

STATISTIK : SALAH SATU INDIKATOR UTAMA PERADABAN *

Maman A. Djauhari **

*When using a statistical model,
careful attention must be given to the uncertainties in the model*

Author's adaptation from Richard Feynman's
"On the reliability of the Challenger space shuttle"

Abstract

Today's statistics has appeared as universal language in the modern communication in Indonesia. It has been a need of civilization such as reading and writing, and also as a civilization indicator of a nation. This paper offers an idea that views statistics as a soft technology that is able to bring this nation to the more civilized nation. An example is presented as the illustration.

1. Pendahuluan

Atas inisiatif *National Statistical Office of Korea* (NSO) pada bulan September 1999 di Taejon dibentuk *International Statistical Forum* (ISF). Tujuannya adalah membangun kerja sama internasional bidang perstatistikan sesuai dengan "*Fundamental Principles of Official Statistics*" yang diterbitkan oleh Komisi PBB Bidang Statistik. Kerja sama tersebut meliputi upaya menginisiasi peraturan tentang penambangan data (*data mining*) di antara anggota ISF. Pada kesempatan ini, penulis mendapat kehormatan dan kepercayaan dari NSO untuk duduk sebagai salah seorang anggota pendiri ISF tersebut.

Dalam acara pendirian ISF itu telah diberikan catatan bahwa di kawasan Asia Timur dan Asia Tenggara banyak negara yang sering melakukan praktik statistik karet (*rubber statistics*). Suatu julukan, ditujukan pula ke arah kita, yang memilukan namun harus kita terima karena, seperti yang tampak pada kasus IDT (Inpres Desa Tertinggal), kasus kompensasi BBM atau BLT (Bantuan Langsung Tunai) dan contoh-contoh lain, masih banyak anggota masyarakat yang mempraktikkan statistik karet. Banyak yang tidak menyadari bahwa perilaku seperti ini disimak dunia. Kita tidak mungkin memperoleh kepercayaan masyarakat internasional kalau data kita tidak andal atau kalau statistik kita tidak sesuai dengan

* Sebagian besar tulisan ini telah disampaikan pada Kuliah Umum di Sekolah Tinggi Ilmu Statistik, Jakarta, 2 Februari 2007.

** Profesor Statistik, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Bandung, Jalan Ganesa 10, Bandung 4013, Email: maman@dns.math.itb.ac.id

tuntutan pergaulan masyarakat global atau kalau kita dianggap mempraktikkan statistik karet.

“*Fundamental Principles of Official Statistics*” yang telah disebut di depan pada dasarnya mencerminkan kehendak agar kebijakan dan kegiatan perstatistikan dan hasilnya mencerminkan wajah budaya/kultur dan tingkat peradaban yang tinggi dari masyarakat pelaksana kegiatan itu. Kualitas kebijakan dan kualitas kegiatan perstatistikan suatu bangsa akan menggambarkan kualitas kultur dan peradaban serta martabat bangsa itu. Acuanya adalah norma-norma pergaulan masyarakat internasional yang berbudaya intelektual.

Pada Bagian 2 akan dikemukakan mengapa bangsa yang berbudaya memerlukan statistik. Selanjutnya, Bagian 3 mengemukakan pandangan tentang statistik dilihat dari konsep fundamental yang akan dapat menghantarkan seseorang ke perbatasan (*frontier*) pengembangan ilmu statistik. Bagian 4 menyajikan contoh statistik sebagai *soft technology* yang dapat memberikan ilustrasi bagaimana statistik harus kita produksi sebagai komoditas ekonomi yang bernilai tinggi. Bagian 5 berisi daftar beberapa masalah aktual yang potensial menjadi produk berkelas dunia. Tulisan ini diakhiri dengan sebuah catatan tambahan di Bagian 6.

2. Mengapa statistik diperlukan

Berbicara tentang statistik, maka di dalam mental model kita, pikiran kita, atau *mind set* kita, akan langsung muncul entitas-entitas

”sumber data dan informasi”, ”data”, ”informasi”, ”*knowledge*”, dan ”alat statistikal” Di bawah semua itu masih ada dua entitas yang menjadi fondasinya, yaitu ”kultur” dan ”matematika” Ketujuh entitas tersebut adalah komponen-komponen utama yang kita perlukan dalam mengelola kerandoman (*randomness*) yang muncul secara alamiah tatkala kita berupaya memahami kelakuan Alam. Kerandoman adalah suatu fenomena alam yang kalau intelek manusia tidak mampu mengontrolnya dengan baik akan dapat menimbulkan kejadian yang lantak oleh manusia disebut bencana alam. Kerandoman tampil dalam bentuk variasi.

Mengelola kerandoman atau variasi menuntut pemahaman tentang konsep ketidak-pastian (*uncertainty*) yang terdapat dalam alat-alat statistikal. Dalam praktek, ketidak-pastian tersebut diukur dengan alat matematikal yang disebut probabilitas. Alat ini terletak di dalam arsenal analisis matematikal. Probabilitas berperan mengukur kualitas alat statistikal; seberapa besar alat statistikal tersebut mengandung unsur ketidakpastian. Inilah yang dipesankan di awal tulisan ini: ”*When using a statistical model, careful attention must be given to the uncertainties in the model*”

Variasi muncul di mana-mana; pada sumber data dan informasi, dan juga di dalam data yang terekam. Informasi yang digali dari data yang terekam, dan *knowledge* yang diperoleh dari informasi hasil penambangan pada tumpukan data yang terekam, dibayang-bayangi oleh ketidakpastian. Oleh karena itu, tatkala

berpikir tentang hal-hal tersebut, secara paralel intelek manusia bekerja memikirkan alat-alat yang diperlukan untuk memahaminya. Untuk mendapatkan data dari sumbernya diperlukan alat. Jika sumber data dan informasi adalah manusia, maka alat yang diperlukan bukan hanya alat statistik dan alat matematikal, melainkan juga alat kultural. Namun, jika sumber informasi itu bukan manusia, maka alat yang diperlukan cukup dua yang disebut pertama. Tatkala data sudah diperoleh dari sumbernya, maka data tersebut merupakan pesan dari alam tentang perilaku parsialnya sejauh yang mampu ditangkap oleh intelek manusia. Alam, setiap saat, memancarkan informasi dan intelek manusia memproduksi alat perekam dan alat statistik. Alat perekam itu dibuat untuk menangkap informasi alam dan mengubahnya menjadi data. Sedangkan alat statistik dibuat untuk mentransformasikan data menjadi informasi. Selanjutnya, melalui kemampuan hermeneutikal, yakni kemampuan menafsirkan segala sesuatu, informasi ini ditransformasikan lagi menjadi *knowledge*.

Di samping kemampuan membuat alat statistik dan kemampuan hermeneutikal, yang diperlukan tatkala data yang terekam hendak ditransformasikan menjadi informasi dan *knowledge*, masih diperlukan kemampuan mengembangkan arsenal alat-alat matematikal. “*There is no Applied Mathematics without Mathematics to Apply*” Demikian tertera dalam *The Art of Modeling in Science and Engineering*

karya Basmadjian (1999) diterbitkan oleh Chapman & Hall / CRC. Jelas bahwa kualitas alat statistik yang diproduksi ditentukan secara mutlak oleh kapasitas intelek manusia pegiat statistik. Semua kemampuan ini, yang berkembang karena hasrat untuk mampu mengelola kerandoman dan memahami ketidak-pastian dengan baik, diperlukan karena tuntutan kultural, peradaban, dan pergaulan antar bangsa yang bermartabat.

3. Apa itu statistik?

Sejak awal paruh kedua abad ke-20 dunia menyaksikan fenomena global di mana statistik menjelma menjadi bahasa universal dalam pergaulan modern masyarakat internasional. Ia telah menjelma menjadi kebutuhan peradaban seperti halnya membaca dan menulis. Ia adalah salah satu indikator utama peradaban sebuah bangsa. Oleh karena itu, perkuliahan modern bidang statistika sejatinya merupakan perkuliahan kultural yang menekankan kepada penghormatan terhadap nilai-nilai luhur kemanusiaan, norma-norma dan etika kehidupan, kejujuran, transparansi, akuntabilitas, kualitas, dan produktivitas.

Statistik, secara populer, sering diartikan sebagai data atau hasil hitungan berdasarkan data. Statistik juga dapat diartikan sebagai *information science* yang telah teruji keunggulannya. Melalui pengertian ini, statistik telah diterima oleh ilmuwan dari semua bidang; mulai dari ilmuwan yang bekerja pada *the very hard sciences* seperti astronomi – dengan

obyek penelitian yang berada amat jauh di luar angkasa – dan fisika teoretis – dengan obyek penelitiannya amat dekat namun tidak tampak seperti atom – hingga ilmuwan yang menekuni *the very soft sciences* seperti seni yang bergelut dengan forma-forma. Secara fundamental, statistik adalah sebuah fungsi terukur (*measurable function*). Distribusinya, yang merupakan kuantifikasi dari kelakuannya secara umum, adalah ukuran (*measure*) yang memungkinkan orang dapat mengukur besarnya ketidak-pastian (*probability*). Adapun parameter distribusi adalah fungsional yang menunjukkan karakteristik parsialnya. Pada tataran fundamental inilah statistisi atau statistikawan bekerja untuk mengembangkan arsenal alat-alat statistik dan atau mencipta statistik yang baru. Apa itu statistisi? Itu adalah sebuah profesi. Apabila kita berpegang pada definisi yang diberikan Jacques Dieudonnees tentang matematisi, maka statistisi adalah mereka yang selama hidupnya pernah membuat statistik sebagai fungsi terukur atau teorema tentang statistik.

Pada permulaan abad ke-20 terjadi perdebatan yang sangat hebat antara Pearson dan Yule. Saking hebatnya, Pearson sempat sangat emosional. Mereka memperdebatkan derajat kebebasan (*degree of freedom*) statistik penguji kebebasan dua variabel kategorik berbasis tabel kontingensi berukuran r baris dan c kolom. Sejak awal Pearson menyatakan bahwa derajat kebebasannya adalah $(rc - 1)$. Yule memberikan konjektur bahwa derajat kebebasan itu adalah $(r - 1)(c - 1)$. Konjektur inilah yang mengundang

‘kemarahan’ Pearson. Dunia menyaksikan; perlu waktu sekitar satu dekade sebelum ada penurunan matematikal yang membuktikan kebenaran konjektur Yule dan mengakhiri perdebatan tersebut.

Perdebatan semacam itu adalah bukti kebenaran filsafat Karl Popper tentang kegiatan ilmiah dan filsafat Rene Descartes tentang alasan keberadaan (*raison d’etre*). Popper menegaskan bahwa kegiatan ilmiah terdiri atas kegiatan verifikasi dan proses pemaknaan (*hermeneutics*). Verifikasi adalah kata lain dari bukti (*proof*). Namun, dalam sains, yang dimaksud adalah bukti matematikal (*mathematical proof*) seperti kata Paul Valery – sastrawan dan budayawan ternama Perancis seangkatan Hadamard dan Einstein: “*Aucune investigation humaine ne se peut appeler vraie science si elle ne passe pas par demonstration mathematique* (Tidak ada satupun penyelidikan manusia yang boleh disebut sains bila tidak melalui pembuktian matematikal)” atau seperti kata Plato: “... *that the reality which scientific thought is seeking must be expressible in mathematical terms, mathematics being the most precise and definite kind of thinking of which we are capable*” Bukti matematikal adalah *raison d’etre* yang merupakan spirit pengembangan sains. Selanjutnya, tidak akan ada pengembangan sains tanpa memiliki kepekaan rasa tidak puas / kecewa (*dissatisfied / disappointment*) terhadap teori atau metode yang sudah ada. Dalam hal ini, lagi-lagi manusia lari mencari solusi di dunia matematikal. Alfred Whitehead mengatakan: “*The study of*

mathematics is apt to commence in disappointment”

Tahun 1997 Kuwahara menyebutkan, dalam jurnal *BioPharm: The Applied Technologies of Biopharmaceutical Development*, bahwa era statistik modern dimulai pertengahan abad ke-19 tatkala Pierce mengembangkan statistik untuk memisahkan data *outlier* dari kelompok besar data lainnya. Siapa Pierce? Ia adalah astronom Perancis yang merekam data benda-benda angkasa dan tertantang untuk memisahkan *anomaly/outlier data* secara ilmiah. Mulai saat itu telah berjuta karya dalam bidang statistik yang dimaksudkan untuk lebih memahami data *outlier*. Sejak saat itu pula ilmu statistik terus berkembang secara eksponensial dalam ragam dan jumlahnya. Namun, itu bukan tanpa tantangan. Sebagai contoh, dalam satu dekade terakhir, homo sapiens berusaha keras agar mampu mengurutkan data yang berupa vektor di ruang berdimensi lebih dari satu. Kemampuan ini diperlukan untuk menyusun teori dan metode penaksiran *robust*, umpamanya, bagi lokasi multivariat dan dispersi multivariat. Contoh lain, umpamanya, hingga kini intelek manusia masih belum mampu memahami dengan baik dispersi multivariat.

Masih banyak masalah statistik yang *unsolvable* secara analitis. Kesulitan terletak pada kompleksitas matematikal yang sangat tinggi. Inilah tantangan dari segi sains dalam pengembangan ilmu statistik. Padahal, dalam budaya intelektual, tingkat peradaban suatu bangsa diukur oleh tingkat kompleksitas cara berpikir dan

tingkat kompleksitas produk berpikir. Sayangnya, paradigma ini belum tumbuh di masyarakat kita. Bahkan di masyarakat kita tumbuh pandangan bahwa yang namanya rumus statistikal atau rumus matematikal adalah sesuatu yang harus dihapal dan keterampilan menggunakannya dianggap sebagai suatu prestasi (*achievement*) penting. Padahal sebenarnya yang namanya rumus statistikal atau matematikal tidak lain merupakan wujud penampakan fenomena alam dalam bentuk objek matematikal yang berhasil diinterpretasikan oleh intelek manusia. Sedangkan dalam budaya intelektual (baca: Barat), seperti dikemukakan Xenophon, filosof Yunani kuno, “penampakan” atau “*appearance*” (*eidōs*) dan “wujud” (*morphe*) identik dengan “*beauty*” atau “indah” yang merupakan realitas ontologis. Oleh karena itulah, dalam budaya tersebut, matematika adalah bentuk paling sublim dari intelektualitas manusia mengingat matematika dapat menjadi penghubung dengan makna transenden. Tidaklah mengherankan bahwa dalam budaya intelektual, matematika adalah komponen utama dalam pengembangan kebudayaan dan peradaban.

Patut kita catat bahwa peradaban manusia memerlukan waktu sekitar dua abad untuk mampu memahami dengan baik perilaku distribusi normal. Rahasia distribusi normal terkuak berkat matematisi Perancis, Paul Levy, yang memperkenalkan konsep tentang fungsi karakteristik di tahun 1920 an. Waktu dua abad sebenarnya tidak lama bila dibandingkan dengan proses

pemahaman manusia terhadap luas lingkaran yang memerlukan waktu jauh lebih lama. Terhitung sejak Ahmos, ilmuwan Mesir kuno pada abad 18 sebelum Masehi – dengan tingkat kultur dan peradaban manusia saat itu – intelek manusia memerlukan waktu sekitar tiga puluh lima abad untuk mengetahui luas lingkaran secara eksak. Ini berkat kerja keras matematisi Perancis, Francois Viete, di abad ke-16. Jejak-jejak sejarah tersebut mengindikasikan akan pentingnya upaya menciptakan kultur dan membangun peradaban yang kondusif agar intelek manusia Indonesia berkembang dengan pesat.

Uraian di atas memperlihatkan bahwa, dalam budaya intelektual, statistik adalah salah satu indikator utama tingkat peradaban. Dengan demikian, pendidikan statistik sebenarnya adalah pendidikan kultural yang: (1) menjunjung tinggi nilai-nilai luhur kultural dan norma-norma serta etika kehidupan bermasyarakat, dan (2) mengutamakan kejujuran, transparansi, akuntabilitas, kualitas, dan produktivitas. Sejatinya, secara umum, pendidikan ditujukan untuk mempersiapkan manusia agar:

1. Tetap menempati posisi teratas di alam di antara semua makhluk Tuhan;
2. Mampu berinteraksi secara harmonis dengan alam.

dan sains telah terbukti mampu menghantarkan manusia meraih tujuan pertama.

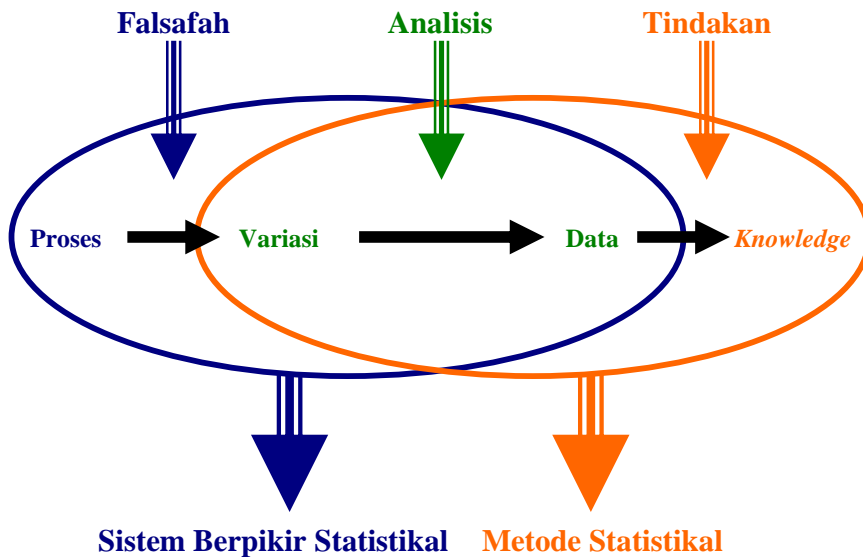
4. Teknologi baru pengontrolan variabilitas proses

Kegiatan perstatistikan menuntut sikap taat azas terhadap konsep ‘berpikir statistikal’ dan penciptaan / penerapan ‘metode statistikal’. Berpikir statistikal adalah suatu falsafah pembelajaran dan falsafah tindakan yang didasarkan pada prinsip-prinsip fundamental berikut:

1. Proses ada di mana-mana dan di setiap saat. Berbagai proses pada umumnya saling berkaitan;
2. Setiap proses menimbulkan variasi;
3. Memahami variasi dan upaya mereduksinya adalah kunci dalam peningkatan kualitas hidup.

Pada gambar 1 diberikan ilustrasi bagaimana konsep ‘berpikir statistikal’ dan penciptaan / penerapan ‘metode statistikal’ berinteraksi.

Metode statistikal yang diciptakan atau yang dipakai seseorang akan sangat tergantung kepada proses berpikir statistikal orang tersebut. Yang terakhir ini terkandung dalam tiga falsafah dasar tentang penanganan data – klasikal, Bayesian, dan analisis data – dan pada kemampuan hermeneutikal.



Gambar 1. Skema berpikir statistikal

Tahun 2005, salah satu produk statistikal yang penulis buat telah diterbitkan dalam *Journal of Quality Technology*, 37(1), 2005. Sejak saat itu, statistik tersebut telah banyak dipakai oleh berbagai industri di dunia. Statistik itu adalah teknologi baru dalam pengontrolan variabilitas proses multivariat (artinya, kualitas proses ditentukan oleh lebih dari satu variabel). *Soft technology* ini memberikan ilustrasi tentang keterkaitan antara konsep 'berpikir statistikal' dan penciptaan/penerapan 'metode statistikal' dalam upaya pengembangan ilmu statistikal.

Montgomery (1991, 1996, 2001, dan 2005) dalam bukunya yang terkenal dan menjadi pegangan para profesional bidang kualitas "*Introduction to Statistical Quality Control, Second – Fifth Editions*" mengemukakan tentang teknologi pengontrolan variabilitas multivariat berdasarkan beberapa buah sampel independen. Metodologinya

sama dengan metodologi pengujian hipotesis tentang homogenitas dispersi multivariat yang dilakukan dengan menguji satu per satu secara berulang. Dengan pendekatan ini, di bawah asumsi kenormalan multivariat, Montgomery menurunkan taksiran parameter distribusi asimtotis *sample generalized variance* dan sampai pada titik-titik kritis *lower control limit* (LCL) dan *upper control limit* (UCL). Sinyal yang menunjukkan bahwa proses berada dalam keadaan *out-of-control* akan muncul pada suatu sampel bila nilai *sample generalized variance* berada di luar interval (LCL, UCL).

Mengamati statistik Montgomery tersebut, timbul perasaan tidak takala penulis menelusuri taksiran parameter distribusi asimtotis *sample generalized variance* yang diturunkan dalam keempat edisi bukunya itu. Kelas distribusinya dapat penulis terima dengan sangat logis. Namun taksiran parameter-

parameternya terasa ganjil, rancu antara fenomena satu sampel dan banyak sampel independen, sehingga memberikan LCL dan UCL yang bias dan mengakibatkan *average run length* (ARL) yang besar tatkala proses berada dalam keadaan *out-of-control*. ARL adalah alat statistikal untuk mengukur kualitas alat penguji hipotesis di atas; nilainya akan besar bila proses dalam keadaan *in-control* dan akan kecil dalam keadaan *out-of-control*.

Ketidak-puasan tersebut semakin bertambah tatkala banyak literatur dan juga *web sites* yang menggunakan LCL dan UCL yang bias tersebut. Hal inilah yang menuntun penulis untuk menyusun pembuktian matematikal bahwa LCL dan UCL Montgomery bersifat bias dan selanjutnya menyajikan titik-titik kritis LCL dan UCL yang tak bias (*unbiased*). Oleh karena itu, teknologi pengontrolan yang dihasilkan bersifat tak bias dan memiliki ARL yang jauh lebih kecil ketimbang ARL teknologi Montgomery bila proses berada dalam keadaan *out-of-control*. Pada saat *in-control*, kedua teknologi itu memberikan ARL yang sama. Bagi yang berminat mempelajari pembuktian matematikal, silakan baca artikel penulis berjudul "*Improved Monitoring of Multivariate Process Variability*" dalam *Journal of Quality Technology*, 37(1), 2005.

5. Masalah terbuka (*open problems*)

Sulit sekali melaksanakan amanat Undang Undang yang berkaitan dengan pengembangan ilmu statistik tanpa mengembangkan riset

fundamental, tanpa kerjasama dengan berbagai pakar bidang sains fundamental, dan tanpa daya juang. Di bawah ini disajikan beberapa masalah aktual yang sangat potensial menjadi produk riset berkualitas tinggi.

1. Redefinisi ukuran dispersi multivariat.

Perhatikan masalah pengontrolan variabilitas proses multivariat di mana kualitas proses ditentukan oleh banyak variabel sehingga perhitungan *sample generalized variance* sulit dilakukan bahkan dengan menggunakan komputer secanggih apapun pada saat ini. Dalam hal seperti ini, maka diperlukan redefinisi ukuran dispersi multivariat.

2. Deteksi intrusi pada sistem komputer.

Informasi yang masuk ke dalam setiap sistem komputer secara alamiah berbentuk vektor berdimensi tinggi. Bagaimana mengontrol, hampir secara *real time*, bahwa informasi yang masuk bukan informasi yang membahayakan? Masalah ini membuka peluang untuk membuat statistik yang lebih baik.

3. Penaksiran *robust* bagi lokasi dan dispersi multivariat

Masalah penaksiran *robust* bagi parameter lokasi dan dispersi pada skema multivariat sangat dibutuhkan oleh industri, pemerintahan, bisnis, kesehatan, dan kegiatan-kegiatan kultural lainnya. Namun demikian, tatkala berhadapan dengan himpunan data

berdimensi tinggi, masalah tersebut masih sangat terbuka.

4. **Depth function** atau pengurutan data multivariat

Depth function adalah fungsi yang harganya menunjukkan seberapa dalam sebuah observasi berdimensi lebih dari satu berada relatif terhadap pusatnya. Fungsi ini dibangun dalam konteks non-parametrik untuk melakukan analisis data multivariat. Masalah utama terletak pada bagaimana mendefinisikan urutan dalam ruang vektor berdimensi lebih dari satu.

6. Bahan refleksi

Sebenarnya, Indonesia telah memiliki perangkat perundang-undangan yang baik. Namun, kita belum secara serius menanganinya. Ini tampak, misalnya, dari pelaksanaan Undang Undang No. 16 Tahun 1997 tentang Statistik yang belum optimal. Sebagai contoh, Pasal 1 Ayat 4 yang menyatakan :

”Kegiatan statistik adalah tindakan yang meliputi upaya penyediaan dan penyebarluasan data, upaya pengembangan ilmu statistik, dan upaya yang mengarah pada berkembangnya Sistem Statistik Nasional”.

belum secara serius dilaksanakan. Tampak nyata “**pengembangan ilmu statistik**” dalam pasal itu belum di tangani dengan serius oleh semua

lembaga yang terkait erat dengan kebijakan dan kegiatan perstatistikan nasional, termasuk perguruan tinggi. Apalagi dalam penciptaan alat-alat statistikal yang baru dan lebih baik.

Apa yang menghambat ? Penulis menduga, hambatannya antara lain terletak pada daya juang yang masih kurang. Daya juang para ilmuwan (calon ilmuwan) di tengah godaan faktor ekonomi. Ini menyangkut pilihan hidup seperti diungkapkan dengan bijak oleh St. Agustinus di abad ke satu Masehi, yang kemudian dikumandangkan lagi oleh Sartre, sebagai berikut: ”*La vie est une series des choix*” Hidup adalah untaian pilihan-pilihan – kata meraka. Di samping itu, perlu ada loncatan perubahan paradigma dalam belajar. Khusus kepada mahasiswa dan akademisi bidang statistik, penulis pesankan: ”berhentilah (hanya) belajar tentang statistik dan mulailah belajar memproduksi statistik”. Yang penulis maksudkan dengan statistik pada pesan tersebut adalah statistik dalam pengertian fundamental.