14:     120226.50
## 15:     112353.83
## 16:     102214.57
## 17:     102298.81
## 18:      93754.59
## 19:      92141.64
## 20:      90117.15
## 21:      85351.35
##      Exposure_Training
```

## Graphs with Count CI

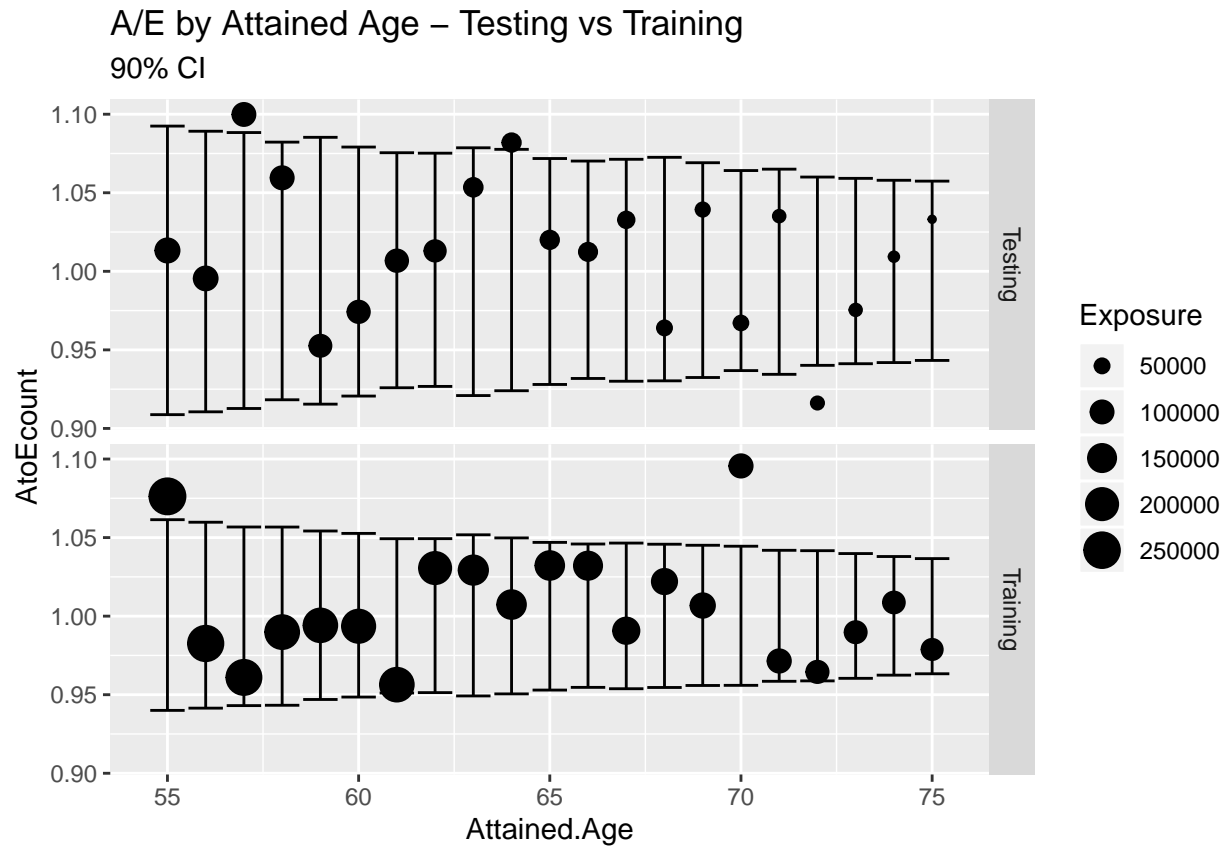
Simple example for 90% CI (5% each tail)

```
plotTable <- dat[Attained.Age %between% c(55,75),
  .(Exposure=sum(Exposure),ExpectedDeaths=sum(ExpectedDeaths),AtoEcount=sum(Deaths)/sum(
    by=. (Set,Attained.Age)) [order(Attained.Age)]

plotTable[,upperCI:=qpois(p=.95,lambda=ExpectedDeaths,lower.tail = TRUE)/ExpectedDeaths]
plotTable[,lowerCI:=qpois(p=.95,lambda=ExpectedDeaths,lower.tail = FALSE)/ExpectedDeaths]

ggplot(data=plotTable,aes(x=Attained.Age,y=AtoEcount,ymin=lowerCI,ymax=upperCI)) +
  geom_point(aes(size=Exposure))+
  geom_errorbar()+
```

```
facet_grid(Set~.)+
labs(title='A/E by Attained Age - Testing vs Training',
      subtitle='90% CI')
```



```
plotTable <- dat[Attained.Age %between% c(55,75),
  .(Exposure=sum(Exposure),ExpectedDeaths=sum(ExpectedDeaths),AtoEcount=sum(Deaths)/sum(
    by=. (Set,Gender,Attained.Age)) [order(Attained.Age)]

plotTable[,upperCI:=qpois(p=.95,lambda=ExpectedDeaths,lower.tail = TRUE)/ExpectedDeaths]
plotTable[,lowerCI:=qpois(p=.95,lambda=ExpectedDeaths,lower.tail = FALSE)/ExpectedDeaths]

ggplot(data=plotTable,aes(x=Attained.Age,y=AtoEcount,ymin=lowerCI,ymax=upperCI,color=Gender)) +
  geom_point(aes(size=Exposure))+
  geom_errorbar()+
  facet_grid(Set~.)+
  labs(title='A/E by Attained Age - Testing vs Training',
        subtitle='90% CI')
```

# A/E by Attained Age – Testing vs Training

90% CI

