DATA SCIENCE

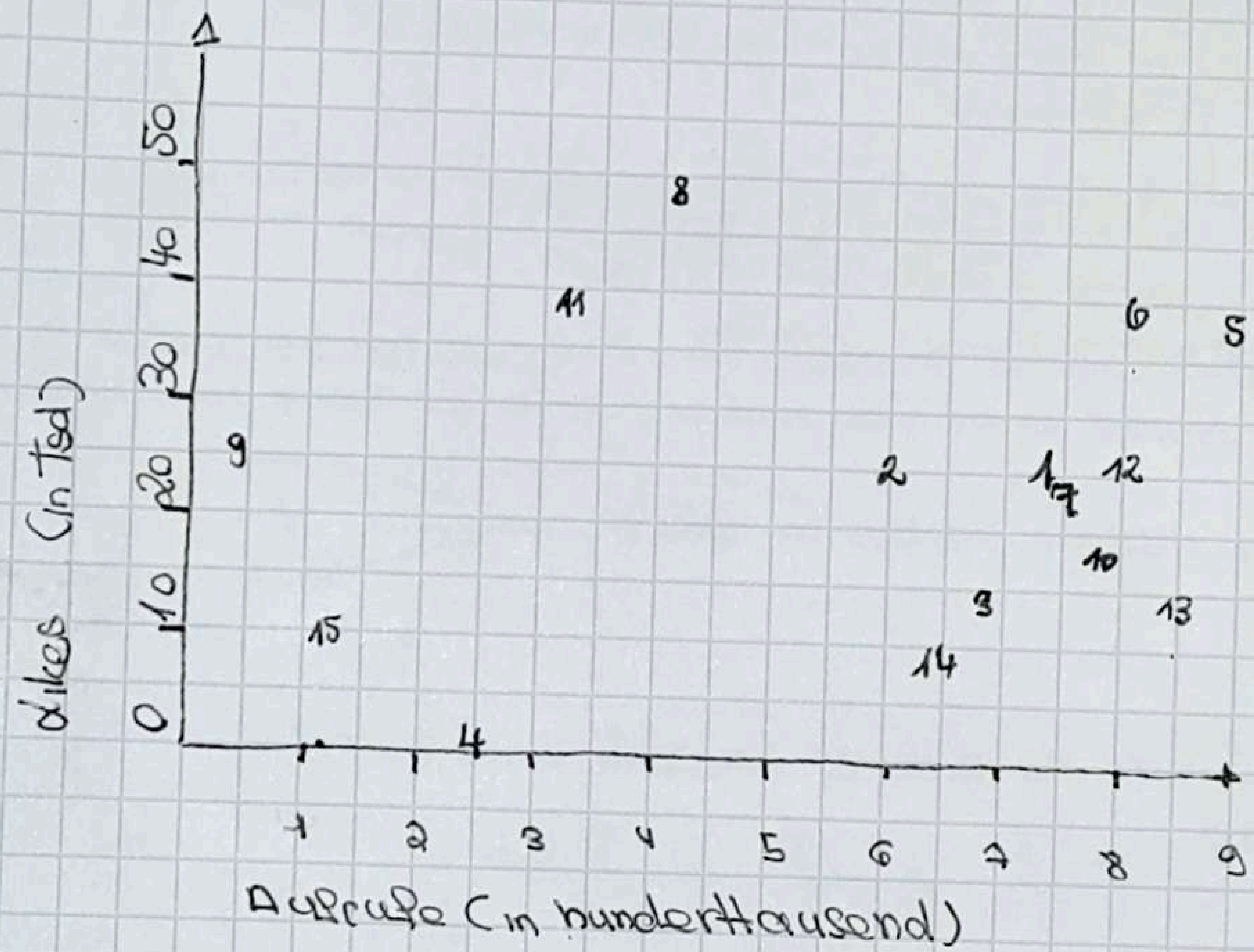
Praktikum 2

Fabian Gottong & Larissa Puchlau

A1

a) statistische Einheiten: "Videos"
Merkmale: "Likes", "Aufrufe"
Skalenniveau: kardinal (Verhältnisskala)

b)



"Likes" und "views" korrelieren.

c) $r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$

Ø Aufrufe: $580,9\bar{3} = \sim 581$
Ø Likes: $23,\bar{6} = \sim 24$

$r = \frac{8797}{13100,48}$

$r = 0,67$

*	NR $*^2$	NR Likes	Aufrufe
153	23409	25 - 24 = 1	734 - 581 = 153
28	784	25 - 24 = 1	609 - 581 = 28
-1078	1.162.084	13 - 24 = -11	679 - 581 = 98
7458	55.621.764	2 - 24 = -22	242 - 581 = -339
4560	20.793.600	39 - 24 = 15	885 - 581 = 304
3712	13.778.944	40 - 24 = 16	813 - 581 = 232
0	0	24 - 24 = 0	757 - 581 = 176
-3956	15.649.936	47 - 24 = 23	409 - 581 = -172
-522	272484	25 - 24 = 1	59 - 581 = -522
-1152	1327104	18 - 24 = -6	773 - 581 = 192
-3556	12645136	38 - 24 = 14	327 - 581 = -254
446	198916	26 - 24 = 2	804 - 581 = 223
-2730	7452900	14 - 24 = -10	854 - 581 = 273
-1020	1040400	9 - 24 = -15	649 - 581 = 68
6454	41654116	10 - 24 = -14	120 - 581 = -461
	171622577		

d) Nicht zwingend, eine nicht-niedrig/überdurchschnittliche Like-Anzahl ist aber wahrscheinlich.

A2

a) $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} = 1 - \frac{6 \sum (R_{x_i} - R_{y_i})^2}{n(n^2-1)}$

$\Rightarrow \rho = 1 - \frac{6 \cdot 8}{9(9^2-1)} = 1 - \frac{48}{720} = 0{,}9\overline{3}$

NR		(x^2)	Test	
1	1	0	4	0
1	2	1	1	1
1	2	1	1	1
2	2	0	0	1
2	2	0	0	1
2	2	0	0	1
3	3	0	0	1
3	1	2	4	1
3	5	4	16	9
		8	26	16

b) Die Zuordnung der nicht-numerischen Werte ist rein subjektiv. Eine andere Bewertung führt zu anderen Ergebnissen

c) Es muss eine ordinale Sortierung geben bzw es müssen objektive Standarts (zB bpm) für die Sortierung gefunden werden

2.3 WAHRSCHEINLICHKEIT

Daten: 1000 zufällige Nutzer, die angerufen werden sollen
x Mitarbeiter die
⌀ 10 Anrufe/ Stunde schaffen

Wahrscheinlichkeitsmodelle
Die Mathematische Darstellung zufälliger Ereignisse.
Quantifizieren die Wahrscheinlichkeit mit d. Ereignisse eintreten

Verteilungen
Diskret • zählbare & abzählbare Ereignisse
stetig • kontinuierlicher Bereich v. Ereignissen → oft glatte Kurve im Modell

→ Hier vermutlich stetig

Binomialverteilung → Anzahl v. Erfolgen in Binären Ereignissen (unabhängig & identisch)
n = Anzahl Versuche
p = Wahrscheinlichkeit jedes Versuchs

Hypergeometrische Verteilung → Anzahl v. Erfolgen in Stichprobe (ohne Zurücklegen)
N = Gesamtzahl d. Elemente
K = Elemente in relevanter Eigenschaft
n = Größe d. Stichprobe

Poissonverteilung → misst Anzahl v. Ereignissen in Intervall (bei seltenen Ereignissen
λ = durchschnittliche Rate v. Ereignissen in Intervall

→ Hier vermutlich Poisson
Wir haben Intervalle, durchschnittliche Raten & Ereignisse

2.3. a)
Wir werden die Poissonverteilung nutzen
$$P(X) = \frac{e^{-\lambda}\lambda^k}{k!}$$
da wir eine zählbare, endliche Menge von Ereignissen mit Intervallen und durchschnittlichen Ereignisraten haben

Parameter:
k (tatsächliche Anzahl v. Ereignissen im Intervall) = X = ?
λ (erwarteter Durchschnitt v. Ereignissen im Intervall) = 10
e (eulersche Zahl) = ~ 2,71828
k (das Produkt aller positiven ganzen Zahlen 1-k) = ?

2.3 b)

$$P(X=8) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-10}\,10^8}{8!} = \sim 0.112599$$

mit 11,26 % Wahrscheinlichkeit mit der ein Mitarbeiter 8 Leute anruft

2.3 c)

hier wird die kummulierte Wahrscheinlichkeit von 0 - 8 anrufen benutzt:

$$\sum_{i=0}^{8} \frac{e^{-\lambda}\lambda^i}{i!} = P(X \leq 8)$$

$$\Rightarrow \frac{e^{-10}\,10^0}{0!} + \frac{e^{-10}\,10^1}{1!} + \frac{e^{-10}\,10^2}{2!} + \frac{e^{-10}\,10^3}{3!} + \frac{e^{-10}\,10^4}{4!} + \frac{e^{-10}\,10^5}{5!} + \frac{e^{-10}\,10^6}{6!} + \frac{e^{-10}\,10^7}{7!} + \frac{e^{-10}\,10^8}{8!}$$

$\Rightarrow$ mit 33,28 % Wahrscheinlichkeit ruft d. Mitarbeiter maximal 8 Personen an

2.3 d)

Wir kummulieren die Wahrscheinlichkeit von $X < 8$

$$\rightarrow \sum_{i=0}^{7} \frac{e^{-10}\,10^i}{i!} \Rightarrow 22.02\,\%$$

Das ziehen wir von 100 % ab

mit 77,78 % Wahrscheinlichkeit ruft d. Mitarbeiter mindestens 8 Personen an