Hotelling, and the teaching of statistics

Hotelling Lecture #1
March 26, 2007
Hotelling: some dates

• 1895 b. MN, 1904 moved to WA
• 1915-19 BA in Journalism, UW Seattle (+Army)
• 1920-21 MS in Math, UW Seattle
• 1924 PhD in Math, Princeton
• 1924-31 Food Research Institute, Math dept, Stanford
• 1931-46 Columbia, Economics (inc. the SRG)
• 1946-73 UNC, Statistics
• 1973 d. Chapel Hill after a 1972 stroke
Hotelling: life and work

Harold Hotelling 1895-1973

Harold Hotelling 1895-1973


Hotelling: the man

- “His relationship with his (and Wald’s) students were extraordinary” (K Arrow)
- “He helped others finish their books, but never took time to finish his own” (WE Deming)
- Friedman met 4 people in his life that he would label a genius: RA Fisher, JW Tukey and LJ Savage and Hotelling (S Stigler)
- “[H] wrote and spoke in verse - one could feel the punctuation…” (I Olkin)
- “..a rare combination of logical ability and near photographic memory” (RA Bradley)
- “It is a pity [H] died in 1973 before being awarded the Nobel Prize he so richly deserved.” (PA Samuelson)
- Monopoly, WW2, NYC real estate,…
Hotelling and Fisher

- H. reviewed each of the first 7 editions of Fisher’s Statistical Methods for Research Workers for JASA, the first in 1927, and the first 2 editions of The Design of Experiments
- H. spent 6 months at Rothamsted, June-Dec, 1929
- F. clearly respected H’s mathematical skills, and wanted to write a book with him, on theory
Who among us has not published a “proof” where a double limit was either falsely or incompletely justified?
APPLICATIONS OF THE THEORY OF ERROR TO THE INTERPRETATION OF TRENDS

BY HOLBROOK WORKING AND HAROLD HOTELLING

J. Amer. Statist. Assoc. 24 (March Suppl.) 73-85 (1929)
with the calculated trend line as a diameter, and asserting that the probability of the true line cutting this hyperbola is less than \( P = e^{-\frac{\chi^2}{2}} \),\(^1\) the values obtained for the \( \chi^2 \) — distribution for \( n = 2 \).

---

\(^1\)This mode of speech is an ellipsis, unless one accepts inverse probability. What is meant is that if a certain line cutting the \( \chi^2 \) hyperbola is the true line, then the probability that the calculated trend line would be obtained by chance is less than \( P \). A similar elliptical interpretation might be given to many common statements involving probable errors.

This footnote, especially the last sentence, is in effect the invention of confidence intervals as we now know them.
THE GENERALIZATION OF STUDENT’S RATIO*

By

HAROLD HOTELLING

Ann. Math. Statist. 2 360-78 (1931)

Motivation: biology, physics, economics..
Aim: to find genes whose expression profiles differ between genotypes?

Extra green dots at day 7 from uninfected wt samples.

Yu Chuan Tai
biostat student
Mary Wildermuth
biologist
Notation and models

We denote by $X_{g,1}, \ldots, X_{g,n}$ the replicate random $k$-vectors representing the observed time series for a single gene. For our Example, $n = 4$ and $k = 6$, and the $X_{g,i,t}$ are differences of log intensities, i.e. log ratios.

Our underlying model is that these $X_{g,i}$ are i.i.d. $N(\mu_g, \Sigma_g)$, and we make different assumptions about $\mu_g$ and $\Sigma_g$.

With our Example, we are interested in testing the null hypothesis $H: \mu_g = 0, \Sigma > 0$, against the alternative $K: \mu_g \neq 0, \Sigma_g > 0$. 
For our empirical Bayes approach, we have priors for $\mu_g$ and $\Sigma_g$ reflecting the indicator status $I = I_g$ of the gene, where $I_g = 1$ if $\mu_g \neq 0$, and $I_g = 0$ if $\mu_g = 0$.

We suppose that $Pr(I_g = 1) = p$, independently for every gene, for a hyperparameter $p$, $0 < p < 1$.

From now on, we drop the subscripts $g$ wherever possible.
With this background, our prior for $\Sigma$ is inverse Wishart with degrees of freedom $\nu$ and matrix parameter $(\nu \Lambda)^{-1}$, where $\Lambda > 0$ is positive definite. Our prior for $\mu$ is given by:

$$\mu \mid \Sigma, I = 1 \sim N(0, \eta^{-1} \Sigma), \eta > 0,$$  
$$\mu \mid \Sigma, I = 0 \sim N(0,0).$$

Finally, the data $X_1, \ldots, X_n$ are supposed i.i.d. given $I, \Sigma$ and $\mu$, with $X_i \mid I, \Sigma, \mu \sim N(\mu, \Sigma)$. 
Some distributional facts

The marginal distribution of the sample covariance matrix $S$ is Wishart, while that of $\bar{X}$ given $I=1$ is a multivariate $t$, and the distribution of $S$ given $I=1$ is a matrix generalized type II beta distribution, also called a Siegel distribution.

The joint distribution of $(\bar{X}, S)$ given $I=1$ is called a Student-Siegel, and is a little complicated. The same forms arise in the case $I=0$, but with different parameters.

These facts can be combined to obtain the simple formula on the next page for the posterior odds against a gene having zero mean vector.
Summary of results

Define the moderated $S$ by

$$\tilde{S} = \left[ E(\Sigma^{-1} \mid S) \right]^{-1} = \frac{(n - 1)S + \nu \Lambda}{n - 1 + \nu},$$

and the moderated $t$-statistic by

$$\tilde{t} = n^{1/2} \tilde{S}^{-1/2} \tilde{X}.$$  

Then the posterior odds against $I=0$ is an increasing function of

$$\tilde{T}^2 = \tilde{t}' \tilde{t}.$$
ANALYSIS OF A COMPLEX OF STATISTICAL VARIABLES INTO PRINCIPAL COMPONENTS

HAROLD HOTELLING

Columbia University

J. Educ. Psychol. 24 417-441, 498-520 (1933)

Analysis of variance, spectral analysis, …..
# RELATIONS BETWEEN TWO SETS OF VARIATES*

**By HAROLD HOTELLING, Columbia University.**

## CONTENTS.

<table>
<thead>
<tr>
<th>Sect.</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Correlation of Vectors. The Most Predictable Criterion and the Tetrad Difference</td>
<td>321</td>
</tr>
<tr>
<td>2. Theorems on Determinants and Matrices</td>
<td>325</td>
</tr>
<tr>
<td>3. Canonical Variates and Canonical Correlations. Applications to Algebra and Geometry</td>
<td>326</td>
</tr>
<tr>
<td>4. Vector Correlation and Alienation Coefficients</td>
<td>332</td>
</tr>
<tr>
<td>5. Standard Errors</td>
<td>336</td>
</tr>
<tr>
<td>6. Examples, and an Iterative Method of Solution</td>
<td>342</td>
</tr>
<tr>
<td>7. The Vector Correlation as a Product of Correlations or of Cosines</td>
<td>349</td>
</tr>
<tr>
<td>8. An Exact Sampling Distribution of ( q )</td>
<td>352</td>
</tr>
<tr>
<td>9. Moments of ( q ). The Distribution for Large Samples</td>
<td>354</td>
</tr>
<tr>
<td>10. The Distribution for Small Samples. Form of the Frequency Curve</td>
<td>359</td>
</tr>
<tr>
<td>11. Tests for Complete Independence</td>
<td>362</td>
</tr>
<tr>
<td>12. Alternants of a Plane and of a Sample</td>
<td>365</td>
</tr>
<tr>
<td>13. The Bivariate Distribution for Complete Independence ((s=t=2, n=4))</td>
<td>369</td>
</tr>
<tr>
<td>14. Theorem on Circularly Distributed Variates</td>
<td>371</td>
</tr>
<tr>
<td>15. Generalization of Section 13 for Samples of Any Size</td>
<td>372</td>
</tr>
<tr>
<td>16. Further Problems</td>
<td>375</td>
</tr>
</tbody>
</table>

*Biometrika* **28** 321-77 (1936)

Two subspace of a Hilbert space: design, spectral analysis,..
TUBES AND SPHERES IN $n$-SPACES, AND A CLASS OF STATISTICAL PROBLEMS.*

By Harold Hotelling.

Amer. J. Math. 61 440-60 (1939)
The Statistical Research Group 1942-45 (see W. Allen Wallis, JASA 75 1980)

Based at Columbia University during WW2, supported by the Applied Mathematics Panel of the National Defence Research Committee, part of the Office of Scientific Research and Development.

H. was the official PI of this project, but Wallis was Director of Research, later M. Friedman Deputy Director and A. Bowker Assistant Director.
The 18 principal members and their stays

H. part-time: “in evidence for a short time one day a week”.

Wald, Freeman also part-time.

Allen Wallis, 45 months
Harold Hotelling, 39 months
Jacob Wolfowitz, 39 months
Edward Paulson, 35 months
Julian Bigelow, 31 months
Milton Friedman, 31 months
Abraham Wald, 30 months
Albert Bowker, 27 months
Harold Freeman, 21 months
Rollin Bennett, 20 months
Jimmie Savage, 19 months
Kenneth Arnold, 18 months
Millard Hastay, 18 months
Abraham Girshick, 17 months
Frederick Mosteller, 12 months
Churchill Eisenhart, 11 months
Herbert Solomon, 11 months
George Stigler, 10 months
Hotelling on the teaching of statistics


This was based on an address to a meeting of the IMS at Hanover, NH. At the time H. was chair of the IMS committee on the Teaching of Statistics (HH, JN, WED, and BC), and in essence this was its first position paper.

Neyman was very affected by the talk. He went back to Berkeley and immediately started writing his First course in probability and statistics, 1950.
Other works by H. on this topic


The place of statistics in the university.

THE PLACE OF STATISTICS IN THE UNIVERSITY

Contents

A. Minor nuisances and inefficiencies in statistical teaching
   6. Lack of coordination among departments. Lack of advanced courses and laboratory facilities
   7. Inefficient decentralization of libraries

B. The major evil: failure to recognize the statistical method as a science, requiring specialists to teach it
   8. Too many teachers not specialists
   9. Results: students ill equipped
  10. Reasons why teachers of statistics are often not specialists
      a. The rapid growth of the subject
      b. Confusion between the statistical method and applied statistics
      c. Failure to recognize the need for continuing research
      d. The system of making appointments to teach statistics within particular departments that are devoted primarily to other subjects

11. Appointments under the existing system are not all bad
12. Unsatisfactory texts
13. Omission of probability theory from texts and teaching
C. Proper qualifications of teachers of statistics
   14. Statistics compared with other subjects
   15. Current research in the statistical method is essential for teachers
   16. Minimum requirements in mathematics for the training of teachers and research men in statistical theory

D. Need for relating theory with applied statistics
   17. An example of the interaction between theory and practice
   18. Supplying opportunities for application in graduate studies of statistics

E. Recommendations on the organization of statistical teaching and research in institutions of higher learning
   19. Research should be encouraged; teaching schedules should not be overloaded
   20. Organization of statistical service in the university
   21. Organization for teaching
   22. The statistical curriculum
   23. Statistical method as part of a liberal education
Two of many things which resonate

“In addition to the pure mathematics and the knowledge of statistical theory, a competent statistician or teacher of statistics needs a really intimate acquaintance with the problems of one or more empirical subjects in which statistical methods are applied. This is quite important.

……..

A specialist in statistics on a university faculty has a threefold function. In addition to the usual duties of teaching and research, there is a need for him to advise his colleagues, and other research workers, regarding the statistical methods appropriate…….The statistician is then very likely to find himself embarked on a co-operative research venture in a field new to him.”
Two of the few things which don’t resonate

“The work of a Department of Statistics should be concerned largely with sampling theory, and should emphasize the unity of statistical methods and theory, regardless of the field of application.”

“Without probability theory, statistical methods are of minor value, for although they may put data into forms from which intuitive inferences are easy, such inferences are very likely to be incorrect.”
Some comments from Deming (1940)

“I take it that they [Hotelling’s recommendations] are not supposed to embody all that there is to the teaching of statistics, because there are many other neglected phases that ought to be stressed.

Most of [the Bureau of Census clerks] are shocked to learn that many of the so-called “modern theories of estimation” are not theories of estimation at all, but rather theories of distribution and are a disappointment to one who is faced with the necessity of making a prediction from his data, i.e. of basing some critical course of action on them…..

histogram…scatter diagram…heterogeneity…randomness

…a statistician must be a scientist……a scientist does not neglect any pertinent information, yet students of statistics are often taught to do just the opposite of this…”
My own views (in summary)

Questions, data and statistics.

Why do we need it?
Where does it come from?
What can we now do that we couldn’t do before?

Practice before theory!
Postscript (-20 years)

The two papers on the teaching of statistics that I have so briefly discussed were reprinted in *Statistical Science* vol 3, #1 (1988) as Golden Oldies. Then followed the biographical note by Adrian C Darnell already mentioned, and after that Comments by David S Moore, James V. Zidek, Kenneth J Arrow, Harold Hotelling Jr, Ralph A Bradley, W. Edwards Deming, Shanti S Gupta, and Ingram Olkin.

Let’s hope this important discussion will continue for many years to come!