

M7222 – 5. CVIČENÍ : GLM05a (*Kontingenční tabulky*)

## Příklad: TABÁK A LÉKAŘI

V roce 1951 byl všem britským lékařům zaslán krátký dotazník o tom, zda kouří tabák. Od té doby byly shromažďovány informace o tom, zda a jak zemřeli.

V následující tabulce jsou uvedeny za 10 let průzkumu počty úmrtí u mužů na ischemické choroby. Dává také údaj, který dostaneme, když vynásobíme celkový počet osob počtem let pozorování (Breslow a Day, 1987).

Age Group	Smokers		Non-smokers	
	Deaths	Person-years	Deaths	Person-years
35-44	32	55407	2	18790
45-54	104	43248	12	10673
55-64	206	28612	28	5710
65-74	186	12663	28	2585
75-84	102	5317	31	1462

```
> heart <- data.frame(age = factor(rep(c("35-44", "45-54", "55-64", "65-74",
  "75-84"), 2)), smoker = factor(c(rep("smoker", 5), rep("non-smoker",
  5))), death = c(32, 104, 206, 186, 102, 2, 12, 28, 28, 31), person = c(52407,
  43248, 28612, 12663, 5317, 18790, 10673, 5710, 2585, 1462))
> heart
```

```
   age      smoker death person
1 35-44      smoker    32 52407
2 45-54      smoker   104 43248
3 55-64      smoker   206 28612
4 65-74      smoker   186 12663
5 75-84      smoker   102  5317
6 35-44 non-smoker     2 18790
7 45-54 non-smoker    12 10673
8 55-64 non-smoker    28  5710
9 65-74 non-smoker    28  2585
10 75-84 non-smoker    31  1462
```

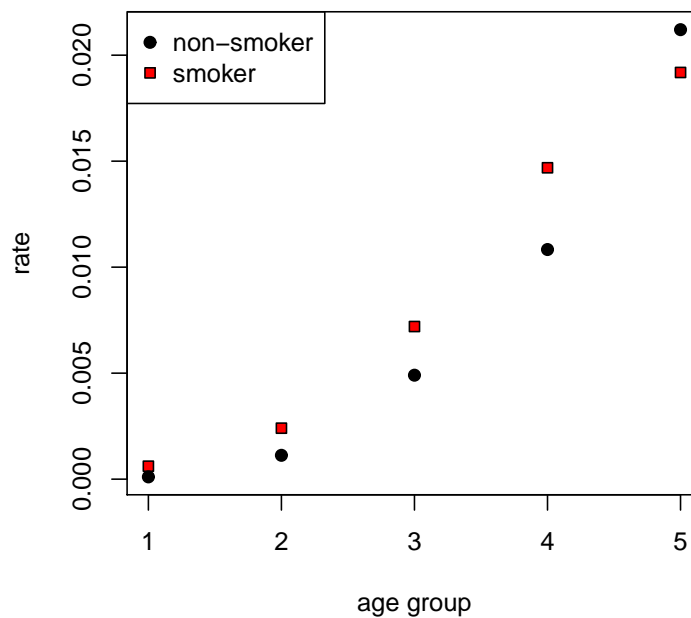
Nebo můžeme totéž vytvořit následujícím způsobem

```
> heart <- data.frame(expand.grid(age = factor(c("35-44", "45-54", "55-64",
  "65-74", "75-84")), smoker = factor(c("smoker", "non-smoker"), levels = c("non-smoker",
  "smoker"))), death = c(32, 104, 206, 186, 102, 2, 12, 28, 28, 31),
  person = c(52407, 43248, 28612, 12663, 5317, 18790, 10673, 5710,
  2585, 1462))
> heart
```

	age	smoker	death	person
1	35-44	smoker	32	52407
2	45-54	smoker	104	43248
3	55-64	smoker	206	28612
4	65-74	smoker	186	12663
5	75-84	smoker	102	5317
6	35-44	non-smoker	2	18790
7	45-54	non-smoker	12	10673
8	55-64	non-smoker	28	5710
9	65-74	non-smoker	28	2585
10	75-84	non-smoker	31	1462

Data graficky znázorníme.

```
> rate <- with(heart, death/person)
> with(heart, plot(as.integer(age), rate, bg = c(1, 2)[as.integer(smoker)],
  xlab = "age group", pch = c(21, 22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), pch = c(21, 22), pt.bg = c(1,
  2)))
```



Obrázek 1: Vstupní data.

Z grafu je patrné, že podíl mrtvých je závislý na čase, takže vytvoříme novou proměnnou Age, která nebude kategoriální, ale budeme ji chápat jako spojitou.

```
> heart <- cbind(heart, Age = as.integer(heart$age))
```

Nejprve budeme uvažovat poissonovský regresní model pouze s regresory `Age` a `smoker`. Protože známe celkový počet (vynásobený léty pozorování), sledujeme závisle proměnnou, která má binomické rozdělení. Chceme-li použít Poissonovo rozdělení, nesmíme zapomenout na `offset`.

$$\text{MODEL 1: } \eta_i = \log(\mu_i) = \log(n_i\pi_i) = \log(n_i) + \beta_0 + \beta_{02} + \beta_1 \text{Age}_i.$$

```
> heart.fit1 <- glm(death ~ smoker + Age, family = poisson, data = heart,
  offset = log(person))
> summary(heart.fit1)
```

Call:

```
glm(formula = death ~ smoker + Age, family = poisson, data = heart,
  offset = log(person))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.5712	-2.7562	0.2857	1.4261	3.7183

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.11833	0.13929	-58.282	< 2e-16 ***
smokersmoker	0.40637	0.10720	3.791	0.00015 ***
Age	0.83583	0.02904	28.777	< 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: 935.067 on 9 degrees of freedom  
 Residual deviance: 69.182 on 7 degrees of freedom  
 AIC: 130.25

Number of Fisher Scoring iterations: 4

Vidíme, že jednotlivé kovariáty jsou statisticky významné.

Vhodnost modelu ověříme pomocí Waldova testu.

```
> library(lmtest)
> waldtest(heart.fit1, test = "Chisq")
```

Wald test

Model 1: death ~ smoker + Age

Model 2: death ~ 1

Res.Df	Df	Chisq	Pr(>Chisq)
1	7		
2	9	-2 846.05	< 2.2e-16 ***

---

Signif. codes: 0 '\*\*\*', 0.001 '\*\*', 0.01 '\*', 0.05 '.', 0.1 ' ', 1

Z výsledku je zřejmé, že oproti nulovému modelu se přidáním kovariát `Age` a `smoker` model výrazně zlepšil, přesto však vidíme velký nepoměr mezi stupni volnosti a reziduální deviací. Přidáme proto interakce mezi `age` a `smoker`.

$$\text{MODEL 2: } \eta_i = \log(\mu_i) = \log(n_i\pi_i) = \log(n_i) + \beta_0 + \beta_{02} + (\beta_1 + \beta_{12})Age_i.$$

```
> heart.fit2 <- glm(death ~ smoker * Age, family = poisson, data = heart,
  offset = log(person))
> summary(heart.fit2)
```

Call:

```
glm(formula = death ~ smoker * Age, family = poisson, data = heart,
  offset = log(person))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8784	-2.1217	-0.2482	1.7183	3.5269

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-8.86708	0.30567	-29.008	< 2e-16 ***
smokersmoker	1.28360	0.32584	3.939	8.17e-05 ***
Age	1.04683	0.07743	13.519	< 2e-16 ***
smokersmoker:Age	-0.24898	0.08359	-2.978	0.00290 **

---

Signif. codes: 0 '\*\*\*', 0.001 '\*\*', 0.01 '\*', 0.05 '.', 0.1 ' ', 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 935.067 on 9 degrees of freedom  
 Residual deviance: 59.893 on 6 degrees of freedom  
 AIC: 122.96

Number of Fisher Scoring iterations: 4

Pořád ještě nejsme spokojeni se vztahem mezi stupni volnosti a reziduální deviací. Přidáme proto novou kovariátu  $I(\text{Age}^2)$ .

$$\text{MODEL 3: } \eta_i = \log(\mu_i) = \log(n_i\pi_i) = \log(n_i) + \beta_0 + \beta_{02} + (\beta_1 + \beta_{12})Age_i + \beta_2Age_i^2.$$

```
> heart.fit3 <- glm(death ~ smoker * Age + I(Age^2), family = poisson,
  data = heart, offset = log(person))
> summary(heart.fit3)
```

Call:

```
glm(formula = death ~ smoker * Age + I(Age^2), family = poisson,
  data = heart, offset = log(person))
```

Deviance Residuals:

1	2	3	4	5	6	7	8	9
0.43820	-0.27329	-0.15265	0.23393	-0.05700	-0.83049	0.13404	0.64107	-0.41058

10

-0.01275

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-10.79176	0.45008	-23.978	< 2e-16 ***
smokersmoker	1.44097	0.37220	3.872	0.000108 ***
Age	2.37648	0.20795	11.428	< 2e-16 ***
I(Age^2)	-0.19768	0.02737	-7.223	5.08e-13 ***
smokersmoker:Age	-0.30755	0.09704	-3.169	0.001528 **

---  
Signif. codes: 0 ,\*\*\*, 0.001 \*\*, 0.01 \*, 0.05 ., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 935.0673 on 9 degrees of freedom  
Residual deviance: 1.6354 on 5 degrees of freedom  
AIC: 66.703

Number of Fisher Scoring iterations: 4

Nyní nastal opačný případ, stupně volnosti jsou vyšší než reziduální deviance. Provedme proto poslední úpravu modelu na tvar

MODEL 4:  $\eta_i = \log(\mu_i) = \log(n_i\pi_i) = \log(n_i) + \beta_0 + \beta_{02} + (\beta_1 + \beta_{12})Age_i + (\beta_2 + \beta_{22})Age_i^2$ .

```
> heart.fit4 <- glm(death ~ smoker * Age + smoker * I(Age^2), family = poisson,
  data = heart, offset = log(person))
> summary(heart.fit4)
```

Call:

```
glm(formula = death ~ smoker * Age + smoker * I(Age^2), family = poisson,
  data = heart, offset = log(person))
```

Deviance Residuals:

1	2	3	4	5	6	7	8	9
0.29593	-0.29864	-0.04481	0.29049	-0.18302	-0.48763	0.25588	0.39375	-0.64870
10								
0.26431								

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-11.26262	0.90564	-12.436	< 2e-16 ***
smokersmoker	1.97405	0.95484	2.067	0.03870 *
Age	2.68289	0.54401	4.932	8.15e-07 ***
I(Age^2)	-0.24213	0.07745	-3.126	0.00177 **
smokersmoker:Age	-0.65712	0.57742	-1.138	0.25510
smokersmoker:I(Age^2)	0.05109	0.08279	0.617	0.53713

---  
Signif. codes: 0 ,\*\*\*, 0.001 \*\*, 0.01 \*, 0.05 ., 0.1 , , 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 935.0673 on 9 degrees of freedom  
Residual deviance: 1.2456 on 4 degrees of freedom

AIC: 68.314

Number of Fisher Scoring iterations: 4

Nyní všechny čtyři modely porovnáme pomocí odchylek deviací.

```
> anova(heart.fit4, heart.fit3, heart.fit2, heart.fit1, test = "Chi")
```

Analysis of Deviance Table

Model 1: death ~ smoker \* Age + smoker \* I(Age^2)

Model 2: death ~ smoker \* Age + I(Age^2)

Model 3: death ~ smoker \* Age

Model 4: death ~ smoker + Age

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi )
1	4	1.246			
2	5	1.635	-1	-0.390	0.532429
3	6	59.893	-1	-58.258	2.299e-14 ***
4	7	69.182	-1	-9.289	0.002306 **

---

Signif. codes: 0 '\*\*\*', 0.001 '\*\*', 0.01 '\*', 0.05 '.', 0.1 ' ', 1

Z výsledku vyplývá, že ideálním modelem je model

$$\text{MODEL 3: } \eta_i = \log(\mu_i) = \log(n_i \pi_i) = \log(n_i) + \beta_0 + \beta_{02} + (\beta_1 + \beta_{12}) \text{Age}_i + \beta_2 \text{Age}_i^2$$

a odtud dostaneme, že pravděpodobnost úmrtí na ischemickou chorobu je

$$\pi_i = \frac{\mu_i}{n_i} = e^{\eta_i - \log(n_i)} = \begin{cases} e^{\beta_0} e^{\beta_1 \text{Age}_i} e^{\beta_2 \text{Age}_i^2} & \text{pro nekuřáky} \\ e^{\beta_0} e^{\beta_{02}} e^{\beta_1 \text{Age}_i} e^{\beta_{12} \text{Age}_i} e^{\beta_2 \text{Age}_i^2} & \text{pro kuřáky} \end{cases}$$

pro danou hodnotu kovaritivity Age.

Odhadněme dle modelu nejprve pravděpodobnosti úmrtí na ischemickou chorobu u kuřáků i nekuřáků v jednotlivých věkových skupinách a výsledky porovnejme s empirickou pravděpodobností *rate*, kterou jsme vypočítali jako podíl *death/person*.

```
> Model <- heart.fit3
> EstProb <- exp(Model$linear.predictor - Model$offset)
> cbind(heart, EmpProb = round(rate, 6), EstProb = round(EstProb, 6))
```

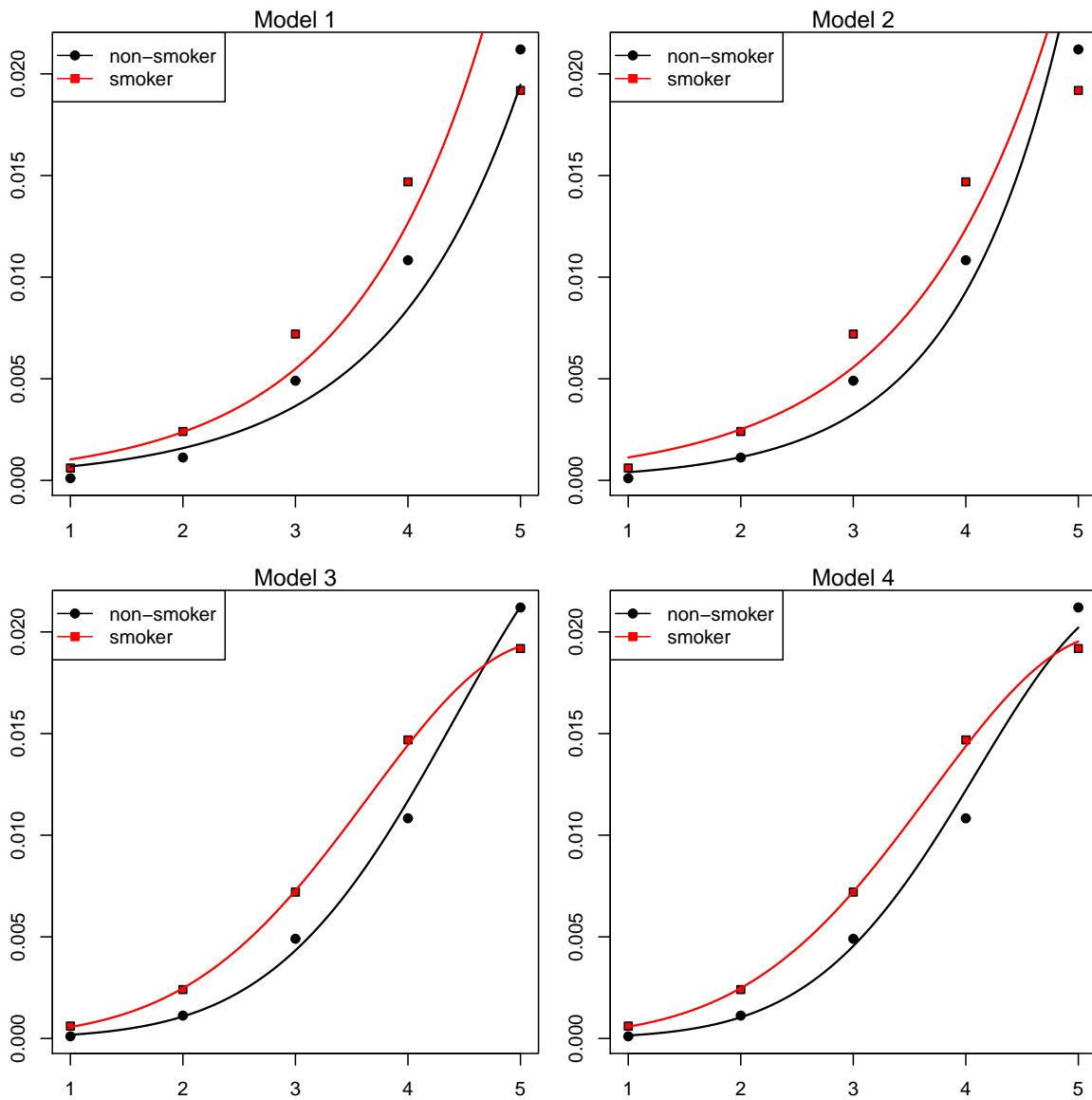
	age	smoker	death	person	Age	EmpProb	EstProb
1	35-44	smoker	32	52407	1	0.000611	0.000565
2	45-54	smoker	104	43248	2	0.002405	0.002470
3	55-64	smoker	206	28612	3	0.007200	0.007277
4	65-74	smoker	186	12663	4	0.014688	0.014438
5	75-84	smoker	102	5317	5	0.019184	0.019292
6	35-44	non-smoker	2	18790	1	0.000106	0.000182
7	45-54	non-smoker	12	10673	2	0.001124	0.001081
8	55-64	non-smoker	28	5710	3	0.004904	0.004333
9	65-74	non-smoker	28	2585	4	0.010832	0.011694
10	75-84	non-smoker	31	1462	5	0.021204	0.021252

Na závěr vyneseme výsledky do grafu.

```

> op <- par(mfrow = c(2, 2), mar = c(2.5, 2.5, 1.5, 0.5) + 0.05)
> TXT <- "Model 1"
> Model <- heart.fit1
> CoefNonSmoker <- coef(Model)[c(1, 3)]
> CoefSmoker <- CoefNonSmoker + c(coef(Model)[2], 0)
> with(heart, plot(Age, rate, bg = c(1, 2)[as.integer(smoker)], xlab = "age group",
  pch = c(21, 22)[as.integer(smoker)]))
> curve(exp(CoefNonSmoker[1] + CoefNonSmoker[2] * x), add = T, lty = 1,
  lwd = 1.5, col = 1)
> curve(exp(CoefSmoker[1] + CoefSmoker[2] * x), add = T, lty = 1, lwd = 1.5,
  col = 2)
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> TXT <- "Model 2"
> Model <- heart.fit2
> CoefNonSmoker <- coef(Model)[c(1, 3)]
> CoefSmoker <- CoefNonSmoker + coef(Model)[c(2, 4)]
> with(heart, plot(Age, rate, bg = c(1, 2)[as.integer(smoker)], xlab = "age group",
  pch = c(21, 22)[as.integer(smoker)]))
> curve(exp(CoefNonSmoker[1] + CoefNonSmoker[2] * x), add = T, lty = 1,
  lwd = 1.5, col = 1)
> curve(exp(CoefSmoker[1] + CoefSmoker[2] * x), add = T, lty = 1, lwd = 1.5,
  col = 2)
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> TXT <- "Model 3"
> Model <- heart.fit3
> CoefNonSmoker <- coef(Model)[c(1, 3, 4)]
> CoefSmoker <- CoefNonSmoker + c(coef(Model)[c(2, 5)], 0)
> with(heart, plot(Age, rate, bg = c(1, 2)[as.integer(smoker)], xlab = "age group",
  pch = c(21, 22)[as.integer(smoker)]))
> curve(exp(CoefNonSmoker[1] + CoefNonSmoker[2] * x + CoefNonSmoker[3] *
  x^2), add = T, lty = 1, lwd = 1.5, col = 1)
> curve(exp(CoefSmoker[1] + CoefSmoker[2] * x + CoefSmoker[3] * x^2),
  add = T, lty = 1, lwd = 1.5, col = 2)
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> TXT <- "Model 4"
> Model <- heart.fit4
> CoefNonSmoker <- coef(Model)[c(1, 3, 4)]
> CoefSmoker <- CoefNonSmoker + coef(Model)[c(2, 5, 6)]
> with(heart, plot(Age, rate, bg = c(1, 2)[as.integer(smoker)], xlab = "age group",
  pch = c(21, 22)[as.integer(smoker)]))
> curve(exp(CoefNonSmoker[1] + CoefNonSmoker[2] * x + CoefNonSmoker[3] *
  x^2), add = T, lty = 1, lwd = 1.5, col = 1)
> curve(exp(CoefSmoker[1] + CoefSmoker[2] * x + CoefSmoker[3] * x^2),
  add = T, lty = 1, lwd = 1.5, col = 2)
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)

```



Obrázek 2: Srovnání jednotlivých modelů.

Protože u všech modelů nebyl úplně dobrý vztah mezi počty stupňů volnosti a reziduální deviací, pro všechny čtyři modely zvolíme ještě variantu s volbou `family=quasipoisson`.

```
> heart.fit1q <- update(heart.fit1, family = quasipoisson)
> summary(heart.fit1q)
```

Call:

```
glm(formula = death ~ smoker + Age, family = quasipoisson, data = heart,
     offset = log(person))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.5712	-2.7562	0.2857	1.4261	3.7183



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-8.11833	0.41357	-19.630	2.22e-07 ***
smokersmoker	0.40637	0.31827	1.277	0.242
Age	0.83583	0.08624	9.692	2.63e-05 ***

---  
Signif. codes: 0 '\*\*\*', 0.001 '\*\*', 0.01 '\*', 0.05 '..', 0.1 ' ', 1

(Dispersion parameter for quasipoisson family taken to be 8.815372)

Null deviance: 935.067 on 9 degrees of freedom  
Residual deviance: 69.182 on 7 degrees of freedom  
AIC: NA

Number of Fisher Scoring iterations: 4

```
> heart.fit2q <- update(heart.fit2, family = quasipoisson)
> summary(heart.fit2q)
```

Call:

```
glm(formula = death ~ smoker * Age, family = quasipoisson, data = heart,
     offset = log(person))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.8784	-2.1217	-0.2482	1.7183	3.5269

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-8.8671	0.9347	-9.487	7.82e-05 ***
smokersmoker	1.2836	0.9964	1.288	0.24509
Age	1.0468	0.2368	4.421	0.00447 **
smokersmoker:Age	-0.2490	0.2556	-0.974	0.36765

---  
Signif. codes: 0 '\*\*\*', 0.001 '\*\*', 0.01 '\*', 0.05 '..', 0.1 ' ', 1

(Dispersion parameter for quasipoisson family taken to be 9.350393)

Null deviance: 935.067 on 9 degrees of freedom  
Residual deviance: 59.893 on 6 degrees of freedom  
AIC: NA

Number of Fisher Scoring iterations: 4

```
> heart.fit3q <- update(heart.fit3, family = quasipoisson)
> summary(heart.fit3q)
```

Call:

```
glm(formula = death ~ smoker * Age + I(Age^2), family = quasipoisson,
     data = heart, offset = log(person))
```

Deviance Residuals:

```

      1      2      3      4      5      6      7      8      9
0.43820 -0.27329 -0.15265  0.23393 -0.05700 -0.83049  0.13404  0.64107 -0.41058
      10
-0.01275

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-10.79176	0.25061	-43.062	1.27e-07 ***
smokersmoker	1.44097	0.20725	6.953	0.000946 ***
Age	2.37648	0.11579	20.524	5.08e-06 ***
I(Age^2)	-0.19768	0.01524	-12.972	4.85e-05 ***
smokersmoker:Age	-0.30755	0.05403	-5.692	0.002335 **

---

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

(Dispersion parameter for quasipoisson family taken to be 0.3100502)

Null deviance: 935.0673 on 9 degrees of freedom  
 Residual deviance: 1.6354 on 5 degrees of freedom  
 AIC: NA

Number of Fisher Scoring iterations: 4

```

> heart.fit4q <- update(heart.fit4, family = quasipoisson)
> summary(heart.fit4q)

```

Call:

```

glm(formula = death ~ smoker * Age + smoker * I(Age^2), family = quasipoisson,
     data = heart, offset = log(person))

```

Deviance Residuals:

```

      1      2      3      4      5      6      7      8      9
0.29593 -0.29864 -0.04481  0.29049 -0.18302 -0.48763  0.25588  0.39375 -0.64870
      10
0.26431

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-11.26262	0.49895	-22.573	2.28e-05 ***
smokersmoker	1.97405	0.52605	3.753	0.019902 *
Age	2.68289	0.29971	8.952	0.000862 ***
I(Age^2)	-0.24213	0.04267	-5.675	0.004758 **
smokersmoker:Age	-0.65712	0.31812	-2.066	0.107758
smokersmoker:I(Age^2)	0.05109	0.04561	1.120	0.325335

---

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

(Dispersion parameter for quasipoisson family taken to be 0.3035241)

Null deviance: 935.0673 on 9 degrees of freedom  
 Residual deviance: 1.2456 on 4 degrees of freedom  
 AIC: NA

Number of Fisher Scoring iterations: 4

Nejprve vytvoříme pomocnou funkci pro kreslení křivek spolu s intervaly spolehlivosti.

```
> PolygPoiss <- function(Model, NewData, x, col = "red", ColPolyg = "gray75",
  border = "gray75", lwd = 2, lty = 1) {
  LinPred <- predict(Model, newdata = NewData, type = "link", se = T)
  y <- exp(LinPred$fit)
  yL <- exp(LinPred$fit - 1.96 * LinPred$se.fit)
  yH <- exp(LinPred$fit + 1.96 * LinPred$se.fit)
  xx <- c(x, rev(x))
  yy <- c(yL, rev(yH))
  polygon(xx, yy, col = ColPolyg, border = border)
  lines(x, y, lty = lty, col = col, lwd = lwd)
}
```

Následně připravíme pomocné datové rámce, které budeme potřebovat pro vykreslování křivek.

```
> AGE <- seq(1, 5, length = 100)
> NewData1 <- data.frame(Age = AGE, smoker = factor(rep("non-smoker",
  length(AGE)), levels = c("non-smoker", "smoker")), person = rep(heart$person[1:n],
  each = nn))
> NewData1 <- data.frame(Age = AGE, smoker = factor(rep("non-smoker",
  length(AGE)), levels = c("non-smoker", "smoker")), person = rep(1,
  length(AGE)))
> NewData2 <- data.frame(Age = AGE, smoker = factor(rep("smoker", length(AGE)),
  levels = c("non-smoker", "smoker")), person = rep(1, length(AGE)))
```

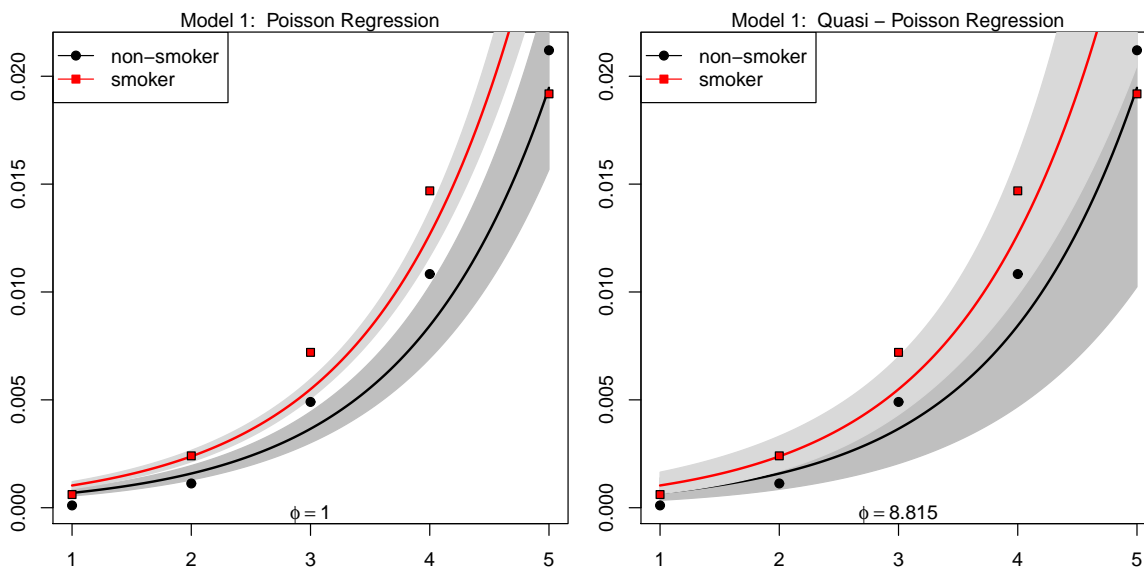
Nyní postupně každou dvojici modelů graficky porovnáme, abychom ukázali, jak se změnila variabilita odhadu střední hodnoty. Začneme s prvním modelem.

```
> par(mfrow = c(1, 2), mar = c(2.5, 2.5, 1.5, 0.5) + 0.05)
> TXT <- "Model 1: Poisson Regression"
> Model <- heart.fit1
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)
> TXT <- "Model 1: Quasi - Poisson Regression"
> Model <- heart.fit1q
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
```

```

> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)

```

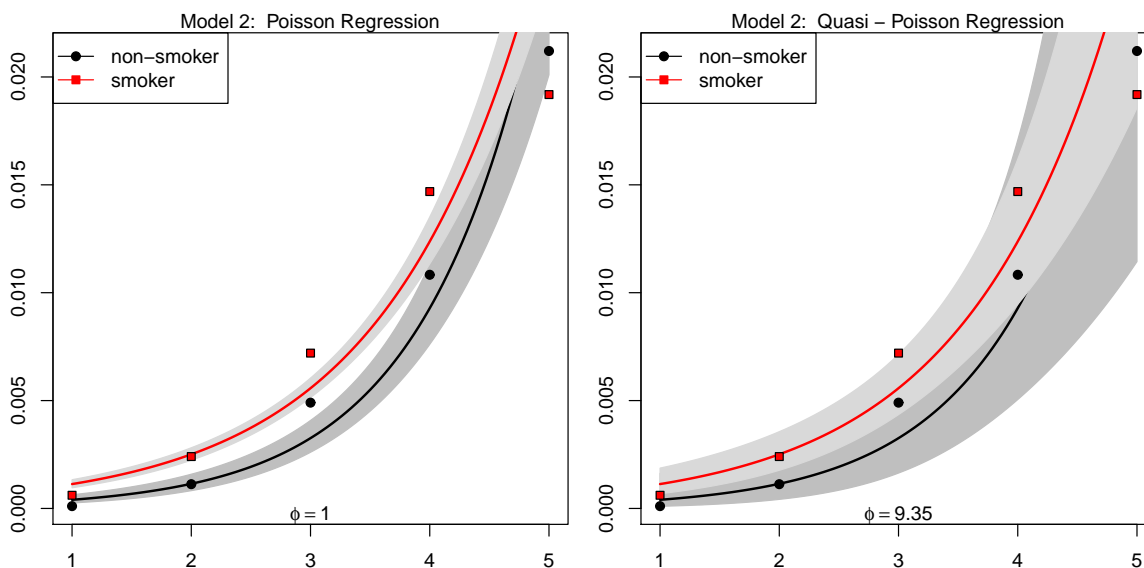


Obrázek 3: MODEL 1: Poissonovská regrese spolu s odhadem trendu a intervalu spolehlivosti – bez a s řešením problému *overdispersion*.

```

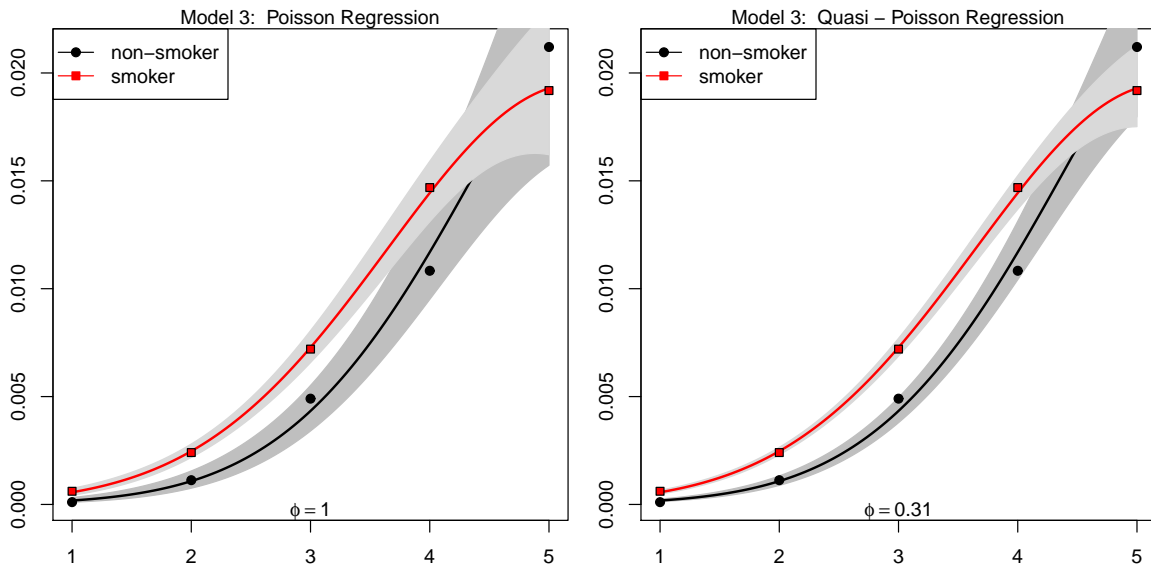
> op <- par(mfrow = c(1, 2), mar = c(2.5, 2.5, 1.5, 0.5) + 0.05)
> TXT <- "Model 2: Poisson Regression"
> Model <- heart.fit2
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)
> TXT <- "Model 2: Quasi - Poisson Regression"
> Model <- heart.fit2q
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)

```



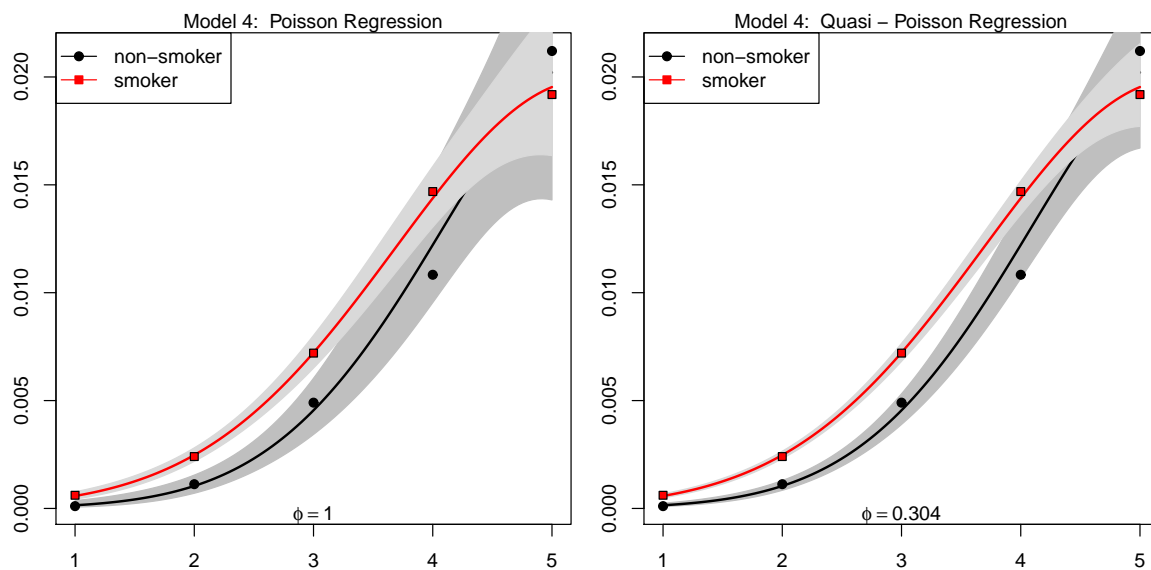
Obrázek 4: MODEL 2: Poissonovská regrese spolu s odhadem trendu a intervalu spolehlivosti – bez a s řešením problému *overdispersion*.

```
> op <- par(mfrow = c(1, 2), mar = c(2.5, 2.5, 1.5, 0.5) + 0.05)
> TXT <- "Model 3: Poisson Regression"
> Model <- heart.fit3
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)
> TXT <- "Model 3: Quasi - Poisson Regression"
> Model <- heart.fit3q
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)
```



Obrázek 5: MODEL 3: Poissonovská regrese spolu s odhadem trendu a intervalu spolehlivosti – bez a s řešením problému *overdispersion*.

```
> op <- par(mfrow = c(1, 2), mar = c(2.5, 2.5, 1.5, 0.5) + 0.05)
> TXT <- "Model 4: Poisson Regression"
> Model <- heart.fit4
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)
> TXT <- "Model 4: Quasi - Poisson Regression"
> Model <- heart.fit4q
> with(heart, plot(Age, rate, type = "n", xlab = "age group"))
> PolygPoiss(Model, NewData1, AGE, col = "black", ColPolyg = "gray75",
  border = "gray75")
> PolygPoiss(Model, NewData2, AGE, col = "red", ColPolyg = "gray85", border = "gray85")
> with(heart, points(Age, rate, bg = c(1, 2)[as.integer(smoker)], pch = c(21,
  22)[as.integer(smoker)]))
> with(heart, legend("topleft", levels(smoker), col = 1:2, lty = 1, pch = c(21,
  22), pt.bg = c(1, 2)))
> mtext(TXT)
> phi <- round(summary(Model)$dispersion, 3)
> mtext(text = bquote(phi == .(phi)), side = 1, line = -1)
```



Obrázek 6: MODEL 4: Poissonovská regrese spolu s odhadem trendu a intervalu spolehlivosti – bez a s řešením problému *overdispersion*.