

PedCare ITS

Bertil Baron

20.09.2021

Deskriptive analysis

tabulate U-Untersuchungen before and after the intervention

```
summary(kv_data$P_04001[kv_data$Intervention==0])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      5080  12960  19292  23675  30698  56419
```

```
summary(kv_data$P_04001[kv_data$Intervention==1])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3971  12043  16876  19120  23491  42144
```

```
summary(kv_data$rate[kv_data$Intervention==0])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      27.04  68.98 102.68 126.01 163.40 300.30
```

```
summary(kv_data$rate[kv_data$Intervention==1])
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      21.14  64.10  89.83 101.77 125.03 224.32
```

Poisson regression model

Poisson with the standardised population as an offset

```
model_u_ges <- glm(P_04001 ~ offset(log(Pop2019)) + Intervention +
                    Zeit + Intervention*Zeit, family=poisson, kv_data)
summary(model_u_ges)
```

```
##
## Call:
## glm(formula = P_04001 ~ offset(log(Pop2019)) + Intervention +
##      Zeit + Intervention * Zeit, family = poisson, data = kv_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -137.75  -66.97  -24.09   42.43  169.01
##
## Coefficients:
##              Estimate Std. Error  z value Pr(>|z|)
## (Intercept)  -4.263e+00  1.613e-03 -2642.451 < 2e-16 ***
## Intervention   2.768e-03  7.977e-03   0.347 0.728640
## Zeit         -3.528e-03  4.508e-05  -78.253 < 2e-16 ***
```

```
## Intervention:Zeit -3.589e-04 1.026e-04 -3.499 0.000467 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 679783 on 104 degrees of freedom
## Residual deviance: 647555 on 101 degrees of freedom
## AIC: 648789
##
## Number of Fisher Scoring iterations: 4
summary(model_u_ges)$dispersion

## [1] 1
round(ci.lin(model_u_ges,Exp=T),3)

##              Estimate StdErr          z      P exp(Est.)  2.5% 97.5%
## (Intercept)    -4.263  0.002 -2642.451 0.000    0.014 0.014 0.014
## Intervention      0.003  0.008   0.347 0.729    1.003 0.987 1.019
## Zeit           -0.004  0.000  -78.253 0.000    0.996 0.996 0.997
## Intervention:Zeit  0.000  0.000   -3.499 0.000    1.000 0.999 1.000
```

Test für overdispersion

Overdispersion: Varianz größer als erwartet → Poisson Regression nicht anwendbar

```
dispersiontest(model_u_ges)

##
## Overdispersion test
##
## data: model_u_ges
## z = 7.7778, p-value = 3.689e-15
## alternative hypothesis: true dispersion is greater than 1
## sample estimates:
## dispersion
## 6542.777
```

Overdispersion liegt vor

Overdispersion: Quasi-Poisson model

In the model above we have not allowed for overdispersion - in order to do this we can use a quasipoisson model, which allows the variance to be proportional rather than equal to the mean

```
model2_u_ges <- glm(P_04001 ~ offset(log(Pop2019)) + Intervention +
                    Zeit + Intervention*Zeit,
                    family=quasipoisson, kv_data)
summary(model2_u_ges)

##
## Call:
## glm(formula = P_04001 ~ offset(log(Pop2019)) + Intervention +
##      Zeit + Intervention * Zeit, family = quasipoisson, data = kv_data)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -137.75   -66.97   -24.09    42.43   169.01
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -4.2631107  0.1330593 -32.039  <2e-16 ***
## Intervention    0.0027675  0.6579104   0.004   0.997
## Zeit          -0.0035279  0.0037182  -0.949   0.345
## Intervention:Zeit -0.0003589  0.0084591  -0.042   0.966
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 6802.23)
##
##      Null deviance: 679783  on 104  degrees of freedom
## Residual deviance: 647555  on 101  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4
summary(model2_u_ges)$dispersion

## [1] 6802.23
round(ci.lin(model2_u_ges,Exp=T),3)

##              Estimate StdErr        z      P exp(Est.)  2.5% 97.5%
## (Intercept)      -4.263  0.133 -32.039 0.000      0.014 0.011 0.018
## Intervention       0.003  0.658   0.004 0.997      1.003 0.276 3.641
## Zeit             -0.004  0.004  -0.949 0.343      0.996 0.989 1.004
## Intervention:Zeit  0.000  0.008  -0.042 0.966      1.000 0.983 1.016

auch nicht gut

deshalb das negative binomial model
model3_u_ges <- glm.nb(P_04001 ~ offset(log(Pop2019)) + Intervention,
                      data = kv_data)
summary(model3_u_ges)

##
## Call:
## glm.nb(formula = P_04001 ~ offset(log(Pop2019)) + Intervention,
##        data = kv_data, init.theta = 3.531502196, link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3458  -0.9393  -0.2721   0.5058   1.9065
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.37395    0.06705 -65.237  <2e-16 ***
## Intervention -0.21372    0.10601  -2.016   0.0438 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for Negative Binomial(3.5315) family taken to be 1)
##
## Null deviance: 113.91 on 104 degrees of freedom
## Residual deviance: 109.92 on 103 degrees of freedom
## AIC: 2247.5
##
## Number of Fisher Scoring iterations: 1
##
## Theta: 3.532
## Std. Err.: 0.466
##
## 2 x log-likelihood: -2241.454
summary(model3_u_ges)$dispersion
```

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

```
round(cil.lin(model3_u_ges,Exp=T),3)
```

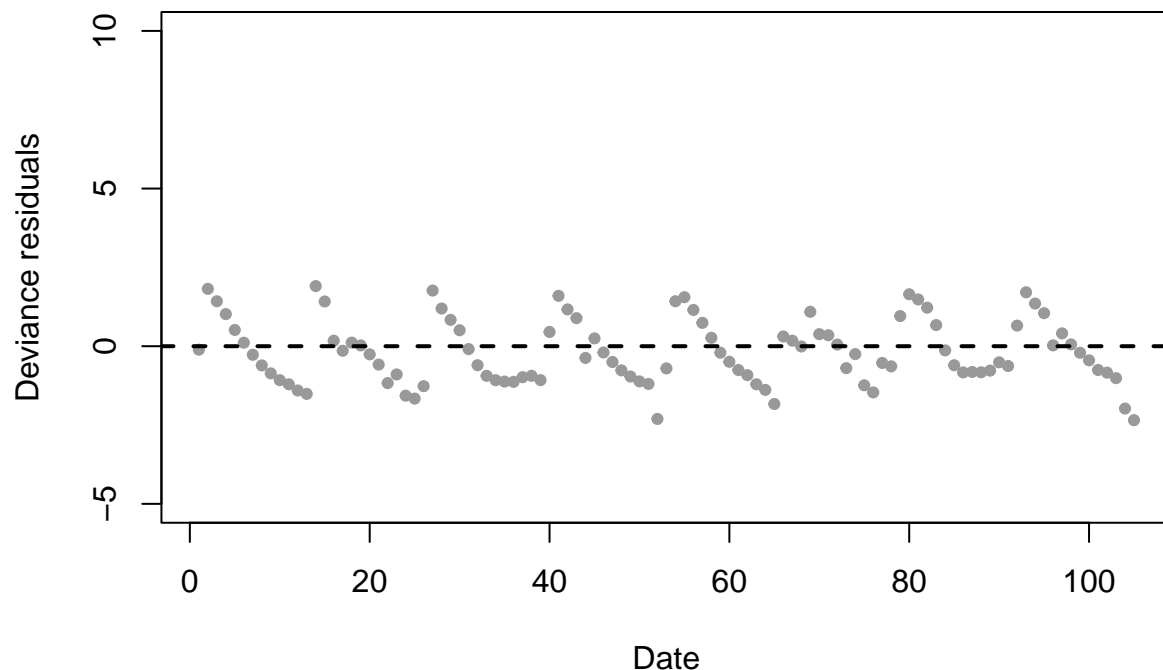
```
## Estimate StdErr z P exp(Est.) 2.5% 97.5%
## (Intercept) -4.374 0.067 -65.237 0.000 0.013 0.011 0.014
## Intervention -0.214 0.106 -2.016 0.044 0.808 0.656 0.994
```

Model checking and autocorrelation

Check the residuals by plotting against time

```
res <- residuals(model3_u_ges,type="deviance")
plot(kv_data$Zeit,res,ylim=c(-5,10),pch=19,cex=0.7,col=grey(0.6),
     main="Residuals over time",ylab="Deviance residuals",xlab="Date")
abline(h=0,lty=2,lwd=2)
```

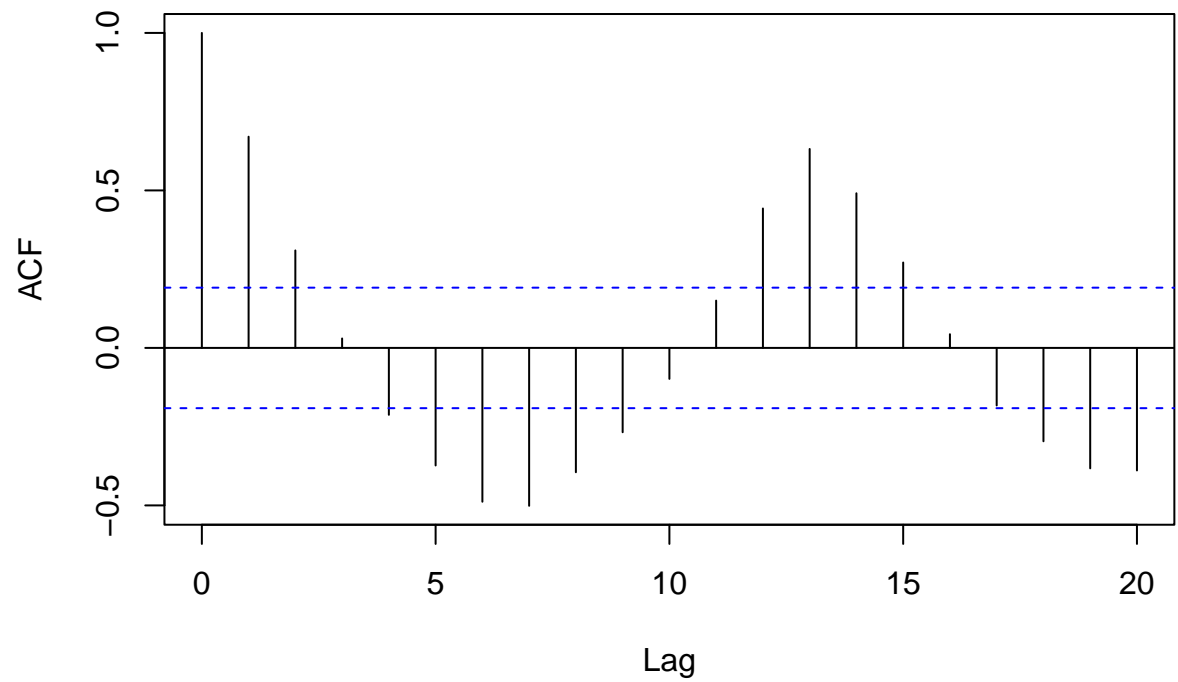
Residuals over time



```
acf(res)
```

Further check for autocorrelation by examining the autocorrelation and partial autocorrelation

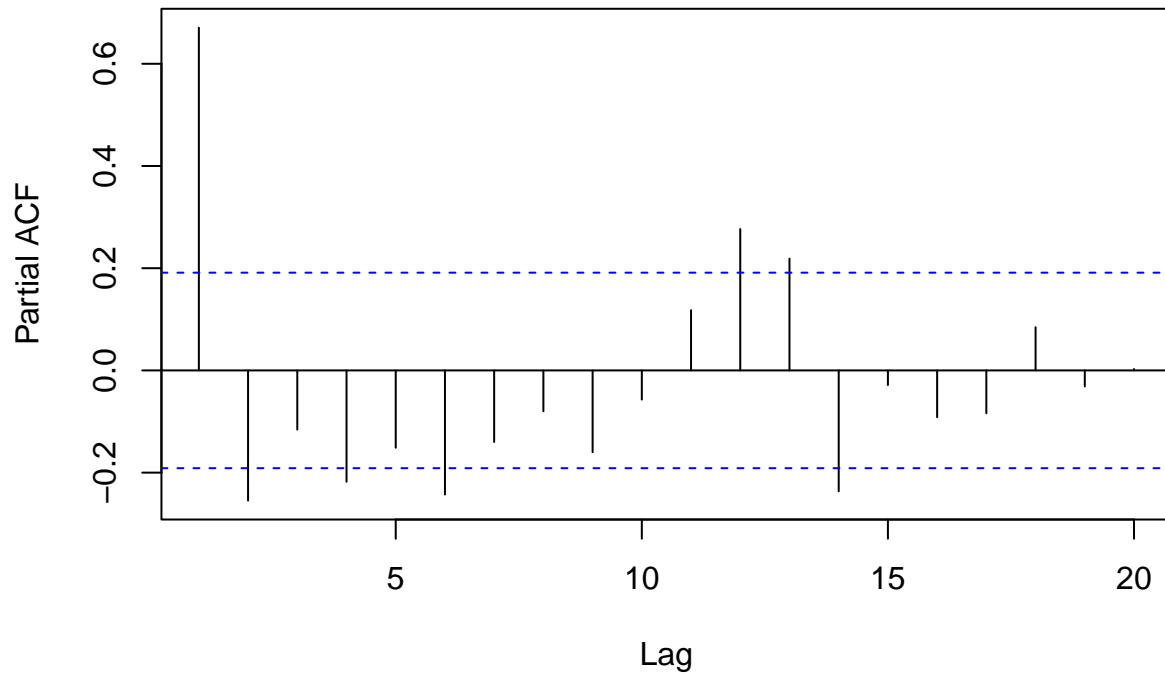
Series res



functions

```
pacf(res)
```

Series res



keine Autokorrelation

Adjustieren für Ferien und Feiertage

```
model4_u_ges <- glm.nb(P_04001 ~ offset(log(Pop2019)) + Intervention +
  Ferien + Feiertag,
  data = kv_data)
summary(model4_u_ges)
```

```
##
## Call:
## glm.nb(formula = P_04001 ~ offset(log(Pop2019)) + Intervention +
##   Ferien + Feiertag, data = kv_data, init.theta = 4.211543651,
##   link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1893  -0.9340  -0.2127   0.6924   1.7570
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.25516    0.06840 -62.213  < 2e-16 ***
## Intervention  -0.20089    0.09721  -2.067   0.0388 *
## Ferien        -0.46878    0.11910  -3.936 8.29e-05 ***
## Feiertag     -0.11865    0.12442  -0.954   0.3403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for Negative Binomial(4.2115) family taken to be 1)
##
## Null deviance: 135.85 on 104 degrees of freedom
## Residual deviance: 109.13 on 101 degrees of freedom
## AIC: 2231.4
##
## Number of Fisher Scoring iterations: 1
##
##
## Theta: 4.212
## Std. Err.: 0.560
##
## 2 x log-likelihood: -2221.388
summary(model4_u_ges)$dispersion
```

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

```
round(cil.lin(model4_u_ges,Exp=T),3)
```

	Estimate	StdErr	z	P	exp(Est.)	2.5%	97.5%
## (Intercept)	-4.255	0.068	-62.213	0.000	0.014	0.012	0.016
## Intervention	-0.201	0.097	-2.067	0.039	0.818	0.676	0.990
## Ferien	-0.469	0.119	-3.936	0.000	0.626	0.495	0.790
## Feiertag	-0.119	0.124	-0.954	0.340	0.888	0.696	1.133

Monate als Dummy-Variablen

```
kv_data <- kv_data %>%
  mutate(Jan = ifelse(Monat == 1, 1, 0),
         Feb = ifelse(Monat == 2, 1, 0),
         Mar = ifelse(Monat == 3, 1, 0),
         Apr = ifelse(Monat == 4, 1, 0),
         Mai = ifelse(Monat == 5, 1, 0),
         Jun = ifelse(Monat == 6, 1, 0),
         Jul = ifelse(Monat == 7, 1, 0),
         Aug = ifelse(Monat == 8, 1, 0),
         Sep = ifelse(Monat == 9, 1, 0),
         Okt = ifelse(Monat == 10, 1, 0),
         Nov = ifelse(Monat == 11, 1, 0),
         Dez = ifelse(Monat == 12, 1, 0))
```

Adjustieren für Saisonalität (Monate als Dummy-Variablen)

```
model6_u_ges <- glm.nb(P_04001 ~ offset(log(Pop2019)) + Intervention +
                      Ferien + Feiertag ,
                      data = kv_data)
summary(model6_u_ges)
```

```
##
## Call:
## glm.nb(formula = P_04001 ~ offset(log(Pop2019)) + Intervention +
## Ferien + Feiertag, data = kv_data, init.theta = 4.211543651,
## link = log)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1893  -0.9340  -0.2127   0.6924   1.7570
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.25516    0.06840 -62.213  < 2e-16 ***
## Intervention -0.20089    0.09721  -2.067   0.0388 *
## Ferien       -0.46878    0.11910  -3.936 8.29e-05 ***
## Feiertag     -0.11865    0.12442  -0.954  0.3403
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(4.2115) family taken to be 1)
##
##      Null deviance: 135.85  on 104  degrees of freedom
## Residual deviance: 109.13  on 101  degrees of freedom
## AIC: 2231.4
##
## Number of Fisher Scoring iterations: 1
##
##              Theta:  4.212
##              Std. Err.:  0.560
##
## 2 x log-likelihood:  -2221.388
```

```
summary(model6_u_ges)$dispersion
```

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

```
round(ci.lin(model6_u_ges,Exp=T),3)
```

```
##              Estimate StdErr      z      P exp(Est.)  2.5% 97.5%
## (Intercept)   -4.255  0.068 -62.213 0.000    0.014 0.012 0.016
## Intervention   -0.201  0.097  -2.067 0.039    0.818 0.676 0.990
## Ferien        -0.469  0.119  -3.936 0.000    0.626 0.495 0.790
## Feiertag      -0.119  0.124  -0.954 0.340    0.888 0.696 1.133
```

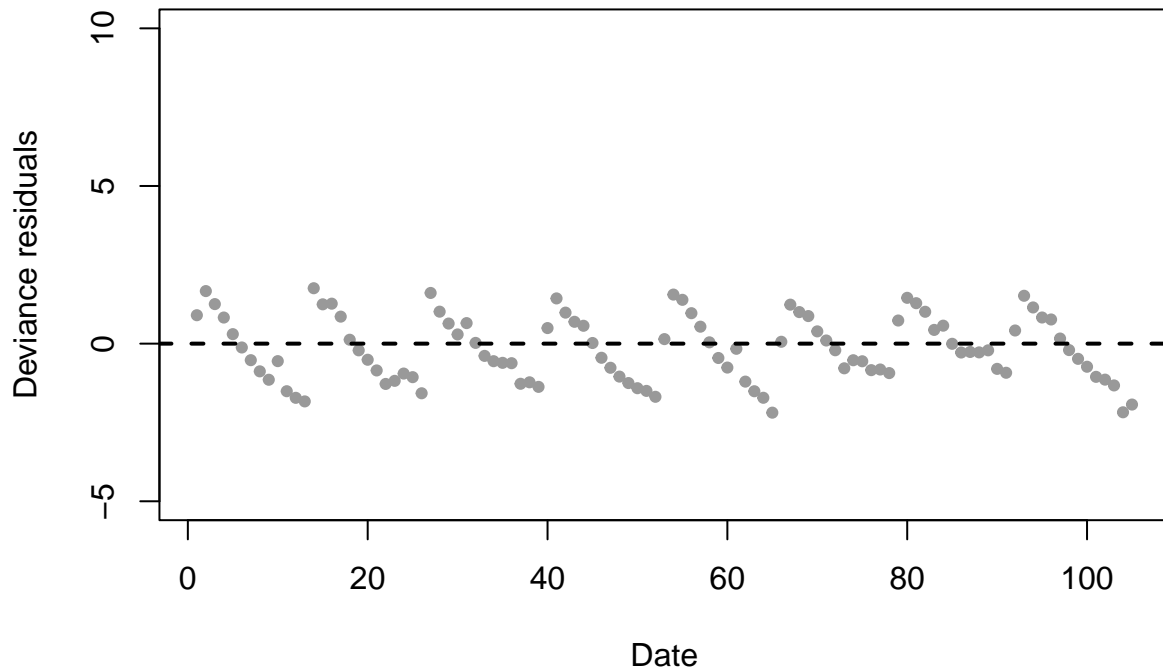
Frage: Reichen uns die Daten aus, um die typische Saisonalität zu modellieren oder wird es nicht zum Teil durch den Pandemie-Effekt modelliert?

Nochmaliges checken auf Autokorrelation

Check the residuals by plotting against time

```
res <- residuals(model6_u_ges,type="deviance")
plot(kv_data$Zeit,res,ylim=c(-5,10),pch=19,cex=0.7,col=grey(0.6),
     main="Residuals over time",ylab="Deviance residuals",xlab="Date")
abline(h=0,lty=2,lwd=2)
```

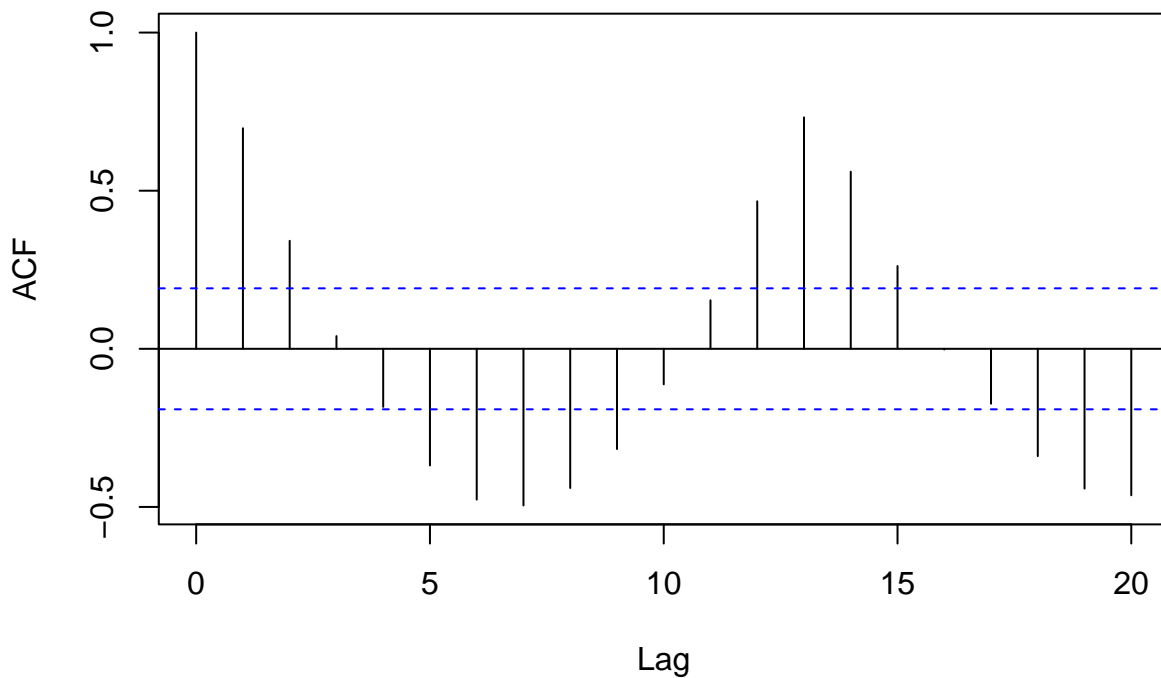

Residuals over time



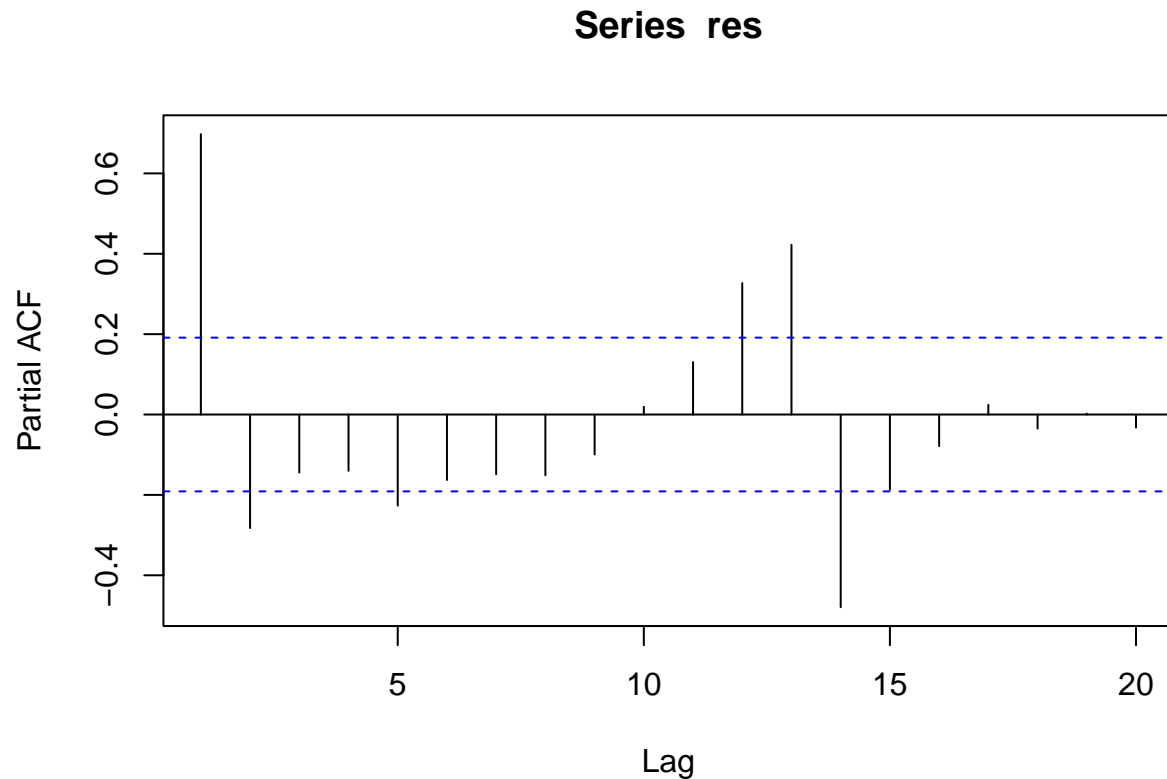
Further check for autocorrelation by examining the autocorrelation and partial autocorrelation functions

```
acf(res)
```

Series res



```
pacf(res)
```



keine Autokorrelation

```
model_u_ges_test <- glm(P_04001 ~ offset(log(Pop2019)) + Intervention +  
  harmonic(Monat,2,12),  
  family=quasipoisson, kv_data)  
summary(model_u_ges_test)
```

Adjustieren für Saisonalität, wenn man Poisson nehmen würde

```
##  
## Call:  
## glm(formula = P_04001 ~ offset(log(Pop2019)) + Intervention +  
##     harmonic(Monat, 2, 12), family = quasipoisson, data = kv_data)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max   
## -154.90  -69.96  -22.01   37.90  196.05   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)    -4.37420    0.07007  -62.428  <2e-16 ***  
## Intervention    -0.21808    0.12213   -1.786   0.0772 .      
## harmonic(Monat, 2, 12)1 -0.04126    0.08075   -0.511   0.6105      
## harmonic(Monat, 2, 12)2  0.02348    0.07828    0.300   0.7649      
## harmonic(Monat, 2, 12)3  0.02026    0.07784    0.260   0.7952      
## harmonic(Monat, 2, 12)4  0.04305    0.07958    0.541   0.5898      
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 6943.638)
##
## Null deviance: 679783  on 104  degrees of freedom
## Residual deviance: 650465  on  99  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
summary(model_u_ges_test)$dispersion
```

```
## [1] 6943.638
round(ci.lin(model_u_ges_test,Exp=T),3)
```

```
##              Estimate StdErr      z      P exp(Est.)  2.5% 97.5%
## (Intercept)      -4.374  0.070 -62.428 0.000    0.013 0.011 0.014
## Intervention      -0.218  0.122  -1.786 0.074    0.804 0.633 1.022
## harmonic(Monat, 2, 12)1 -0.041  0.081  -0.511 0.609    0.960 0.819 1.124
## harmonic(Monat, 2, 12)2  0.023  0.078   0.300 0.764    1.024 0.878 1.194
## harmonic(Monat, 2, 12)3  0.020  0.078   0.260 0.795    1.020 0.876 1.189
## harmonic(Monat, 2, 12)4  0.043  0.080   0.541 0.589    1.044 0.893 1.220
```

Löschen der Wochen im März 2020 vor dem, da unsicher, inwiefern die Intervention da schon gewirkt hat

```
kv_data1 <- kv_data %>%
  filter(! Kalenderwoche %in% c("2020-10", "2020-11"))
```

```
model6_u_ges <- glm.nb(P_04001 ~ offset(log(Pop2019)) + Intervention +
  Intervention +
  Ferien + Feiertag,
  data = kv_data1)
summary(model6_u_ges)
```

anwenden des vorherigen finalen Models

```
##
## Call:
## glm.nb(formula = P_04001 ~ offset(log(Pop2019)) + Intervention +
## Intervention + Ferien + Feiertag, data = kv_data1, init.theta = 4.278966959,
## link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2127  -0.9316  -0.2204   0.6889   1.7135
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.23377    0.06923 -61.153  < 2e-16 ***
## Intervention -0.21800    0.09705  -2.246   0.0247 *
## Ferien       -0.47816    0.11835  -4.040 5.34e-05 ***
## Feiertag     -0.12753    0.12352  -1.032   0.3018
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(4.279) family taken to be 1)
##
##      Null deviance: 135.81  on 102  degrees of freedom
## Residual deviance: 106.99  on  99  degrees of freedom
## AIC: 2188.9
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  4.279
##             Std. Err.:  0.575
##
## 2 x log-likelihood:  -2178.882
summary(model6_u_ges)$dispersion
```

```
## [1] 1
round(ci.lin(model6_u_ges,Exp=T),3)
```

	Estimate	StdErr	z	P	exp(Est.)	2.5%	97.5%
## (Intercept)	-4.234	0.069	-61.153	0.000	0.014	0.013	0.017
## Intervention	-0.218	0.097	-2.246	0.025	0.804	0.665	0.973
## Ferien	-0.478	0.118	-4.040	0.000	0.620	0.492	0.782
## Feiertag	-0.128	0.124	-1.032	0.302	0.880	0.691	1.121

keine Veränderung

Modellierung 3-phasig: Phase 2 Lockdown ab Kalenderwoche 12; Phase 3

schrittweise Öffnungen (Kita) ab Kalenderwoche 27

```
kv_data3 <- kv_data %>%
  mutate(Intervention2 = case_when(Zeit < 65 ~ "phase1",
                                    Zeit >= 65 & Zeit < 79 ~ "phase2",
                                    Zeit >= 79 ~ "phase3"))
model8_u_ges <- glm.nb(P_04001 ~ offset(log(Pop2019)) + Intervention2 +
  Zeit + Intervention2*Zeit +
  Ferien + Feiertag + Jan + Feb + Mar + Apr +
  Mai + Jun + Jul + Aug + Okt + Nov + Dez,
  data = kv_data3)
summary(model8_u_ges)

##
## Call:
## glm.nb(formula = P_04001 ~ offset(log(Pop2019)) + Intervention2 +
##      Zeit + Intervention2 * Zeit + Ferien + Feiertag + Jan + Feb +
##      Mar + Apr + Mai + Jun + Jul + Aug + Okt + Nov + Dez, data = kv_data3,
##      init.theta = 28.57245179, link = log)
##
## Deviance Residuals:
```

```

##      Min      1Q      Median      3Q      Max
## -2.10600 -0.84714 -0.08351  0.53402  2.40007
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -4.8201242   0.0924499 -52.138 < 2e-16 ***
## Intervention2phase2 -3.4719820   1.1481841  -3.024 0.002495 **
## Intervention2phase3  0.6555132   0.6207093   1.056 0.290936
## Zeit             -0.0001348   0.0014356  -0.094 0.925190
## Ferien           -0.2386374   0.0658219  -3.626 0.000288 ***
## Feiertag         -0.3240602   0.0624463  -5.189 2.11e-07 ***
## Jan              1.0385446   0.0967578  10.733 < 2e-16 ***
## Feb              0.2984469   0.0986158   3.026 0.002475 **
## Mar             -0.2862783   0.0999710  -2.864 0.004188 **
## Apr              1.1214010   0.1138718   9.848 < 2e-16 ***
## Mai              0.4143079   0.1096521   3.778 0.000158 ***
## Jun            -0.0654351   0.1164552  -0.562 0.574190
## Jul              0.9186642   0.1006006   9.132 < 2e-16 ***
## Aug              0.2528050   0.1007700   2.509 0.012116 *
## Okt              0.9547684   0.0927714  10.292 < 2e-16 ***
## Nov              0.3584908   0.1021237   3.510 0.000448 ***
## Dez            -0.1945457   0.1107342  -1.757 0.078940 .
## Intervention2phase2:Zeit 0.0454713   0.0161253   2.820 0.004804 **
## Intervention2phase3:Zeit -0.0085349   0.0066916  -1.275 0.202150
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(28.5725) family taken to be 1)
##
##      Null deviance: 920.4  on 104  degrees of freedom
## Residual deviance: 105.6  on  86  degrees of freedom
## AIC: 2053.5
##
## Number of Fisher Scoring iterations: 1
##
##              Theta: 28.57
##              Std. Err.: 3.93
##
## 2 x log-likelihood: -2013.451
summary(model8_u_ges)$dispersion

## [1] 1
round(ci.lin(model8_u_ges,Exp=T),3)

##              Estimate StdErr      z      P exp(Est.)  2.5% 97.5%
## (Intercept)      -4.820   0.092 -52.138 0.000    0.008 0.007 0.010
## Intervention2phase2 -3.472   1.148  -3.024 0.002    0.031 0.003 0.295
## Intervention2phase3  0.656   0.621   1.056 0.291    1.926 0.571 6.502
## Zeit              0.000   0.001  -0.094 0.925    1.000 0.997 1.003
## Ferien            -0.239   0.066  -3.626 0.000    0.788 0.692 0.896
## Feiertag         -0.324   0.062  -5.189 0.000    0.723 0.640 0.817
## Jan              1.039   0.097  10.733 0.000    2.825 2.337 3.415

```

## Feb	0.298	0.099	3.026	0.002	1.348	1.111	1.635
## Mar	-0.286	0.100	-2.864	0.004	0.751	0.617	0.914
## Apr	1.121	0.114	9.848	0.000	3.069	2.455	3.837
## Mai	0.414	0.110	3.778	0.000	1.513	1.221	1.876
## Jun	-0.065	0.116	-0.562	0.574	0.937	0.746	1.177
## Jul	0.919	0.101	9.132	0.000	2.506	2.057	3.052
## Aug	0.253	0.101	2.509	0.012	1.288	1.057	1.569
## Okt	0.955	0.093	10.292	0.000	2.598	2.166	3.116
## Nov	0.358	0.102	3.510	0.000	1.431	1.172	1.748
## Dez	-0.195	0.111	-1.757	0.079	0.823	0.663	1.023
## Intervention2phase2:Zeit	0.045	0.016	2.820	0.005	1.047	1.014	1.080
## Intervention2phase3:Zeit	-0.009	0.007	-1.275	0.202	0.992	0.979	1.005