3. **Inference**: interpretation of results, plotting results, confidence intervals, hypothesis tests (Wald,LRT).
4. **Asymptotic distribution** of maximum likelihood estimators and tests.
5. Checking the **adequacy of the model** (deviance, AIC), **choose between models** (nested=LRT or AIC, not nested=AIC), how well it fits the data (residuals, qqplots - but very little focus in our course).

⊕: writing this out in more detail in class.

Comparing R print-outs from LM, GLM, LMM and GLMM

Below we have fit a model to a data set, and then printed the `summary` of the model. For each of the print-outs you need to know (be able to identify and explain) every entry. In particular identify and explain:

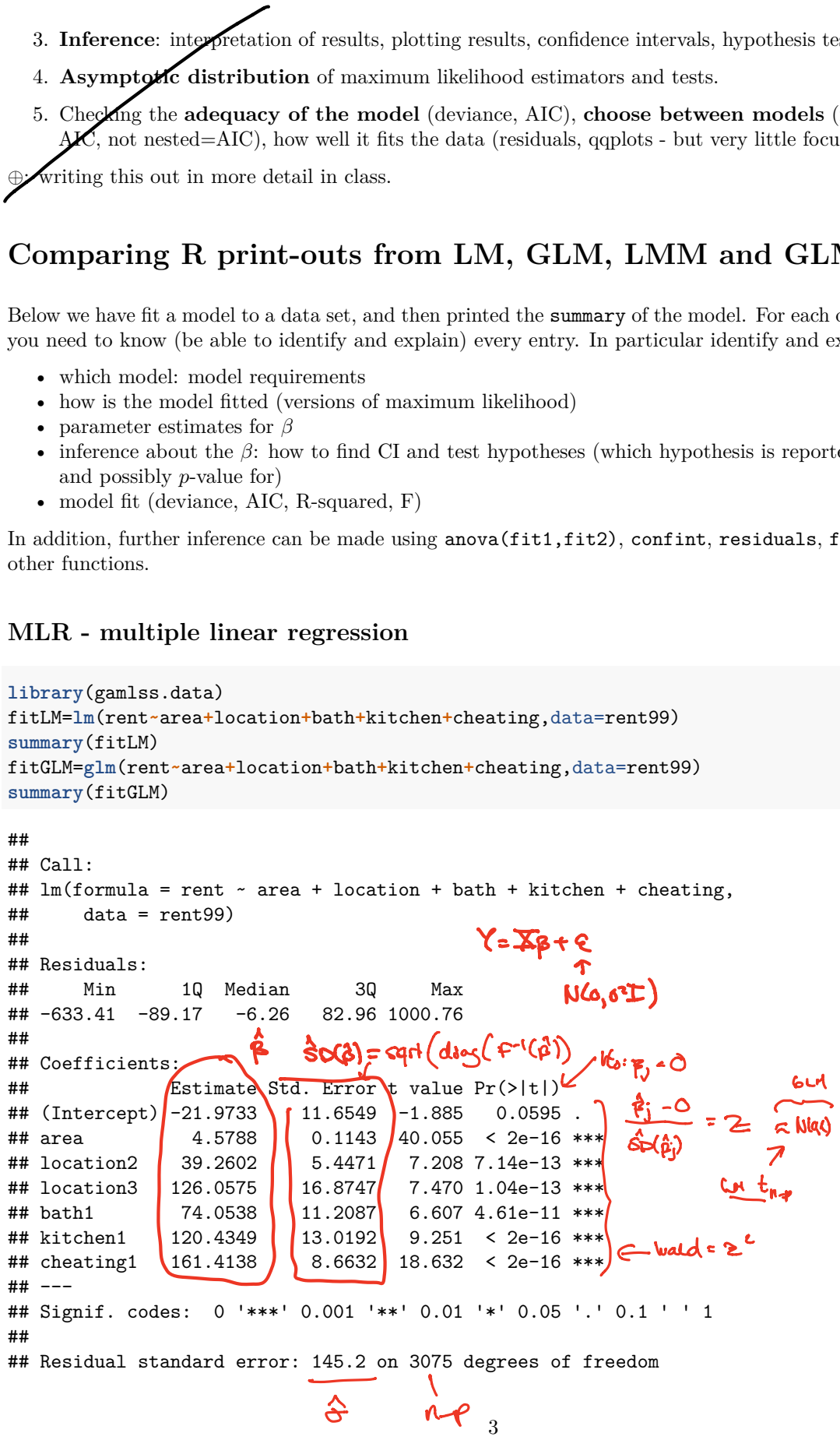
- which model: model requirements
- how is the model fitted (versions of maximum likelihood)
- parameter estimates for β
- inference about the β: how to find CI and test hypotheses (which hypothesis is reported test statistic, and possibly p-value for)
- model fit (deviance, AIC, R-squared, F)

In addition, further inference can be made using `anova(fit1,fit2)`, `confint`, `residuals`, `fitted`, `AIC` and other functions.

MLR - multiple linear regression

```
library(gamlss.data)
fitLM=lm(rent~area+location+bath+kitchen+cheating,data=rent99)
summary(fitLM)
fitGLM=glm(rent~area+location+bath+kitchen+cheating,data=rent99)
summary(fitGLM)
```

```
##
## Call:
## lm(formula = rent ~ area + location + bath + kitchen + cheating,
##     data = rent99)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -633.41  -89.17   -6.26   82.96 1000.76
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -21.9733    11.6549  -1.885   0.0595 .
## area          4.5788     0.1143  40.055  < 2e-16 ***
## location2    39.2602     5.4471   7.208 7.14e-13 ***
## location3   126.0575    16.8747   7.470 1.04e-13 ***
## bath1        74.0538    11.2087   6.607 4.61e-11 ***
## kitchen1    120.4349    13.0192   9.251  < 2e-16 ***
## cheating1   161.4138     8.6632  18.632  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 145.2 on 3075 degrees of freedom
```



```
## Multiple R-squared:  0.4504, Adjusted R-squared:  0.4494
## F-statistic:   420 on 6 and 3075 DF,  p-value: < 2.2e-16
##
##
## Call:
## glm(formula = rent ~ area + location + bath + kitchen + cheating,
##     data = rent99)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -633.41   -89.17    -6.26    82.96  1000.76
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -21.9733    11.6549  -1.885   0.0595 .
## area          4.5788     0.1143  40.055  < 2e-16 ***
## location2    39.2602     5.4471   7.208 7.14e-13 ***
## location3   126.0575    16.8747   7.470 1.04e-13 ***
## bath1        74.0538    11.2087   6.607 4.61e-11 ***
## kitchen1    120.4349    13.0192   9.251  < 2e-16 ***
## cheating1   161.4138     8.6632  18.632  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 21079.53)
##
##     Null deviance: 117945363  on 3081  degrees of freedom
## Residual deviance:  64819547  on 3075  degrees of freedom
## AIC: 39440
##
## Number of Fisher Scoring iterations: 2
```

GLM - Binomial regresion with logit-link

```
library(investr)
fitgrouped=glm(cbind(y, n-y) ~ ldose, family = "binomial", data = investr::beetle)
summary(fitgrouped)
```

```
##
## Call:
## glm(formula = cbind(y, n - y) ~ ldose, family = "binomial", data = investr::beetle)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5941  -0.3944   0.8329   1.2592   1.5940
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -60.717      5.181  -11.72   <2e-16 ***
## ldose         34.270      2.912   11.77   <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: 284.202  on 7  degrees of freedom
## Residual deviance:  11.232  on 6  degrees of freedom
## AIC: 41.43
##
## Number of Fisher Scoring iterations: 4
```

GLM - Poisson regression with log-link

```
crab=read.table("https://www.math.ntnu.no/emner/TMA4315/2017h/crab.txt")
colnames(crab)=c("Obs","C","S","W","Wt","Sa")
crab=crab[,-1] #remove column with Obs
crab$C=as.factor(crab$C)
model3=glm(Sa~W+C,family=poisson(link=log),data=crab,contrasts=list(C="contr.sum"))
summary(model3)
```

```
##
## Call:
## glm(formula = Sa ~ W + C, family = poisson(link = log), data = crab,
##     contrasts = list(C = "contr.sum"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.0415  -1.9581  -0.5575   0.9830   4.7523
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.92089    0.56010  -5.215 1.84e-07 ***
## W            0.14934    0.02084   7.166 7.73e-13 ***
## C1           0.27085    0.11784   2.298   0.0215 *
## C2           0.07117    0.07296   0.975   0.3294
## C3          -0.16551    0.09316  -1.777   0.0756 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 632.79  on 172  degrees of freedom
## Residual deviance: 559.34  on 168  degrees of freedom
## AIC: 924.64
##
## Number of Fisher Scoring iterations: 6
```

LMM - random intercept and slope

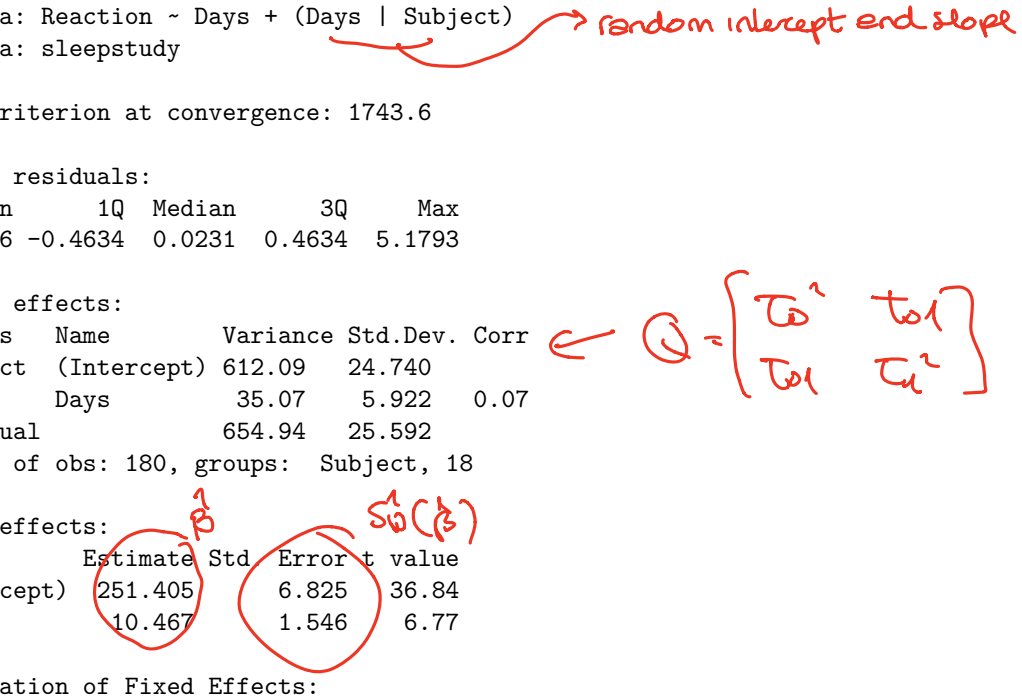
```
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 3.4.2
```

```
## Loading required package: Matrix
```

```
fm1 <- lmer(Reaction ~ Days + (Days | Subject), sleepstudy)
summary(fm1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: Reaction ~ Days + (Days | Subject)
##    Data: sleepstudy
##
## REML criterion at convergence: 1743.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9536 -0.4634  0.0231  0.4634  5.1793
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 612.09   24.740
##           Days         35.07    5.922   0.07
##  Residual             654.94   25.592
## Number of obs: 180, groups:  Subject, 18
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  251.405      6.825   36.84
## Days          10.467      1.546    6.77
##
## Correlation of Fixed Effects:
##      (Intr)
## Days -0.138
```



GLMM - random intercept Poisson

```
library("AED")
data(RIKZ)
library(lme4)
fitRI=glmer(Richness~NAP +(1|Beach),data=RIKZ,family=poisson(link=log))
summary(fitRI)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: Richness ~ NAP + (1 | Beach)
##    Data: RIKZ
##
##      AIC      BIC   logLik deviance df.resid
##    220.8    226.2   -107.4    214.8       42
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9648 -0.6155 -0.2243  0.2236  3.1869
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  Beach  (Intercept) 0.2249   0.4743
```

```
## Number of obs: 45, groups:  Beach, 9
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.66233    0.17373   9.569  < 2e-16 ***
## NAP         -0.50389    0.07535  -6.687 2.28e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##     (Intr)
## NAP 0.013
```

Exam and exam preparation

We take look at the information posted at Blackboard Exam at Blackboard and the relevant exams are found on the bottom of each module page.

Dates for supervision are also found at the exam page on Bb.

After TMA4315 - what is next?

For the 4th year student

- TMA4250 Spatial statistics
- TMA4268 Statistical learning
- TMA4275 Survival analysis
- TMA4300 Computational statistics
- KLMED8005 Analysis of repeated measurements
- SMED8002 Epidemiology 2
- TDT4300 Datavarehus og datagruvedrift
- TDT4173 Maskinlæring og case-based reasoning (Big overlap with TMA4268)
- NEVR3004 Nevrale nettverk

For the 5th year student

- Computational statistics 2 Phd course

Course evaluation in TMA4315

Please answer the course evaluation (anonymous): https://kvass.svt.ntnu.no/TakeSurvey.aspx?SurveyID=tma4315h2017