

Løsningsforslag M4STI1 2017F

Opgave 1

Jeg skal bruge effektiviteten for de fire delprocesser, som er oplyst i opgaveteksten:

$$p_A = 0.9; \quad p_B = 0.85; \quad p_C = 0.92; \quad p_D = 0.87$$

Æbletræet bærer $n = 975$ æbler, hvoraf $n_M = 850$ er plukkemodne, så $n_U = n - n_M = 125$ er umodne.

- a. Æbleplukkerobotten genkender 90% af de modne æbler, så det forventes, at den på det pågældende træ vil genkende

$$n_A = n_M \cdot p_A = 850 \cdot 0.9 = \mathbf{765}$$

Det vil lykkes robotten at gribe fat om 85% af de genkendte æbler, så

$$n_B = n_A \cdot p_B = 765 \cdot 0.85 = \mathbf{650.25}$$

Robotten plukker 92% af de æbler, den har grebet fat om, så

$$n_C = n_B \cdot p_C = 650.25 \cdot 0.92 = \mathbf{598.23}$$

Det lykkes robotten at pakke 87% af de æbler, den har plukket, så

$$n_D = n_C \cdot p_D = 598.23 \cdot 0.87 = \mathbf{520.46}$$

Den samlede effektivitet er

$$e = \frac{n_D}{n_M} = \frac{520.46}{850} = 0.6123 = \mathbf{61.23\%}$$

- b. De modne æbler, der er tilbage på træet efter plukningen, er dem, som robotten ikke har genkendt, plus dem, den har genkendt, men ikke har kunnet gribe. De æbler, som robotten har genkendt og grebet, men ikke har kunnet plukke er faldet på jorden, så de sidder ikke tilbage på træet.

Dermed kan vi beregne antallet af modne æbler, der ikke er blevet genkendt, som

$$n_{A^c} = n_M \cdot (1 - p_A) = 850 \cdot (1 - 0.9) = 85$$

Antal modne æbler, der er blevet genkendt, men ikke grebet, er

$$n_{B^c} = n_A \cdot (1 - p_B) = 765 \cdot (1 - 0.85) = 114.75$$

Dermed er det forventede antal modne æbler tilbage på træet efter robottens plukning:

$$n_{M2} = n_{A^c} + n_{B^c} = 85 + 114.75 = \mathbf{199.75}$$

MatLab kode til opgave 1

```
%% Opgave 1: Effektivitet af æbleplukningens delprocesser
```

```
clc; clear; close all;
```

```
%% Basisoplysninger
```

```
Antal_Aebler = 975
```

```
Antal_Modne = 850
```

```
P_A = 0.90
```

```
P_B = 0.85
```

```
P_C = 0.92
```

```
P_D = 0.87
```

```
%% a
```

```
Antal_Genkendt = Antal_Modne*P_A
```

```
Antal_Grebet = Antal_Genkendt*P_B
```

```
Antal_Plukket = Antal_Grebet*P_C
```

```
Antal_Pakket = Antal_Plukket*P_D
```

```
Effektivitet = Antal_Pakket/Antal_Modne
```

```
%% b
```

```
% De modne æbler, der er tilbage på træet efter plukningen, er dem, som  
% robotten ikke har genkendt, plus dem, den har genkendt, men ikke har  
% kunnet gribe. De æbler, som robotten har genkendt og grebet, men ikke har  
% kunnet plukke er faldet på jorden, så de sidder ikke tilbage på træet
```

```
Antal_Ikke_Genkendt = Antal_Modne*(1 - P_A)
```

```
Antal_Ikke_Grebet = Antal_Genkendt*(1 - P_B)
```

```
Antal_Modne_Paa_Traeet = Antal_Ikke_Genkendt + Antal_Ikke_Grebet
```

Opgave 2

Basisoplysninger:

Billedbehandlingssystemet bedømmer 7% af æbler, der i virkeligheden er modne, som umodne, så

$$P(m^c | M) = 0.07$$

Desuden bedømmer det 12% af æbler, der i virkeligheden er umodne, som modne, så

$$P(m | M^c) = 0.12$$

Som i opgave 1 har vi et æbletræ med $n = 975$ æbler, hvoraf $n_M = 850$ er plukkemodne, så $n_U = n - n_M = 125$ er umodne.

- a. Sandsynligheden for at et tilfældigt udvalgt æble fra træet er modent, er

$$P(M) = \frac{n_M}{n} = \frac{850}{975} = \mathbf{0.8718}$$

Sandsynligheden for at et tilfældigt udvalgt æble fra træet er umodent, er

$$P(M^c) = \frac{n_U}{n} = \frac{125}{975} = \mathbf{0.1282}$$

Sandsynligheden for at et modent æble bliver bedømt som modent er

$$P(m | M) = 1 - P(m^c | M) = 1 - 0.07 = \mathbf{0.93}$$

Sandsynligheden for at et umodent æble bliver bedømt som umodent er

$$P(m^c | M^c) = 1 - P(m | M^c) = 1 - 0.12 = \mathbf{0.88}$$

- b. De æbler, som systemet bedømmer som modne, er dels modne æbler, der er bedømt korrekt, og dels umodne æbler, der er bedømt forkert. Vi kan bruge reglen om den totale sandsynlighed:

$$P(m) = P(m|M) \cdot P(M) + P(m | M^c) \cdot P(M^c) = 0.93 \cdot 0.8718 + 0.12 \cdot 0.1282 = \mathbf{0.8262}$$

- c. Alle æblerne i kassen er blevet bedømt modne af systemet. De umodne i kassen er dem, der er blevet fejlbedømt som modne, selvom de i virkeligheden var umodne. D.v.s. antal umodne i kassen er $50 \cdot P(M^c|m)$. Vi kan beregne $P(M^c|m)$ med Bayes' formel:

$$P(M^c|m) = \frac{P(m | M^c) \cdot P(M^c)}{P(m)} = \frac{0.12 \cdot 0.1282}{0.8262} = 0.01862$$

Dermed er det forventede antal umodne æbler i kassen $50 \cdot 0.01862 = \mathbf{0.9311}$

- d. Forkert bedømte æbler er dels modne æbler, der er bedømt som umodne, dels umodne æbler, der er bedømt som modne.

$$\text{Antal æbler, der er fejlbedømt som umodne: } f_U = n_M \cdot P(m^c | M) = 850 \cdot 0.07 = 59.5$$

$$\text{Antal æbler, der er fejlbedømt som modne: } f_M = n_U \cdot P(m | M^c) = 125 \cdot 0.12 = 15.0$$

$$\text{Antal fejlbedømte æbler: } f = f_U + f_M = 59.5 + 15.0 = \mathbf{74.5}$$

MatLab kode til opgave 2

```
% Opgave 2: Effektivitet af robottens billedbehandlingssystem
clc; clear; close all;

%% Basisoplysninger
Antal_Aebler = 975
Antal_Modne = 850
Antal_Umodne = Antal_Aebler - Antal_Modne
P_mc_givet_M = 0.07 % P(mc|M)
P_m_givet_Mc = 0.12 % P(m|Mc)

%% a
P_M = Antal_Modne / Antal_Aebler % P(M)
P_Mc = 1 - P_M % P(Mc)
P_m_givet_M = 1 - P_mc_givet_M % P(m|M)
P_mc_givet_Mc = 1 - P_m_givet_Mc % P(mc|Mc)

%% b
% Det er sandsynligheden for, at det tilfældige æble er modent og systemet
% bedømmer korrekt, plus sandsynligheden for at æblet er umodent og
% systemet fejlbedømmer det som modent.
% Loven om den totale sandsynlighed:
%  $P(m) = P(m|M) \cdot P(M) + P(m|Mc) \cdot P(Mc)$ 
P_m = P_m_givet_M * P_M + P_m_givet_Mc * P_Mc

%% c
% Alle æblerne i kassen er bedømt modne. Vi skal bruge sandsynligheden for
% at et æble, der er bedømt modent i virkeligheden er umodent:  $P(Mc|m)$ .
% Vi skal bruge Bayes' formel:
%  $P(Mc|m) = P(m|Mc) \cdot P(Mc) / P(m)$ 
P_Mc_givet_m = P_m_givet_Mc * P_Mc / P_m
Antal_I_Kassen = 50
Antal_Umodne_I_Kassen = Antal_I_Kassen * P_Mc_givet_m

%% d
% Forkert bedømte er de modne, der bedømmes umodne plus de umodne, der
% bedømmes modne:
Antal_Fejlbedømte = Antal_Modne * P_mc_givet_M + Antal_Umodne * P_m_givet_Mc
```

Opgave 3

- a. Resultatet af den lineære regression vises her:

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	36.846	2.7736	13.285	9.8415e-07
x1	-1.5257	0.1788	-8.5329	2.7373e-05

Number of observations: 10, Error degrees of freedom: 8

Root Mean Squared Error: 5.05

R-squared: 0.901, Adjusted R-Squared 0.889

F-statistic vs. constant model: 72.8, p-value = 2.74e-05

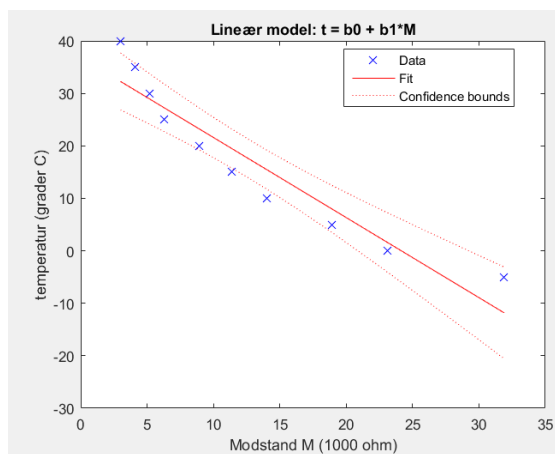
Regressionsligningen, der udtrykker temperatur t som en lineær funktion af modsand M ser sådan ud:

$$t_{lin} = 36.846 - 1.5257 \cdot M$$

Når der måles en modstand på 10.0 kΩ svarer det til en temperatur på

$$t_{lin}(10.0) = 36.846 - 1.5257 \cdot 10.0 = \mathbf{21.59^\circ C}$$

- b. Det er en god model, bedømt ud fra statistikkerne. R-squared er 0.901, så modellen forklarer 90% af variationen i data. Adjusted R-squared er 0.889, hvilket også er tilfredsstillende højt. P-værdierne for begge koefficienter er tæt på nul, så for begges vedkommende kan vi afvise nulhypotesen om, at de i virkeligheden er 0. Regressionsligningen har altså både en skæring med 2-aksen og en hældningskoefficient. Anova-analysen giver en F-værdi på 72.8. Hertil svarer en p-værdi på 2.74e-05, altså tæt på 0. Det er således nærmest umuligt, at der ikke er en korrelation mellem M og t . Når man plotter data sammen med regressionsligningen er det dog tydeligt, at sammenhængen mellem M og t ikke er lineær:



- c. Resultatet af lineær regression med den **polynomielle** model, hvor jeg har brugt Wilkinson notation:

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	46.869	1.8017	26.013	3.1734e-08
x1	-3.376	0.27717	-12.18	5.7524e-06
x1^2	0.055853	0.0081069	6.8895	0.00023351

Number of observations: 10, Error degrees of freedom: 7

Root Mean Squared Error: 1.94

R-squared: 0.987, Adjusted R-Squared 0.984

F-statistic vs. constant model: 272, p-value = 2.32e-07

Dermed er regressionsligningen:

$$t_{pol} = 46.869 - 3.376 \cdot M + 0.055853 \cdot M^2$$

Når der måles en modstand på 10.0 kΩ svarer det til en temperatur på

$$t_{pol}(10.0) = 46.869 - 3.376 \cdot 10.0 + 0.055853 \cdot (10.0)^2 = \mathbf{18.69\text{ }^{\circ}\text{C}}$$

Resultatet af lineær regression med den **logaritmiske** model:

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	61.707	0.68954	89.49	2.7131e-13
x1	-19.39	0.28767	-67.403	2.6123e-12

Number of observations: 10, Error degrees of freedom: 8

Root Mean Squared Error: 0.673

R-squared: 0.998, Adjusted R-Squared 0.998

F-statistic vs. constant model: 4.54e+03, p-value = 2.61e-12

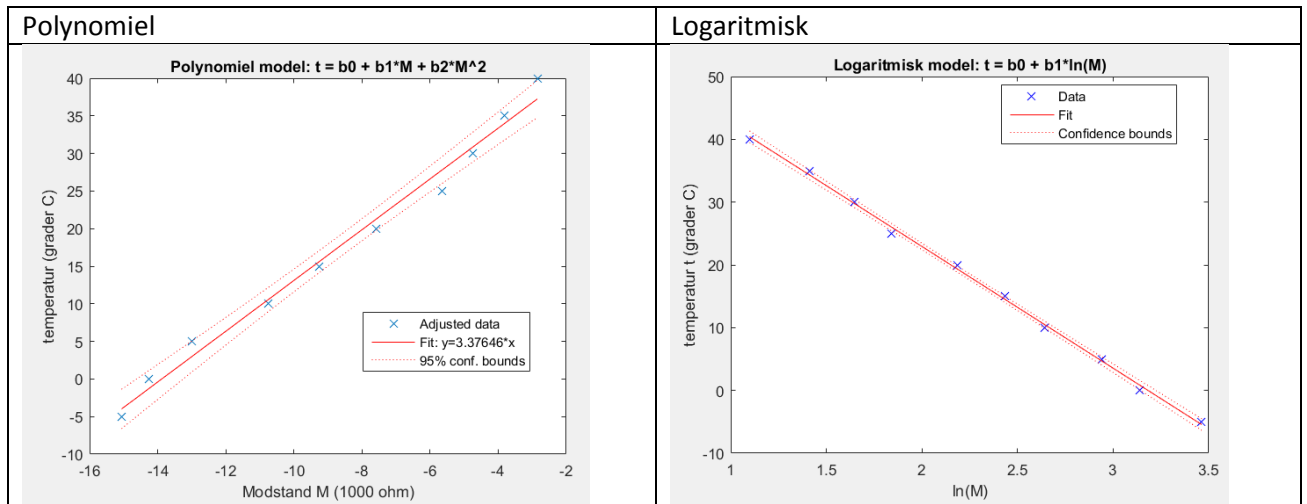
Dermed er regressionsligningen:

$$t_{log} = 61.707 - 19.39 \cdot \ln(M)$$

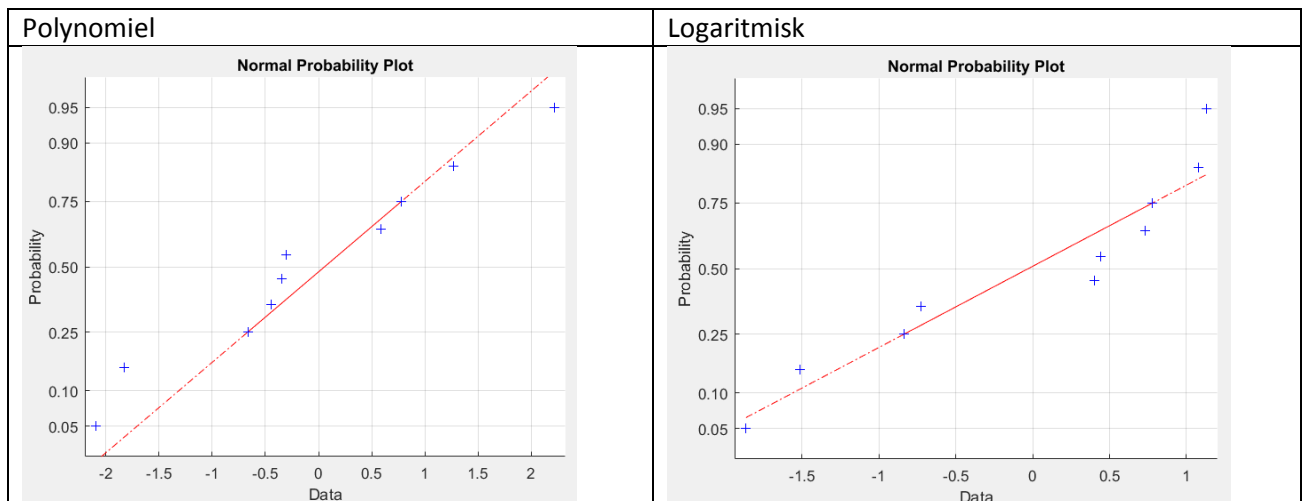
Når der måles en modstand på 10.0 kΩ svarer det til en temperatur på

$$t_{log}(10.0) = 61.707 - 19.39 \cdot \ln(10.0) = \mathbf{17.06\text{ }^{\circ}\text{C}}$$

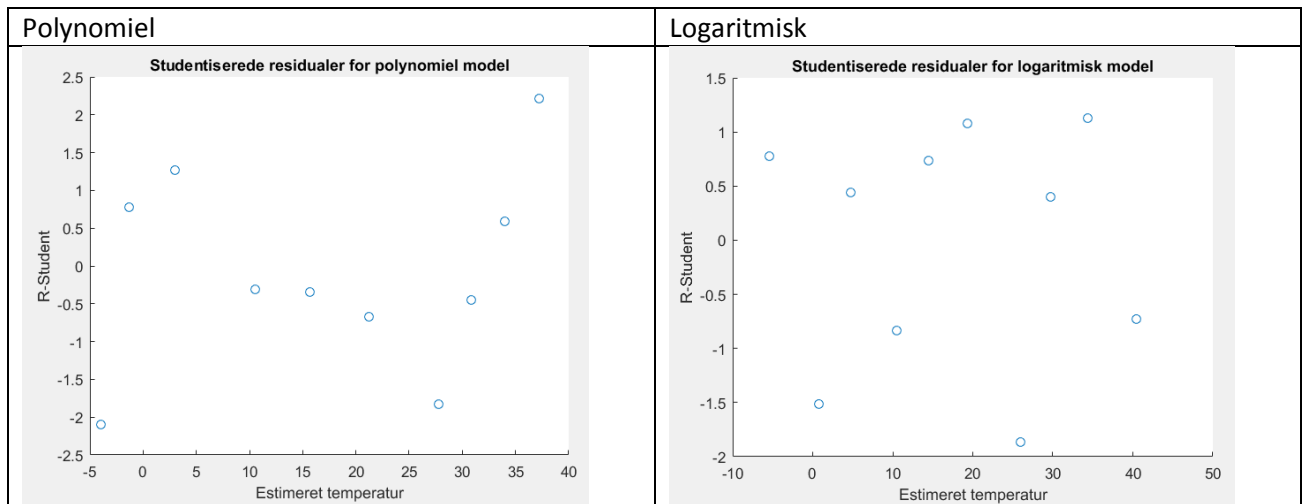
d. Scatter plots:



Normalfordelingsplots af residualer:



Residualplots:



- e. Begge modeller er gode og forklarer næsten al variationen i data (R^2 er hhv. 0.987 og 0.998. Adjusted R^2 er næsten samme værdi). Alle koefficienter er stærkt signifikante i begge modeller. Begge modeller er gode fits af data og kan forsvares.

Jeg foretrækker den logaritmiske model, fordi den trods alt har højest R^2 og Adjusted R^2 . Desuden er jeg mest tryk ved residualanalysen for den logaritmiske model: I normalfordelingsplottet for den logaritmiske model ser residualerne ud til at ligge pænere på en ret linje, så antagelsen om normalfordelte residualer er mere overbevisende. Desuden ser residualerne helt tilfældige ud i residualplottet for den logaritmiske model. I residualplottet for den polynomielle model ser de også tilfældige ud, i hvert fald er der hverken kurve- eller tragt-formet tendens, men der er en tendens til at residualerne ligger som en slange, så et residual muligvis afhænger af det forrige.

Alt i alt: Jeg foretrækker den logaritmiske model.

MatLab kode til opgave 3

```
%% Opgave 3: NTC termistor
clc; clear; close all;

%% Indlæs og behandl data
D = xlsread('Data_M4STI1_2017F.xlsx', 'A:B')
t = D(:,1) % temperatur t
M = D(:,2) % modstand M
```



```

%% a: Lineær regression
mdl1 = fitlm(M,t)

figure(1)
plot(mdl1)
title('Lineær model:  $t = b_0 + b_1 \cdot M$ ');
xlabel('Modstand M (1000 ohm)');
ylabel('temperatur (grader C)');

coeff = table2array(mdl1.Coefficients)
b0 = coeff(1,1)
b1 = coeff(2,1)

% Regressionsligning:
%  $t_{lin} = 36.8460 - 1.5257 \cdot M$     %  $t_{lin}$ : temperatur beregnet med lineær funk.

t_lin_10 = b0 + b1*10.0
% Den forventede temperatur, der svarer til en målt modstand på 10.0
% kilo-ohm er 21.5889 ~ 21.6 grader Celsius

%% b
% Det er en god model, bedømt ud fra statistikkerne. R-squared er 0.901, så
% modellen forklarer 90% af variationen i data. Adjusted R-squared er
% 0.889, hvilket også er tilfredsstillende højt. P-værdierne for begge
% koefficienter er tæt på nul, så for begges vedkommende kan vi afvise
% nulhypotesen om, at de i virkeligheden er 0. Regressionsligningen har
% altså både en skæring med 2-aksen og en hældningskoefficient.
% Anova-analysen giver en F-værdi på 72.8. Hertil svarer en p-værdi på
% 2.74e-05, altså tæt på 0. Det er således nærmest umuligt, at der ikke er
% en korrelation mellem M og t.
% Når man plotter data sammen med regressionsligningen er det dog tydeligt,
% at sammenhængen mellem M og t ikke er lineær.

%% c
% 2-gradspolynomium
M2 = M.^2
mdl2 = fitlm([M, M2],t)

% Alternativt kan Wilkinson notation bruges:
%  $mdl2 = fitlm(M, t, 'y \sim x_1 + x_1^2')$ 

% Funktionsudtryk:
t_pol = 46.869 - 3.376*M + 0.055853*M.^2 %  $t_{pol}$ : temp. beregnet med
polynomium
t_pol_10 = 46.869 - 3.376*10.0 + 0.055853*(10.0)^2
%  $t_{pol\_10} = 18.69$ 

```

```

% logaritmisk transformation
log_M = log(M)
mdl3 = fitlm(log_M,t)

% Funktionsudtryk:
% t_log = 61.707 - 19.39*log(M)
t_log_10 = 61.707 - 19.39*log(10.0)
% t_log_10 = 17.06

%% d
% Scatterplots polynomiel model
figure(2)
plot(mdl2)
title('Polynomiel model:  $t = b_0 + b_1M + b_2M^2$ ');
xlabel('Modstand M (1000 ohm)');
ylabel('temperatur (grader C)');

% Scatterplots logaritmisk model
figure(3)
plot(mdl3)
title('Logaritmisk model:  $t = b_0 + b_1\ln(M)$ ');
xlabel('ln(M)');
ylabel('temperatur t (grader C)');

% Residualplots polynomiel model:

res_pol = mdl2.Residuals.Studentized
t_pol = 46.869 - 3.376*M + 0.055853*M.^2

figure(4)
normplot(res_pol)

figure(5)
scatter(t_pol, res_pol)
title('Studentiserede residualer for polynomiel model');
xlabel('Estimeret temperatur');
ylabel('R-Student');

% Residualplots logaritmisk model:

res_log = mdl3.Residuals.Studentized
t_log = 61.707 - 19.39*log(M)

figure(6)
normplot(res_log)

figure(7)

```

```

scatter(t_log, res_log)
title('Studentiserede residualer for logaritmisk model');
xlabel('Estimeret temperatur');
ylabel('R-Student');

%% e:
% Begge modeller er gode og forklarer næsten al variationen i data
% ( $R^2$  er hhv. 0.987 og 0.998. Adjusted  $R^2$  er næsten samme værdi).
% Alle koefficienter er stærkt signifikante i begge modeller. Begge
% modeller er gode fits af data og kan forsvares.
% Jeg foretrækker den logaritmiske model, fordi den trods alt har højest
%  $R^2$  og Adjusted  $R^2$ . Desuden er jeg mest tryk ved residualanalysen for
% den logaritmiske model: I normalfordelingsplottet for den logaritmiske
% model ser residualerne ud til at ligge pænere på en ret linje, så
% antagelsen om normalfordelte residualer er mere overbevisende. Desuden
% ser residualerne helt tilfældige ud i residualplottet for den
% logaritmiske model. I residualplottet for den polynomielle model ser de
% også tilfældige ud, i hvert fald er der hverken kurve- eller tragt-formet
% tendens, men der er en tendens til at residualerne ligger som en slange,
% så et residual muligvis afhænger af det forrige.
% Alt i alt: Jeg foretrækker den logaritmiske model.

```

Opgave 4

- a. Nedenfor vises det parallelle boxplot for de to metoder til at vurdere æblers modenhed, til venstre metoden Hue_{ab} og til højre a*:



Det parallelle boxplot viser meget ensartede boxplots for de to metoder. Kasserne, kostene og kvartilerne er nogenlunde de samme, måske ligger kassen for a* en anelse lavere og har lidt mere variation. Begge boxplots har en outlier, men det er bare et udtryk for, at der er umodne æbler blandt de 24 udvalgte. Det lader til at begge metoder bedømmer dem som umodne, så det er ikke noget problem.

- b. Hvert æble måles med de to metoder, så vi skal lave en parret t-test. Vores nulhypotese er, at der ikke er forskel på metoderne, og derfor er forskellen δ på modenhedsscoren med de to metoder lig med 0 på populationsniveau. Vi beregner forskellen på de to målemetoder for de 24 æbler: $d = \text{Hue}_{ab} - a^*$

Hypoteser:

$$H_0: \delta = 0$$

$$H_a: \delta \neq 0$$

- c. Teststatistikken er:

$$d_0 = \frac{\bar{d}}{s_d/\sqrt{n}}$$

hvor \bar{d} er gennemsnittet af forskellen på målt Hue_{ab} og a^* for de 24 æbler, dvs. stikprøvemiddelværdien, s_d er stikprøve-standardafvigelsen, dvs. standardafvigelsen på vores 24 målinger af forskellen på Hue_{ab} og a^* , og $n = 24$ er stikprøvestørrelsen.

d_0 er t-fordelt med $n - 1$ frihedsgrader.

- d. Da vi har en to-sidet test forkaster vi nulhypotesen, hvis den numeriske værdi af teststatistikken er tilstrækkeligt langt fra 0. Vi forkaster nulhypotesen, hvis:

$$|d_0| > t_{\alpha/2, n-1}$$

Den kritiske grænse $t_{\alpha/2, n-1}$ beregnes i MatLab, hvor signifikansniveauet er 5% ($\alpha=0.05$):

$$t_{\alpha/2, n-1} = \text{tinv}(\alpha/2, n-1) = \mathbf{2.0687}$$

Teststatistikken beregnes til

$$d_0 = \frac{2.1250}{3.4429/\sqrt{24}} = \mathbf{3.0237}$$

Da teststatistikken er større end den kritiske region, forkaster vi nulhypotesen. Der er altså forskel på metoderne på 5% signifikansniveau.

- e. 95% konfidensintervallets bredde omkring stikprøvemiddelværdien $\bar{d} = 2.1250$ beregnes som

$$b = t_{\alpha/2, n-1} \cdot \frac{s_d}{\sqrt{n}} = 2.0687 \cdot \frac{3.4429}{\sqrt{24}} = 1.4538$$

Derfor er konfidensintervallet:

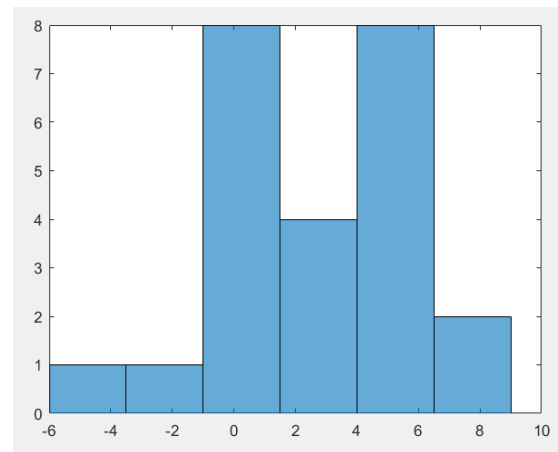
$$[\bar{d} - b; \bar{d} + b] = [2.1250 - 1.4538; 2.1250 + 1.4538] = [\mathbf{0.6712}; \mathbf{3.5788}]$$

- f. Selv om forskellen på de to metoder ser ud til at være minimal på de parallelle boxplots, så viser hypotesetesten, at forskellen er signifikant på 5% signifikansniveau. 95% konfidensintervallet viser, at forskellen på metoderne ligger imellem 0.6712 og 3.5788 med 95% sikkerhed. Da 0 ikke ligger i konfidensintervallet er dette en alternativ måde at vise, at der er forskel på de to metoder.
- g. Vores antagelse om, at teststatistikken d_0 følger en t-fordeling med $n - 1$ frihedsgrader, kommer af den centrale grænseværdisætning. Den holder, uanset hvilken fordeling stikprøven kommer fra, hvis stikprøvestørrelsen n er tilstrækkeligt stor, typisk over 30. Da $n = 24$ skal vi undersøge, om stikprøven kommer fra en 'pæn' fordeling, helst symmetrisk med et enkelt toppunkt og hurtigt uddøende haler. Det kan vi gøre med et stem-and-leaf plot, et histogram og et normalfordelingsplot:

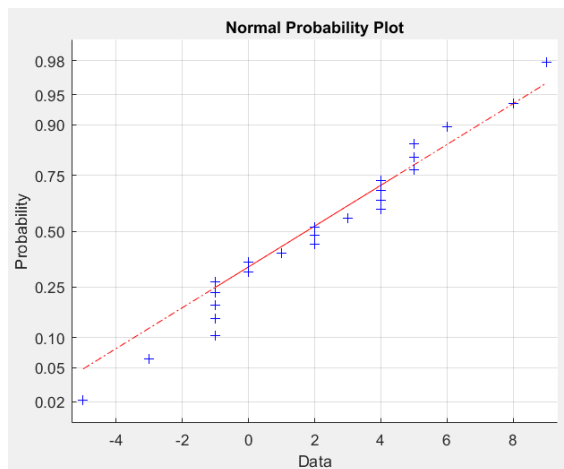
Stem-and-leaf plot:

```
-1 |  
0 | 0 0  
1 | 0  
2 | 0 0 0  
3 | 0  
4 | 0 0 0 0  
5 | 0 0 0  
6 | 0  
7 |  
8 | 0  
9 | 0  
key: 36|5 = 36.5  
stem unit: 1.0  
leaf unit: 0.1
```

Histogram:



Normalfordelingsplot:



Stem-and-leaf plot og histogram viser en nogenlunde symmetrisk fordeling med hurtigt uddøende haler. Der er lidt tvivl om den har flere toppunkter, men det kan skyldes tilfældigheder. Normalfordelingsplottet er nogenlunde lineært. Heldigvis betyder den ret høje stikprøvestørrelse på 24, at den centrale grænseværdisætning holder, også selv om fordelingen ikke er så 'pæn', som vi kunne ønske. Vi kan altså roligt gå ud fra, at antagelsen holder.

MatLab kode for opgave 4

```
% Opgave 4: Hypotesetest af metoder til vurdering af æblers modenhed
clc; clear; close all;

%% Indlæs og behandl data
D = xlsread('Data_M4STI1_2017F.xlsx','D:F')
Hue_ab = D(:,2)
a_stjerne = D(:,3)

%% a: Boxplot
figure(1)
boxplot([Hue_ab, a_stjerne], 'labels',{'Hue_ab','a*'})
title('Boxplot af to metoder til at vurdere æblers modenhed');
ylabel('Modenhedsscore [0-100]');

% Det parallelle boxplot viser meget ensartede boxplots for de to metoder.
% Kasserne, kostene og kvartilerne er nogenlunde de samme, måske ligger
% kassen for a_stjerne en anelse lavere og har lidt mere variation.
% Begge boxplots har en outlier, men det er bare et udtryk for, at der er
% umodne æbler blandt de 24 udvalgte. Det lader til at begge metoder
% bedømmer dem som umodne, så det er ikke noget problem.

%% b-d Hypotesetest med to afhængige stikprøver med ukendt varians
% Baggrundsoplysninger
alfa = 0.05
n = 24

% Det samme æble er målt med begge metoder. Derfor er målingerne ikke
% uafhængige.
% Vi foretager derfor hypotesetest på forskellen på de to målemetoder.
% Vi beregner d som forskellen på metoderne og bruger d som en stikprøve.
d = Hue_ab - a_stjerne

% Hvis metoderne er statistisk ens vil forskellen i gennemsnit være 0.
% Vi lader delta betegne forskellen i populations-middelværdi for de to
% metoder.

%% b: Skridt 1, Opstilling af hypoteser
% H0:  $\delta = \delta_0 = 0$ 
% Ha:  $\delta \neq 0$  ( $\neq$  betyder 'forskellig fra', så det er en to-sidet test)

%% c: Skridt 2, formel for teststatistikken
%  $d_0 = d_{streg} / (s_d / \sqrt{n})$ 
% hvor d_streg er gennemsnit af forskellen på de to stikprøver, dvs.
```

```

% d_streg er middelværdien af d, s_d er standardafvigelsen af d og n er
% stikprøvestørrelsen.
% Teststatistikken d_0 er t-fordelt med n-1 frihedsgrader.

%% d: Skridt 3-5
% Skridt 3: kritisk region
% Da det er en to-sidet test forkaster vi nulhypotesen, hvis
%  $|d_0| > t_{\alpha/2}$ 
% hvor  $t_{\alpha/2}$  er den værdi, hvor  $P(t > t_{\alpha/2}) = \alpha/2$ 
t_alfahalve = tinv(1 - alfa/2, n-1)

% Skridt 4: Beregning af teststørrelsen
d_streg = mean(d) % Stikprøvemiddelværdi
s_d = std(d,0) % Stikprøve-standardafvigelse
d_0 = d_streg/(s_d/sqrt(n))

% Alternativt kan stikprøve-standardafvigelsen beregnes 'manuelt':
%  $s_d = \sqrt{(n*d'^2 - \sum(d,1)^2)/(n*(n-1))}$ 

% Skridt 5: Konklusion
% Vi kan se, at d_streg = 2.1250, så i gennemsnit bedømmer metoden Hue_ab
% æblerne 2 enheder højere end a_stjerne. Er denne forskel statistisk
% signifikant? Da teststatistikken d_0 = 3.0237 er større end den øvre
% kritiske grænse på t_alfahalve = 2.0687 kan vi forkaste nulhypotesen om,
% at der ikke er forskel på metoderne.
% Vi kan beregne p-værdien:
pvalue = 2*(1 - tcdf(d_0, n-1))

% P-værdien er 0.0060, så hvis der ikke var forskel på metoderne, så ville
% vi tilfældigvis få en stikprøve med den observerede forskel, eller mere,
% i 0.6% af tilfældene. Med andre ord: Vores hypotesetest viser, at der er
% forskel på de to metoder. Testen siger ikke noget om, hvilken metode, der
% er bedst.

%% e: 95% konfidensinterval
ki_bredde = t_alfahalve*s_d/sqrt(n)
ki = [d_streg - ki_bredde, d_streg + ki_bredde]

%% f: Diskussion
% Selv om forskellen på de to metoder ser ud til at være minimal på de
% parallelle boxplots, så viser hypotesetesten, at forskellen er
% signifikant på 5% signifikansniveau. 95% konfidensintervallet viser, at
% forskellen på metoderne ligger imellem 0.6712 og 3.5788 med 95%
% sikkerhed. Da 0 ikke ligger i konfidensintervallet er dette en alternativ
% måde at vise, at der er forskel på de to metoder.

```



```

%% g: Antagelser
% Vores antagelse om, at teststatistikken  $d_0$  følger en t-fordeling med  $n-1$ 
% frihedsgrader, kommer af den centrale grænseværdisætning. Den holder
% uanset hvilken fordeling stikprøven kommer fra, hvis stikprøvestørrelsen
%  $n$  er tilstrækkeligt stor, typisk over 30. Da  $n=24$  skal vi undersøge om
% stikprøven kommer fra en 'pæn' fordeling, helst symmetrisk med et enkelt
% toppunkt og hurtigt uddøende haler

stemleafplot(d,-1)

figure(2)
histogram(d,6)

figure(3)
normplot(d)

% Stem-and-leaf plot og histogram viser en nogenlunde symmetrisk fordeling
% med hurtigt uddøende haler. Der er lidt tvivl om den har flere toppunkter,
% men det kan skyldes tilfældigheder.
% Normalfordelingsplottet er nogenlunde lineært. Heldigvis betyder den ret
% høje stikprøvestørrelse på 24, at den centrale grænseværdisætning holder,
% også selv om fordelingen ikke er så 'pæn', som vi kunne ønske.
% Vi kan altså roligt gå ud fra, at antagelsen holder.

```