**Логистическая регрессия**

На основе базы данных списка отелей, создана переменная:

«1” – отель имеет рейтинг 5/5 и по крайней мере 30 отзывов; т.е. действительно отличная гостиница;

«0” – во всех остальных случаях.

Необходимо построить модель логистической регрессии, чтобы выяснить, зависят ли “1”-ы от параметров цена и тип комнаты;

Оценить параметры модели и ее общую точность.

> myData<-read.csv("C:/data/airbnb\_task.csv")

> attach(myData)

> fit<-glm(factor(y)~price+room\_type, data=myData, family=quasibinomial())

#иногда можно использовать квазибиномиальное семейство моделей, если, например, допускается очень высокая вариация в данных

> summary(fit)

Call:

glm(formula = factor(y) ~ price + room\_type, family = quasibinomial(),

data = myData)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.3932 -0.3826 -0.3711 -0.3647 2.6097

Coefficients:

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

(Intercept) -2.5126128 0.0558458 -44.992 <2e-16 \*\*\*

price -0.0008550 0.0003922 -2.180 0.0293 \*

room\_typePrivate room -0.1096534 0.0653740 -1.677 0.0935 .

room\_typeShared room -0.8331273 0.4561424 -1.826 0.0678 .

---

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

(Dispersion parameter for quasibinomial family taken to be 0.998869)

Null deviance: 8663.8 on 17650 degrees of freedom

Residual deviance: 8653.5 on 17647 degrees of freedom

AIC: NA

Number of Fisher Scoring iterations: 5