Uitwerkingen tentamen Statistiek 1 voor Scheikunde (2DS00) van maandag 7 oktober 2002

N.B. Ieder onderdeel telt voor 2 punten, dus in totaal 40 punten.

1.

a) Aangezien de massa van een lege pyknometer zonder meetonnauwkeurigheid bepaald kan worden, moet dit in de berekening als een constante worden meegenomen. Hieruit volgt dat

$$0,01^2 \geq Var(\rho) = \frac{1}{0,05^2} Var(m_{vol} - m_{leeg}) = \frac{1}{0,05^2} Var(m_{vol}) \Leftrightarrow$$

$$\sigma_{m_{vol}} \leq 0,01 * 0,05 = 0,0005 \text{ g/l} = 5.10^{-4} \text{ g/l}$$

b)

$$0,01^2 \geq Var(\rho) = \frac{1}{0,05^2} Var(\overline{m}_{vol}) = \frac{1}{0,05^2} \frac{Var(\overline{m}_{vol})}{5} \Leftrightarrow$$

$$\sigma_{m_{vol}} \leq 0,01 * 0,05 * \sqrt{5} = 0,001 \text{ g/l} = 1.10^{-3} \text{ g/l}$$

c)

$$0,01^2 \geq Var(\rho) = \frac{1}{0,05^2} Var(m_{vol} - m_{leeg}) = \frac{1}{0,05^2} \left(Var(m_{vol}) + Var(m_{leeg}) \right) \Leftrightarrow$$

$$\sigma_{m_{vol}} \leq \frac{0,01 * 0,05}{\sqrt{2}} = 0,0004 \text{ g/l} = 4.10^{-4} \text{ g/l}$$

d)

$$\rho = \frac{m_{vol} - m_{leeg}}{V} = f(m_{vol}, m_{leeg}, V); \; \frac{\partial f}{\partial m_{vol}} = \frac{1}{V}; \; \frac{\partial f}{\partial m_{vol}} = -\frac{1}{V}; \; \frac{\partial f}{\partial V} = \frac{m_{vol} - m_{leeg}}{V^2}$$

$$Var(\rho) = \left(\frac{\partial f}{\partial m_{vol}} (0,302; 0,0352; 0,050) \right)^2 * Var(m_{vol}) +$$

$$\left(\frac{\partial f}{\partial m_{leeg}} (0,302; 0,0352; 0,050) \right)^2 * Var(m_{leeg}) + \left(\frac{\partial f}{\partial V} (0,302; 0,0352; 0,050) \right)^2 * Var(V) =$$

$$\frac{0,001^2}{0,050^2} + \frac{0,0009^2}{0,050^2} + \left(\frac{(0,302 - 0,0352)}{0,050^2} \right)^2 * 0,00001^2 = 7,25.10^{-4} \Rightarrow \sigma_\rho = 0,03$$

Het kan ook via de formule voor foutenvoortplanting van variatiecoëfficiënten:

$$\frac{Var(\rho)}{\mu_\rho^2} = \frac{Var(m_{vol} - m_{leeg})}{\left(\mu_{m_{vol} - m_{leeg}}\right)^2} + \frac{Var(V)}{\mu_V^2} = \frac{Var(m_{vol}) + Var(m_{leeg})}{\left(\mu_{m_{vol}} - \mu_{m_{leeg}}\right)^2} + \frac{Var(V)}{\mu_V^2} \Leftrightarrow$$

$$Var(\rho) = 5,34^2 \left(\frac{0,001^2 + 0,0009^2}{(0,302 - 0,0352)^2} + \frac{0,00001^2}{0,050^2} \right) = 7,25.10^{-4} \Rightarrow \sigma_\rho = 0,03$$

2. Er zijn hier veel goede analyses mogelijk (afhankelijk van de gekozen termen in het model, zoals interactietermen of eventueel kwadratische termen). De opgave wordt beoordeeld door te kijken of de standaard onderdelen van een regressie-analyse (zoals aangegeven in het stappenplan in het collegedictaat) goed gevolgd wordt.

3.

a) Merk eerst op dat $-\pi^2 / R^2 \approx -0,39$. Het is nu handig om als onafhankelijke variable $-\pi^2 / R^2 * t = -0,39t$ te nemen. De schattingen voor a en b volgen nu een gewone lineaire regressie uit te voeren met $\ln(\theta)$ als onafhankelijke variabele of via de optie **Exponential model**.

Regression Analysis - Linear model: Y = a + b*X

Dependent variable: LOG(Theta)
Independent variable: -0,4*Tijd

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	0,670242	0,014265	46,985	0,0000
Slope	0,121765	0,000936544	130,015	0,0000

Regression Analysis - Exponential model: Y = exp(a + b*X)

Dependent variable: Theta
Independent variable: -0,4*Tijd

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	0,670242	0,014265	46,985	0,0000
Slope	0,121765	0,000936544	130,015	0,0000

Hieruit volgen de schattingen $a = 0,124$ en $b = e^{0,67} = 1,95$.

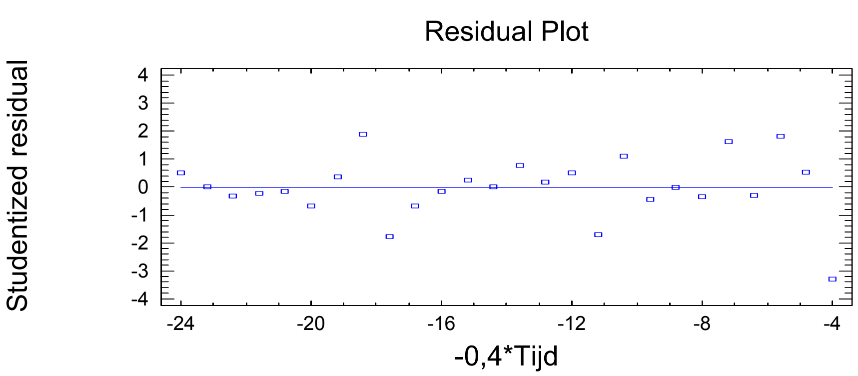
b)

Unusual Residuals

Row	X	Y	Predicted Y	Residual	Studentized Residual
1	-4,0	0,109751	0,183183	-0,0734322	-3,29

Influential Points

Row	X	Y	Predicted Y	Studentized Residual	Leverage



Er is één uitbijter dat geen invloedrijk punt is. Deze waarneming is de waarneming met de laagste temperatuur. Aangezien de waarneming geen invloedrijk punt is, zal weglaten van deze waarneming geen grote invloed hebben op de parameterschattingen. Bekend is dat de formule niet goed werkt bij korte tijden, dus lage temperaturen. Daarom is het in dit geval geen probleem om de waarneming weg te laten.

c) Hier moet zowel gekeken worden of het model significant is als de gebruikelijke modelveronderstellingen gecontroleerd worden via residuenplots, density trace en normal probability plots + bijbehorende toetsen. Zie het stappenplan in het collegedictaat voor details.

d) Het onderstaande betrouwbaarheidsinterval is gebaseerd op de oorspronkelijke waarnemingen. Bij de regressie-analyse was –0,4*tijd als onafhankelijke variabele gebruikt en $\ln(\theta)$ als afhankelijke variabele. Dus t = 75 wordt –0,4*75 = -30. Statgraphics geeft dus een betrouwbaarheidsinterval voor $\ln(\theta)$, niet voor de temperatuur. Het verband tussen temperatuur en $\ln(\theta)$ wordt gegeven door $T = 100 - 75 * e^{\ln(\theta)}$.

Predicted Values

		95,00% Prediction Limits		95,00% Confidence Limits	
X	Predicted Y	Lower	Upper	Lower	Upper
-30,0	-2,9827	-3,05044	-2,91497	-3,01573	-2,94967

Het gevraagde betrouwbaarheidsinterval is dus (96,1;96,3).

4. De eerste stap is het bepalen van startwaarden voor de niet-lineaire regressie. De Radke-Prausnitz isotherm is om te schrijven als een verband tussen $1/q$ en $1/C_e$: $\frac{1}{q} = \frac{1}{K}\frac{1}{C_e} + \frac{1}{F}\frac{1}{C_e^N}$. Het probleem is dat dit een polynoom is waarvan de orde een parameter is. Daarom proberen we eerst of een vereenvoudiging werkt door $N = 0$ te nemen. Een grafiek waarin $1/q$ tegen $1/C_e$ uitgezet wordt geeft een redelijk rechte lijn te zien. Om een startwaarde voor K te vinden passen we daarom een lineaire regressie toe met $1/q$ als afhankelijke variabele en $1/C_e$ als onafhankelijke variabele.

Regression Analysis - Linear model: Y = a + b*X

Dependent variable: 1/q
Independent variable: 1/Ce

Parameter	Estimate	Standard Error	T Statistic	P-Value
Intercept	3,19779	0,225876	14,1573	0,0000
Slope	0,979643	0,0584557	16,7587	0,0000

Dit levert $1/0{,}979 \approx 1{,}02$ als startwaarde voor K. Een verfijning is mogelijk door niet de gehele data set te gebruiken, maar de grote waarden van C_e (dus de zeer kleine waarden van $1/C_e$) weg te laten (i.h.b. door de uitkomsten van experiment 1 niet mee te nemen). Om tot startwaarden voor de parameters F

en N te komen, vullen we de startwaarde voor K in en gebruiken $\ln\left(\frac{1}{q}-\frac{1}{K}\frac{1}{C_e}\right)$ als afhankelijke variabele en $-\ln\left(C_e\right)$ als onafhankelijke variabele (de natuurlijke logaritme ln wordt in Statgraphics met LOG aangeduid).

```
Regression Analysis - Linear model: Y = a + b*X
-----------------------------------------------------------------------------
Dependent variable: LOG(1/q-0,979/Ce)
Independent variable: -LOG(Ce)
-----------------------------------------------------------------------------
                               Standard          T
Parameter       Estimate         Error       Statistic        P-Value
-----------------------------------------------------------------------------
Intercept        1,38985       0,026749        51,9591         0,0000
Slope          0,0881115      0,0066066        13,3369         0,0000
-----------------------------------------------------------------------------
```

Aangezien $\ln\left(\frac{1}{q}-\frac{1}{K}\frac{1}{C_e}\right)=-\ln F-N\ln C_e$, lezen we hieruit af dat Hieruit lezen we af dat $-\ln F\approx 1,39\Rightarrow F\approx e^{-1,39}\approx 0,25$ en $N\approx\sim 0,088$.

```
Nonlinear Regression
--------------------
Dependent variable: q
Independent variables:
   Ce

Function to be estimated: 1/(1/(K*Ce)+1/(F*(Ce^N)))
Initial parameter estimates:
   K = 1,02
   F = 0,25
   N = -0,088

Estimation method: Marquardt
Estimation stopped due to convergence of residual sum of squares.
Number of iterations: 6
Number of function calls: 27

Estimation Results
-----------------------------------------------------------------------------
                                                      Asymptotic 95,0%
                                      Asymptotic      Confidence Interval
Parameter                Estimate  Standard Error        Lower          Upper
-----------------------------------------------------------------------------
K                        0,789725       0,0661977      0,655934       0,923515
F                        0,270585       0,0112221      0,247904       0,293266
N                       0,0734117      0,00821371     0,0568111      0,0900122
-----------------------------------------------------------------------------

Analysis of Variance
-----------------------------------------------------
Source               Sum of Squares     Df  Mean Square
-----------------------------------------------------
Model                       4,29653      3      1,43218
Residual                 0,00711365     40  0,000177841
-----------------------------------------------------
Total                       4,30365     43
Total (Corr.)              0,743387     42
```

De uitkomsten van deze niet-lineaire regressie analyse moeten nu volgens de gebruikelijke stappen bekeken worden (zie dictaat): residuenplots, normal probability plot, invloed van startwaarden enz. Er zijn enkele verdachte waarnemingen.