57. Die folgende Tabelle gibt die Entwicklung der Weltbevölkerung (in Milliarden) seit dem Jahr 1950 wieder:

Jahr t	1950	1960	1970	1980	1990	2000
Bevölkerung f(t)	2,5	3	3,6	4,4	5,3	?

Man finde eine Trendfunktion der Form $g(t)=c\cdot e^{at}$ und extrapoliere die Bevölkerungszahl für das Jahr 2000. (Hinweis: Man bestimme die Ausgleichsgerade für die Wertepaare (t, $\ln g(t)$) nach der Methode der kleinsten Quadrate.)

Geg.: komplizierte Funktion $f: I \to \mathbb{R}$
Ges.: polynomielle Ersatzfunktion $p: I \to \mathbb{R}$, die f möglichst gut annähert

Prinzip der kleinsten Quadrate: $Q=\sum_{i=1}^{n}\left(f(x_i)-p(x_i)\right)^2=\min$

$f(x)$ ist gegeben durch $f(x_i)=y_i$ für $i=1,...,n$

Gesucht wird Ersatzfunktion $p(x)=a+bx$, sodass $Q(a,b)=\sum_{i=1}^{n}\left(f(x_i)-p(x_i)\right)^2=\min$

$Q(a,b)=\sum_{i=1}^{n}(y_i-a-bx_i)^2=\min \Rightarrow$ notwendige Bedingung

1) $\frac{\partial Q}{\partial a}=2\cdot\sum_{i=1}^{n}(y_i-a-bx_i)^2\cdot(-1)=0$

2) $\frac{\partial Q}{\partial b}=2\cdot\sum_{i=1}^{n}(y_i-a-bx_i)^2\cdot(x_i)=0$

add 1) $\sum_{i=1}^{n}(y_i-a-bx_i)=0 \to \sum_{i=1}^{n}y_i-na-b\sum_{i=1}^{n}x_i=0$

aus der Statistik ist gegeben: $\overline{x}=\frac{1}{n}\sum x_i \Rightarrow n\overline{x}=\sum x_i$ sowie $\overline{y}=\frac{1}{n}\sum y_i \Rightarrow \overline{y}n=\sum y_i$

eingesetzt in $\sum_{i=1}^{n}y_i-na-b\sum_{i=1}^{n}x_i=0 \Rightarrow \overline{y}n-na-b\overline{x}n=0$, nun durch n dividieren und wir bekommen $\overline{y}-a-b\overline{x}=0 \Rightarrow \overline{y}=a+b\overline{x}$, das heißt $(\overline{x},\overline{y})$ liegt auf $y=a+bx$

add 2) $\sum_{i=1}^{n}(y_i-a-bx_i)\cdot(-x_i)=0 \Rightarrow \sum_{i=1}^{n}(y_i-a-bx_i)\cdot(x_i)=\sum x_iy_i-a\sum x_i-b\sum x_i$

wir können nun teilweise Einsetzungen machen: $\overline{y}=a+b\overline{x} \Rightarrow a=\overline{y}-b\overline{x}$ und aus der Statistik $\overline{x}=\frac{1}{n}\sum x_i \Rightarrow n\overline{x}=\sum x_i$, also eingesetzt ergibt es dann:

$$\sum x_iy_i-\underbrace{a}_{\overline{y}-b\overline{x}}\underbrace{\sum x_i}_{n\overline{x}}-b\sum x_i=\sum x_iy_i-(\overline{y}-b\overline{x})\cdot(n\overline{x})-b\sum x_i=\sum x_iy_i-n\overline{xy}+bn\overline{x}^2-b\sum x_i$$

nun können wir uns das b ausdrücken:

$$0=\sum x_i y_i - n\overline{xy} + bn\overline{x}^2 - b\sum x_i^{\,2}$$
$$b\sum x_i^{\,2} - bn\overline{x}^2 = \sum x_i y_i - n\overline{xy}$$
$$b\left(\sum x_i^{\,2} - n\overline{x}^2\right) = \sum x_i y_i - n\overline{xy}$$
$$b=\frac{\sum x_i y_i - n\overline{xy}}{\sum x_i^{\,2} - n\overline{x}^2}$$

b ist der Regressionskoeffizient von y auf x, damit sind a und b bestimmt.
also zusammengefasst haben wir:

$$\overline{x}=\frac{1}{n}\sum x_i\,,\ \overline{y}=\frac{1}{n}\sum y_i\,.\ a=\overline{y}-b\overline{x}\,,\ b=\frac{\sum x_i y_i - n\overline{xy}}{\sum x_i^{\,2} - n\overline{x}^2}$$

Wobei unsere Ausgangsfunktion $Q(a,b)=\sum_{i=1}^{n}\left(f(x_i)-p(x_i)\right)^2=\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2=\min$ lautet und unsere gesuchte Ausgleichsgerade die Form $p(x)=a+bx$ hat.

Bis jetzt haben wir mit der Formel laut Skriptum $Q(a,b)=\sum_{i=1}^{n}\left(f(x_i)-p(x_i)\right)^2=\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2=\min$ gearbeitet, wir haben jedoch $g(t)=c\cdot e^{at}$ gegeben und sollen für die Wertpaare $(t,\ \ln g(t))$ eine Ausgleichsgerade suchen, also ist unsere Trendfunktion $p(x_i)=\ln\left(g(t)\right)=\ln(c)+at\cdot\ln(e)=\ln(c)+at$

Das heißt also bei uns gilt $a=\ln c$, $b=a$ und $y=t$, wobei wir als Variable t haben (nicht x), also $Q=\sum_{i=1}^{n}\left(f(t_i)-p(t_i)\right)^2=\sum_{i=1}^{n}\left(y_i-\ln c-at_i\right)^2$.

Zu erst habe ich „unsere“ y-Warte berechnet, indem ich den natürlichen Logarithmus $\ln(\ )$ von den $g(t)$-Werten berechnet habe (siehe Tabelle unten).

Dann muss man die Formel $Q=\sum_{i=1}^{n}\left(y_i-\ln c-at_i\right)^2$ wie oben vorgeführt umformen und bekommen $a=\frac{\sum t_i y_i - n\overline{ty}}{\sum t_i - n\overline{t}^2}=0{,}01885$ (das war in der ursprünglichen Formel das b, die Berechnung siehe Tabelle unten)

Nun berechnen wir uns $\ln c = y_i - at = -35{,}86171162$ (das war in der ursprünglichen Formel das a, Berechnung siehe Tabelle unten)

Also eingesetzt in unsere angenommene Trendfunktion haben wir schließlich:
$p(x_i)=\ln\left(g(t)\right)=-35{,}86171162+0{,}01855\cdot t$
Wir wollen jedoch nur $g(t)$ wissen und müssen den Logarithmus weg bekommen
$g(t)=e^{-35{,}86171162+0{,}01855\cdot t}$

Nun nur noch ganz einfach den Wert 2000 einsetzen und ausrechnen:

$$g(2000) = e^{-35,86171162+0,01855\cdot 2000} = 6,39027$$

Hier in der Tabelle die restlichen Werte:

x=t	1950,0000	1960,0000	1970,0000	1980,0000	1990,0000
g(t)	2,5000	3,0000	3,6000	4,4000	5,3000
y=ln(g(t))	0,9163	1,0986	1,2809	1,4816	1,6677
Zwischenwerte:					
t_i^2	3802500,0000	3841600,0000	3880900,0000	3920400,0000	3960100,0000
t_i*y_i	1786,7669	2153,2801	2523,4397	2933,5770	3318,7366

n	5
a (wie formel oben)	0,01885824
ln(c) (wie formel oben)	-35,86171162
g(2000)	6,390272897

Zwischenwerte:					
sum(t_i^2)	19405500,0000	**(t quer) $\bar{t}$**	1970,0000	**$n\bar{t}^2$**	19404500,0000
sum(t_i*y_i)	12715,8003	**$\bar{y}$**	1,2890	**$n\bar{t}\bar{y}$**	12696,9420