Der Informatikkurs hat in einem Projekt die Häufigkeit des Auftretens von Primzahlen in Centurien (Blöcken von 100 Zahlen) untersucht. In den beiden ersten Centurien (Zahlen 1-200) finden sich sehr viele (nämlich 25 und 21) Primzahlen, in den ersten zwanzig Centurien insgesamt 303 Primzahlen.

Ausschnitt aus der Ergebnisliste: (vgl. http://lernportal.ziemke-koeln.de/informatik/gk11/java/kryptologie/primfinder2.htm)

Suche im Zahlenbereich von 1 bis 2000 Suche starten

Primzahlen-Suche:

#1 | Anzahl: 25 | Primzahlen: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97,
#2 | Anzahl: 21 | Primzahlen: 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 173, 179, 181, 191, 193, 197, 199,
#3 | Anzahl: 16 | Primzahlen: 211, 223, 227, 229, 233, 239, 241, 251, 257, 263, 269, 271, 277, 281, 283, 293,
#4 | Anzahl: 16 | Primzahlen: 307, 311, 313, 317, 331, 337, 347, 349, 353, 359, 367, 373, 379, 383, 389, 397,
#5 | Anzahl: 17 | Primzahlen: 401, 409, 419, 421, 431, 433, 439, 443, 449, 457, 461, 463, 467, 479, 487, 491, 499,

Die einzelnen Ergebnisse:

Centurium #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
abs. Häuf.	25	21	16	16	17	14	16	14	15	14
Centurium #	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
abs. Häuf.	16	12	15	11	17	12	15	12	12	13

a) (1) Bestimmen Sie die durchschnittliche und die mittlere Anzahl Primzahlen in den ersten zwanzig Centurien. Benennen Sie ggf. auch Ausreißer.
(2) Berechnen Sie auch die Standardabweichung s als Maß der Streuung um den Mittelwert.
(3) (Zusatz) Stellen Sie zum Mittelwert und zur Varianz s^2 jeweils Berechnungsterme auf.
Hinweis: Statt vieler Summanden nur die ersten und letzten drei notieren (z. B. 1+2+4+…+12+14+15).

b) Beantworten Sie begründend und kurz z. B. unter Einbeziehung des Boxplots:
(1) Was ist die maximale Anzahl an Primzahlen, die in 75% der Centurien auftritt?
(2) Welches ist die größte minimale Anzahl Primzahlen in 50% der Centurien?
(3) (Zusatz) Welches ist die kleinste maximale Anzahl Primzahlen in 60% der Centurien?

c) Für eine Prognose soll die lineare Regression eingesetzt werden.
(1) Durch welchen Punkt M wird die Gerade verlaufen? Bestimmen Sie die Koordinaten.
(2) Führen Sie die Regression durch. Geben Sie den Term rg(x) der Regressionsgeraden an (Koeffizienten nur mit jeweils zwei Dezimalen!).
(3) Beschreiben Sie kurz die zur Regression verwendete Idee der kleinsten Quadrate.
Fertigen Sie dazu für den Datenpunkt P4 (4/16) und die Gerade eine realistische Skizze an, nachdem Sie Punktwolke und Regressionsgerade am GTR angezeigt haben.
(4) Bestimmen oder berechnen Sie die Kantenlänge dy des Quadrates auf zwei Dezimale genau.

d) (1) Bestimmen Sie, wie viele Primzahlen vermutlich im 30. Centurium zu erwarten sind.
Protokollieren Sie Ihren Lösungsweg.
(2) Bestimmen und begründen Sie, wie gut diese lineare Regression die vorhandenen Daten repräsentiert.
(3) Bewerten Sie Ihre Regression kritisch bezüglich der Aussagegüte für große Primzahlen.

e) Tatsächlich sind im 30. Centurium genau elf Primzahlen. Weitere Anzahlen:

Centurium #	**25**	**30**	**50**	**100**	**150**	**200**	**250**	**300**
abs. Häuf.	10	11	15	9	8	12	11	7

Wegen der Häufung bei kleinen Primzahlen wird vorgeschlagen, für weitere Prognosen die Anzahlen der ersten beiden Centurien zu ignorieren und obige weitere Anzahlen mit in die lineare Regression einzubeziehen.
(1) Notieren Sie den Term rg(x) und berechnen Sie den Gütegrad dieser neuen Regression.
Hinweis: Speichern Sie den Regressionsterm als Y2 für den folgenden Vergleich.
(2) Vergleichen Sie die Korrelation mit der zuvor ermittelten.
Bewerten Sie das vorgeschlagene Vorgehen unter Einbeziehung Ihrer Berechnungen.
(3) Erkunden Sie, ob andere Regressionsmodelle hier sinnvoll eingesetzt werden können.
(4) Formulieren und begründen Sie eine Aussage über die Primzahl-Häufigkeit für große Primzahlen in späteren Centurien. Hinweis: Ein Primzahl-Finder für die Probe ist hier:
http://lernportal.ziemke-koeln.de/informatik/gk11/java/kryptologie/primfinder2.htm