Datenanalyse + Wissenserwerb

Übungen Korrelation / Regression

Aufgabe 2

Gegeben sind die folgenden vier Punkte:

```
> restart:
> Px:=array(1..4,[0,2,3,5]);
```

$$Px := \begin{bmatrix} 0 & 2 & 3 & 5 \end{bmatrix} \quad (1.1)$$

```
> Py:=array(1..4,[8,3,1,-2]);
```

$$Py := \begin{bmatrix} 8 & 3 & 1 & -2 \end{bmatrix} \quad (1.2)$$

a.) Bestimme die Mittelwerte xMittel, yMittel, die Varianzen σ_x^2, σ_y^2, und die Kovarianz c_{xy}

$$\sigma_x^2 \quad (1.3)$$
$$\sigma_y^2$$
$$c_{xy}$$

```
> mittelwert :=proc(p::array)
  local s,i, size;
  size:=linalg[vectdim](p);
  s:=0;
  for i to size do
    s:=s+p[i];
  end do;
  return s/size;
  end:
> varianz :=proc(p::array)
  local mw, size;
  mw:=mittelwert(p);
  size:=linalg[vectdim](p);
  return (1/(size-1))*sum( (p[k]-mw)^2 ,k=1..size );
  end:
> covarianz :=proc(px::array, py::array)
  local mwx,mwy, size;
  size:=linalg[vectdim](px);
  mwx:=mittelwert(px);
  mwy:=mittelwert(py);
  return (1/(size-1))*sum( (px[k]-mwx)*(py[k]-mwy) ,k=1..size )
  ;
  end:
```

```
> mittelwertX:=mittelwert(Px);
```

$$mittelwertX := \frac{5}{2} \quad (1.4)$$

```
> mittelwertY:=mittelwert(Py);
```

$$mittelwertY := \frac{5}{2} \quad (1.5)$$

```
> sigma2[x]:=varianz(Px);
```

$$\sigma 2_x := \frac{13}{3} \quad (1.6)$$

```
> sigma2[y]:=varianz(Py);
```

$$\sigma 2_y := \frac{53}{3} \quad (1.7)$$

```
> c[xy]:=covarianz(Px,Py);
```

$$c_{xy} := -\frac{26}{3} \quad (1.8)$$

b.) Die Gleichung der Ausgleichsgeraden y(x) = ax + b

```
> steigung :=proc(px::array,py::array)
  local size, mwx, mwy;
  mwx:=mittelwert(px);
  mwy:=mittelwert(py);
  size:=linalg[vectdim](px);
  return ( sum(px[k]*py[k] , k=1..size) - size * mwx * mwy ) /
  (sum(px[k]^2, k=1..size) - size * mwx^2 );
  end:
```

```
> a:=steigung(Px,Py);
```

$$a := -2 \quad (1.9)$$

```
> b:=mittelwert(Py)-a*mittelwert(Px);
```

$$b := \frac{15}{2} \quad (1.10)$$

```
> y:=a*x+b;
```

$$y := -2\,x + \frac{15}{2} \quad (1.11)$$

c.) Der Korrelationskoeffizient r_{xy}

$$r_{xy} \quad (1.12)$$

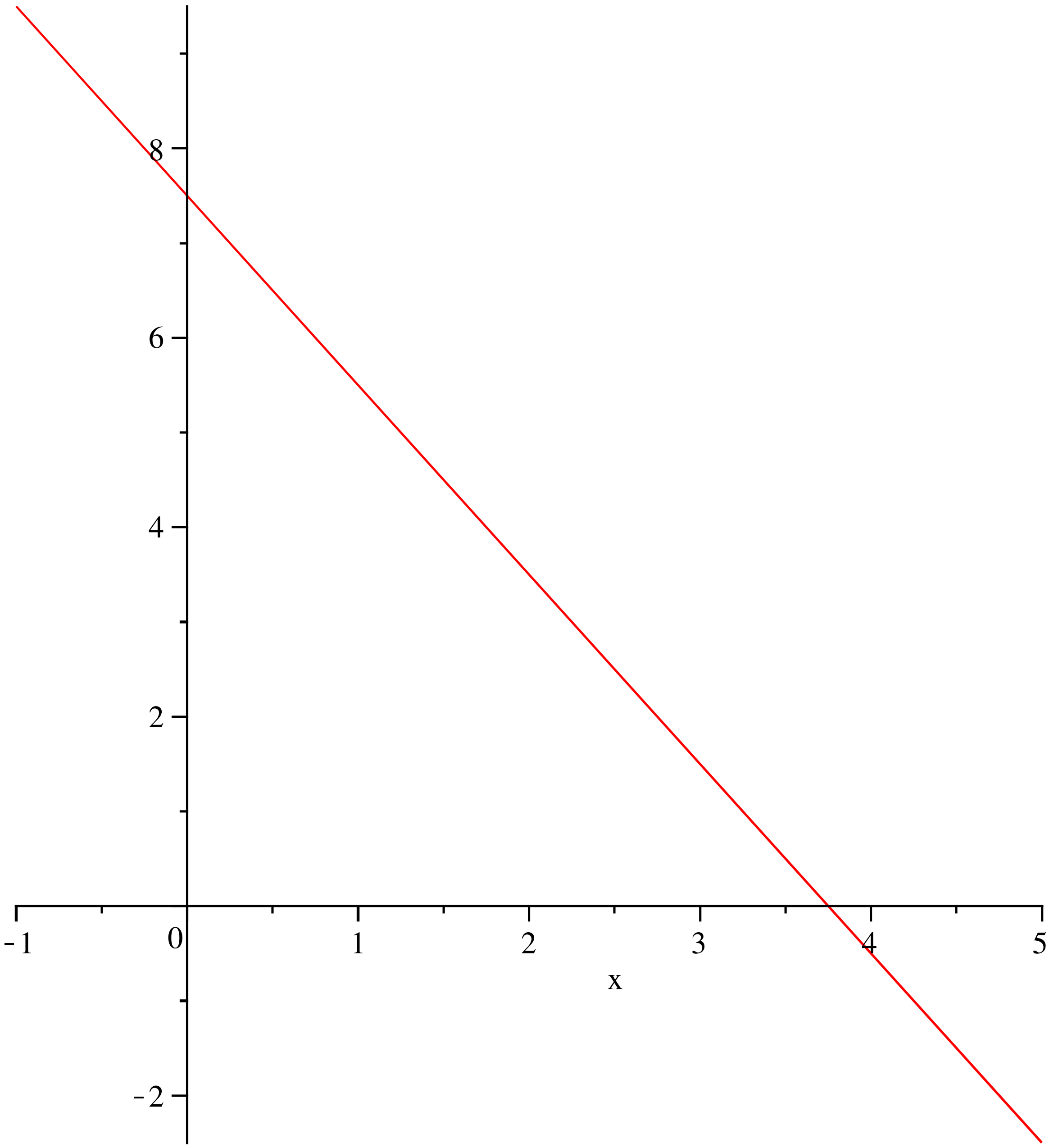
```
> r[xy]:=evalf(covarianz(Px,Py)/(sqrt(varianz(Px)*varianz(Py)))
  );
```

$$r_{xy} := -0.9905211133 \quad (1.13)$$

Demzufolge eine sehr gute Korrelation

d.) Skizziere die Punkte, sowie die Ausgleichsgerade

```
> plot(y,x=-1..5);
```



e.) Ist die Annahme eines linearen Zusammenhangs zwischen den x- und y-Werten sinnvoll?

Ist sinnvoll, da die Korrelation sehr gut ist.

Alternative zu c.)

Wir prüfen noch zur Sicherheit mit Maple, ob es dieselbe Ausgleichsgerade ermittelt, wie unsere berechnete. Die leastsquare Methode von Maple akzeptiert keine Arrays, also definieren wir die Punkte erneut.

```
> restart:
> with(stats);
```

$$[anova, describe, fit, importdata, random, statevalf, statplots, transform] \quad \textbf{(1.14)}$$

```
> Px:=[0,2,3,5];
```

$$Px := [0, 2, 3, 5] \quad \textbf{(1.15)}$$

```
> Py:=[8,3,1,-2];
```

$$Py := [8, 3, 1, -2] \tag{1.16}$$

```
> fit[leastsquare[[x,y], y=a*x+b, {a,b}]]([Px,Py]);
```

$$y = -2\,x + \frac{15}{2} \tag{1.17}$$

Aufgabe 3

Entwickle (ganz analog, wie beim Vorgehen zu der Geraden) eine Methode, um die Parabel y(x) = $a\,x^2 + b\,x + c$ im Sinne der kleinsten Quadrate optimal an die Punke (x_i, y_i) i = 1,...,n zu approximieren.

```
> restart:
> with(stats):
> XValues:=[0,2,3,5];
```

$$XValues := [0, 2, 3, 5] \tag{2.1}$$

```
> YValues:=[8,3,1,-2];
```

$$YValues := [8, 3, 1, -2] \tag{2.2}$$

```
> Parabel:= fit[leastsquare[[x,y], y=a*x^2+b*x+c, {a,b,c}]](
  [XValues, YValues]);
```

$$Parabel := y = \frac{1}{6}\,x^2 - \frac{17}{6}\,x + 8 \tag{2.3}$$

Aufgabe 4

Für eine Kondensatorenentladung gilt das Spannungs-Gesetz:

$U_t = U_0\,\mathrm{e}^{-\frac{1}{R\,C}\,t}$

a.) Bestimme $\tau = R\,C$ und U_0 aufgrund der gegebenen Messreihe

$$\tau = R\,C \qquad U_0 \tag{3.1}$$

```
> restart:
> t[s]:=array(1..5,[1,4,7,10,15]);
```

$$t_s := \begin{bmatrix} 1 & 4 & 7 & 10 & 15 \end{bmatrix} \tag{3.2}$$

```
> U[t]:=array(1..5,[80.2,45.5,24.5,13.9,4.7]);
```

$$U_t := \begin{bmatrix} 80.2 & 45.5 & 24.5 & 13.9 & 4.7 \end{bmatrix} \tag{3.3}$$

Die Gleichung U(t) muss nun durch etwas Substitution auf eine normale Gerandengleichung y= ax + b gebracht werden. Dies erreichen wir, wenn dir die Gleichung U(t) mit dem Logarithmus naturalis multiplizieren.

> $\ln(U_t) = \ln(U_0) - \frac{1}{R\,C}\,t$ dies entspricht $y = b + a\,x$

$$\ln(U_t) = \ln(U_0) - \frac{t}{R\,C} \tag{3.4}$$

$$y = b + a\,x$$

```
> x:=t[s];
```

$$x := t_s \quad (3.5)$$

```
> y:=array(1..5,[ln(80.2),ln(45.5),ln(24.5),ln(13.9),ln(4.7)]);
```

$$y := \left[\begin{array}{ccccc} 4.384523515 & 3.817712326 & 3.198673118 & 2.631888840 & 1.547562509 \end{array}\right] \quad (3.6)$$

> $S(a, b) = \sum_{i=1}^{n} \left(a\,x_i + b - y_i\right)^2$ muss minimal sein.

Dieses Minimum erhält man indem die partiellen Ableitungen nach a resp. b beide Null gesetzt werden. $\frac{\partial}{\partial a} S(a, b) = 0$, $\frac{\partial}{\partial b} S(a, b) = 0$

$$S(a, b) = n\,b^2 + \sum_{i=1}^{n} \left(a^2 t_{s_i}^2 + 2\,a\,t_{s_i} b - 2\,a\,t_{s_i} y_i - 2\,b\,y_i + y_i^2\right) \quad (3.7)$$

$$\frac{\partial}{\partial a} S(a, b) = 0$$

$$\frac{\partial}{\partial b} S(a, b) = 0$$

```
> S:=sum((a*x[k]+b-y[k])^2,k=1..5);
```

$$S := (a + b - 4.384523515)^2 + (4\,a + b - 3.817712326)^2 + (7\,a + b - 3.198673118)^2 + (10\,a + b - 2.631888840)^2 + (15\,a + b - 1.547562509)^2 \quad (3.8)$$

```
> Sa:=diff(S,a);
```

$$Sa := 782\,a + 74\,b - 183.1568214 \quad (3.9)$$

```
> Sb:=diff(S,b);
```

$$Sb := 74\,a + 10\,b - 31.16072062 \quad (3.10)$$

```
> solve({Sa=0,Sb=0},[a,b]);
```

$$[[a = -0.2023571296, b = 4.613514821]] \quad (3.11)$$

```
> a:=-.2023571296;
```

$$a := -0.2023571296 \quad (3.12)$$

```
> b:=4.613514821;
```

$$b := 4.613514821 \quad (3.13)$$

Ja, aber das ist nur die halbe Wahrheit. Wir müssen nun die Substitution wieder rückgängig machen, damit wir auf die ursprüngliche Gleichung U(t) rechnen.
a und b müssen also zurücktransformiert werden um U_0 und τ herauszufinden.

$$U_0 \quad (3.14)$$

$$\tau$$

```
> U[0]:=exp(b);
```

$$U_0 := 100.8379549 \quad (3.15)$$

```
> tau:=-1/a;
```

$$\tau := 4.941758178 \quad (3.16)$$