AdvStDaAn, Worksheet, Week 7

Michael Lappert

27 April, 2022

Contents

Exercise 1

Exercise 1.a)

Turbine Data (cf. Exercise 1 on Worksheet Week 4) Does the GLM that you have fitted in part 1(b) model the data adequately?

```
path <- file.path('Datasets', 'turbines.dat')
df <- read.table(path, header=TRUE)

# Fitted model in w4, 1.b)
glm1.1 <- glm(cbind(Fissures, Turbines-Fissures) ~ Hours, family = binomial, data = df)
summary(glm1.1)
```

```
##
## Call:
## glm(formula = cbind(Fissures, Turbines - Fissures) ~ Hours, family = binomial,
##     data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5055  -0.7647  -0.3036   0.4901   2.0943
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.9235966  0.3779589 -10.381   <2e-16 ***
## Hours        0.0009992  0.0001142   8.754   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
##       Null deviance: 112.670  on 10   degrees of freedom
## Residual deviance:   10.331  on  9   degrees of freedom
## AIC: 49.808
##
## Number of Fisher Scoring iterations: 4
```

Because the response is binomially distributed with m > 1, we can test on overdispersion:

```
1-pchisq(10.331, 9) # if resulting value > 0.05 -> no overdispersion
```

```
## [1] 0.3243594
```

Because the p-value is > than the significance level of 5% we have no evidence against the null hypothesis that $\phi = 1$ -> no overdispersion.

Or altenatively:

```
qchisq(0.95, df=9) # if resulting value > Residual deviance -> no overdispersion
```
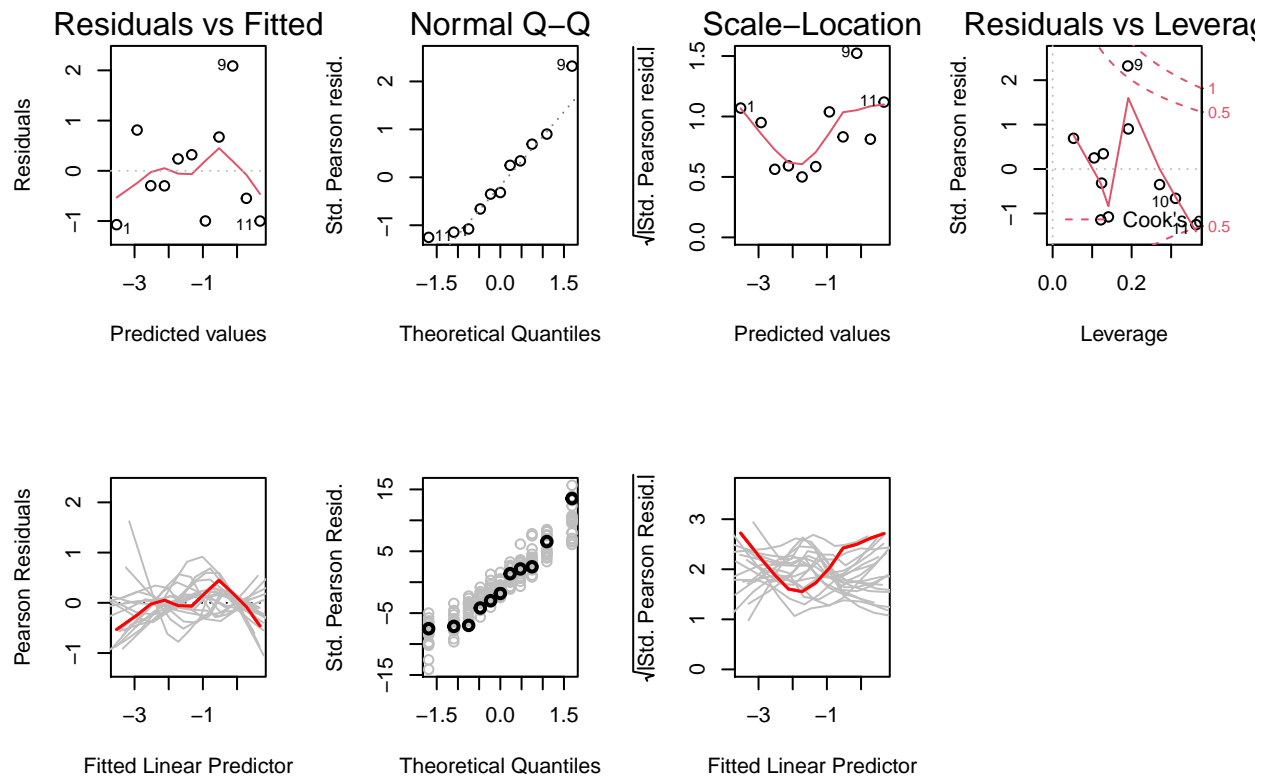
```
## [1] 16.91898
```

Because the residual deviance is smaller than $q_{0.95}^{\chi_9^2}$ the null hypothesis that $\phi = 1$ cannot be rejected -> no overdispersion.

Question 1.a)

Is the conclusion in the two cells abvoe right? Are these two different methods to come to the same result (looking for overdispersion)?

Performing a residual and sensitivity analysis for the fitted model:

```
par(mfrow=c(2,4))
plot(glm1.1)
plot.glmSim(glm1.1, SEED = 1)
```

Residuals vs Fitted

Residuals

Predicted values

Normal Q–Q

Std. Pearson resid.

Theoretical Quantiles

Scale–Location

$\sqrt{|\text{Std. Pearson resid.}|}$

Predicted values

Residuals vs Leverage

Std. Pearson resid.

Cook's d

Leverage

Pearson Residuals

Fitted Linear Predictor

Std. Pearson Resid.

Theoretical Quantiles

$\sqrt{|\text{Std. Pearson Resid.}|}$

Fitted Linear Predictor

Interpretation:

1. Tukey-Anscombe plot: The smoother shows a banana form, however in the stochastic fluctuation the smoother is not extreme.
   => The assumption of constant expactation is not violated.
2. Q-Q plot: The data scattersnot fully around the straight line but is within the stochastic fluctuation.
   => The assumption of Gaussian distributed errors is violated.
3. Scale-location plot: The smoother has a banana form and stays within the stochastic fluctuation.
   => There is no evidence against the assumption of constant variance of the residuals.
4. Residuals vs. Leverage: All observations have Cook's Distance <1 and therewith no too influential points are present. Leverage points > 2 * 2 [nr. of coefficients] / 11 [nr. of observations] = 0.3636364.
   => No too influential (dangerous) observations

CONCLUSION: The model does fit the data adequately.

3