

MCMC στα Γενικευμένα Γραμμικά Μοντέλα

- Response Variable (Y): Εξαρτημένη Μεταβλητή.
- Explanatory Variables (X_j): Επεξηγηματικές Μεταβλητές.
- Random Component (Τυχαία Συνιστώσα).
 - $Y | (X_1, \dots, X_p) \sim \text{ΚΑΤΑΝΟΜΗ}(\boldsymbol{\theta})$.
 - $\boldsymbol{\theta}$: διάνυσμα παραμέτρων,
 - $\mu = \mu(\boldsymbol{\theta})$: the mean of the above distribution (function of $\boldsymbol{\theta}$).
- Systematic Component (Συστηματική Συνιστώσα).
 - $n = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$.
 - n : γραμμικός προσδιορισμός του μοντέλου (linear predictor).
- Link Function (Συνδετική Συνάρτηση). Ο τρόπος σύνδεσης της δεσμευμένης μέσης τιμής της τυχαίας συνιστώσας με το γραμμικό προσδιορισμό. $g(\mu) = n = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$.

MCMC στα Γενικευμένα Γραμμικά Μοντέλα

- Συνήθως έχουμε ένα τυχαίο δείγμα μεγέθους n $(Y_i, X_{1i}, \dots, X_{pi}; i=1, \dots, n)$, με την βοήθεια του οποίου εκτιμούμε τα β_0, \dots, β_p .
- Τότε $Y_i | (X_{1i}, \dots, X_{pi}) \sim \text{ΚΑΤΑΝΟΜΗ}(\boldsymbol{\theta}_i)$ (για λόγους απλότητας δεν θα βάζουμε την δέσμευση στις επόμενες διαφάνειες) και $g(\mu_i) = \eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$, $i = 1, 2, \dots, n$, όπου $\mu_i = E[Y_i | (X_{1i}, \dots, X_{pi})] = \mu_i(\boldsymbol{\theta}_i)$.
- Η g συνήθως επιλέγεται με τέτοιο τρόπο ώστε να ικανοποιούνται τυχών περιορισμοί στο μ .
- Στην Μπεϋζιανή Στατιστική τα β_0, \dots, β_p είναι τυχαίες μεταβλητές οπότε αφού τους τοποθετήσουμε εκ των προτέρων κατανομές, σκοπός μας είναι ο υπολογισμός των εκ των υστέρων κατανομών τους.

Κανονικό Μοντέλο

- Τυχαία συνιστώσα:
 - $Y_i \sim N(\mu_i, \sigma^2)$ συνεχής μεταβλητή.
 - $\mu_i \rightarrow$ μέση τιμή
- Επεξηγηματικές Μεταβλητές:
 - X_j συνεχείς ή διακριτές, $j=1, \dots, p$.
- Συνδετική συνάρτηση:
 - $\mu_i = \eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$.
 - Άρα $g(\mu_i) = \mu_i$.

Μοντέλο Bernoulli

- Τυχαία συνιστώσα:
 - $Y_i \sim \text{Bernoulli}(p_i)$ δίτιμη μεταβλητή (0/1).
 - $p_i \rightarrow$ μέση τιμή
- Επεξηγηματικές Μεταβλητές:
 - X_j συνεχείς ή διακριτές, $j=1, \dots, p$.
- Συνδετική συνάρτηση:
 - $\log(p_i/(1-p_i)) = \ln = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$.
 - Άρα $g(p_i) = \text{logit}(p_i)$.

Διωνυμικό Μοντέλο

- Τυχαία συνιστώσα:
 - $Y_i \sim \text{Binomial}(n_i, p_i)$: # επιτυχιών σε n_i (ανεξάρτητες) επαναλήψεις.
 - Y_i : άθροισμα n_i ανεξάρτητων Bernoulli τ.μ. με μέση τιμή p_i .
- Επεξηγηματικές Μεταβλητές:
 - X_j συνεχείς ή διακριτές, $j=1, \dots, p$.
- Συνδετική συνάρτηση:
 - $\log(p_i/(1-p_i)) = \eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$.
 - Άρα $g(p_i) = \text{logit}(p_i)$.

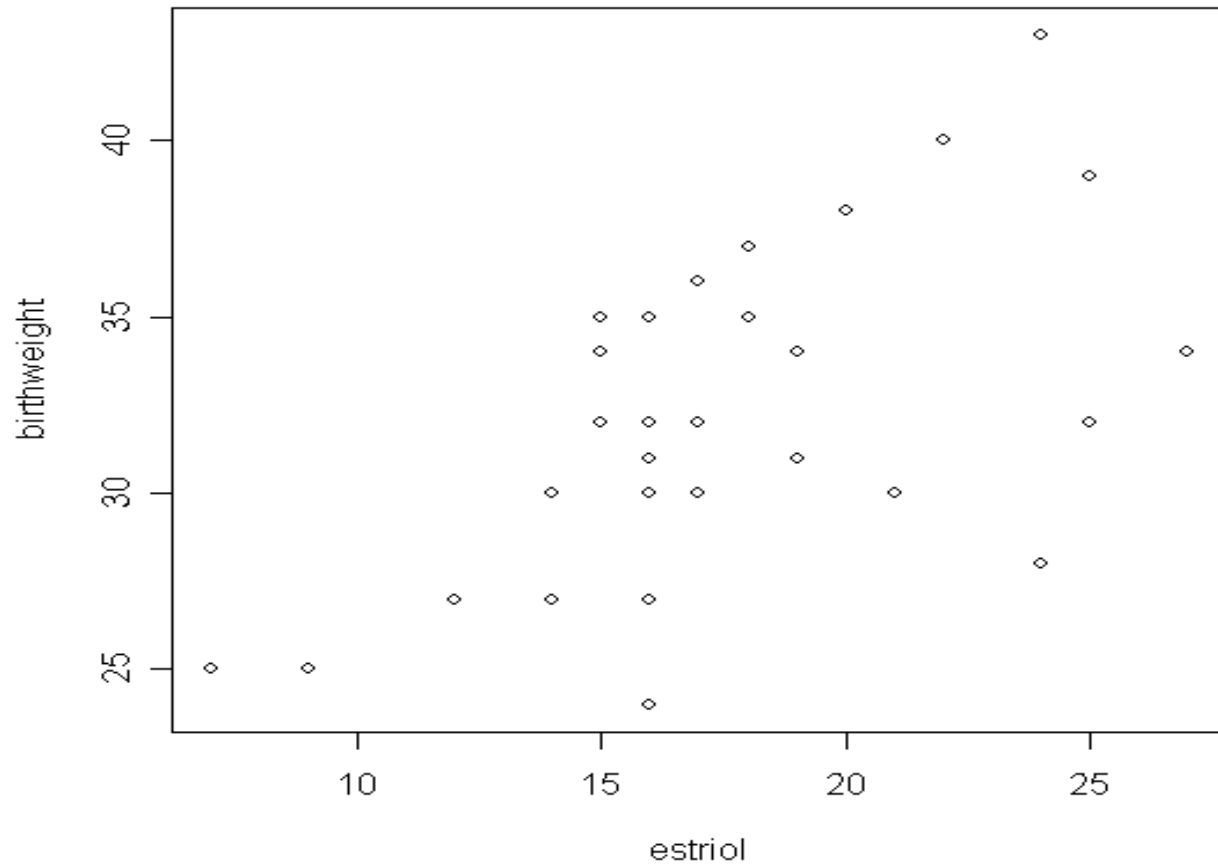
Μοντέλο Poisson

- Τυχαία συνιστώσα:
 - $Y_i \sim \text{Poisson}(\lambda_i)$: # επιτυχιών σε ένα χρονικό/χωρικό διάστημα.
 - $\lambda_i \rightarrow$ μέση τιμή (και διασπορά).
- Επεξηγηματικές Μεταβλητές:
 - X_j συνεχείς ή διακριτές, $j=1, \dots, p$.
- Συνδετική συνάρτηση:
 - $\log(\lambda_i) = \eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$.
 - Άρα $g(\lambda_i) = \log(\lambda_i)$.

Παράδειγμα 1^ο

- Green & Touchston (1963, Am. Jour. Of Obstetrics & Gynecology).
- Y : βάρος γέννησης παιδιού
- X : επίπεδο οιστριόλης (estriol) των εγκύων γυναικών.
- $n=31$.

Παράδειγμα 1^ο



Παράδειγμα 1^ο

- Τυχαία συνιστώσα: $\text{birth}_i \sim N(\mu_i, \sigma^2)$.
- Επεξηγηματική μεταβλητή: Estriol.
- Συνδετική συνάρτηση:
 - $\mu_i = \alpha + \beta \times \text{estriol}_i$
- Priors (Diffuse).
 - $\alpha \sim N(0, 10^4)$
 - $\beta \sim N(0, 10^4)$
 - $\tau = \sigma^{-2} \sim \text{Gamma}(10^{-4}, 10^{-4})$

Παράδειγμα 1^ο

```
model estriol1;
{

# definition of likelihood function
#
for (i in 1:n) {
birth[i]~dnorm( mu[i], tau ); # random component
mu[i]<-a+b*estriol[i];      # systematic component
                             # & link function
}

# prior distributions
#
a~dnorm( 0, 1.0E-04 ); # normal prior for a
b~dnorm( 0, 1.0E-04 ); # normal prior for b
tau~dgamma( 1.0E-04 , 1.0E-04 ); # gamma prior for precision
s2<-1/tau;
}
```

Παράδειγμα 1^ο

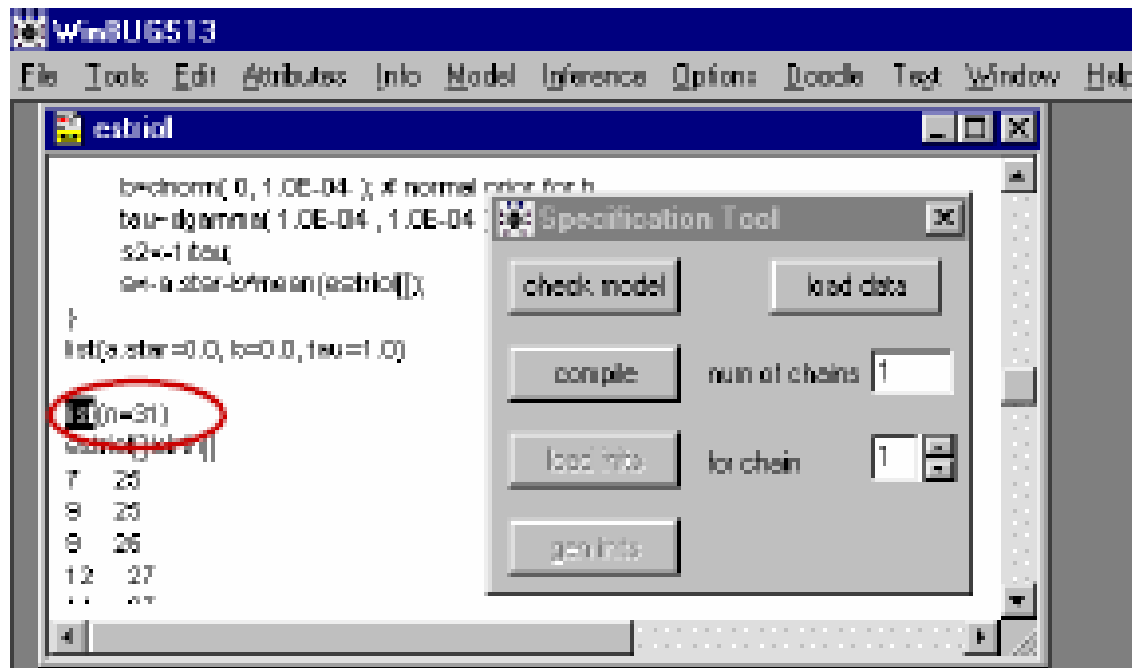
```
list(a=0.0, b=0.0, tau=1.0) # initial
values
list(n=31) # data
estriol[] birth[]
7 25
9 25
9 25
12 27
14 27
.....
20 38
22 40
25 39
24 43
END
```

Hit Return

Παράδειγμα 1^ο

Φόρτωση Δεδομένων:

ΜΑΥΡΙΖΟΥΜΕ ΤΟ **list** Η ΤΗΝ 1^η ΓΡΑΜΜΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ



The screenshot shows a software window titled 'winBUG513' with a menu bar (File, Tools, Edit, Attributes, Info, Model, Inference, Options, Load, Test, Window, Help). The main window contains a code editor with the following code:

```
lccnorm( 0, 1.0E-04 ); # normal prior for h
tau~dgamma( 1.0E-04 , 1.0E-04 )
s2~t(tau)
cc~a.star~b.mean(estrio());
}
list(a.star=0.0, b=0.0, tau=1.0)
list(n=31)
estrio[1:n] {}
```

The line `list(n=31)` is circled in red. Below the code, a table of data is visible:

7	25
8	25
9	26
12	27
..	..

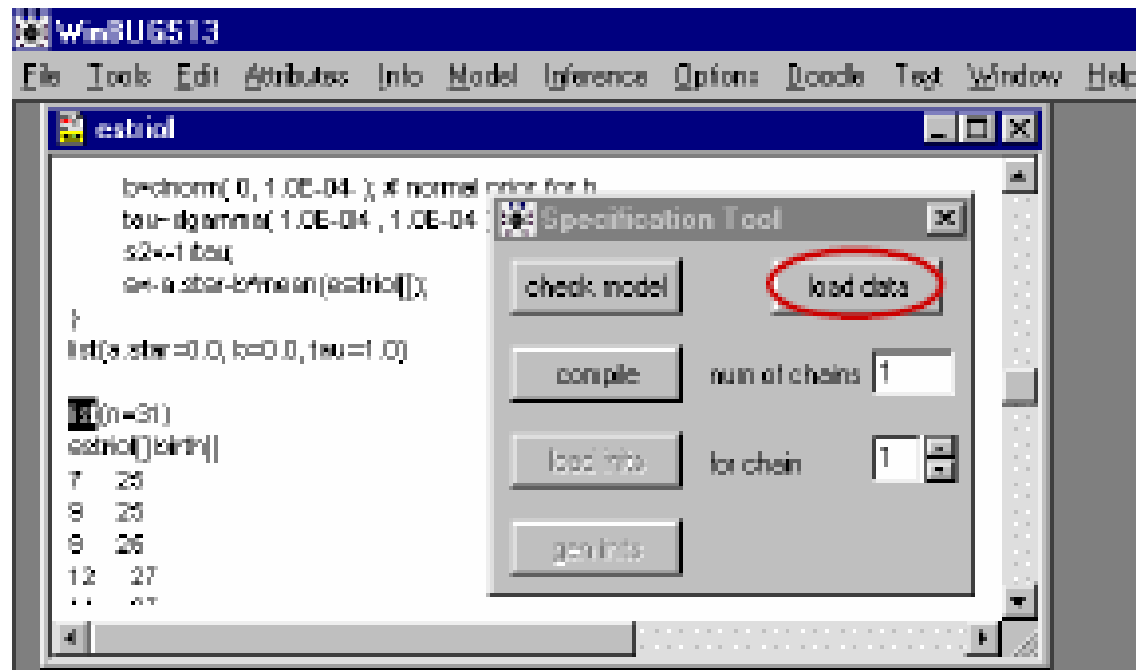
Overlaid on the right side of the code editor is a 'Specification Tool' dialog box with the following controls:

- Buttons: check model, load data, compile, load info, gen info
- num of chains: 1
- for chain: 1

Παράδειγμα 1^ο

Φόρτωση Δεδομένων:

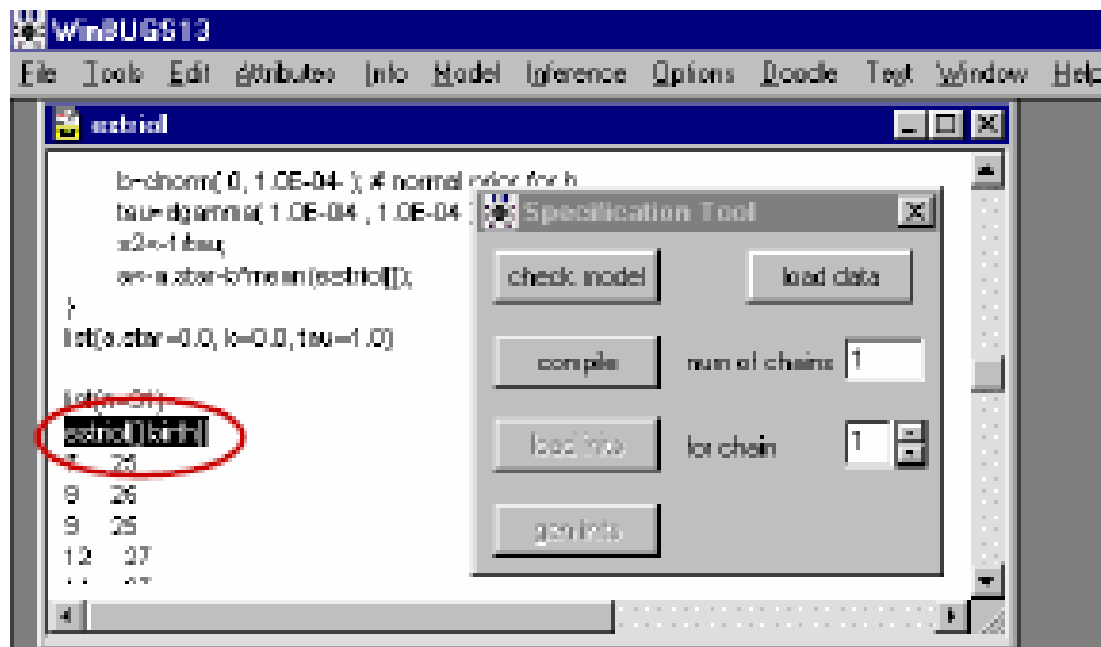
ΕΠΙΛΕΓΟΥΜΕ ΤΟ ΚΟΥΤΙ "LOAD DATA"



Παράδειγμα 1^ο

Φόρτωση Δεδομένων:

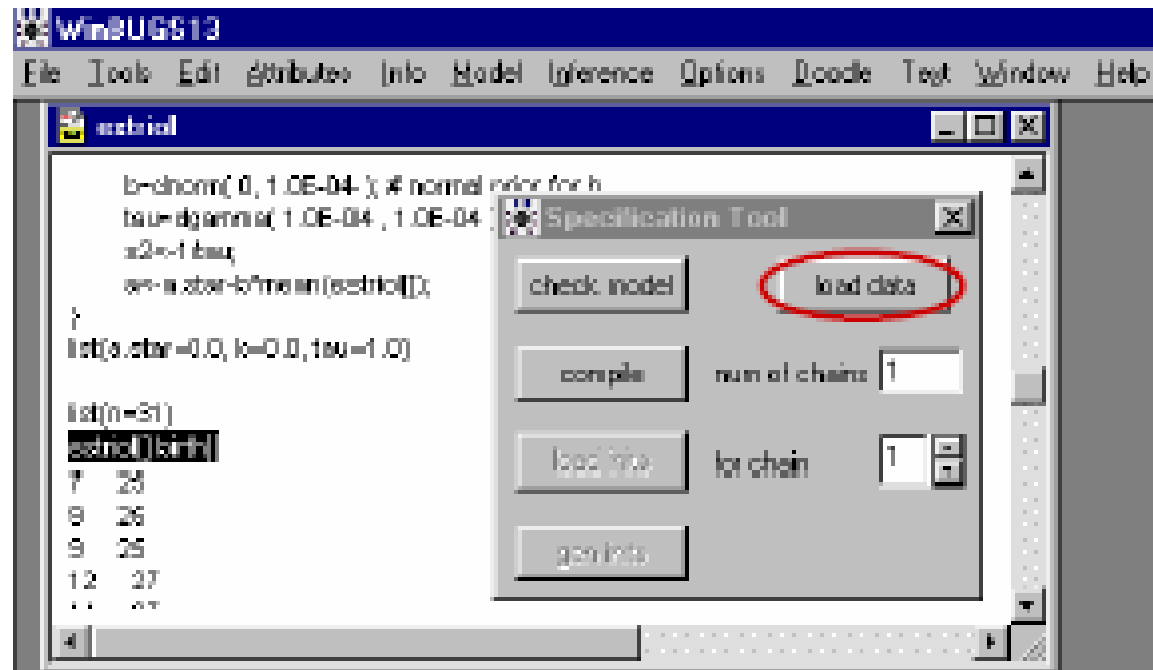
ΜΑΥΡΙΖΟΥΜΕ ΕΠΙΠΛΕΟΝ ΔΕΔΟΜΕΝΑ ΓΙΑ ΦΟΡΤΩΜΑ



Παράδειγμα 1^ο

Φόρτωση Δεδομένων:

ΞΑΝΑΕΠΙΛΕΓΟΥΜΕ ΤΟ ΚΟΥΤΙ "LOAD DATA"



Παράδειγμα 1^ο

- Χρησιμοποιούμε 1000 επαναλήψεις για burn-in και $m=1000$.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
a	21.49	2.757	0.09319	16.4	21.44	26.84	1	1000
b	0.6092	0.152	0.00493	0.3262	0.6101	0.8987	1	1000
s2	15.64	4.432	0.1353	9.314	14.84	27.31	1	1000

Παράδειγμα 2^ο

- Τυχαία συνιστώσα:
 - $\text{death.pneumonia}_i \sim B(p_i)$ (δίτιμη 0/1)
- Επεξηγηματικές μεταβλητές:
 - X_1, \dots, X_{14} .
- Συνδετική συνάρτηση:
 - $\log(p_i/(1-p_i)) = \eta_i = \beta_0 + \beta_1 X_{1_i} + \dots + \beta_{14} X_{14_i}$.
- Priors (Diffuse).
 - $\beta_0 \sim N(0, 10^4), \dots, \beta_{14} \sim N(0, 10^4)$.

Παράδειγμα 2^ο

X1	X2	X3	X4	X5	X6	X7
Total APACHE II Score	Age	Systolic Blood Pressure	Chest X-ray congestive heart failure score	Blood Urea Nitrogen	APACHE II coma score	Serum Albumin
X8	X9	X10	X11	X12	X13	X14
Shortness of breath	Respiratory distress	Septic complications	Prior respiratory failure	Recently hospitalized	Ambulatory score	temperature

Παράδειγμα 2^ο

```
model pneumonia;
```

```
{
```

```
  for (i in 1:n) {
```

```
    Y[i] ~ dbern(p[i]);
```

```
    logit(p[i]) <- beta0 + X1[i] * beta1 + X2[i] * beta2 + X3[i] * beta3 + X4[i] * beta4 +  
      X5[i] * beta5 + X6[i] * beta6 + X7[i] * beta7 + X8[i] * beta8 + X9[i] * beta9 + X10[i] *  
      beta10 + X11[i] * beta11 + X12[i] * beta12 + X13[i] * beta13 + X14[i] * beta14;
```

```
  }
```

```
  beta0 ~ dnorm(0, 1.0E-04);
```

```
  beta1 ~ dnorm(0, 1.0E-04);
```

```
  beta2 ~ dnorm(0, 1.0E-04);
```

```
  beta3 ~ dnorm(0, 1.0E-04);
```

```
  beta4 ~ dnorm(0, 1.0E-04);
```

```
  beta5 ~ dnorm(0, 1.0E-04);
```

```
  beta6 ~ dnorm(0, 1.0E-04);
```

```
  beta7 ~ dnorm(0, 1.0E-04);
```

```
  beta8 ~ dnorm(0, 1.0E-04);
```

```
  beta9 ~ dnorm(0, 1.0E-04);
```

```
  beta10 ~ dnorm(0, 1.0E-04);
```

```
  beta11 ~ dnorm(0, 1.0E-04);
```

```
  beta12 ~ dnorm(0, 1.0E-04);
```

```
  beta13 ~ dnorm(0, 1.0E-04);
```

```
  beta14 ~ dnorm(0, 1.0E-04);
```

```
}
```

Παράδειγμα 2^ο

```
# initial values
list(beta0 = 0.0, beta1 = 0.0, beta2 = 0.0 beta3 = 0.0, beta4 = 0.0, beta5 = 0.0, beta6 =
  0.0,
beta7 = 0.0, beta8 = 0.0, beta9 = 0.0, beta10 = 0.0, beta11 = 0.0, beta12 = 0.0, beta13
  = 0.0, beta14 = 0.0)

# data
list(n=2532)
Y[] X1[] X2[] X3[] X4[] X5[] X6[] X7[] X8[] X9[] X10[] X11[] X12[] X13[] X14[]
0 -0.416523 0.5028569 0 -0.5170125 0.7042943 -0.8563629 -0.6164393 1 0 0 0 1 -
  0.5987173 -1.127952
0 -0.2482658 1.625088 0 -0.5170125 0.7572356 -0.8563629 -0.6164393 1 0 0 0 1 -
  0.5987173 0.2333229
0 1.939078 -0.7440667 1 -0.5170125 1.233707 0.8852523 2.002393 1 1 0 0 1 -0.5987173
  2.518391
1 0.5930203 0.5028569 1 -0.5170125 2.080767 -0.8563629 -0.6164393 1 0 0 0 0 -
  0.5987173 1.205686
0 -0.7530375 0.3781645 0 -0.5170125 -0.5662958 -0.8563629 -0.6164393 0 0 0 0 0 -
  0.5987173 1.254343
0 -0.2482658 -0.7440667 0 1.052623 -0.3545308 -0.8563629 -0.6164393 1 0 0 0 0
  0.7731852 0.9140019
.....
0 0.2565059 -0.868759 0 -0.5170125 0.9160594 -0.8563629 -0.6164393 1 0 0 0 0
  0.7731852 -0.1070183
0 -0.0800086 0.876934 0 -0.5170125 0.0689993 0.8852523 1.129449 1 0 0 0 0 -
  0.5987173 -0.3987027
1 0.2565059 -0.494682 0 1.052623 0.1748818 0.8852523 2.002393 1 1 0 0 1 2.145088 -
  0.5445865
END
```

Παράδειγμα 2^ο

- Χρησιμοποιούμε 1000 επαναλήψεις για burn-in και $m=2000$.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	-3.059	0.1627	0.01056	-3.396	-3.05	-2.766	1000	2001
beta1	0.3654	0.1054	0.004737	0.1441	0.3664	0.5676	1000	2001
beta10	2.151	0.6403	0.01893	0.7995	2.167	3.391	1000	2001
beta11	0.808	0.3467	0.007954	0.1168	0.8264	1.431	1000	2001
beta12	0.4639	0.1536	0.004595	0.1618	0.4651	0.7493	1000	2001
beta13	0.2124	0.05942	0.00157	0.09662	0.2115	0.3367	1000	2001
beta14	-0.2715	0.06287	0.001616	-0.3952	-0.2714	-0.1501	1000	2001
beta2	0.3135	0.06936	0.001611	0.1765	0.3135	0.4463	1000	2001
beta3	1.001	0.1834	0.00496	0.6426	0.9995	1.351	1000	2001
beta4	0.1539	0.05773	0.00111	0.0392	0.1543	0.2648	1000	2001
beta5	0.2702	0.06488	0.001667	0.1432	0.2699	0.4004	1000	2001
beta6	0.4281	0.1002	0.004193	0.2335	0.4283	0.6236	1000	2001
beta7	0.231	0.06	0.001439	0.1123	0.2304	0.3449	1000	2001
beta8	0.6681	0.1708	0.009972	0.3575	0.6572	1.016	1000	2001
beta9	0.5089	0.1713	0.005449	0.1752	0.5126	0.8394	1000	2001

Παράδειγμα 3^ο

- Διεξήγαμε ένα πείραμα με σκοπό να μετρήσουμε την πιθανότητα ανάπτυξης ενός συγκεκριμένου όγκου σε ποντίκια, ανάλογα με την δόση ενός συγκεκριμένου φαρμάκου την οποία χορηγούμε σε αυτά και είναι ενάντια στην ανάπτυξη του συγκεκριμένου όγκου. Για τον σκοπό αυτό συλλέξαμε ένα δείγμα ποντικιών και το χωρίσαμε σε τρεις ομάδες . Στην πρώτη ομάδα, που αποτελείται από 14 ποντίκια, δεν χορηγήσαμε καθόλου φάρμακο και παρατηρήθηκαν 4 ποντίκια που είχαν αναπτύξει όγκο. Στην δεύτερη ομάδα, που αποτελείται από 34 ποντίκια, χορηγήσαμε μια συγκεκριμένη δόση φαρμάκου και παρατηρήθηκαν 4 ποντίκια που είχαν αναπτύξει όγκο. Τέλος στην τρίτη ομάδα, που αποτελείται από 34 ποντίκια, χορηγήσαμε την διπλάσια δόση φαρμάκου από αυτή που χορηγήσαμε στην δεύτερη ομάδα και παρατηρήθηκαν 2 ποντίκια που είχαν αναπτύξει όγκο. Σκοπός μας είναι η εκτίμηση της αναλογίας κατά την οποία η πιθανότητα ανάπτυξης του όγκου αυξάνεται ή μειώνεται ως συνάρτηση του επιπέδου δόσης του φαρμάκου, που χορηγείται στα ποντίκια.

Παράδειγμα 3^ο

- Τυχαία συνιστώσα:
 - $Y_i \sim \text{Binomial}(n_i, p_i)$: # ποντικιών που ανέπτυξαν όγκο στην ομάδα i .
- Επεξηγηματικές μεταβλητές:
 - X (δόση φαρμάκου).
- Συνδετική συνάρτηση:
 - $\log(p_i/(1-p_i)) = \eta_i = \beta_0 + \beta_1 X_i$.
- Priors (Diffuse).
 - $\beta_0 \sim N(0, 10^4)$.
 - $\beta_1 \sim N(0, 10^4)$.

Παράδειγμα 3^ο

```
model rats;
{
  for(i in 1:N) {
    y[i] ~dbin(p[i],n[i])
    logit(p[i]) <-beta0 + beta1*dose[i]
  }
  beta0 ~ dnorm(0,1.0E-6)
  beta1 ~ dnorm(0,1.0E-6)
}

# initial values
list(beta0=0, beta1=0)

# data
list(N=3, n=c(14, 34, 34), y=c(4, 4, 2), dose=c(0,1,2))
```


Παράδειγμα 3^ο

- Χρησιμοποιούμε 1000 επαναλήψεις για burn-in και $m=1000$.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
beta0	-1.022	0.5663	0.02543	-2.2	-0.97	0.02332	1001	1000
beta1	-0.9855	0.4891	0.02406	-1.974	-0.9926	-0.027	1001	1000
p[1]	0.2781	0.1051	0.004713	0.09976	0.2749	0.5058	1001	1000
p[2]	0.1238	0.03948	0.001078	0.06003	0.1207	0.2101	1001	1000
p[3]	0.05655	0.03394	0.001349	0.0125	0.04974	0.1445	1001	1000

Παράδειγμα 4^ο

- Σε μία έρευνα που πραγματοποιήθηκε (Hendriksen et al. 1984), 572 ηλικιωμένοι που ζούσαν σε χωριά στην Δανία χωρίστηκαν τυχαία σε δύο ομάδες. 287 στην ομάδα ελέγχου (control group C) στην οποία έλαβαν την συνήθης φροντίδα στο σπίτι, και 285 στην ομάδα πειράματος (experiment group E) στην οποία πέραν της συνήθης φροντίδας στο σπίτι λήφθηκαν υπόψιν οι φαρμακευτικές και κοινωνικές ανάγκες των ηλικιωμένων (in home geriatric assessment - **IHGA**). Σκοπός ήταν να ελεγχθεί αν το πρόγραμμα IHGA μειώνει των αριθμό εισαγωγών στο νοσοκομείο των ηλικιωμένων για τους επόμενα 2 χρόνια.

Group	Number of Hospitalizations									n	Mean	SD
	0	1	2	3	4	5	6	7				
Control	138	77	46	12	8	4	0	2		287	0.944	1.24
Experimental	147	83	37	13	3	1	1	0		285	0.768	1.01

Παράδειγμα 4^ο

- Με μια πρώτη ματιά φαίνεται το IHGA να έχει ρίξει τον μέσο όρο του αριθμού εισαγωγών στο νοσοκομείο κατά $(0.944 - 0.768) = 0.176$, το οποίο δηλώνει περίπου $100((0.768-0.944)/0.944) = 19\%$ μείωση από την ομάδα ελέγχου. Σαν μια πρώτη ανάλυση (**Model 0**) ας υποθέσουμε ότι τα δεδομένα μας ακολουθούν Κανονική κατανομή και θέλουμε να συγκρίνουμε τους μέσους (z-test).

$$(C_i | \mu_C, \sigma_C^2) \stackrel{\text{IID}}{\sim} N(\mu_C, \sigma_C^2), i = 1, \dots, n_C,$$

$$(E_j | \mu_E, \sigma_E^2) \stackrel{\text{IID}}{\sim} N(\mu_E, \sigma_E^2), j = 1, \dots, n_E,$$

Παράδειγμα 4^ο

```
> C <- c( rep( 0, 138 ), rep( 1, 77 ), rep( 2, 46 ),
          rep( 3, 12 ), rep( 4, 8 ), rep( 5, 4 ), rep( 7, 2 ) )

> print( n.C <- length( C ) )

[1] 287                # sample size in the control group

> mean( C )

[1] 0.9442509         # control group mean

> sqrt( var( C ) )

[1] 1.239089          # control group
                        # standard deviation (SD)

> table( C )

      0  1  2  3  4  5  7      # control group
138 77 46 12 8 4 2      # frequency distribution
```

Παράδειγμα 4^ο

```
> E <- c( rep( 0, 147 ), rep( 1, 83 ), rep( 2, 37 ),
          rep( 3, 13 ), rep( 4, 3 ), rep( 5, 1 ), rep( 6, 1 ) )

> print( n.E <- length( E ) )

[1] 285                # sample size in the
                       # experimental group

> mean( E )

[1] 0.7684211         # experimental group mean

> sqrt( var( E ) )

[1] 1.008268          # experimental group SD

> table( E )

  0  1  2  3  4  5  6      # experimental group
147 83 37 13 3 1 1      # frequency distribution
```


Παράδειγμα 4^ο

- Η ανάλυση του Model 0 που μόλις κάναμε είναι ισοδύναμη με την Μπεϋζιανή Ανάλυση του ίδιου μοντέλου με diffuse priors για τους μέσους και τυπικές αποκλίσεις των 2 ομάδων.

	Posterior		
	Mean	SD	95% Interval
Treatment effect ($\mu_E - \mu_C$)	-0.176	0.0944	(-0.361, 0.009)

- Παρατηρούμε όμως (επόμενες 2 διαφάνειες) ότι η υπόθεση της Κανονικής κατανομής δεν ήταν σωστή, καλύτερα να χρησιμοποιήσουμε Poisson.

Παράδειγμα 4^ο

```
> par( mfrow = c( 1, 2 ) )           # to plot two histograms

> hist( C, nclass = 8, probability = T,
      xlab = 'Days Hospitalized', ylab = 'Density',
      xlim = c( 0, 7 ), ylim = c( 0, 0.8 ) )

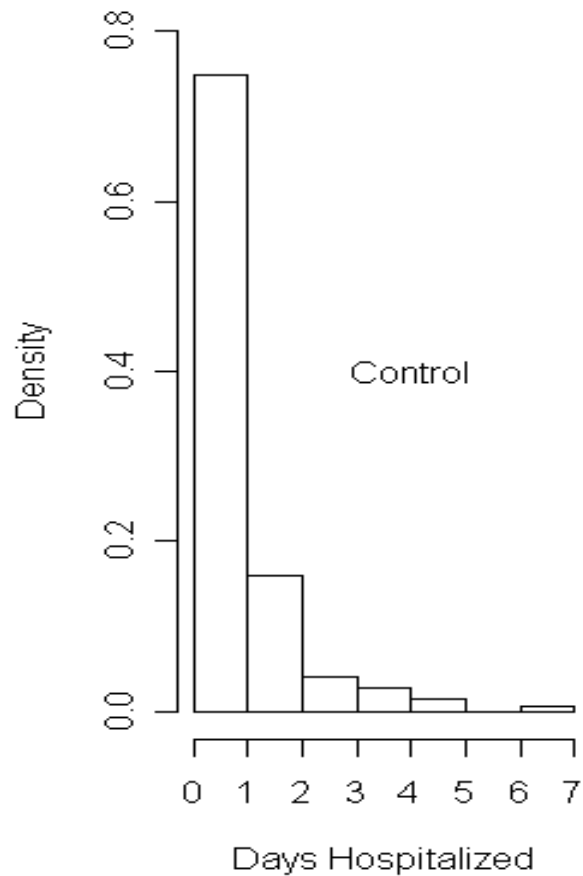
> text( 4, 0.4, 'Control' )

> hist( E, nclass = 8, probability = T,
      xlab = 'Days Hospitalized', ylab = 'Density',
      xlim = c( 0, 7 ), ylim = c( 0, 0.8 ) )

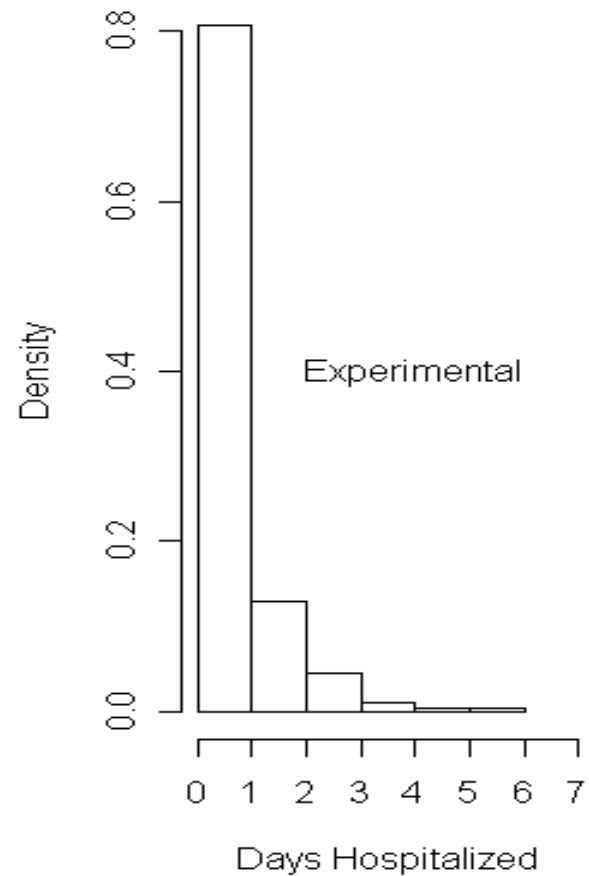
> text( 4, 0.4, 'Experimental' )
```


Παράδειγμα 4^ο

Histogram of C



Histogram of E



Παράδειγμα 4^ο

- Θα χρησιμοποιήσουμε λοιπόν Poisson με παραμέτρους λ_C και λ_E αντίστοιχα. Ως εκ των προτέρων κατανομές θα χρησιμοποιήσουμε την διακεχυμένη $\Gamma(0.001, 0.001)$ για να εκφράσουμε την πλήρη άγνοια μας για τις τιμές των λ_C και λ_E η οποία (όπως φαίνεται και στις παρακάτω δύο διαφάνειες) είναι επίπεδη στην περιοχή όπου η πιθανοφάνειες είναι μη αμελητέες.

Παράδειγμα 4^ο

```
> sqrt( var( C ) / n.C )
```

```
[1] 0.07314114
```

```
> sqrt( var( E ) / n.E )
```

```
[1] 0.05972466
```

```
> c( mean( C ) - 3.0 * sqrt( var( C ) / n.C ),  
     mean( C ) + 3.0 * sqrt( var( C ) / n.C ) )
```

```
[1] 0.7248275 1.1636743
```

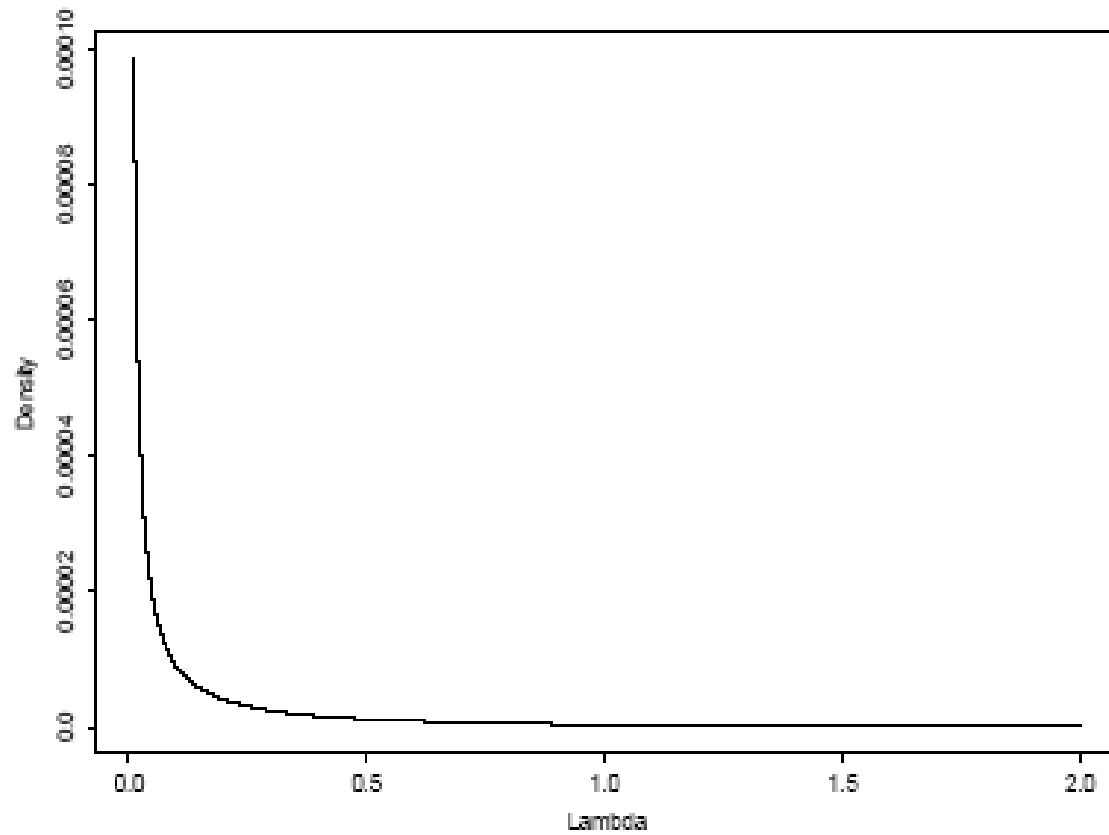
```
> c( mean( E ) - 3.0 * sqrt( var( E ) / n.E ),  
     mean( E ) + 3.0 * sqrt( var( E ) / n.E ) )
```

```
[1] 0.5892471 0.9475950
```

```
> lambda.grid <- seq( 0.01, 2.0, 0.01 )
```

```
> plot( lambda.grid, 0.001 * dgamma( lambda.grid, 0.001 ),  
        type = 'l', xlab = 'Lambda', ylab = 'Density' )
```

Παράδειγμα 4^ο



Παράδειγμα 4^ο

```
model poisson1;
{
lambda.C ~ dgamma(0.001, 0.001)
lambda.E ~ dgamma(0.001, 0.001)

for (i in 1:n.C){
    C[i] ~ dpois(lambda.C)
}

for (j in 1:n.E){
    E[j] ~ dpois(lambda.E)
}

effect <- lambda.E - lambda.C
}
```


Παράδειγμα 4^ο

- Χρησιμοποιούμε 2000 επαναλήψεις για burn-in και $m=8000$.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
effect	-0.174	0.07709	8.639E-4	-0.325	-0.1805	-0.024	2001	8000

Παράδειγμα 4^ο

λ_C Model	Posterior Mean	Posterior SD	Central 95% Interval
Gaussian	0.944	0.0731	(0.801, 1.09)
Poisson	0.943	0.0577	(0.832, 1.06)

λ_E Model	Posterior Mean	Posterior SD	Central 95% Interval
Gaussian	0.768	0.0597	(0.651, 0.885)
Poisson	0.769	0.0521	(0.671, 0.875)

$\Delta = \lambda_E - \lambda_C$ Model	Posterior Mean	Posterior SD	Central 95% Interval
Gaussian	-0.176	0.0944	(-0.361, 0.009)
Poisson	-0.174	0.0774	(-0.325, -0.024)

Τα δύο μοντέλα μας δίνουν σχεδόν πανομοιότυπα αποτελέσματα, με το Poisson μοντέλο να καταλήγει σε εκτιμήτριες με μικρότερο SD (το posterior SD για το treatment effect Δ είναι 22% μεγαλύτερο στο Gaussian Model).

Παράδειγμα 4^ο

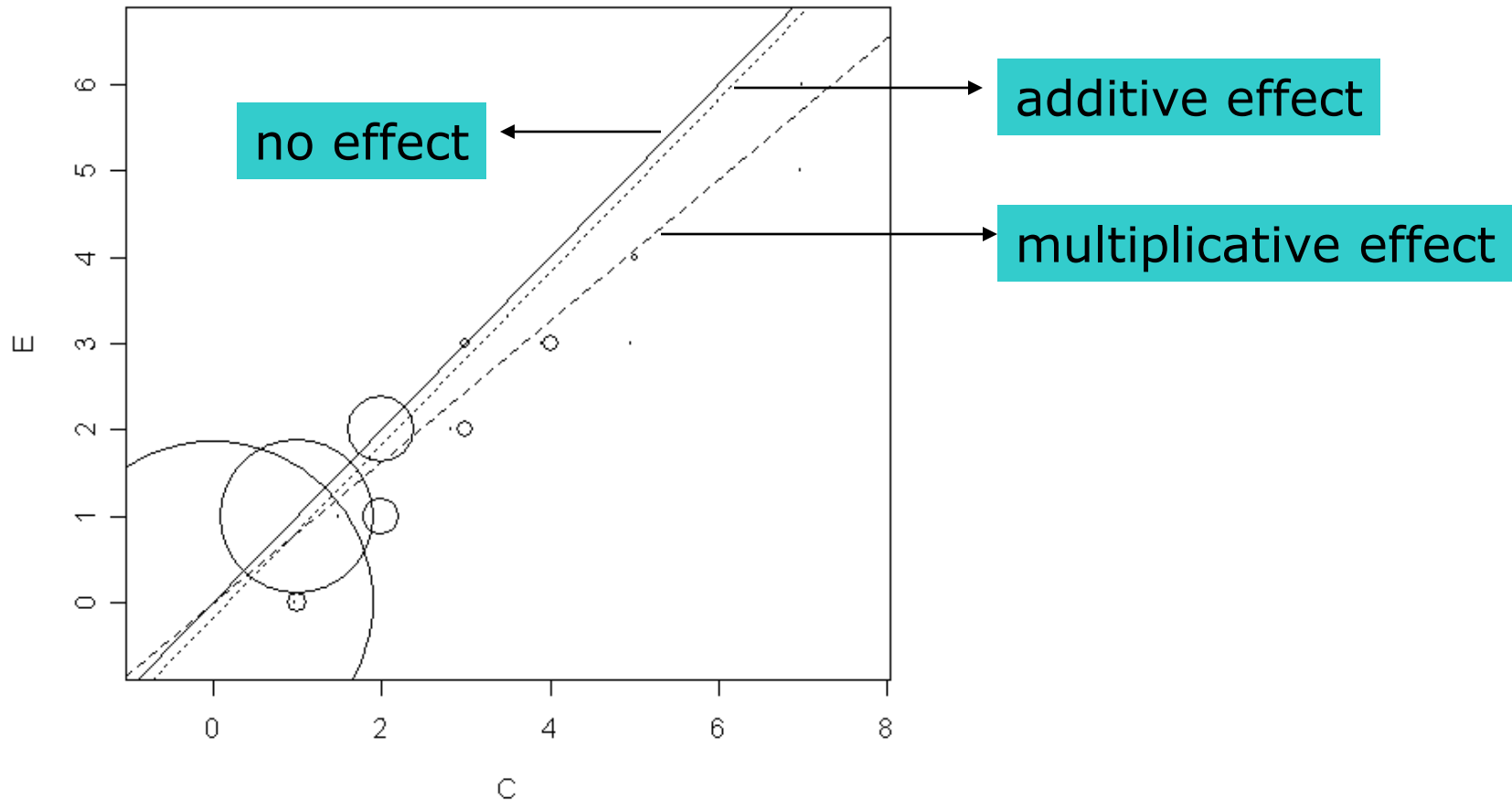
- Η προηγούμενη παρατήρηση, είναι σχεδόν ανάλογη με αυτήν που είχαμε κάνει για τα NB10 δεδομένα. Όταν ενδιαφερόμαστε για την παράμετρο θέσης, το Gaussian μοντέλο είναι “συντηρητική επιλογή” (το μοντέλο Poisson, κατακτά την επιπλέον ακρίβεια, από το γεγονός ότι η διασπορά και ο μέσος είναι ίσοι, γεγονός που επιτρέπει στην χρήση δευτέρων ροπών, πέραν από των πρώτων για την εκτίμηση του λ).
- Και στα δύο μοντέλα που χρησιμοποιήσαμε μέχρι τώρα υποθέσαμε ότι επίδραση από την θεραπευτική αγωγή ήταν αθροιστική (**additive model**), δηλαδή $E = C + \text{effect}$. Με άλλα λόγια πέραν από μία τυχαία απόκλιση, η επίδραση του IHGA είναι να προσθέσει ή να αφαιρέσει μια σταθερά στον αριθμό εισαγωγών στο νοσοκομείο κάθε ηλικιωμένου.
- Μια εναλλακτική μοντελοποίηση είναι να θεωρήσουμε ότι η θεραπευτική αγωγή που κάναμε επιδρά πολλαπλασιαστικά στο μοντέλο (**multiplicative model**), δηλαδή $E = (1 + \text{effect}) C$. Σε αυτή την περίπτωση, πέραν από μία τυχαία απόκλιση, η επίδραση του IHGA είναι να πολλαπλασιάζει τον αριθμό εισαγωγών στο νοσοκομείο κάθε ηλικιωμένου με μια σταθερά μεγαλύτερη ή μικρότερη της μονάδας. Με την βοήθεια της R συνάρτησης `aarplot`, μπορούμε κάποιες φορές να δούμε ποιο από τα 2 μοντέλα είναι καλύτερο.

Παράδειγμα 4^ο

- > `abline(0,1)` # $E = C$ (no effect)
- > `abline(0,0.816,lty=2)` # $E = 0.816 C$
 # (multiplicative)
- > `abline(-0.174,1,lty=3)` # $E = C - 0.174$ (additive)

Παράδειγμα 4^ο

Q-Qplot του E vs. C, με τις ακτίνες των κύκλων ανάλογες του αριθμού των παρατηρήσεων στο συγκεκριμένο σημείο.



Παράδειγμα 4^ο

- Ένας τρόπος να γράψουμε ένα πολλαπλασιαστικό μοντέλο είναι εκφράζοντας τα δεδομένα υπό μορφή παλινδρόμησης της εξαρτημένης μεταβλητής Y πάνω σε μια **εικονική μεταβλητή (dummy variable)** X , η οποία παίρνει την τιμή 1, αν ο ηλικιωμένος είναι στην ομάδα πειράματος (E) και 0 αν είναι στην ομάδα ελέγχου (C).

i	1	2	...	287	288	289	...	572
x_i	0	0	...	0	1	1	...	1
y_i	1	0	...	2	0	3	...	1

Παράδειγμα 4^ο

- Τότε για $i = 1, \dots, n=572$, το πολλαπλασιαστικό μοντέλο είναι
$$\begin{aligned}(y_i | \lambda_i) &\stackrel{\text{indep}}{\sim} \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \gamma_0 + \gamma_1 x_i \\ (\gamma_0, \gamma_1) &\sim \text{diffuse}\end{aligned}$$

- Για το παραπάνω μοντέλο οι ηλικιωμένοι τις ομάδας ελέγχου έχουν:

$$\log(\lambda_i) = \gamma_0 + \gamma_1(0) = \gamma_0, \text{ δηλαδή } \lambda_C = e^{\gamma_0}$$

- ενώ οι ηλικιωμένοι της ομάδας πειράματος

$$\log(\lambda_i) = \gamma_0 + \gamma_1(1) = \gamma_0 + \gamma_1,$$

$$\text{δηλαδή } \lambda_E = e^{\gamma_0 + \gamma_1} = e^{\gamma_0} e^{\gamma_1} = \lambda_C e^{\gamma_1} \approx \lambda_C (1 + \gamma_1),$$

για γ_1 κοντά στο 0.

Παράδειγμα 4^ο

```
model poisson2;
{
  gamma.0 ~ dnorm(0.0, 1.0E-4)
  gamma.1 ~ dnorm(0.0, 1.0E-4)

  for (i in 1:n){
    log(lambda[i]) <- gamma.0 + gamma.1 * x[i]
    y[i] ~ dpois(lambda[i])
  }
  lambda.C <- exp(gamma.0)
  lambda.E <- exp(gamma.0+gamma.1)
  mult.effect <- exp(gamma.1)
}
```


Παράδειγμα 4^ο

initial values

```
list(gamma.0 = 0.0, gamma.1 = 0.0)
```


Παράδειγμα 4^ο

- Χρησιμοποιούμε 2000 επαναλήψεις για burn-in και $m=8000$.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
mult.effect	0.816	0.0743	0.001146	0.676	0.811	0.967	2001	8000

Παράδειγμα 4^ο

λ_C Model	Posterior Mean	Posterior SD	Central 95% Interval
additive	0.943	0.0577	(0.832, 1.06)
multiplicative	0.945	0.0574	(0.837, 1.06)

λ_E Model	Posterior Mean	Posterior SD	Central 95% Interval
additive	0.769	0.0521	(0.671, 0.875)
multiplicative	0.768	0.0518	(0.671, 0.872)

effect Model	Posterior Mean	Posterior SD	Central 95% Interval
additive	-0.174	0.0774	(-0.325, -0.024)
multiplicative	-0.184	0.0743	(-0.324, -0.033)

↘ $\gamma_1 = 1 - \text{mult.effect}$

Πανομοιότυπα αποτελέσματα αφού οι επιδράσεις και στις 2 περιπτώσεις είναι κοντά στο 0.

Παράδειγμα 4^ο

- Ακόμα βέβαια δεν έχουμε ελέγξει αν το μοντέλο Poisson είναι λογικό για τα δεδομένα μας. Ένας τρόπος να ελέγξουμε αν το μοντέλο Poisson είναι λογικό είναι υπολογίζοντας τον **συντελεστή μεταβλητότητας** (σ^2/μ) για τα δεδομένα μας. Θυμίζουμε ότι όταν αν μια τ.μ. ακολουθεί την κατανομή Poisson(λ), τότε ο μέσος και η διασπορά συμπίπτουν και άρα ο συντελεστής μεταβλητότητας είναι 1.

```
> var( C ) / mean( C )
```

```
[1] 1.62599
```

```
> var( E ) / mean( E )
```

```
[1] 1.322979
```

Παράδειγμα 4^ο

- Παρατηρούμε λοιπόν ότι τα δεδομένα μας παρουσιάζουν **έξτρα Poisson μεταβλητότητα**, γεγονός το οποίο είναι λογικό αν το σκεφτούμε ως εξής. Αν υποθέσουμε ότι το μοντέλο Poisson είναι λογικό τότε είναι σαν να θεωρούμε ότι στην ομάδα ελέγχου C ο καθένας έχει τον ίδιο ρυθμό λ_c εισαγωγής στο νοσοκομείο. Όμοια και στην ομάδα πειράματος ο ρυθμός είναι σταθερός λ_E . Στην πραγματικότητα όμως είναι πιο λογικό να θεωρήσουμε ότι κάθε άτομο έχει τον δικό του ρυθμό εισαγωγής, στο νοσοκομείο ο οποίος μπορεί να εξαρτάται από παράγοντες όπως η ηλικία, η κατάσταση υγείας, κλπ. Όλους αυτούς του έξτρα παράγοντες που επηρεάζουν τον ρυθμό εισαγωγής μπορούμε να τους εισάγουμε στο μοντέλο μας υπό την μορφή ενός τυχαίου σφάλματος e (**random effect**).

$$\begin{aligned}(y_i | \lambda_i) &\stackrel{\text{indep}}{\sim} \text{Poisson}(\lambda_i) \\ \log(\lambda_i) &= \gamma_0 + \gamma_1 x_i + e_i \\ e_i &\stackrel{\text{IID}}{\sim} N(0, \sigma_e^2) \\ (\gamma_0, \gamma_1, \sigma_e^2) &\sim \text{diffuse.}\end{aligned}$$

Lognormal Mixture of
Poisson Distributions
OR
Random Effects Poisson
Regression (REPR)

Παράδειγμα 4^ο

- Στο παραπάνω μοντέλο ο ρυθμός εισαγωγής κάθε ασθενή είναι διαφορετικός. Για κάθε ηλικιωμένο δημιουργούμε ένα λ από την λογαριθμοκανονική κατανομή, και εν συνεχεία ο αριθμός εισαγωγών γ προέρχεται από μια Poisson με το συγκεκριμένο λ για τον ασθενή αυτόν. Η συγκεκριμένη αυτή διαδικασία θα έχει ως αποτέλεσμα η διασπορά του γ να είναι μεγαλύτερη του μέσου.

Παράδειγμα 4^ο

```
model poisson3;
{
gamma.0 ~ dnorm(0.0, 1.0E-4)
gamma.1 ~ dnorm(0.0, 1.0E-4)
tau.e ~ dgamma (0.001, 0.001)

for (i in 1:n){
  e[i] ~ dnorm(0.0, tau.e)
  log(lambda[i]) <- gamma.0 + gamma.1 * x[i] + e[i]
  y[i] ~dpois(lambda[i])
}
lambda.C <-exp(gamma.0)
lambda.E <-exp(gamma.0+gamma.1)
mult.effect <- exp(gamma.1)
sigma.e<-1.0/sqrt(tau.e)
}

# initial values

list(gamma.0 = -0.05, gamma.1 = -0.21, tau.e = 3.2)
```


Παράδειγμα 4^ο

- Για το συγκεκριμένο μοντέλο, υπάρχουν πέραν των σ_e^2 , γ_0 και γ_1 , n -τυχαίες επιδράσεις (random effects) e_i που πρέπει να προσομοιώσουμε. Για αυτό τον λόγο πέρα των αρχικών τιμών που θα ορίσουμε θα πρέπει να αφήσουμε το WinBugs να δημιουργήσει επιπλέον αρχικές τιμές.
- Στο συγκεκριμένο μοντέλο είναι κρίσιμη η επιλογή των αρχικών τιμών. Αν τρέξετε το μοντέλο με παράλογες αρχικές τιμές όπως `list(gamma.0 = -100.0, gamma.1 = 200.0, tau.e = 0.001)`, τότε θα πάρετε το μήνυμα "Rejection1", όταν θα προσπαθήσετε να προσομοιώσετε τιμές.

Παράδειγμα 4^ο

- Για τα γ_0 και γ_1 μπορούμε να χρησιμοποιήσουμε τα αποτελέσματα του πολλαπλασιαστικού μοντέλου Poisson για αρχικές τιμές. Το λ_C είχε posterior mean 0.945 και το λ_E είχε posterior mean 0.768.

$$\lambda_C = e^{\gamma_0} \Rightarrow \gamma_0 = \log(0.945) = -0.05$$

$$\lambda_E = \lambda_C e^{\gamma_1} \Rightarrow \gamma_1 = \log(0.768 / 0.945) = -0.21$$

Παράδειγμα 4^ο

- Για την αρχική τιμή του τ_e κάνουμε τους ακόλουθους υπολογισμούς (συμβολίζουμε τις τυχαίες επιδράσεις με n_i για να μην υπάρχει σύγχυση με τον αριθμό e).

$$V(y_i) = V[E(y_i | n_i)] + E[V(y_i | n_i)],$$

$$\text{όπου } (y_i | n_i) \sim \text{Poisson}(e^{\gamma_0 + \gamma_1 x_i + n_i}).$$

$$\text{Άρα } E(y_i | n_i) = V(y_i | n_i) = e^{\gamma_0 + \gamma_1 x_i + n_i}.$$

$$\text{Τότε } V[E(y_i | n_i)] = e^{2(\gamma_0 + \gamma_1 x_i)} V(e^{n_i}) \text{ και}$$

$$E[V(y_i | n_i)] = e^{\gamma_0 + \gamma_1 x_i} E(e^{n_i}).$$

$$\text{Αλλά } e^{n_i} \sim \text{LogNormal}(0, \sigma_n^2), \text{ οπότε } E(e^{n_i}) = e^{\frac{1}{2}\sigma_n^2}$$

$$\text{και } V(e^{n_i}) = e^{\sigma_n^2} (e^{\sigma_n^2} - 1). \text{ Αντικαθιστώντας προκύπτει:}$$

$$V(y_i) = e^{2(\gamma_0 + \gamma_1 x_i)} e^{\sigma_n^2} (e^{\sigma_n^2} - 1) + e^{\gamma_0 + \gamma_1 x_i} e^{\frac{1}{2}\sigma_n^2}.$$

Παράδειγμα 4^ο

- Θέτοντας $x_i = 0$ για την C ομάδα, της οποίας η διασπορά είναι 1.54, και χρησιμοποιώντας την τιμή $\gamma_0 = -0.05$ από το προηγούμενο μοντέλο, λύνουμε την προηγούμενη εξίσωση και καταλήγουμε στο

$$\sigma_n^2 \approx 0.31 \Rightarrow \tau_e \approx 3.2.$$

Παράδειγμα 4^ο

- Χρησιμοποιούμε 1000 επαναλήψεις για burn-in και $m=5000$.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Sigma.e	0.6813	0.0748	0.00518	0.5361	0.6778	0.8314	1001	5000

Το MC error για τον εκ των υστέρων μέσο του σ_e είναι 0.005182. Για να πάρουμε μια τιμή γύρω στο 0.001 πρέπει να αυξήσουμε τον αριθμό των επαναλήψεων m κατά ένα παράγοντα $(0.005182/0.001)^2 = 26.9$, δηλαδή να κάνουμε $26.9 \times 5000 = 134000$ επαναλήψεις.

Παράδειγμα 4^ο

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
Sigma.e	0.6747	0.0745	0.001114	0.5272	0.6748	0.8219	1001	134000

- Παρατηρούμε ότι υπάρχουν σαφείς αποδείξεις ότι το σ_e είναι διάφορο του μηδενός. Ο εκ των υστέρων μέσος του εκτιμάται στο 0.675 (με MC error 0.001 ύστερα από 134000 επαναλήψεις) και η τυπική απόκλιση στο 0.074. Άρα καταλήγουμε στο συμπέρασμα ότι η επέκταση που κάναμε στο μοντέλο μας προσθέτοντας τα random effects ήταν σωστή.
- Ένας διαφορετικός τρόπος περιγραφής της έξτρα Poisson μεταβλητότητας θα ήταν να χρησιμοποιήσουμε δύο Αρνητικές Διωνυμικές κατανομές για τις δύο ομάδες, ή να χρησιμοποιήσουμε την Γενικευμένη/Lagrangian Poisson κατανομή. Τα αποτελέσματα είναι παραπλήσια.

Παράδειγμα 4^ο

- Σύγκριση των αποτελεσμάτων για τον πολλαπλασιαστικό παράγοντα επίδρασης e^{γ_1} για τα fixed-effects και random-effects Poisson μοντέλα παλινδρόμησης (FEPR vs. REPS).

Model	Posterior Mean	Posterior SD	Central 95% Interval
FEPR	0.816	0.0735	(0.683, 0.969)
REPR	0.830	0.0921	(0.665, 1.02)

- Όπως είδαμε το σ_e έχει μέσο 0.68 ± 0.07 , και άρα έχουμε έξτρα Poisson μεταβλητότητα, γεγονός που μεταφράζεται σε αύξηση της αβεβαιότητας μας για τον πολλαπλασιαστικό παράγοντα (μεγαλύτερο SD στο REPR). Βάση λοιπόν του REPR μοντέλου το IHGA μειώνει τον μέσο αριθμό εισαγωγών για 2 χρόνια κατά $100(1-0.083)\% = 17\% \pm 9\%$ περίπου.