Forum Statistika dan Komputasi, April 2003; p:1-7
ISSN 0853-8115
Vol 8. No.1

SEJARAH PERKEMBANGAN STATISTIKA DAN APLIKASINYA

Sony Sunaryo[1], Setiawan[1], Anik Djuraidah[1], Asep Saefuddin[2]

[1] Dosen pada Jurusan Statistika FMIPA ITS, Surabaya
[2] Dosen pada Departemen Statistika FMIPA IPB, Bogor

Abstrak

Statistika diawali sebagai ilmu untuk mengumpulkan angka (data). Pada abad 17 statistika deskriptif mulai berkembang, begitu juga ilmu peluang yang awalnya dilahirkan dari meja judi sudah mulai muncul . Ilmu peluang ini melandasi berkembangnya statistika induktif yang terjadi pada pertukaran abad 19 dan 20 dengan Karl Pearson sebagai pelopornya. Statistika induktif berkembang pesat setelah R. A. Fisher memperkenalkan metode Maximum Likelihood pada tahun 1922. Dengan adanya perkembangan teknologi komputer, metode eksplorasi data dan bootstrap mulai berkembang pada tahun 1970. Metode ini sebagai awal dari analisis data tanpa model peluang yang populer dengan data driven. Seiring dengan perkembangan statistika induktif, statistika mulai diterapkan pada berbagai bidang seperti ekonomi, industri, pertanian, sosiologi, psikologi, dan lain-lain.Di bidang ekonomi aplikasi statistika pada ekonometrika, sedangkan di bidang industri aplikasi yang sangat terkenal adalah metode Quality Control dan metode Six-Sigma.Pada abad 21 diperkirakan metode data mining akan banyak digunakan dalam bidang terapan. Perkembangan ini akan berpengaruh terhadap model pendidikan dan pengajaran statistika dewasa ini.Di Indonesia penggunaan statistika dipelopori dengan dibukanya program pendidikan statistika di bawah naungan Jurusan Statistika IPB (S1 sejak tahun 1967 dan S2 sejak tahun 1975). Peran Jurusan Statistika IPB baik lewat mata kuliah pelayanan pada jurusan lain di lingkungan IPB, maupun para lulusannya yang sudah tersebar di bergai bidang pekerjaan memberikan dampak positif bagi penggunaan statistika sebagai alat bantu analisis.Sekarang selain IPB ada PTN dan PTS yang telah membuka jurusan statistika.

AWAL PERKEMBANGAN STATISTIKA SECARA UMUM

Perkembangan statistika diawali sebagai suatu ilmu yang membahas cara-cara mengumpulkan angka sebagai hasil pengamatan menjadi bentuk yang lebih mudah dipahami. Menurut Spiegel (1961) statistika berasal dari kata "status" yang berarti negara. Sehingga pada awalnya statistika berkaitan dengan ilmu untuk angka-angka (keterangan) atas perintah raja suatu negara, yang ingin mengetahui kekayaan negaranya, jumlah penduduk, hewan piaraan, hasil pertanian, dan modal. Contoh tertua mengenai hal ini dapat diambil dari zaman Kaisar Agustus yang membuat pernyataan bahwa seluruh dunia harus dikenai pajak, sehingga setiap orang harus melapor kepada statistikawan terdekat (pengumpul pajak). Peristiwa lain di dalam sejarah yang dapat dikemukakan ialah sewaktu William si Penakluk memerintahkan mengadakan pencacahan jiwa dan kekayaan di seluruh wilayah Inggris untuk pengumpulan pajak dan tugas militer. Semua pengamatan dicatat di dalam sebuah buku yang dikenal dengan *Domesday Book*.

Dari keperluan semacam ini timbullah teknik pencatatan angka-angka pengamatan dalam bentuk daftar dan grafik. Bagian statistika yang membicarakan cara mengumpulkan dan menyederhanakan angka-angka pengamatan ini dikenal sebagai statistika deskriptif. Statistika deskriptif dapat berkembang tanpa memerlukan dasar matematika yang kuat, selain kecermatan dalam teknik berhitung.

Sejak tahun 1700-an analisis data yang dilakukan secara deskriptif berdasarkan tabel-tabel frekuensi, rataan, dan ragam untuk sampel (contoh) ukuran besar. Tahun 1800-an merupakan awal penggunaan grafik-grafik untuk penyajian data, seperti histogram, sejalan dengan penemuan sebaran (kurva) Normal. Florence Nightengale (1820-1920) adalah seorang perawat yang terkenal dengan inovasi di bidang ilmu perawatan merupakan pelopor dalam penyajian data secara grafik. Selama perang Crimean, Nightengale mengumpulkan data dan membuat sistem pencatatan. Dari data tersebut dapat ditentukan tingkat mortalitas yang dapat menunjukkan hasil perbaikan kondisi kesehatan yang cenderung menurunkan tingkat kematian. Selanjutnya data tersebut disajikan dalam bentuk grafik yang merupakan suatu inovasi statistika waktu itu.

Dalam statistika deskriptif tidak ada perbedaan antara data yang diperoleh dari sampel dengan populasinya, kemudian apa yang dihitung dari sampel digunakan untuk menandai populasi. Pada taraf selanjutnya orang tidak puas hanya mengumpulkan angka-angka pengamatan saja. Mereka juga tidak puas bahwa yang diperoleh dari sampel digunakan untuk mencirikan populasi. Timbullah usaha-usaha untuk memperbaiki kesimpulan dalam melakukan ramalan-ramalan populasi berdasarkan angka-angka statistik yang dikumpulkan dari sampel tersebut. Bagian ilmu yang membahas cara-cara mengambil kesimpulan berdasarkan angka-angka pengamatan ini dinamakan statistika induktif. Perkembangan statistik induktif tidak lepas dari pengetahuan mengenai peluang, maka ada baiknya kita lihat terlebih dahulu sejarah perkembangan ilmu peluang yang mendasari statistika induktif.

SEJARAH PERKEMBANGAN ILMU PELUANG

Pengetahuan mengenai peluang (*probability*) ini diawali oleh adanya pertanyaan seorang bangsawan Perancis yang juga penjudi bernama Chevalier de Mere kepada Pascal (1623-1662). Penjudi tersebut ingin mengetahui bagaimana pola pembagian uang taruhan pada suatu perjudian apabila permainannya terpaksa dihentikan sebelum selesai. Pertanyaan ini kemudian menjadi bahan pertukaran pikiran antara Pascal dan Fermat (1601-1665) melalui surat. Dari surat-menyurat antara kedua pemikir inilah kemudian timbul dasar-dasar cabang matematika yang dinamakan hitung peluang (***the theory of probability***) pada tahun 1654.

Pada tahun 1657 seorang ilmuwan Jerman Christian Huygens (guru Leibniz) menerbitkan buku *De Ratiocinilis in Ludo Aleae* yang berisi tentang risalat perjudian dan sejak saat itu teori peluang mulai terkenal. Perkembangan pesat terjadi pada abad 18 yang dipelopori oleh Jacob Bernoulli (1654-1705) dan Abraham de Moivre (1667-1754).

Kurva Normal dan persamaannya ditemukan oleh Abraham de Moivre pada tahun 1733. Dia pertama-tama menyatakan sifat-sifat dari kurva Normal yang kemudian dikembangkan oleh dua orang astronom matematika yaitu Pierre de Laplace (1749-1827) berasal dari Perancis dan Gauss (1777-1855) yang berasal dari Jerman secara terpisah sehingga diperoleh fungsi normal dan aplikasinya. Terbitan kurva Normal oleh de Moivre ditemukan Karl Pearson pada tahun 1924 di suatu perpustakan yang digunakan untuk pengembangan statistika induktif untuk ukuran sampel besar. Adolph Quetelet (1796-1874) mempopulerkan sebaran Normal ini pada bermacam-macam data biologi dan sosial.

Thomas Bayes memberikan landasan teori statistika Bayesian (*Bayesian Statistics*) yang pada mulanya menuliskan gagasan tersebut dalam jurnal *Philosophical Transaction* pada tahun 1764. Dewasa ini *Bayesian* sering dipakai oleh para teoritikus genetika kuantitatif secara ekslusif dan juga pada ilmu-ilmu keteknikan, kesehatan, dan lain-lain.

S. D. Poisson dikenal sebagai penemu Sebaran Poisson (*Poisson Distribution*) telah memberikan landasan teori untuk *rare event* yang dituangkan dalam tulisannya *Recherches sur la probabilite* pada tahun 1837. Teori Poisson banyak digunakan dalam dunia industri, manajemen, transportasi, biologi dan lain-lain Pada tahun 1812 Pierre de Laplace memperkenalkan ide baru dan teknik matematika dalam bukunya *Theorie Analytique des Probabilities.* Laplace mulai menerapkan peluang pada banyak permasalahan saintifik dan praktis, tidak hanya pada permainan judi

Jadi walaupun hitung peluang diawali di atas meja judi, ilmu ini telah menjadi pengetahuan yang sangat bermanfaat bagi perikemanusiaan. Di dalam statistika, teori peluang yang melandasi inferensia statistika (statistika induktif) yang menjadi cikal bakal statistika modern.

SEJARAH PERKEMBANGAN STATISTIKA INDUKTIF

Awal perkembangan statistika induktif terjadi pada peralihan abad ke 19 ke abad 20 dengan Karl Pearson (1857-1936) sebagai pelopornya. Masa ini merupakan titik awal perkembangan statistika modern. Pada abad 19 Karl Pearson menerapkan statistika pada biologi untuk masalah hereditas dan proses evolusi biologi yang diterbitkan dalam jurnal *Biometrika.* Dari tahun 1893 sampai 1912 Karl Pearson telah menulis 18 paper yang berjudul kontribusi matematika ke teori evolusi yang berbasiskan analisis regresi dan koefisien korelasi. Pearson menciptakan istilah *standard deviation* (simpangan baku) pada tahun 1893. Dalam statistika deskriptif Pearson juga memperkenalkan ukuran penyimpangan terhadap distribusi data yang simetrik yang disebut koefisien kemiringan dan kurtosis. Pada tahun 1900 Karl Pearson menemukan uji Khi-kuadrat untuk tabel kontingensi 2 arah. Dalam menarik kesimpulan tentang korelasi dan uji khi-kuadrat, Pearson menggunakan sampel besar (n>1000). Analisis data yang digunakan Pearson mengasumsikan data menyebar Normal. Sehingga pada *Biometrika* volume 1 yang terbit tahun 1901 sebagian besar penelitiannya menggunakan ukuran contoh besar.

Sebelum tahun 1912 sedikit sekali penemuan di bidang pengujian hipotesis sampai akhirnya W. S. Gosset (1876-1937) memperkenalkan uji t-student untuk sampel kecil. Gosset adalah seorang mahasiswa (*student*) dari Karl Pearson pada awalnya adalah seorang ahli kimia yang bekerja di perusahaan bir *Guinness* di Dublin. Gosset menemukan uji-t untuk menangani sampel-sampel kecil untuk ***quality control*** di perusahaan bir tersebut. Dia menerbitkan papernya dengan nama *Student* pada jurnal *Biometika* 1908 untuk menghindari larangan dari perusahaan bagi karyawannya yang menulis di dalam sebuah jurnal. Bentuk sebaran secara matematis yang digunakan Gosset tersebut sebenarnya telah ditemukan oleh astronom Jerman Jakob Luroth pada tahun 1875. Gosset menggunakan data hasil pengukuran terhadap tinggi dan jari tengah tangan kiri 3000 narapidana yang dipublikasikan pada volume pertama Biometrika. Dengan metode Monte Carlo dipilih 750 sampel yang berukuran 4 dan diperoleh distribusi data yang mendekati distribusi teoritiknya. Sebaran t-student banyak dipakai sebagai acuan dalam menduga parameter rataan ukuran contoh kecil (n<30).

Metode estimasi parameter populasi yang digunakan sebelum tahun 1912 adalah metode kuadrat terkecil yang dikemukakan oleh Gauss dan metode deviasi mutlak terkecil yang dikemukakan Laplace. Kedua metode ini digunakan untuk mengestimasi parameter dalam model linier. Kemudian Karl Pearson memperkenalkan metode momen untuk estimasi parameter pada tabel frekuensi.

Statistika induktif mulai berkembang pesat setelah R. A. Fisher (1890 -1962) menulis paper yang sangat terkenal pada tahun 1922 yaitu *On the Mathematical Foundations of Theoritical Statistics* (Mallows, 1998). Fisher memperkenalkan istilah *specification* untuk mengidentifikasi 3 problem yang muncul pada reduksi data, yaitu :

1. **Spesifikasi** dari jenis populasi, yaitu bentuk matematis dari populasi yang mencakup parameter yang tidak diketahui.
2. **Estimasi**, yaitu pemilihan metode statistik untuk mengestimasi parameter dari populasi
3. **Sebaran**, yaitu sebaran statistik dari contoh atau sampel.

Tulisan tersebut memperkenalkan suatu metode yang terkenal dengan nama *Maximum Likelihood* yang digunakan untuk estimasi dan pengujian hipotesis.

Pada tahun 1925 terbit buku *Statistical Methods for Research Workers* karangan Fisher yang berisi rancangan percobaan dan analisis varians di bidang biologi. Fisher yang cara berfikirnya dipengaruhi aliran statistika yang dianut Karl Pearson, yaitu penarikan kesimpulan didasarkan pada model peluang (*model-driven*) merupakan promotor penggunaan cara-cara stastistika di dalam bidang-bidang ilmu pertanian, biologi dan genetika. Untuk jasanya ini Fisher dianugerahi gelar *Baronet* oleh Ratu Inggris, sehingga ia berhak menggunakan nama Sir Ronald Fisher. Kontribusi Fisher membuat cakupan metode pengembangan yang sesuai untuk sampel kecil, seperti Gosset, penemuan presisi sebaran dari beberapa statistik sampel dan penemuan analisis varians. Fisher merekomendasikan *maximum likelihood*, yang digunakan untuk estimasi dan pengujian hipotesis. Fisher dianggap penemu statistika modern karena kontribusinya yang sangat penting dan dianggap sebagai pemikir ulung tempaan abad kedua puluh.

Pada era Fisher, seorang pemikir Rusia Jerzy Neyman (1894-1981) juga dipandang sebagai penemu besar dari statistika modern karena kontribusinya dalam mengembangkan teori peluang, uji hipotesis, selang kepercayaan, dan matematika statistik. Neyman bekerjasama dengan Egon Pearson (anak Karl Pearson) mengembangkan teori-teori untuk pengujian hipotesis, salah satu teorinya yang terkenal adalah **Teorema Neymann-Pearson** (1936). Selain itu Neyman juga mengembangkan teori sampling survey pada tahun 1934.

Pada tulisan Fisher (1915) mengemukakan representasi geometrik data peubah-ganda dua (bivariate) untuk menurunkan distribusi sampling bersama dari penduga matriks varian-kovarians. Pada tahun 1928 Wishart menggunakan metode yang sama untuk menurunkan distribusi bersama dari penduga matriks varians-kovarians untuk sebaran Normal ganda (**Multivariate Normal**) yang akhirnya populer dengan **distribusi Wishart**.

Pada tahun 1936, Fisher membuka area baru penelitian yang disebut fungsi diskriminan yang pada awalnya untuk menjawab pertanyaan yang diajukan oleh ahli antropologi yaitu untuk menentukan jenis kelamin dari pemilik tulang rahang tengkorak yang mempunyai ukuran-ukuran tertentu. Fungsi diskriminan untuk membedakan kedua jenis kelamin ini populer dengan nama fungsi linear diskriminan Fisher (LDF) untuk 2 grup dan merupakan salah satu metode pada analisis peubah ganda.

Calyampudi Radhakrishnan Rao (1920 -) adalah mahasiswa bimbingan dari Fisher. Rao bekerja di museum antropologi sambil menyelesaikan Ph.D tahun 1948. Tahun 1946 Rao mengembangkan fungsi diskriminan linear Fisher untuk klasifikasi dengan banyak grup. Selain itu Rao juga berkontribusi dalam mengembangkan matematika statistik dengan teorinya yang terkenal pertidaksamaan Rao-Cramer dan teorema Rao-Blackwell yang dikemukakan secara terpisah oleh Rao pada tahun 1945 dan Blackwell pada tahun 1947. Salah satu buku karangan Rao yang terkenal adalah *Linear Statistical Inference* yang telah diterjemahkan ke dalam 6 bahasa.

Prasantha Chandra Mahalanobis (1893-1972) berkontribusi dalam mengembangkan analisis peubah ganda. Salah satu kontribusinya yang besar adalah *jarak Mahalanobis* (D-statistik) yang merupakan ukuran jarak untuk data dengan variabel banyak yang digunakan dalam analisis klasifikasi. Mahalanobis juga pendiri jurnal statistik India yang sangat terkenal bernama *Sankya*. Pada tahun 1931 Mahalanobis mendirikan *Indian Statistical Institute* dengan salah satu divisinya bernama *National Sample Survey* (NSS) yang bertugas mengumpulkan data sosioekonomik dan demografi di seluruh India. Divisi ini membuat Mahalanobis mempunyai peranan penting dalam perencanaan ekonomi di India dan akhirnya NSS sekarang berfungsi sebagai bagian penting dari *Ministry of Planning*.

Pada tahun 1931 Hotelling memperkenalkan statistik T^2 yang merupakan generalisasi dari statistik t-student untuk menguji hipotesis nilai tengah pada data peubah-ganda. Distribusi tak nol dari T^2 adalah sama dengan *Mahalanobis-Distance* yang mempunyai tujuan berbeda ditemukan oleh Bose dan Roy pada tahun 1938.

Seperti telah disebutkan bahwa aliran yang dianut Karl Pearson, Gosset dan R.A. Fisher mendasarkan kesimpulan pada jenis sebaran dari populasi sehingga dikenal dengan metode parametrik. Jika asumsi sebaran populasi tidak dipenuhi maka perlu dilakukan transformasi data, salah satunya adalah transformasi pangkat yang ditemukan oleh Box dan Cox tahun 1964, sehingga dikenal dengan nama **transformasi Box-Cox**. Di lain pihak pada tahun 1945 Frank Wilcoxon memperkenalkan metode **statistik non parametrik** yang bebas dari sebaran populasi yang sekarang dikenal dengan Uji Tanda Peringkat. Kemudian berkembang metode-metode non parametrik untuk analisis padanan dari metode parametrik. Seperti tahun 1952 W. H. Kruskal dan W. A. Wallis memperkenalkan uji non parametrik yang berpadanan dengan uji kesamaan mean pada analisis varians yang dikenal dengan nama Uji Kruskal-Wallis. Pada tahun 1958 Kaplan mengunakan metode non parametrik untuk pendugaan sekuensial.

Pada akhirnya dengan adanya perkembangan teknologi komputer metode eksplorasi data berkembang pesat sekitar tahun 1970. J. W. Tukey (1915-) mempunyai kontribusi besar dalam pengembangan metode eksplorasi baik secara grafis maupun numerik. Beberapa penemuannya adalah diagram dahan-daun dan diagram kotak garis.

Pada pertengahan 1970 Efron memperkenalkan ***Metode Bootstrap*** untuk menduga parameter dari sebaran yang tidak diketahui bentuknya. Bootstraping ini merupakan teknik modifikasi dari Jacknife yang diperkenalkan oleh Queneuille pada tahun 1948. Berhubung metode ini pada awalnya tidak membobotkan model peluang, tetapi berbasis pada data, bootstrap dikenal sebagai ***data driven approach***. Pada dekade 80-an perkembangan metode non parametrik mulai sering digunakan seperti pada regresi nonparametrik, estimasi distribusi dengan kernel, dan neural network.

SEJARAH PERKEMBANGAN APLIKASI STATISTIKA

Pada abad 20 statistika berkembang menjadi ilmu yang matang. Selain di bidang pertanian, statistika berkembang pada bidang psikologi, ekonomi, sosiologi, industri, dan lain-lain. Perkembangan statistika di bidang pertanian cukup banyak terutama dalam penggunaan rancangan percobaan yang memang sudah diawali pada masa Fisher.

Perkembangan statistika di bidang ekonomi yang dikenal dengan istilah ekonometrika dimulai tahun 1920 dipelopori Ragnar Frisch dan Jan Tinbergen. Ekonometrika adalah cabang dari ilmu ekonomi yang merupakan integrasi antara ekonomi, matematika dan statistika. Walaupun demikian *powerful* perkembangan ekonometrika kurang mendapat sambutan hangat dari ekonom-ekonom besar yang kurang "sreg" dengan pemodelan termasuk John Maynard Keynes. Keynes memandang skeptis terhadap buku Tinbergen yang berjudul *Statistical Testing for Business Cycle Theory*. Baru di akhir tahun 1940 dan awal 1950 ekonometrika mulai berkembang lagi yang dipelopori oleh Chernoff, Haavelmo, Koopmans, Rubin dan Simon yang bekerja pada *Cowles Commision for Research in Economics*.

Peran statistika cukup besar dalam ekonometrika terutama dalam hal metode estimasi parameter model ekonometrika yang pada umumnya terdiri dari beberapa persamaan yang saling terkait (sistem persamaan simultan dan seemingly unrelated regression). Untuk mengestimasi parameter sistem persamaan simultan Hendri Theil tahun 1956 menemukan suatu metode 2SLS (*two stage least squares*). Kemudian pada tahun 1962 Zellner menemukan suatu metode SUR (*Seemingly Unrelated Regression*) untuk mengestimasi parameter model sistem persamaan regresi. Selanjutnya Theil bersama Zellner menemukan metode 3SLS (*three stage least squares*) untuk mengestimasi sistem persamaan simultan yang pada prinsipnya merupakan integrasi antara metode 2SLS dengan metode SUR.

Model ekonometrika pada umumnya dibangun berdasarkan data yang bersifat time series, sehingga memunculkan model distribusi lag maupun *autoregressive* yang dikembangkan oleh Nerlove pada tahun 1972. Pada umumnya model-model tersebut terjadi pelanggaran asumsi klasik (*autocorrelation, heteroscedasticity*), sehingga belakangan muncul suatu model yang dikenal dengan ARCH (*autoregressive and conditional heteroscedasticity*).

Di bidang peternakan, industri peternakan di USA berterimakasih kepada K. Pearson karena melalui Teori Normal, kemajuan genetic (*genetic progress*) pada produksi susu meningkat 40 % pada tahun 1950-an. Untuk analisis genetika kuantitatif umumnya dipergunakan statistika tingkat tinggi seperti *Bayesian Statistics*.

Di bidang industri peran statistika yang menonjol adalah dalam bidang pengendalian kualitas dan penerapan rancangan percobaan factorial sebagian (*fractional factorial*) yang berusaha meminimumkan jumlah percobaan yang relatif mahal. Hal ini dapat kita kenal seperti pada **metode Taguchi** yang ditemukan oleh Dr. Geinichi Taguchi dari Jepang sekitar tahun 1980 yang disebut juga **off line quality control**. Walter Shewhart (1891-1967) seorang fisikawan Amerika, yang bekerja sebagai *engineer* dan konsultan tahun 1924 menunjukkan ide pengendalian kualitas secara statistik (*statistical quality control*) melalui penggunaan *control chart* atau *run-chart*. Pendekatan ini dapat memberikan tanda jika proses produksi menyimpang dari target, yang disebut juga **on line quality control**. Pada tahun 1930 William Edward Deming (1900-1993), yang berguru pada Shewhart untuk masalah manajemen tertarik terhadap aplikasi statistika untuk merencanakan pengendalian kualitas dan metode perbaikan proses untuk industri. Ide-idenya disarikan dalam 14 point untuk managemen (yang dikenal sebagai 14 Prinsip Deming), misalnya salah satu point menyatakan "jangan memilih supplier karena harga, tetapi pertimbangkan kualitas dan pilih supplier yang menggunakan *quality control*". Untuk memperbaiki proses Deming memperkenalkan langkah-langkah tersistem yang dikenal dengan nama **PDCA (Plan, Do, Chek, and Action)**. Karena metode dan idenya ini, pada tahun 1950 para pimpinan bisnis dan industri yang tergabung ke dalam JUSE (*Japan Union of Scientist and Engineering*) di Jepang mengundang Deming untuk mengajarkan metode baru tersebut. Penggunaan metode Deming secara luas di Jepang berpengaruh terhadap pemulihan industri dan ekonomi Jepang yang hancur setelah perang dunia II, yang akhirnya terjadi *economic booming* bagi Jepang pada abad ke 20. Meskipun ide-ide Deming sukses di Jepang, namun secara luas diabaikan di Amerika. Metode Deming baru diterapkan pada awal 1980 ketika perusahaan-perusahaan di Amerika merasa perlu meningkatkan efektifitas agar dapat bersaing dengan pasar asing. Dengan berkembangnya penggunaan statistika di bidang industri yang menunjukkan manfaatnya, maka timbul ilmu baru yang merupakan gabungan statistika dan managemen yang dikenal dengan metode SIX SIGMA yang mengusahakan produk dengan konsep *zero defect*. Metode ini banyak diterapkan di industri industri besar seperti *Motorola* (awal pemakai Six Sigma), *General Electric Company (GE)*, Kodak, dan lain lain.

TREND PERKEMBANGAN STATISTIKA DI ABAD 21

Karl *Pearson*, Fisher, Neyman dan Wald selama setengah abad telah meletakkan dasar statistika yang berbasis matematika, sehingga penelitian-penelitian dan kuliah-kuliah statistika di Perguruan Tinggi umumnya didasarkan pada beberapa pedoman atau dasar yang ditemukan oleh tokoh-tokoh tersebut. Penggunaan statistika secara luas, terkadang timbul kontroversi diantara para ahli tentang pemilihan model data, penggunaan *prior probability* dan interpertasi hasil. Hasil analisis terhadap data yang sama dengan lain konsultan statistika dimungkinkan terjadi perbedaan kesimpulan. Statistika induktif dapat dipakai untuk menangani masalah dimana perolehan data dirasakan perlu efesiensi atau perlu biaya mahal, sehingga umumnya dapat diatasi dengan analisis dengan sampel-sampel ukuran kecil.

Di era millenium dengan dominasi teknologi informasi, data base yang besar, interaksi dengan komputer dan informasi yang kompleks, maka menurut C.R. Rao dalam tulisannya 14 Nopember 2001 (berjudul Has Statistics a Future ? If So in What Form ?) statistika yang berdasarkan pada model-model probabilistik tidak mencukupi, sehingga metode-metode yang akan muncul diarahkan untuk menjawab tantangan zaman yang diberi nama data mining. Istilah data mining (penambangan data) ini menurut Nasoetion (2002) awalnya berasal dari para ahli ilmu komputer yang dalam sehari-harinya bekerja dalam dunia kecerdasan buatan. Untuk pekerjaan ini mereka membangkitkan dan mengumpulkan data dalam ukuran sangat besar dan mencoba menemukan pola-pola keteraturan data yang dapat diterangkan. Pada tahun 1990-an metode data driven yang tidak terlalu ketat dengan asumsi sebaran mulai digunakan untuk analisis berbagai data, terutama untuk eksplorasi data atau "data mining". Berhubung data mining ini sangat computer intensive, maka diusulkan diberi nama Statistical Methods Mining oleh beberapa statistikawan USA. Menurut David M. Rpcke dari University of California metode data mining mempunyai dua prinsip dasar yaitu data cleaning dan data segmentation. Data cleaning untuk mendeteksi data pencilan sedangkan data segmentation untuk pengelompokan data, sehingga akan diketahui pola dari data.

Pada abad 21 diperkirakan metode data mining merupakan metode yang akan banyak digunakan dalam berbagai bidang terapan. Pada metode data mining spesifikasi permasalahan didasarkan pada bidang ilmunya lebih diutamakan daripada pendugaan parameter sehingga masalah tersebut dapat diformulasikan dengan benar untuk memperoleh solusi yang tepat melalui eksplorasi data. Hal ini berbeda dengan periode Fisher yang lebih mementingkan mencari metode pendugaan dan pendekatan sebaran yang tepat, sehingga spesifikasi permasalahan lebih diutamakan pada pendugaan parameter dan asumsi sebaran. Jadi Fisherian Statistics itu sebenarnya model driven yang agak beda dengan data mining yang lebih bersifat data driven. Akan tetapi pada pelaksanaannya, kedua "driven" tersebut harus dikuasai oleh statistikawan di abad millenium ini. Situasi ini akan berpengaruh terhadap model pendidikan dan pengajaran statistika dewasa ini. Emanuel Parzen (Department of statistics Texas A & M University College) baru-baru ini menulis tentang "Data Mining, Statistical Methods Mining and History of Statistics". Dalam tulisannya tersebut dibahas juga masalah pendidikan statistika menghadapi masa depan dimana data mining akan berkembang, seperti bagaimana cara mengajar matematik statistik untuk non matematik statistik, materi yang berhubungan dengan komputer seperti teknik simulasi, analisis numerik, analis data dan struktur data perlu ditingkatkan bagi para mahasiswa.

Belum adanya standard analisis untuk eksplorasi data dalam data ukuran besar inilah diperkirakan, metode data mining akan banyak dikembangkan dan diteliti oleh para pakar statistika.

SEKILAS SEJARAH STATISTIKA DI INDONESIA

Dilihat dari sejarah pendidikan statistika di Indonesia, Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor dapat berbangga, karena jurusan yang dirintis dan didirikan oleh Prof. Dr. Andi Hakim Nasoetion (Alm) tahun 1972 adalah Jurusan Statistika tertua di Indonesia. Awalnya dimulai dari Unit Biometrika di bawah Fakultas Pertanian IPB yang kemudian berubah menjadi Pusat Pengolahan Data Statistika dan Komputasi dan akhirnya menjadi Departemen Statistika dan Komputasi di bawah Fakultas Pertanian. Pada waktu FMIPA disyahkan di IPB pada tahun 1982, namanya berubah menjadi Jurusan Statistika di bawah FMIPA. Jadi boleh dikatakan Departemen Statistika adalah the founding father of FMIPA IPB.

Pada dekade 60 dan 70-an statistika dikenal sebagai "tongkat pembimbing di daerah ketidaktahuan". Pada dasarnya fungsi tersebut tidak akan hilang, karena statistika tetap berperan di dalam proses penelitian mulai dari rancangan dan analisis, sampai ke penarikan kesimpulan. Di wilayah dimana dunia penuh dengan ketidakpastian, keragaman dan proses acak itulah statistika sangat diperlukan. Tanpa bantuan statistika tidak mustahil kita terjebak oleh kesimpulan yang tidak sepatutnya (misleading conclusion). Selain itu, statistika juga perlu menerawang ke masa depan. Statistika sebagai tongkat pembantu ke masa depan itu wajar saja bila saat ini berkembang moto "Statistika adalah alat bantu untuk memecahkan masalah masa depan", problem solver of the future.

"Statistics is not just for statistician", memang demikian adanya. Model-model statistika sangat membantu pemahaman proses pembelajaran dalam dunia pendidikan dan psikologi, regresi dan analisis deret waktu sudah sering membuka tabir kesulitan riset dalam keteknikan, kimia, ekonomi, biologi dan ilmu-ilmu kesehatan. Dewasa ini statistika sering diminta bantuan oleh ahli-ahli hukum kriminalitas, khususnya dengan berkembangnya "statistics for forensic and DNA fingerprinting".

Semua kisah sukses statistika di dunia itu adalah titik cerah bagi masa depan Jurusan Statistika FMIPA-IPB. Saat ini Jurusan Statistika FMIPA IPB sudah mengasuh tidak saja program S1 (sarjana), tetapi juga Program Pascasarjana S2 (magister sains), dan bahkan doktor (S3). Pendidikan tersebut diramu dengan kegiatan riset yang bekerja sama dengan disiplin ilmu lainnya. Dengan demikian, statistika secara keilmuan menjadi satu kesatuan yang tidak terpisahkan mulai dari jenjang S1 ampai S2 dan masuk ke ranah-ranah keilmuan lainnya yang sudah barang tentu sangat memerlukan statistika.

Jurusan Statistika IPB yang merupakan pelopor pendidikan statistika di Indonesia, selain menjalankan program pendidikan statistika pada jenjang S1, S2 bahkan S3 juga mengembangkan program pelayanan mata kuliah metode statistika pada program studi lain di lingkungan IPB. Dampak dari ini, semua alumni IPB dibekali pola berfikir statistika sehingga mereka sudah terbiasa dengan keteraturan berfikir sehingga menjadi "pioner" penggunaan berfikir secara statistika di lingkungan kerjanya di berbagai instansi (khususnya DEPTAN).

Penyelenggaraan Program pendidikan S1 (Sarjana) di IPB dimulai sejak tahun 1967, sedangkan program pendidikan pascasarjana (S2) dimulai sejak tahun 1975 dengan jumlah lulusan kurang lebih 150 (Magister Sains). Dibukanya Program Doktor (S3) sejak empat tahun yang lalu dimaksudkan selain untuk pengembangan statistika di Indonesia juga untuk memperkokoh peran Jurusan Statistka IPB dalam pembangunan bangsa menyongsong Indonesia baru. Dapat dibayangkan selama lebih kurang 35 tahun jurusan statistika berkiprah, tentu alumninya (S1 dan S2) sudah tersebar di berbagai instansi, baik sebagai peneliti, pengambil kebijakan, statistisi profesional, maupun tenaga pengajar di PTN maupun PTS.

Program-program untuk meningkatkan profesionalisme dan akademik lainnya dikembangkan dengan membuka kerjasama akademik dengan program studi sejenis di berbagai Universitas di Indonesia. Setiap tahun jurusan Statistika IPB melakukan program pelatihan untuk dosen PTN di Indonesia. Kerjasama dengan instansi lain, khususnya DEPTAN dalam pelatihan statistika.

Sampai saat ini di Indonesia selain IPB telah ada PTN dan PTS lain yang telah membuka jurusan Statistika secara mandiri tanpa dibawah naungan jurusan Matematika. PTN yang telah membuka jurusan statistika secara mandiri adalah UNPAD, ITS dan UGM. Sedang PTS nya adalah UNISBA, UII Yogyakarta dan salah satu PTS di kota Malang. Selain itu untuk BPS telah membuka pendidikan jurusan statistika untuk keperluan di instansinya yang dulu bernama AIS dengan pendidikan jenjang D3, sedang sekarang bernama STIS dengan jenjang pendidikan setara S1.

DAFTAR PUSTAKA

Gani, J. 1982. *The Making of Statisticians.* Springer- Verlag. New York.

Mallows, C. (1998). 1997, *Fisher Memorial Lecture. The Zeroth Problem.* ASA 52(1) : 1-9.

Nasoetion, A. H. dan Rambe, A. 1984. *Teori Statistika untuk ilmu-ilmu Kuantitatif.* Bhratara Karya Aksara. Jakarta.

Parzen. Emanuel. 2002. *Data Mining, Statistical Methods Mining and History of Statistics".* (Department of statistics Texas A & M University College) termuat di internet.

Rao, C. R. and Szekely, G. J. 2000. *Statistics for The 21st Century. Methodologies for Applications of the future.* Marcell Dekker. New York.

Rao, C. R. 2001 *Has Statistics a Future ? If So in What Form ?* termuat di internet.

Saefuddin, A. 2000. *Profil Jurusan Statistika.* FMIPA-IPB.

Spiegel, Murray .R. 1961. *Theory and Problem of Statistics.* Mc Graw-Hill. New York.

⊗