

Veri Mühendisliği...

Zeynep Kalaylioglu

İstatistik Topluluğu Presentasyonu,

21 Mart 2015

ODTÜ

Lütten



- **2014'ÜN EN İYİ TOP 3 MESLEĞİ (kaynak: careercast.com, Forbes)**

1. Matematikçi
2. Akademisyen
- 3. İstatistikçi**

***Kriterler:** Duygusal/fiziksel çalışma ortamı; gelir; kariyerde gelecek vaadetmesi; stres.*

- **Google Chief Economist:**

- «Ucuz ve her yerde olan şey ne? Veri. Nadiren bulunan ne? Verileri değerlendirecek analitik yetenek.»
- «Gelecek 10 yılın en çekici mesleği İstatistikçiler olacak»

- **İstatistiğin tanımları (A Career in Statistics – Beyond the numbers (2011) Gerald J. Hahn ve Necip Doğanaksoy – ODTÜ kütüphanesi online kaynak)**

- Veriden öğrenmenin bilimi
- Belirsizliğin bilimi



The theory of probabilities is at bottom nothing but common sense reduced to calculus.

Laplace, Théorie analytique des probabilités, 1820

Yakın zamanda bölümümüzde gerçekleştirilen katkılardan bir demet:

1. Daphnie pulex'in (su biti) tuzlu suda yaşayabilmesi için ortamda balık olması gerektiği gösterildi (biyolog)
2. Türk kadınları arasında, erken menstruasyona girmenin meme kanseri riskini artırdığı bulundu. Menopoz sonrası obezite ile meme kanseri arasında ilişki olduğu gösterildi (kanserepidemiyolog)
3. PINK1 adlı antijene karşı vücudun direnç sisteminin ürettiği antikorların Behcet Hastalığı ile ilişkili olduğu gösterildi (genetikçi)
4. 4. yy Roma sikkelerinin sahtelerinden ayırtedilebilmesi için kimyasal elementlere dayanan bir strateji geliştirildi. (arkeolog)

5. Kısa vade rüzgar yönü tahmin modeli (elektrik piyasasında kısa vade rüzgar enerji üretmini tahmin etmek için) geliştirildi (sürdürülebilir enerji)

6. Hücre yüzeyindeki reseptörden alınan sinyalleri hücre çekirdeğindeki DNA'ya taşıyan MAPK/ERK adlı protein zincirinin network hareketi tahmin edildi. (genetikçi)

7. Bireysel emeklilik müşterilerinin, başka bir sigorta şirketine transfer olma nedenleri ortaya kondu (sigortacı)

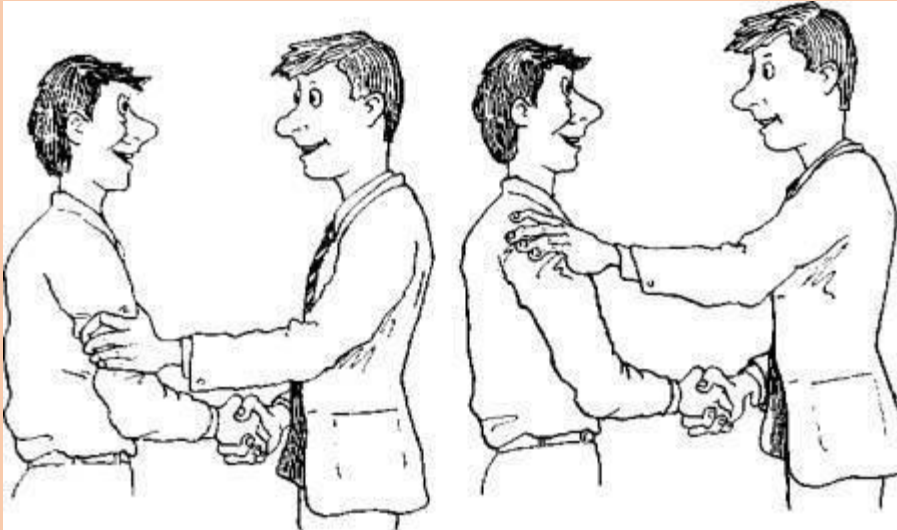
8. Uydu görüntüleri ile, belli bölgelerde, spesifik bitkilerin tespiti yapıldı (ziraatçı)

Görüldüğü gibi:

«*Best thing about being a statistician is that you get to play in everyone's backyard*»
(John Tukey: mathematician/statistician; bit; box-plot; software; EDA)

İstatistikçilerin en büyük avantajı

İstatistikçiler genelde diğer alandan birlikte çalıştıkları iş arkadaşları ile yaptıkları, AKIL YÜRÜTME GEREKTİREN, tartışmaları kazanırlar çünkü hem istatistik biliyorlar....
hem de sözkonusu alanı.



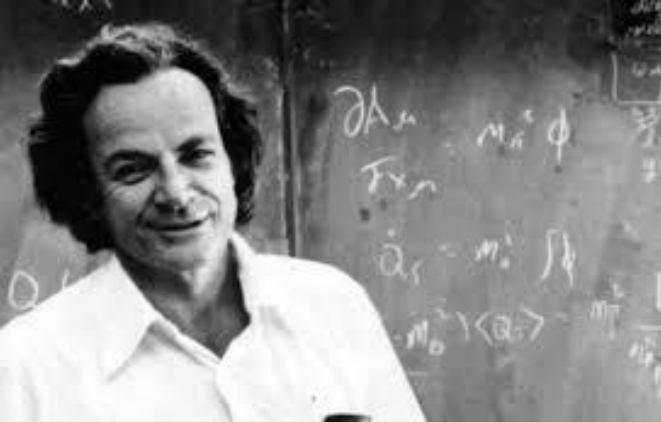
Yine de alçakgönüllülüğü elden bırakmayın

Bu bulgulara nasıl ulaşıldı??????????

- **Araştırma yöntemi:**

- Niteliksel araştırma (eğitim);
- Niceliksel araştırma (ekonomi;finans; işletme; biyoloji;sosyoloji;sağlık;ilaç;iklim;meteoroloji;uzay;atmosfer;sigorta; fizik (atom hareketleri))

- **Richard Feynman:**



«Bana tüm söyleyebileceğiniz geçmişle ilgili ise bilgi tam olarak değerli değil, yarın da ne olacağını söyleyebilmeniz gerekir»

İSTATİSTİKSEL TAHMİN

«Gözlem, gerçeğin yargııcıdır»

İSTATİSTİKSEL ÇIKARIM

«Az güvenilirlikle birşeyler söylemek, hiçbirşey söyleyememekten iyidir»

İSTATİSTİKSEL GÜVEN ARALIĞI

Bu bulgulara nasıl ulaşıldı??????????

- **Niceliksel araştırma yöntemi basamakları:**

Çalışılacak populasyon ve bu populasyondan örneklem seçmek

Verileri toplamak: Hangi veriler toplanmalı; Nasıl toplanmalı; Yöntem ne olmalı

Verilerin analizi: Analizden verinin taabi tutulması gereken önışlemler; uygun analiz yöntemleri; kullanılan temel matematiksel istatistik kuramları/yöntemleri; istatistiksel software

ÖRNEK: HIV’li ve enjekte ilaç kullanan insanlarda Hepatit C virüsü ve karaciğer rahatsızlığı

Konu: HIV, HTLV,HCV virüsü bulunan ve enjekte ilaç kullanan insanlarda karaciğer rahatsızlığından dolayı ölüm riski ve bu riskin bağlı olduğu etkenler.

Çalışılan nüfus: HIV/HTLV virüsü olan 7542 kişi.

Örneklem: (STAT 365 & 366)

7542 kişiden 84’ü karaciğer probleminden dolayı ölmüş durumda, hepsi örnekleme alınıyor.

Geri kalanların arasından hayatta olanlar: HIV/HTLV durumları (4 kategori); cinsiyet (2 kategori) ;ırk (2 kategori) a göre kategorilere ayrılıyor -> 16 kategori. Stratified sampling ile toplam 380 hayatta olan insan alınıyor.

ÖRNEK: HIV'li ve enjekte ilaç kullanan insanlarda Hepatit C virüsü ve karaciğer rahatsızlığı

Veriler: Enjekte ilaç kullanma süresi; yaş; HCM miktarı

İstatistiksel Analiz:

Betimsel istatistikler (STAT 155)

Kuvvet analizi (STAT 156)

Kategorik veri analizi (STAT 356 & 482)

ÖRNEK: HIV’li ve enjekte ilaç kullanan insanlarda Hepatit C virüsü ve karaciğer rahatsızlığı

Genel doğrusal modeller **(STAT 363 & 364)**

Lojistik regresyon **(STAT 462)**

Yaşamsal analiz **(STAT 567)**

Varyans tahminleri (Bootstrap) **(STAT 361)**

Matematiksel İstatistik **(STAT 271 & 272)**

Software **(STAT 291 & 292)**

SON SÖZ-1

Teknoloji ilerledikçe verilerin tam ve doğru şekilde ölçülmesi mümkün hale geldi ve gelişmeye devam ediyor.

Bütün veriler ilk bakışta kompleks (karmaşık) görünürken, keşfedilmeyi bekleyen önemli bilgiler içeriyorlar. Ve bu bilgiler keşfedildikleri takdirde *(araç: istatistik-olabilirlik fonksiyonu)* yaşamı daha iyi anlamamıza daha «iyi» sürmemize yolaçacak (sağlıklı, doğayla barışık)

Ve bütün bu resimde İstatistikçilerin oynayacağı rol ÇOK önemli.

SON SÖZ-2

Tukey'nin istatistik öğrencilerine önerileri

- Teoriyi iyi öğrenin çünkü teori her zaman eninde sonunda uygulama haline gelir.
- Matematik öğrenin. Çünkü diğer **tüm** bilimlerin en temel kısmı her zaman **matematiktir**. (*Yani matematik kafalı birisi, sosyolojiyi de anlar, hukuku da anlar, tarihi, felsefeyi, arkeolojiyi, biyolojiyi vs. anlar*)
- Unutmayın, **bütün** disiplinlerin en özgürü istatistikçilerdir, hangi alanda çalışmak istiyorlarsa orda çalışırlar, bunun bir avantaj olduğunun farkında olun.

Dinlediğiniz için teşekkürler

