

TUGAS BIOSTATISTIK LANJUT

Mahasiswa: Ezra Hans Soputra

NIM: 131520230501

Prodi: Magister Epidemiologi

Lecturer: Dwi Agustian, dr., MPH, PhD

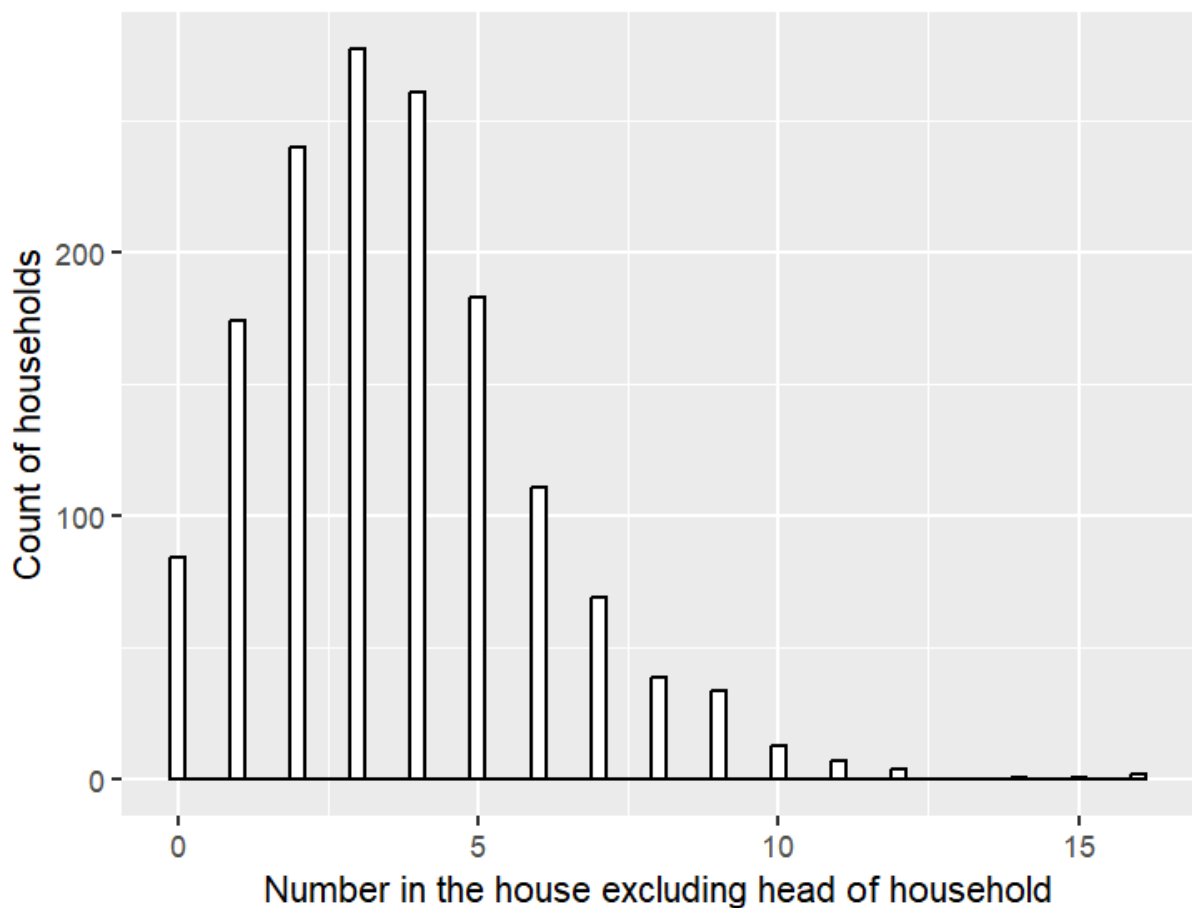
#Explaratory Data analysis

names(fHH1)

```
> names(fHH1)
[1] "No"      "location" "age"      "total"    "numLT5"   "roof"
```

#visualisasi

```
ggplot(fHH1, aes(total)) +
  geom_histogram(binwidth = .25, color = "black",
    fill = "white") +
  xlab("Number in the house excluding head of household") +
  ylab("Count of households")
```



```
#kategorisasi umur kepala keluarga
```

```
cuts = cut(fHH1$age,  
          breaks=c(15,20,25,30,35,40,45,50,55,60,65,70))
```

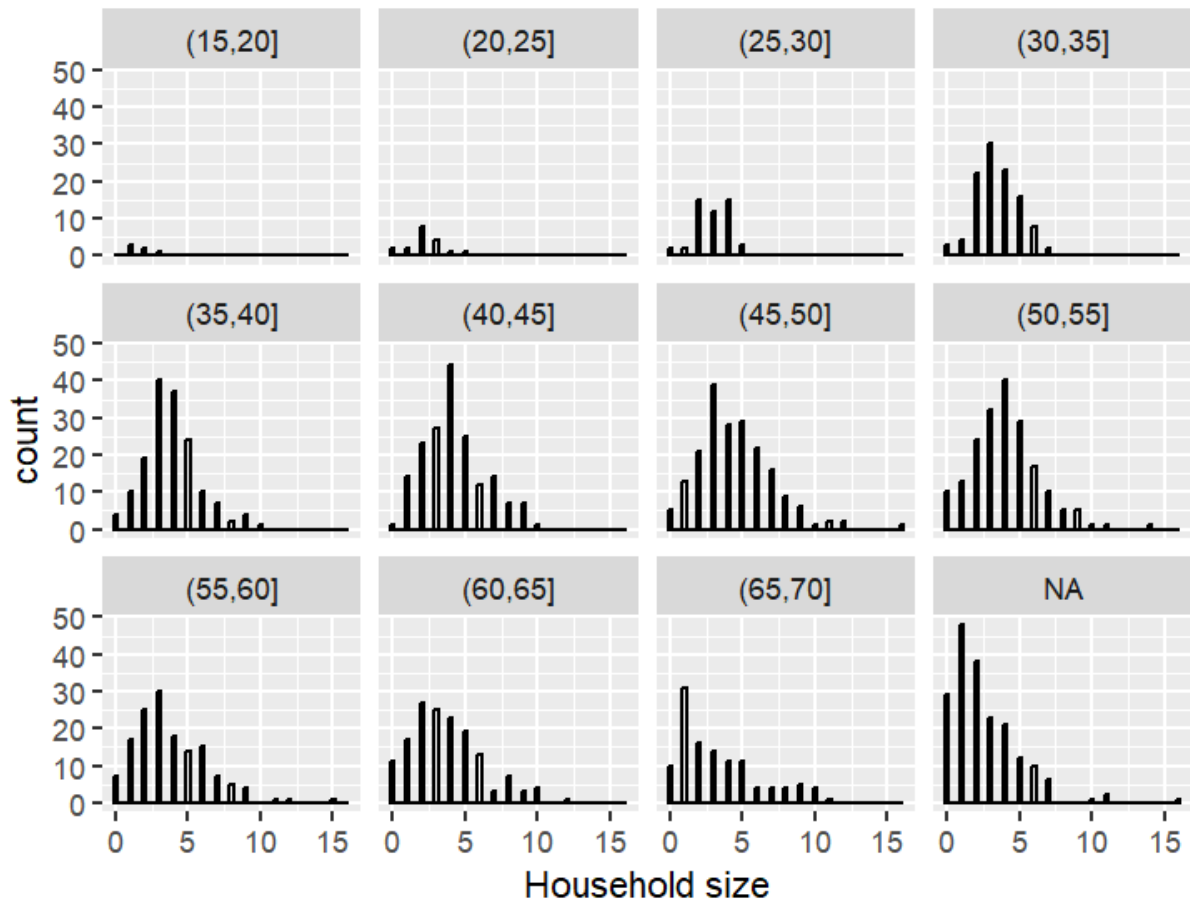
```
cuts      Factor w/ 11 levels "(15,20]","(20,25]..."
```

```
ageGrps <- data.frame(cuts,fHH1)
```

cuts	No	location	age	total	numLTS	roof
(60,65]	1	CentralLuzon	65	0	0	Predominantly Strong Material
NA	2	MetroManila	75	3	0	Predominantly Strong Material
(50,55]	3	DavaoRegion	54	4	0	Predominantly Strong Material
(45,50]	4	Visayas	49	3	0	Predominantly Strong Material
NA	5	MetroManila	74	3	0	Predominantly Strong Material
(55,60]	6	Visayas	59	6	0	Predominantly Strong Material
(50,55]	7	MetroManila	54	5	0	Predominantly Strong Material
(40,45]	8	Visayas	41	5	0	Predominantly Strong Material
(45,50]	9	Visayas	50	6	0	Predominantly Strong Material

```
#visualisasi per golongan umur kepala keluarga
```

```
ggplot(data = ageGrps, aes(x = total)) +  
  geom_histogram(binwidth = .25, color = "black",  
                fill = "white") +  
  facet_wrap(cuts) +  
  xlab("Household size")
```

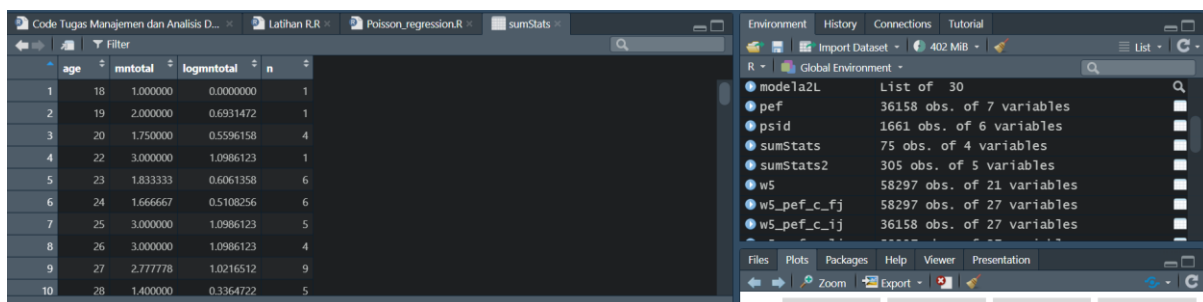


Checking linearity assumption: Empirical log of the means plot

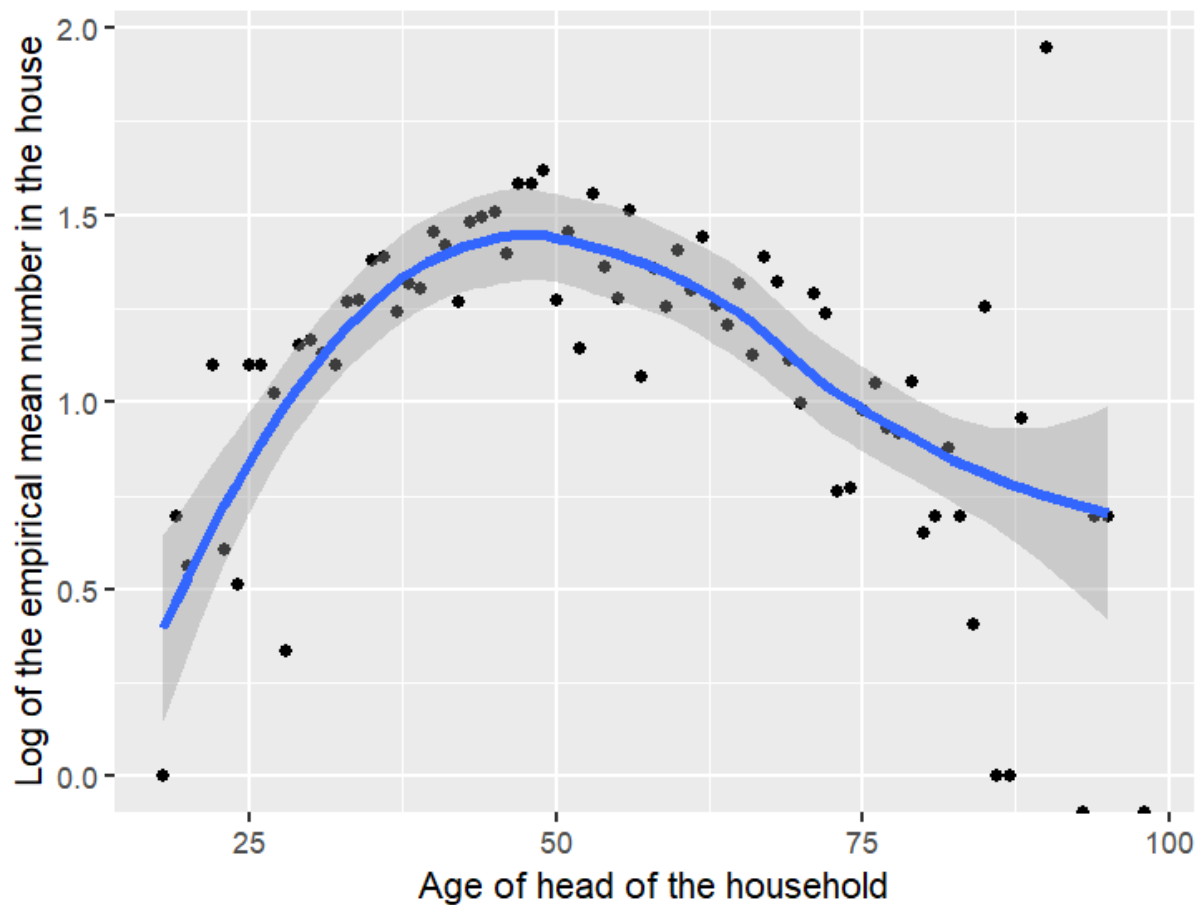
```
sumStats <- fHH1 %>% group_by(age) %>%
```

```
  summarise(mnttotal = mean(total),
```

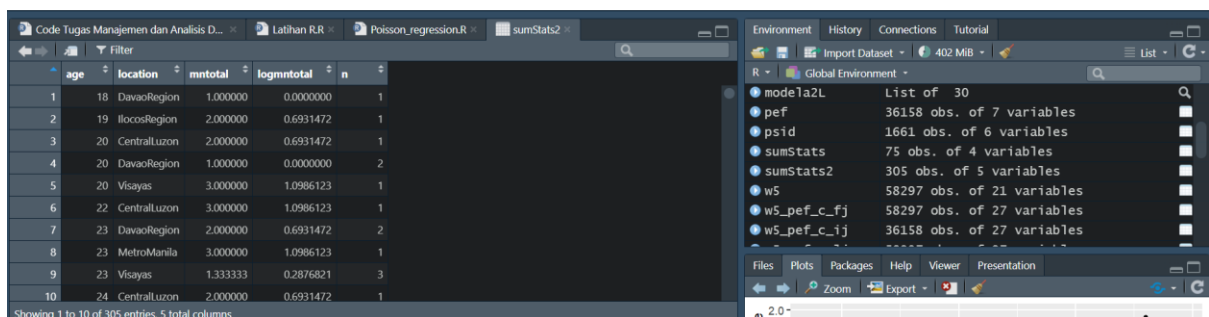
```
            logmnttotal = log(mnttotal), n=n())
```



```
ggplot(sumStats, aes(x=age, y=logmnttotal)) +
  geom_point()+
  geom_smooth(method = "loess", size = 1.5)+
  xlab("Age of head of the household") +
  ylab("Log of the empirical mean number in the house")
```



```
sumStats2 <- fHH1 %>% group_by(age, location) %>%
  summarise(mnttotal = mean(total),
            logmnttotal = log(mnttotal), n=n())
```



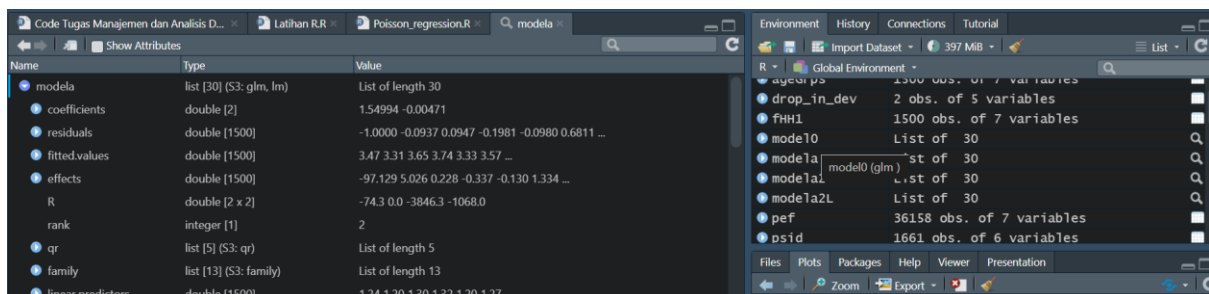
```
ggplot(sumStats2, aes(x=age, y=logmnttotal, color=location,
                      linetype = location, shape = location)) +
```

```
geom_point()+
geom_smooth(method = "loess", se=FALSE)+
xlab("Age of head of the household") +
ylab("Log empirical mean household size")
```



#poisson regression

```
modela = glm(total ~ age, family = poisson, data = fHH1)
```



#coefficient value

```
coef(summary(modela))
```

```
> coef(summary(modela))
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.549942225	0.0502754106	30.829032	1.070156e-208
age	-0.004705881	0.0009363388	-5.025832	5.012548e-07

Note:

$\beta_0 = 1.5499$, $\beta_1 = -0.004705$

#summary of the model

summary(modela)

```
> #summary of the model
> summary(modela)

Call:
glm(formula = total ~ age, family = poisson, data = fHH1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.5499422  0.0502754  30.829  < 2e-16 ***
age          -0.0047059  0.0009363  -5.026  5.01e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 2362.5  on 1499  degrees of freedom
Residual deviance: 2337.1  on 1498  degrees of freedom
AIC: 6714

Number of Fisher Scoring iterations: 5
```

Significance : <0.001

#residual deviance & dispersion

```
cat(" Residual deviance = ", summary(modela)$deviance, " on ",
    summary(modela)$df.residual, "df", "\n",
    "Dispersion parameter = ", summary(modela)$dispersion)
```

```
> #residual deviance & dispersion
> cat(" Residual deviance = ", summary(modela)$deviance, " on ",
+     summary(modela)$df.residual, "df", "\n",
+     "Dispersion parameter = ", summary(modela)$dispersion)
Residual deviance = 2337.089  on 1498 df
Dispersion parameter = 1
```

```
# Wald type CI by hand
#ekstraksi dari Beta0
beta0hat <- summary(modela)$coefficients[1,1]
beta0hat
```

```
> # Wald type CI by hand
> #ekstraksi dari Beta0
> beta0hat <- summary(modela)$coefficients[1,1]
> beta0hat
[1] 1.549942
```

```
#ekstraksi dari Beta1
beta1hat <- summary(modela)$coefficients[2,1]
beta1hat
```

```
> #ekstraksi dari Beta1
> beta1hat <- summary(modela)$coefficients[2,1]
> beta1hat
[1] -0.004705881
```

```
#Kalkulasi rate ratio
exp(beta1hat)
[1] 0.9953052
exp(-0.0047) = [1] 0.995311
```

```
#ekstraksi dari SE Beta1
beta1se <- summary(modela)$coefficients[2,2]
beta1hat - 1.96*beta1se # lower bound
beta1hat + 1.96*beta1se # upper bound
exp(beta1hat - 1.96*beta1se)
exp(beta1hat + 1.96*beta1se)
```

```

> #ekstraksi dari SE Beta1
> beta1se <- summary(modela)$coefficients[2,2]
> beta1hat - 1.96*beta1se    # lower bound
[1] -0.006541105
> beta1hat + 1.96*beta1se    # upper bound
[1] -0.002870657
> exp(beta1hat - 1.96*beta1se)
[1] 0.9934802
> exp(beta1hat + 1.96*beta1se)
[1] 0.9971335

```

CI for betas using profile likelihood

confint(modela)

```

> # CI for betas using profile likelihood
> confint(modela)
Waiting for profiling to be done...
                2.5 %      97.5 %
(Intercept)  1.451170100  1.648249185
age          -0.006543163 -0.002872717

```

#calculating CI for relative risk

exp(confint(modela))

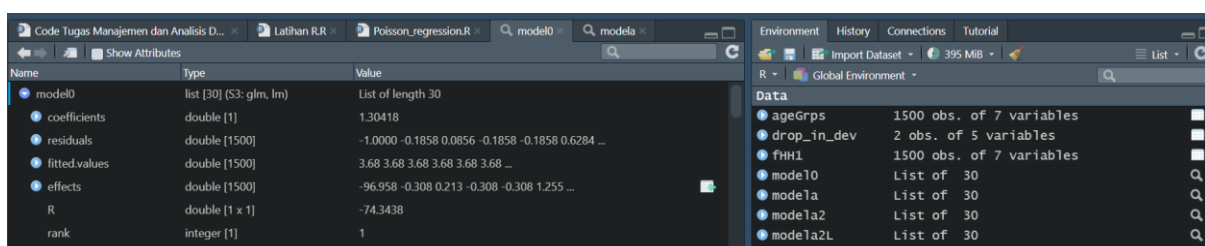
```

> #calculating CI for relative risk
> exp(confint(modela))
Waiting for profiling to be done...
                2.5 %      97.5 %
(Intercept)  4.2681057  5.1978713
age          0.9934782  0.9971314

```

model0 is the null/reduced model

model0 <- glm(total ~ 1, family = poisson, data = fHH1)



The screenshot shows the R Studio interface with two panes visible: 'Environment' and 'Data'. The 'Environment' pane on the left lists several objects: 'model0' (list [30]), 'coefficients' (double [1]), 'residuals' (double [1500]), 'fitted.values' (double [1500]), 'effects' (double [1500]), 'R' (double [1 x 1]), and 'rank' (integer [1]). The 'Data' pane on the right lists datasets: 'ageGrps' (1500 obs. of 7 variables), 'drop_in_dev' (2 obs. of 5 variables), 'fHH1' (1500 obs. of 7 variables), 'model0' (List of 30), 'modela' (List of 30), 'modela2' (List of 30), and 'modela2L' (List of 30).


```
# perform testing for comparing models
```

```
drop_in_dev <- anova(model0, modela, test = "Chisq")
```

```
drop_in_dev
```

```
> # perform testing for comparing models
> drop_in_dev <- anova(model0, modela, test = "Chisq")
> drop_in_dev
Analysis of Deviance Table

Model 1: total ~ 1
Model 2: total ~ age
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1499      2362.5
2      1498      2337.1  1    25.399 4.661e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# second order variable (quadratic term)
```

Significance: <0.001

```
# second order variable (quadratic term)
```

```
# creating variable (age square)
```

```
fHH1 <- fHH1 %>% mutate(age2 = age*age)
```

```
# model with quadratic term
```

```
modela2 = glm(total ~ age + age2, family = poisson,
```

```
data = fHH1)
```

Name	Type	Value
model1a2	list [30] (S3: glm, lm)	List of length 30
coefficients	double [3]	-0.332530 0.070887 -0.000708
residuals	double [1500]	-1.0000 0.1040 -0.0426 -0.2894 0.0664 0.5033 ...
fitted.values	double [1500]	3.61 2.72 4.18 4.22 2.81 3.99 ...
effects	double [1500]	-98.044 4.305 -11.058 -0.541 0.144 1.053 ...
R	double [3 x 3]	-74.3 0.0 0.0 -384 -98.044 4.305 -11.058 -0.541 0.144 1.053 ...
rank	integer [1]	3

```
# comparing models
```

```
drop_in_dev <- anova(model1a, model1a2, test = "Chisq")
```

```
drop_in_dev
```

```
Analysis of Deviance Table

Model 1: total ~ age
Model 2: total ~ age + age2
    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      1498      2337.1
2      1497      2200.9  1    136.15 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Adding covariate (location)
```

```
model1a2L = glm(total ~ age + age2 + location,
```

```
family = poisson, data = fHH1)
```

Name	Type	Value
model1a2L	list [30] (S3: glm, lm)	List of length 30
coefficients	double [7]	-0.384334 0.070363 -0.000703 -0.019387 0.060982 0.054480 ...
residuals	double [1500]	-1.0000 0.1088 0.0399 -0.3230 0.0714 0.4309 ...
fitted.values	double [1500]	3.39 2.71 3.85 4.43 2.80 4.19 ...
effects	double [1500]	-98.133 4.325 -11.048 -2.225 -0.377 -0.863 ...
R	double [7 x 7]	-7.43e+01 0.00e+00 0.00e+00 0.00e+00 0.00e+00 0.00e+00 -3.85e+03 ...

```
# comparing models
```

```
drop_in_dev <- anova(modela2, modela2L, test = "Chisq")
```

```
drop_in_dev
```

```
> # comparing models
> drop_in_dev <- anova(modela2, modela2L, test = "Chisq")
> drop_in_dev
Analysis of Deviance Table

Model 1: total ~ age + age2
Model 2: total ~ age + age2 + location
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      1497      2200.9
2      1493      2187.8  4    13.144  0.01059 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```