

PENGUNAAN *STEM AND LEAF* DAN *BOXPLOT* UNTUK ANALISIS DATA

Moh Yamin Darsyah

Program Studi Statistika Universitas Muhammadiyah Semarang
yamindarsyah@gmail.com

ABSTRAK

Banyak sekali data dalam kehidupan yang dapat dijelaskan dengan statistika. Informasi yang dihasilkan oleh suatu data dapat dijelaskan dengan beberapa metode statistik, salah satunya metode grafik. Beberapa jenis metode grafik yang dapat digunakan sesuai dengan kebutuhan data yang akan ditampilkan, diantaranya *stem and leaf* dan *boxplot*. Aplikasi *stem and leaf* dan *boxplot* digunakan untuk membandingkan nilai mahasiswa pada empat fakultas pada salah satu PTN. Data yang digunakan merupakan nilai 7 mata pelajaran ujian nasional SMA, sampel data diambil dari 100 mahasiswa per fakultas. Untuk mengetahui penyebaran data di masing-masing fakultas digunakan *stem and leaf* dan *boxplot*. Sedangkan untuk membandingkan nilai antar fakultas digunakan *boxplot*. Data nilai mahasiswa pada 7 mata pelajaran SMA di fakultas P tersebar secara merata, cenderung simetris, namun ada 2 data yang *outlier*, fakultas Q tersebar secara merata, cenderung simetris dan tidak ada data yang *outlier*, di fakultas R nilai kurang tersebar secara merata, cenderung condong dan tidak ada data yang *outlier*. Sedangkan nilai di fakultas S tersebar secara merata, cenderung simetris, namun ada 1 data yang *outlier*. Berdasarkan analisis yang dilakukan, dihasilkan bahwa berdasarkan nilai mean dan median, penyebaran, kesimetrisan data serta ada tidaknya *outlier* bisa dikatakan bahwa dari nilai 100 sampel mahasiswa, fakultas Q merupakan fakultas terbaik berdasarkan nilai 7 mata pelajaran SMA.

Kata kunci : *stem and leaf*, *boxplot* dan *outlier*

PENDAHULUAN

Fenomena dalam kehidupan sehari-hari merupakan sumber data. Sebagai statistisi, tentu saja harus pandai-pandai untuk bisa mengolah dan memanfaatkan data-data tersebut sehingga dapat menghasilkan suatu informasi yang akurat. Dalam menganalisis data harus tahu metode yang tepat untuk bisa menggambarkan dan mengolah data sesuai dengan kebutuhan.

Dalam ilmu Statistika, data seringkali disajikan secara visual. Penyajian secara visual ini dilakukan untuk lebih mempermudah interpretasi hanya dengan melihat pola dari grafik saja. Contoh penyajian data secara visual antara lain histogram, diagram lingkaran, *stem and leaf* dan *boxplot*.

Penggunaan diagram-diagram tersebut tergantung pada jenis data yang digunakan dan tujuan dari analisis data

tersebut. Aplikasinya pada studi kasus melihat perbandingan jumlah siswa per fakultas bisa menggunakan diagram lingkaran. Untuk melihat perkembangan IP mahasiswa tiap semester bisa digunakan diagram batang. Sedangkan *stem and leaf* dan *boxplot* dapat digunakan untuk melihat penyebaran data. Berdasarkan data yang akan dianalisis, tujuan dari penelitian ini bagaimana membandingkan penyebaran nilai pada mata pelajaran ujian nasional SMA pada empat fakultas dengan menggunakan *stem and leaf* dan *boxplot*.

Boxplot adalah salah satu cara dalam statistik deskriptif untuk menggambarkan secara grafik dari data numeris. *Boxplot* juga dapat menyampaikan informasi variasi dan penempatan atau lokasi pada data yang telah ditetapkan, terutama untuk mendeteksi dan menggambarkan perubahan variasi dan lokasi antar kelompok data yang berbeda

Boxplot mempunyai komponen-komponen diantaranya adalah:

- 1) nilai observasi terkecil,
- 2) kuartil terendah atau kuartil pertama (Q1), yang memotong 25 % dari data terendah
- 3) median (Q2) atau nilai pertengahan

4) kuartil tertinggi atau kuartil ketiga (Q3), yang memotong 25 % dari data tertinggi

5) nilai observasi terbesar.

Median adalah suatu nilai yang membagi kumpulan data menjadi dua bagian yang kurang lebih sama banyaknya, sehingga setengahnya atau 50% dari data tersebut akan memiliki nilai yang lebih kecil daripada median (Aunuddin, 1989).

Stem And Leaf adalah diagram dahan-daun disusun baris per baris secara vertikal, dan cukup efektif untuk menggambarkan pola sebaran data yang berukuran kecil (Aunuddin, 1989). Untuk digit angka dan puluhan di pisahkan sesuai incrementnya menjadi dahan dan daun masing-masing. "Dahan" digunakan untuk mengelompokkan nilai dan setiap "daun" menunjukkan skor individu dalam setiap kelompok.

Stem And Leaf dan *Boxplot* bisa digunakan untuk menyajikan data kuantitatif dalam format grafis, mirip dengan histogram, untuk membantu dalam memvisualisasikan bentuk distribusi (Montgomery,2005).

METODOLOGI PENELITIAN

Sumber data yang dipergunakan di dalam penelitian ini adalah 4 data. Keempat data ini adalah data jumlah nilai 7 mata pelajaran di SMA yang diambil dengan sampling di 4 buah fakultas, yaitu Fakultas P, Fakultas Q, Fakultas R dan Fakultas S. Masing-masing fakultas diambil sampel sebanyak 100 mahasiswa. Pada keempat data tersebut akan dilakukan analisis ukuran pemusatan data dan ukuran penyebaran data untuk membandingkan nilai mahasiswa di keempat fakultas. Langkah-langkah analisis yang digunakan adalah :

1. Memasukkan data nilai mahasiswa 4 fakultas ke dalam Minitab.
2. Membuat *stem and leaf* dan *boxplot* dari data nilai mahasiswa masing-masing fakultas.
3. Menganalisis *stem and leaf* dan *boxplot* yang dihasilkan dari data

nilai mahasiswa masing-masing fakultas.

4. Membuat *boxplot* nilai mahasiswa per fakultas dalam satu gambar.
5. Membandingkan nilai mahasiswa per fakultas berdasarkan *boxplot* yang dihasilkan.

ANALISIS DAN PEMBAHASAN

Untuk bisa membandingkan nilai 7 mata pelajaran SMA untuk mahasiswa di 4 fakultas berdasarkan *stem and leaf* dan *boxplot*, dilakukan tahap-tahap analisis seperti yang telah dijelaskan sebelumnya.

a. Fakultas P

Berdasarkan data nilai 7 mata pelajaran SMA milik 100 sampel mahasiswa di fakultas P, didapatkan *stem and leaf* seperti di bawah ini :

Stem-and-Leaf Display: Fakultas P

Stem-and-leaf of Fakultas P N = 100

Leaf Unit = 1.0

```

2  4  44
5  4  677
10 4  88899
16 5  000111
25 5  222223333
44 5  44444444445555555555
(18) 5  6666666666667777777
38 5  888888999999999
24 6  000011111
15 6  223333333
7  6  44455
2  6  77

```

Gambar 1. *Stem and Leaf* Nilai Mahasiswa Fakultas P

Hasil *stem and leaf* pada Gambar 1 di atas menggunakan *leaf unit* 1.0 dan *increment* sebesar 2. Berdasarkan *stem and leaf* tersebut, bisa dilihat bahwa nilai mahasiswa di Fakultas P yang tertinggi adalah 67 dan terendah adalah 44. Angka 18 yang diberi tanda kurung menyatakan bahwa median dari data berada pada baris tersebut. Median nilai mahasiswa di fakultas P berada pada di anta data ke-50 dan ke-51, yaitu 56. *Stem and leaf*

tersebut memperlihatkan bahwa data terpusat di sekitar median. Jika *stem and leaf* ini diputar 180° berlawanan arah jarum jam, maka bentuk *stem and leaf* ini akan menyerupai kurva distribusi normal. Hal ini menunjukkan bahwa data nilai 7 mata pelajaran SMA mahasiswa di fakultas P mengikuti distribusi normal. Data nilai mahasiswa fakultas P juga bisa digambarkan dengan statistika deskriptif pada tabel 1.

Tabel 1. Tabel Statistika Deskriptif Nilai Mahasiswa Fakultas P

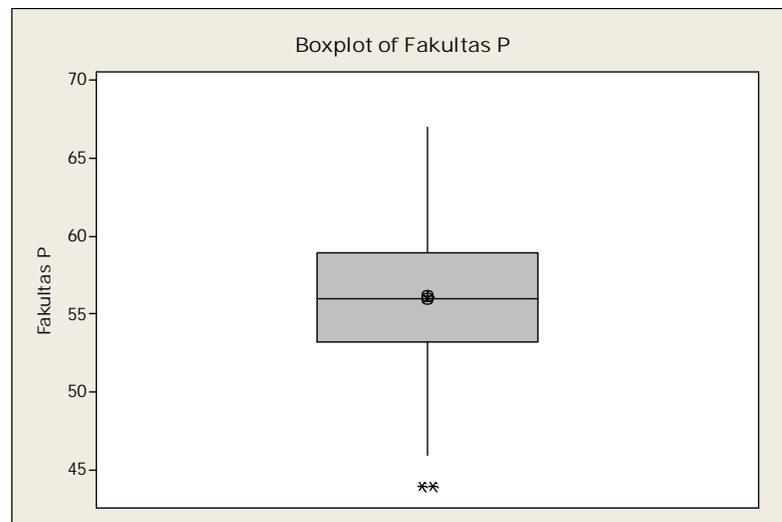
Variabel	Mean	Q1	Median	Q3	Minimum	Maximum	Range
Fakultas P	56.19	53.25	56	59	44	67	23

Tabel tersebut memperlihatkan bahwa nilai rata-rata nilai mahasiswa di fakultas P adalah sebesar 56.19. Kuartil bawah yang membawahi 25% data

berada pada nilai 53.25. Sedangkan untuk nilai median, sama seperti hasil uraian *stem and leaf* di atas, sebesar 56. Untuk kuartil atas yang membawahi

75% data, terletak pada nilai 69. Jika data ini digambarkan secara visual dalam

bentuk *boxplot* dihasilkan gambar seperti pada gambar 2.



Gambar 2. *Boxplot* Nilai Mahasiswa Fakultas P

Dapat dilihat bahwa *boxplot* tersebut menggambarkan bahwa data nilai mahasiswa di fakultas P cenderung simetris dan data berpusat di median. Hal ini dapat dilihat dari jarak median ke kuartil 1 dan jarak median ke kuartil 2 yang cenderung sama panjangnya. Nilai mean juga berada di sekitar nilai median data. Selain dapat dilihat pada *boxplot* yaitu adanya titik pada garis median, hal ini dapat dilihat pada tabel 1. Nilai mean sebesar 56,19 dan nilai median sebesar

56. Namun, dari *boxplot* tersebut memperlihatkan bahwa terdapat 2 nilai yang *outlier*. Seperti yang bisa dilihat di gambar 1 dan gambar 2, yang menjadi *outlier* adalah data yang paling kecil yaitu nilai 44.

b. Fakultas Q

Sedangkan untuk data nilai 7 mata pelajaran SMA milik 100 sampel mahasiswa di fakultas Q, didapatkan *stem and leaf* seperti pada gambar 3.

Stem-and-Leaf Display: Fakultas Q

Stem-and-leaf of Fakultas Q N = 100
Leaf Unit = 1.0

```

1  4  7
3  4  89
10 5 1111111
20 5 2222333333
27 5 4445555
48 5 66666666677777777777
(14) 5 88888899999999
38 6 0000000011111111
22 6 22222333
14 6 4445555
7  6 66777
2  6 9
1  7 0

```

Gambar 3. *Stem and Leaf* Nilai Mahasiswa Fakultas Q

Sama dengan *stem and leaf* pada fakultas P, fakultas Q ini juga menggunakan *increment* sebesar 2. Bisa dilihat bahwa nilai tertinggi mahasiswa fakultas Q adalah 70 dan terendah adalah 47. Angka 14 diberi tanda kurung menyatakan untuk bahwa median dari data berada pada baris tersebut. Median nilai mahasiswa di fakultas Q juga berada pada di antara data ke-50 dan ke-

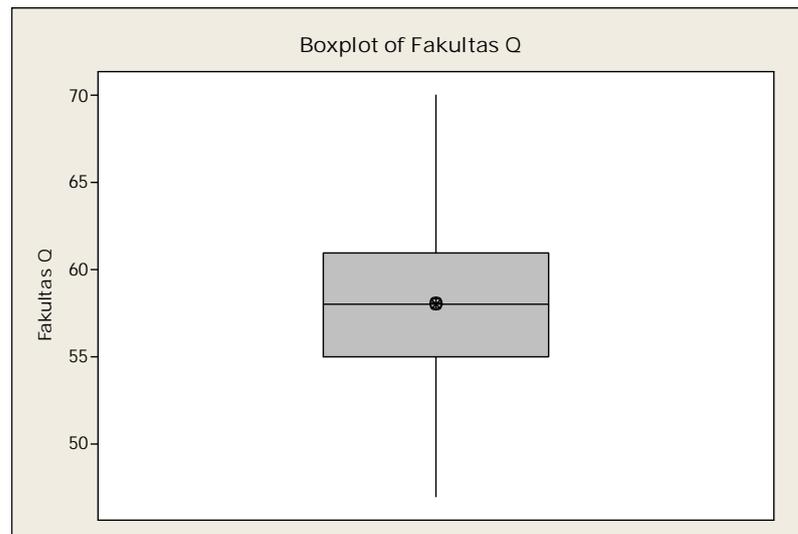
51, yaitu 58. *Stem and leaf* tersebut juga memperlihatkan bahwa data terpusat di sekitar median. Jika *stem and leaf* ini diputar 180° berlawanan arah jarum jam, maka bentuk *stem and leaf* ini akan menyerupai kurva distribusi normal. Data nilai mahasiswa di fakultas Q menghasilkan statistika deskriptif pada tabel 2.

Tabel 2. Tabel Statistika Deskriptif Nilai Mahasiswa Fakultas Q

Variabel	Mean	Q1	Median	Q3	Minimum	Maximum	Range
Fakultas Q	58.08	55	58	61	47	70	23

Berdasarkan tabel 1, dapat diketahui bahwa nilai rata-rata mahasiswa di fakultas Q sebesar 58.08, dengan nilai median 58, sama dengan nilai median dari *stem and leaf* sebelumnya.

Sedangkan nilai kuartil bawah dan kuartil atasnya masing-masing sebesar 55 dan 61. Pola penyebaran data nilai mahasiswa fakultas Q, dapat dilihat pada gambar 4.



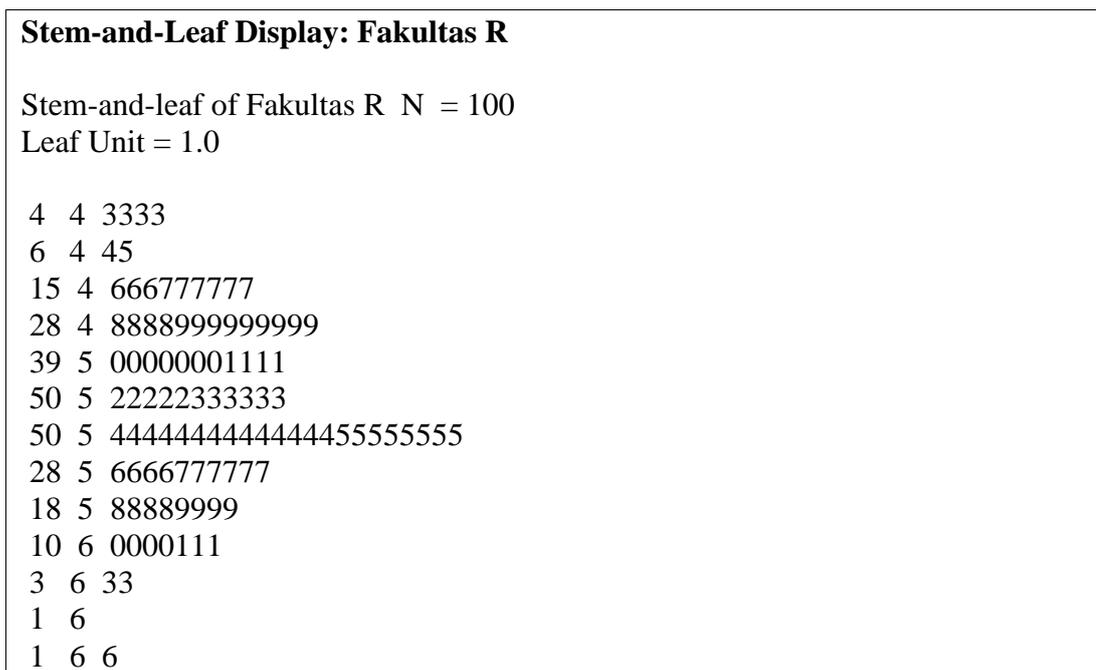
Gambar 4. *Boxplot* Nilai Mahasiswa Fakultas Q

Berdasarkan *boxplot* tersebut, dapat dilihat bahwa data nilai mahasiswa di fakultas Q cenderung berbentuk simetris. Hal ini dapat dilihat dari jarak median ke kuartil 1 dan jarak median ke kuartil 3 yang cenderung sama panjangnya, sama seperti pada fakultas P. Mean dari nilai mahasiswa di fakultas Q hampir berimpit dengan nilai median. Selain bisa dilihat secara visual dari dua titik yang berimpit dari median, hal ini juga bisa dilihat dari tabel 2 yang menunjukkan nilai mean sebesar 58.08 dan median sebesar 58.

Nilai mahasiswa di fakultas Q tidak ada yang terpisah jauh dengan yang lain, atau bisa dikatakan tidak ada nilai yang *outlier*.

c. Fakultas R

Setelah menggambarkan nilai mahasiswa di fakultas P dan fakultas Q, selanjutnya akan dilihat penyebaran nilai di fakultas R. Nilai mahasiswa fakultas R menghasilkan *stem and leaf* dengan *increment* sebesar 2 seperti pada gambar 5.



Gambar 5. *Stem and Leaf* Nilai Mahasiswa Fakultas R

Gambar 5 yang menggambarkan *stem and leaf* nilai mahasiswa di fakultas R memperlihatkan bahwa nilai terendah dan tertinggi di fakultas R adalah 43 dan 66. Berbeda dengan *stem and leaf* kedua fakultas sebelumnya yang memiliki frekuensi yang diberi tanda kurung, *stem and leaf* nilai fakultas Q ini tidak memiliki frekuensi yang diberi tanda kurung. Namun, ada dua baris yang menyatakan frekuensi sebesar 50. Angka 50 yang di atas menyatakan frekuensi kumulatif dari atas, sedangkan angka 50 yang berada di bawah

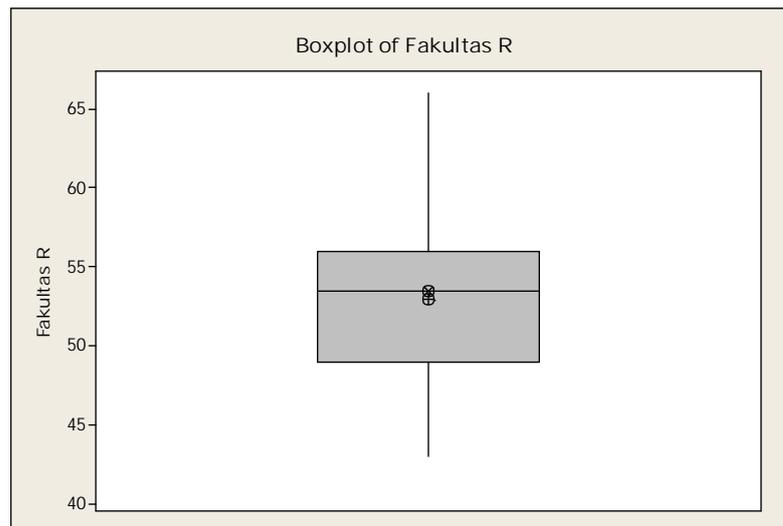
menyatakan frekuensi kumulatif dari bawah. Selain itu, *stem and leaf* juga memperlihatkan bahwa data berpusat di sekitar median, namun banyak data yang berada di terlalu berpusat di atas median, yang menyebabkan data terlihat condong ke kanan. Median dari data ini berada di antara data yang ke-50 dan ke-51. Berarti median dari data adalah sebesar $\frac{53 + 54}{2} = 53.5$. Untuk mendapatkan gambaran umum tentang nilai mahasiswa di fakultas R, dapat dilihat pada tabel 3.

Tabel 3. Tabel Statistika Deskriptif Nilai Mahasiswa Fakultas R

Variabel	Mean	Q1	Median	Q3	Minimum	Maximum	Range
Fakultas R	52.91	49	53.5	56	43	66	23

Tabel 3 memperlihatkan bahwa median data nilai mahasiswa di fakultas R yaitu 53.5, yang berarti sama dengan median dari *stem and leaf*. Sedangkan nilai rata-rata di fakultas R sebesar

52.91, dengan kuartil bawah sebesar 49 dan kuartil atas 56. Untuk melihat pola penyebaran data secara grafis, dapat dilihat pada *boxplot* Gambar 6.



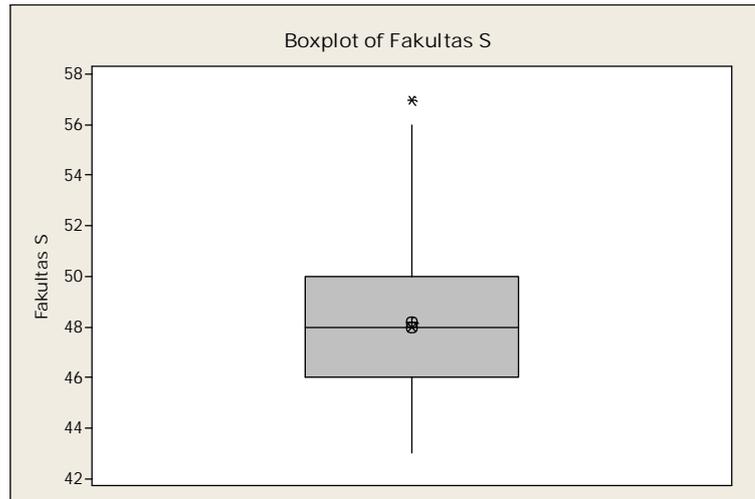
Gambar 6. *Boxplot* Nilai Mahasiswa Fakultas R

Berdasarkan *boxplot* tersebut, dapat dilihat bahwa data nilai mahasiswa di fakultas R cenderung tidak berbentuk simetris. Hal ini dapat dilihat dari jarak median ke kuartil 1 dan jarak median ke kuartil 3 yang tidak sama panjangnya. Data tersebar kurang merata, terutama data yang berada di antara median dan kuartil 3, data terlihat terpusat pada rentang tersebut. Mean dari nilai mahasiswa di fakultas R tidak sama dengan nilai median, seperti yang terlihat pada tabel 3. Nilai mean sebesar

52.91 dan median sebesar 53.5. Perbedaan ini bisa dilihat pada *boxplot*, dimana terdapat sebuah titik yang berada dekat dengan garis median. Nilai mahasiswa di fakultas R tidak ada yang terpisah jauh dengan yang lain, atau bisa dikatakan tidak ada nilai yang *outlier*.

d. Fakultas S

Seperti pada 3 fakultas yang dibahas sebelumnya, nilai mahasiswa di fakultas S ini juga digambarkan dalam bentuk *stem and leaf* dengan *increment* sebesar 2.



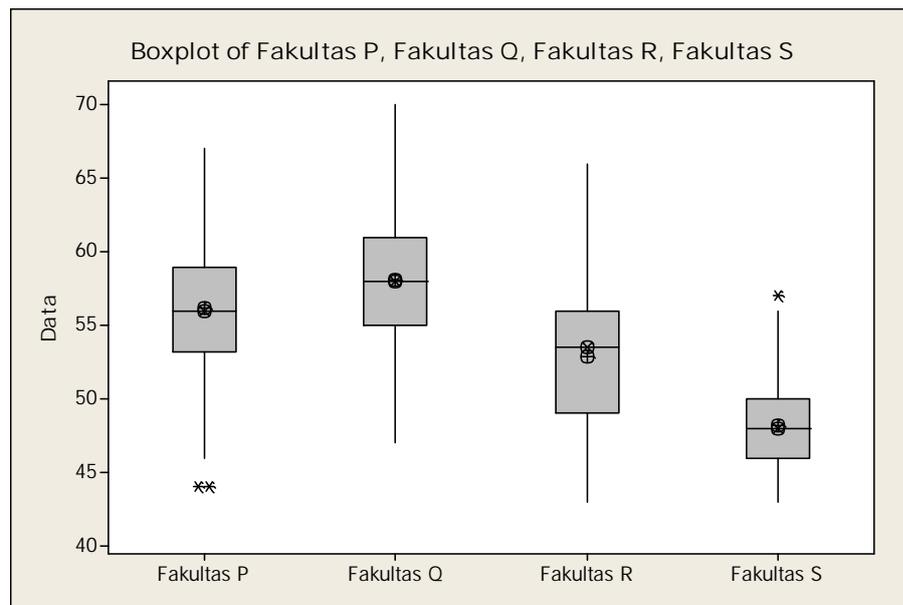
Gambar 8. *Boxplot* Nilai Mahasiswa Fakultas S

Gambar *boxplot* pada gambar 8 memperlihatkan bahwa data nilai mahasiswa di fakultas S cenderung berbentuk simetris. Hal ini dapat dilihat dari jarak median ke kuartil 1 dan jarak median ke kuartil 3 yang cenderung sama panjangnya. Titik mean dari nilai mahasiswa di fakultas S hampir berimpit dengan titik nilai median. Nilai mean sebesar 48.18 dan nilai median sebesar 48. Namun, ada sebuah nilai di fakultas S yang berbeda terlalu jauh dengan nilai yang lain sehingga menjadi *outlier*. Nilai

yang menjadi *outlier* adalah nilai maksimum dari mahasiswa di fakultas S, yaitu 57.

e. Perbandingan Keempat Fakultas

Setelah sebelumnya dibahas tentang penyebaran data di masing-masing fakultas, maka selanjutnya akan dibahas tentang perbandingan penyebaran data keempat fakultas. Jika *boxplot* dari nilai mahasiswa di Fakultas P, Fakultas Q, Fakultas R dan Fakultas S dijadikan dalam satu grafik, maka akan dihasilkan seperti pada Gambar 9.



Gambar 9. Perbandingan *Boxplot* Keempat Fakultas

Berdasarkan gambar 9 dapat dilihat bahwa median dan median dari nilai mahasiswa masing-masing fakultas tidak sama. Yang memiliki mean dan median tertinggi adalah fakultas Q, kemudian disusul oleh fakultas P dan R. Sehingga yang memiliki nilai mean dan median terendah adalah fakultas S. Dari nilai ini dapat dikatakan bahwa dari 100 sampel mahasiswa, fakultas Q menjadi fakultas yang terbaik dalam nilai 7 mata pelajaran SMA jika dilihat dari mean dan mediannya.

Boxplot tersebut memperlihatkan penyebaran data masing-masing fakultas. Penyebaran data fakultas S cenderung terpusat. Kemudian dilanjutkan fakultas P dan Q. Sedangkan fakultas R memiliki penyebaran data yang paling tinggi dibandingkan dengan

fakultas lain. Jika dilihat dari nilai *outlier*, fakultas Q dan fakultas P menjadi fakultas yang tidak memiliki *outlier* dan fakultas S hanya memiliki 1 *outlier*, yaitu nilai maksimumnya sebesar 57. Fakultas P meskipun penyebarannya datanya cukup terpusat dan mean dan mediannya cukup tinggi, namun memiliki dua data yang *outlier*, yaitu nilai minimumnya sebesar 44. Jika bisa diambil kesimpulan dari 100 mahasiswa per fakultas, dapat dikatakan bahwa fakultas Q adalah fakultas yang paling baik berdasarkan nilai 7 mata pelajaran SMA jika dilihat dari ukuran mean dan median, penyebaran dan nilai outlier.

KESIMPULAN

Berdasarkan hasil analisis yang dilakukan sebelumnya, dapat diambil kesimpulan bahwa nilai mahasiswa pada 7 mata pelajaran SMA di fakultas P tersebar secara merata, cenderung simetris, namun ada 2 data yang *outlier* pada nilai minimum. Tidak berbeda dengan fakultas P, data fakultas Q tersebar secara merata, cenderung simetris, namun tidak ada data yang *outlier*. Sedangkan data fakultas R kurang tersebar secara merata, cenderung condong dan tidak ada data yang *outlier*. Di fakultas terakhir, yaitu fakultas S, data nilai mahasiswa pada 7

mata pelajaran SMA tersebar secara merata, cenderung simetris, namun ada 1 data yang *outlier* di nilai maksimumnya. Berdasarkan nilai mean dan median, penyebaran, kesimetrisan data serta ada tidaknya *outlier* bisa dikatakan bahwa dari nilai 100 sampel mahasiswa, fakultas Q merupakan fakultas terbaik berdasarkan nilai 7 mata pelajaran SMA.

DAFTAR PUSTAKA

Aunuddin. 1989 . *Analisis Data*.

Montgomery, Douglas C. 2005. *Introduction to : Statistical Quality Control (5 th ed.)*. Canada : John Wiley & Sons, Inc.