x/delay	good	medium	strong	
\wo alcohol	120	60	20	200
\w alcohol	60	100	40	200
	180	160	60	400

Es wurden gleich viele Leute für „mit Alkohol" und „ohne Alkohol" erfasst
↳ Vergleichbarkeit ist damit gegeben.

Die Durchschnittsreaktionszeit ist: bei „ohne Alkohol" bei „gut"
bei „mit Alkohol" bei „mittel"
↳ Alkohol scheint einen Einfluss zu haben

Conditional relative frequencies

x/delay	good	medium	strong	
\wo alcohol	$\frac{2}{3}$	$\frac{3}{8}$	$\frac{1}{3}$	$\frac{33}{24}$
\w alcohol	$\frac{1}{3}$	$\frac{5}{8}$	$\frac{2}{3}$	$\frac{39}{24}$
	1	1	1	$\frac{72}{24} = 3$

Die Reaktionszeit ist mit Alkohol sehr viel höher.
Die Spalte „mittel" fungiert nahezu als Spiegelachse

χ^2 - Test H_0: Es gibt keinen Zusammenhang zwischen Reaktionszeit und Alkoholeinfluss
H_1: Es gibt einen Zusammenhang zwischen Reaktionszeit und Alkoholeinfluss

absolute erwartete Häufigkeiten

x/delay	good	medium	strong	
\wo alcohol	90	80	30	200
\w alcohol	90	80	30	200
	180	160	60	400

Fehler:

x/delay	good	medium	strong	
\wo alcohol	10	5	$\frac{10}{3}$	$\frac{55}{3}$
\w alcohol	10	5	$\frac{10}{3}$	$\frac{55}{3}$
	20	10	$\frac{20}{3}$	$\frac{110}{3} \approx 36{,}\overline{6} = \chi_p^2$

Freiheitsgrad: 2
Kritischer Wert $(\alpha = 0{,}005) = 10.597 = c$ $\chi_p^2 > c$ ⇒ H_0 verwerfen ⇒ H_1 annehmen
⇒ Es gibt einen Zusammenhang zw. Reaktionszeit & Alkoholeinfluss

2) Fahrenheit: Interval

Soziöökonomischer Status: Nominal

Kelvin: Ratio

cm: Ratio

Schulnoten: Ordinal

Beschreibung: Feature

Datum: Interval

Ratio:
Dioptrien
Geschwindigkeit

Interval:
Raumnummern
pH-Level

3) d_1-Distance

$$d = \sum_{i=1}^{n} |p_i - q_i|$$

$$= |-1{,}66 - 0{,}29| + |0{,}3 - 0{,}89| + |-0{,}08 - 0{,}82| + |0{,}10 - 0{,}97| + |-1{,}17 - 0{,}53| +$$

$$|-0{,}05 - 0{,}83| + |0{,}84 - 1{,}06| + |-0{,}66 - 0{,}67| + |0{,}42 - 0{,}86| + |-0{,}99 - 0{,}51|$$

$$= \underline{1{,}95} + 0{,}59 + 0{,}9 + 0{,}87 + 1{,}7 +$$

$$0{,}88 + 0{,}22 + 1{,}33 + 0{,}44 + 1{,}5$$

$$= 10{,}38$$

$$d_\infty = \underline{\max_i |d_i - q_i|}$$

$$= 1{,}95$$

$$\text{mean}(A) = \frac{-1{,}66 + 0{,}3 + (-0{,}08) + 0{,}10 + (-1{,}17) + (-0{,}05) + 0{,}84 + (-0{,}66) + 0{,}42 + (-0{,}99)}{10}$$

$$= \frac{2{,}95}{10}$$

$$= -0{,}295$$

$$\text{mean}(B) = \frac{0,29+0,89+0,82+0,97+0,53+0,83+1,06+0,67+0,86+0,51}{10}$$

$$= \frac{7,43}{10}$$

$$= 0,743$$

Rechenwege wurden weggelassen, da sich die Formeln nur wiederholen. Die Zwischenergebnisse wurden nicht mit allen Nachkommastellen notiert. Zur Rechnung im Taschenrechner wurden die exakt(eren) Werte genutzt.

Series	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
A'	-1,365	0,595	0,215	0,395	-0,875	0,245	1,135	-0,365	0,715	-0,695
B'	-0,453	0,147	0,077	0,227	-0,213	0,087	0,317	-0,073	0,117	-0,233

$d_1(A'B') \approx 4,656$

$\sigma(A') = 0,7930847$

$\sigma(B') = 0,5699427$

$d_\infty(A'B') \approx 0,912$

Series	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
A''	-1,72	0,75	0,27	0,498	-1,103	0,3089	1,431	-0,46	0,901	-0,876
B''	-0,19	0,257	0,135	0,398	-0,373	0,152	0,556	-0,12	0,205	-0,408

$d_1(A''B'') \approx 4,911067$

$d_\infty(A''B'') \approx 0,926311$

ex4

December 8, 2023

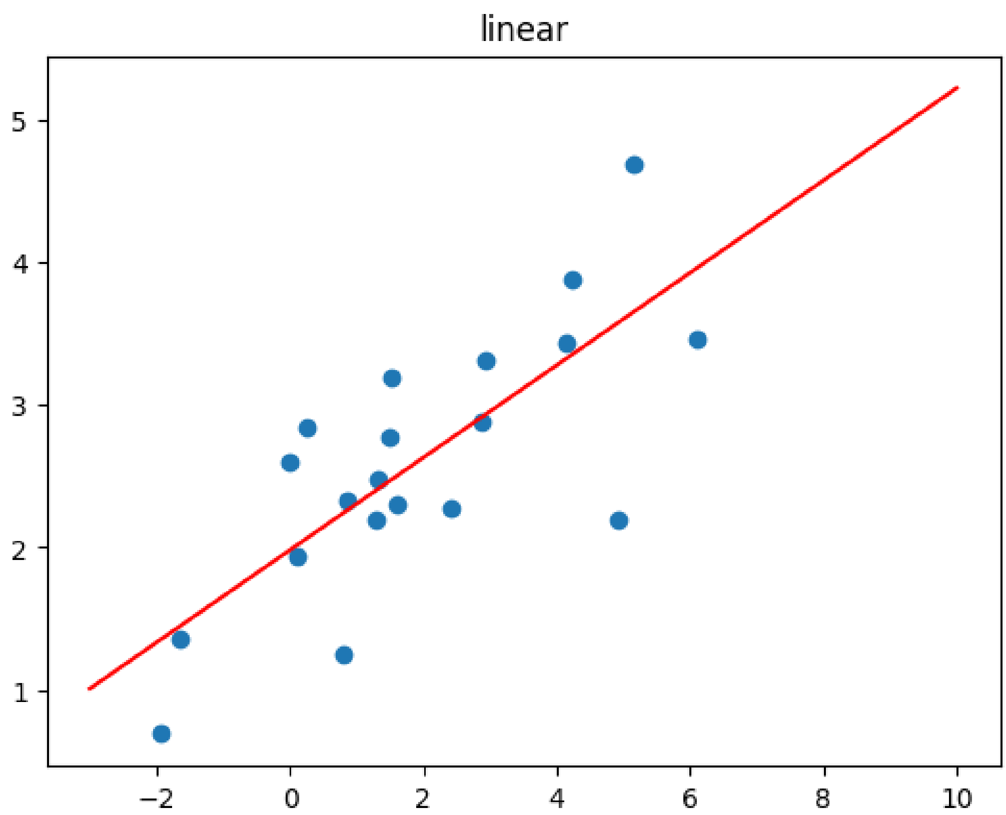
[ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

[ ]:
```
data = pd.read_csv("linreg_data.csv")[["X", "Y"]].values
x = data[:, 0]
X = np.vstack([np.ones_like(x), x]).T
y = data[:, 1]

Bs = np.linalg.inv(X.T@X) @ (X.T @ y)
lin = np.linspace(-3, 10, 100)
preds = np.vstack([np.ones_like(lin), lin]).T @ Bs

plt.scatter(x, y)
plt.plot(lin, preds, color="red")
plt.title("linear")
Bs
```

[ ]:
```
array([1.98091071, 0.32441823])
```

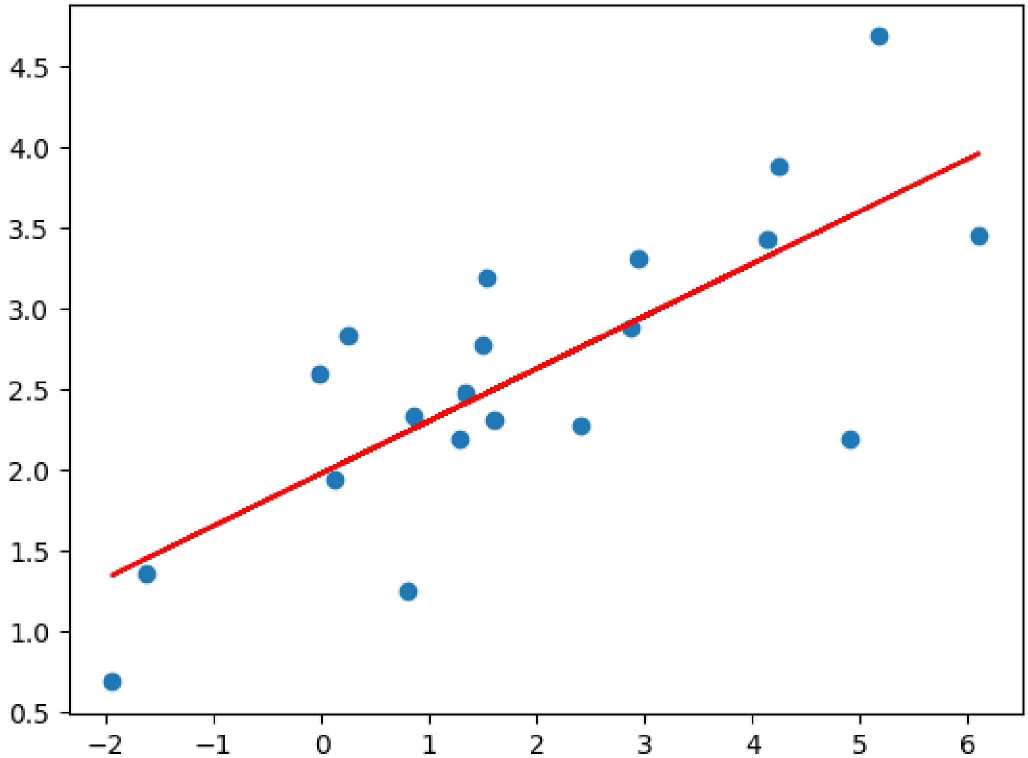


[ ]:
```python
from scipy.stats import linregress

slope, intercept, r_value, p_value, std_err = linregress(x, y)

plt.scatter(x, y)
plt.plot(x, intercept + x * slope, color="red")
```

[ ]:
```
[<matplotlib.lines.Line2D at 0x17c32f5b0>]
```



[ ]:
```
Yx2 = intercept + 2 * slope
Yx2

#4. Replace observed values with regression function values
```

[ ]: 2.629747164010065