STA 22101S: Applied Statistics II
Fridays, 2-5 pm, SS 1083 Spring, 2014

Course description: This course teaches methods of applied statistics, with the applications studied motivating the sets of methods taught. The topics covered will include

- planning of studies
- generalized linear models
- semi-parametric regression
- generalized estimating equations
- mixed linear and non-linear models

Grading: The grade in the course will be based on regular homework (60%) and a final exam (40%).

Text: The course text is *Statistical Models* by A. C. Davison (Cambridge University Press), Chapters 8 through 10. Highly recommended is *Principles of Applied Statistics* by D.R. Cox and C.A. Donnelly (CUP).
Additional resources will be provided as needed; I often refer to the 4th addition of *Modern Applied Statistics with* `S` by W.N. Venables and B.D. Ripley (Springer), *Applied Statistics* by D.R. Cox and E.J. Snell (Chapman & Hall) and *Elements of Statistical Learning*, by T. Hastie, R. Tibshirani, and J. Friedman (Springer).

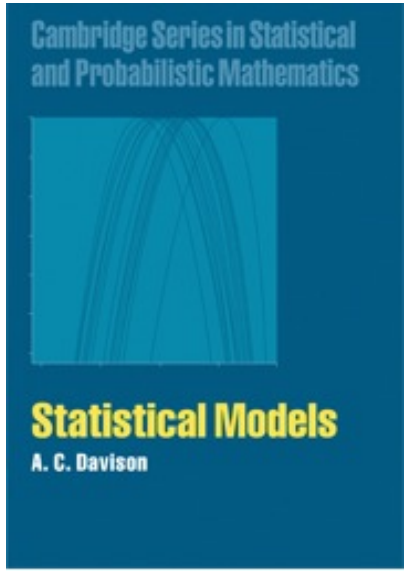
Course web page(s): I am using Blackboard to manage the course list and grades, but the course information is all on the web page
`http://www.utstat.utoronto.ca/reid/2201S14.html`. The Blackboard page for STA2201S will lead you to this page via the first announcement.

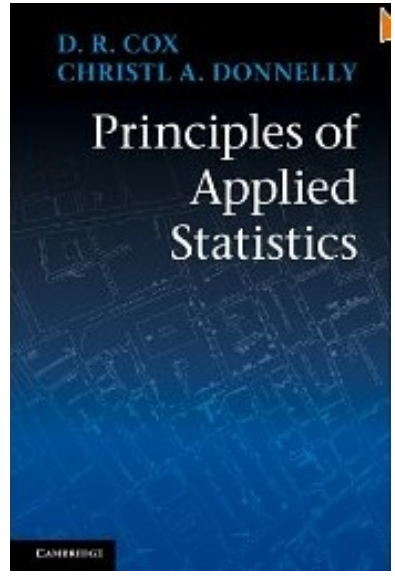
Computing: You are welcome to use the statistical computing package of your choice, but I will refer exclusively to the `R` computing package. There are some `R` resources listed on the course webpage.

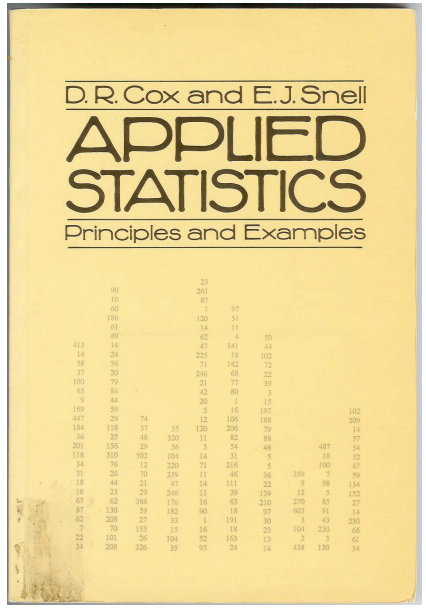
Contact: Nancy Reid: SS 6002A, `reid@utstat.utoronto.ca`, 978-5046.
Office Hours: Thursday 1 to 3, or by appointment.

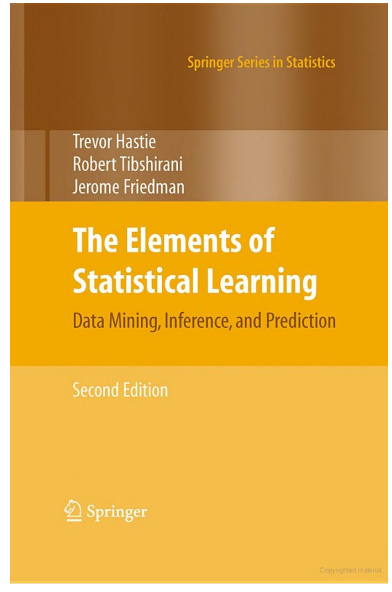
Topics in more detail

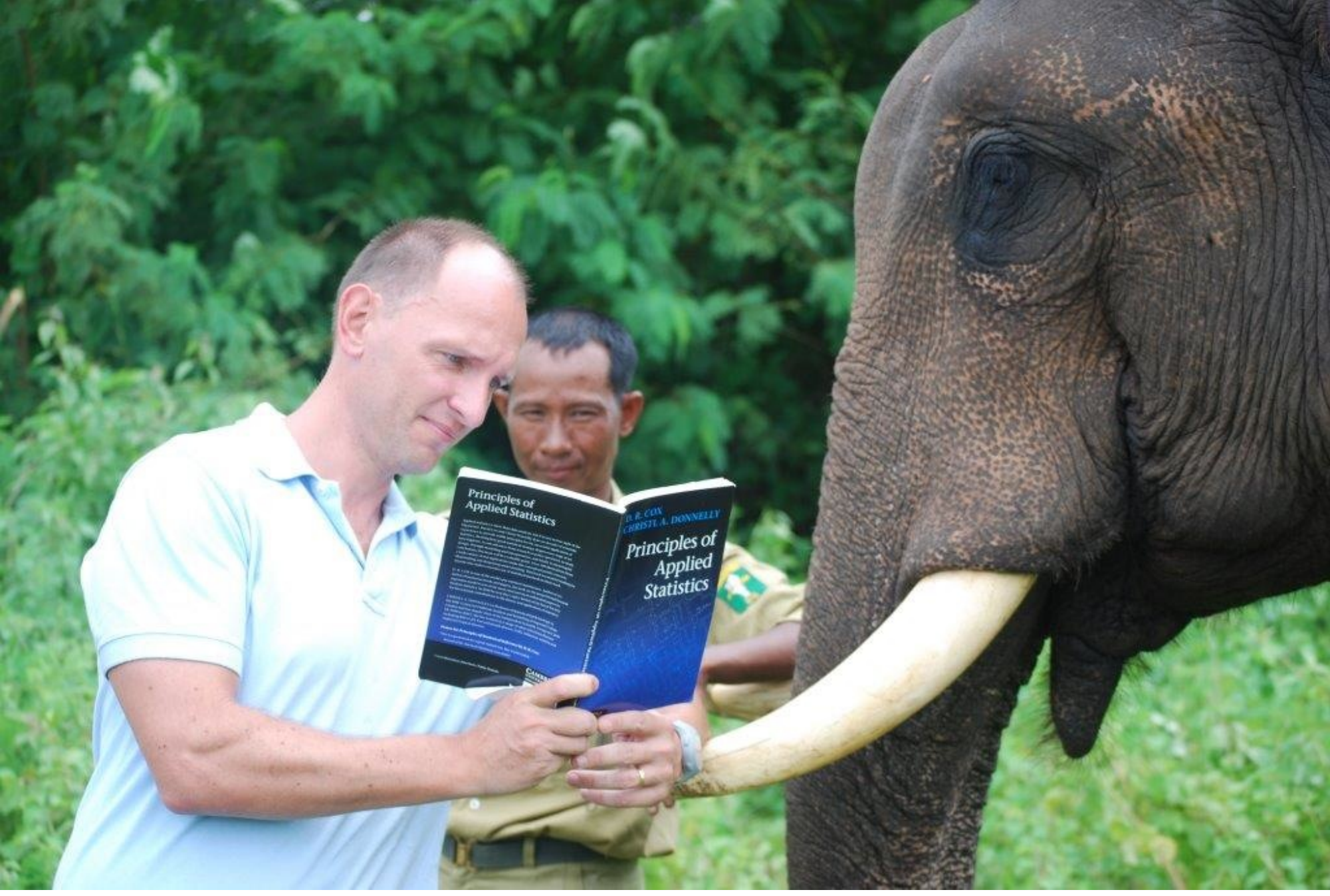
- planning of studies: randomization, causality, missing data, types of studies SM 9.1, 5.5, CD
- generalized linear models: `glm`, analysis of deviance, nonlinear least squares, log-linear models, logistic regression, model choice, generalized linear mixed models, quasi-likelihood SM 10.1-6, CD 7
- semi-parametric regression: local polynomial fitting, cross-validation, mean-variance trade-off, smoothing splines, penalized methods SM 10.7, HTF 3.4, 5.1-6
- longitudinal data, generalized estimating equations Liang & Zeger, 1986
- survival data and proportional hazards model SM 10.8


Cambridge Series in Statistical and Probabilistic Mathematics
Statistical Models
A. C. Davison


D. R. COX
CHRISTL A. DONNELLY
Principles of Applied Statistics


D. R. Cox and E. J. Snell
APPLIED STATISTICS
Principles and Examples


Springer Series in Statistics
Trevor Hastie
Robert Tibshirani
Jerome Friedman
The Elements of Statistical Learning
Data Mining, Inference, and Prediction
Second Edition
Springer


Principles of
Applied Statistics
D. R. COX
CHRISTL A. DONNELLY
Principles of
Applied Statistics

... topics

STA2201H Methods of Applied Statistics II

The course will focus on generalized linear models (GLM) and related methods, such as generalized additive model involving nonparametric regression, generalized estimating equations (GEE) and generalized linear mixed models (GLMM) for longitudinal data. This course is designed for Master and PhD students in Statistics, and is REQUIRED for the Applied paper of the PhD Comprehensive Exams in Statistics. We deal with a class of statistical models that generalizes classical linear models to include many other models that have been found useful in statistical analysis, especially in biomedical applications. The course is a mixture of theory and applications and includes computer projects featuring R (S+) or/and SAS programming.

Topics: Brief review of likelihood theory, fundamental theory of generalized linear models, iterated weighted least squares, binary data and logistic regression, epidemiological study designs, counts data and log-linear models, models with constant coefficient of variation, quasi-likelihood, generalized additive models involving nonparametric smoothing, generalized estimating equations (GEE) and generalized linear mixed models (GLMM) for longitudinal data.

- GLM, GAM, GEE, GLMM

- methods describe how to fit the models
- understanding the methods indicates what information is available, when they are useful

... topics

STA2201H Methods of Applied Statistics II

The course will focus on generalized linear models (GLM) and related methods, such as generalized additive model involving nonparametric regression, generalized estimating equations (GEE) and generalized linear mixed models (GLMM) for longitudinal data. This course is designed for Master and PhD students in Statistics, and is REQUIRED for the Applied paper of the PhD Comprehensive Exams in Statistics. We deal with a class of statistical models that generalizes classical linear models to include many other models that have been found useful in statistical analysis, especially in biomedical applications. The course is a mixture of theory and applications and includes computer projects featuring R (S+) or/and SAS programming.

Topics: Brief review of likelihood theory, fundamental theory of generalized linear models, iterated weighted least squares, binary data and logistic regression, epidemiological study designs, counts data and log-linear models, models with constant coefficient of variation, quasi-likelihood, generalized additive models involving nonparametric smoothing, generalized estimating equations (GEE) and generalized linear mixed models (GLMM) for longitudinal data.

- where do the models come from?
- how are they to be interpreted?
- where is the data? how was it collected?

- principles of applied statistics

... topics

STA2201H Methods of Applied Statistics II

The course will focus on generalized linear models (GLM) and related methods, such as generalized additive model involving nonparametric regression, generalized estimating equations (GEE) and generalized linear mixed models (GLMM) for longitudinal data. This course is designed for Master and PhD students in Statistics, and is REQUIRED for the Applied paper of the PhD Comprehensive Exams in Statistics. We deal with a class of statistical models that generalizes classical linear models to include many other models that have been found useful in statistical analysis, especially in biomedical applications. The course is a mixture of theory and applications and includes computer projects featuring R (S+) or/and SAS programming.

Topics: Brief review of likelihood theory, fundamental theory of generalized linear models, iterated weighted least squares, binary data and logistic regression, epidemiological study designs, counts data and log-linear models, models with constant coefficient of variation, quasi-likelihood, generalized additive models involving nonparametric smoothing, generalized estimating equations (GEE) and generalized linear mixed models (GLMM) for longitudinal data.

- how does this advance science
- case studies, examples, items from the news, ...

Preliminaries CD Ch. 1

- we need statistics when we have “unexplained and haphazard variation”
- distinguish between natural variability and measurement error
- one is of interest, the other needs to be accommodated
- example: blood pressure varies over time scales of minutes, hours, days, even in healthy individuals.

  Measurements of blood pressure are also imprecise, but this variability is not of especial interest, although we need to be aware of it

Components of investigation CD Ch. 1

- “the ideal sequence”
  - formulation of research questions
  - search for relevant data
  - design and implementation of investigations to obtain data
  - analysis of data
  - interpretation of the results
- “The essence of our discussion will be on the achievement of individually secure investigations. These are studies which lead to unambiguous conclusions ... Yet virtually all subject-matter issues are tackled sequentially ... typically the important and challenging issue of synthesizing information of very different kinds, so crucial for understanding, has to be carried out informally”
- Examples: Northern Hemisphere temperature time series; investigations of bovine tuberculosis; evidence for HIV as the cause of AIDS

... components of investigation CD Ch. 1

- very focused research question – ideal
- research questions emerge as the study develops – “consequent reformulation of the detailed statistical model used for analysis... usually causes no conceptual problem ... Major changes of focus... ideally need confirmation in supplementary investigations”
- “An extreme case of departure from the ideal sequence ... a large body of administrative data become available, and there is a perception that it must contain interesting information about something... the term ‘data mining’ is often used in such context... how much effort should be spend on such issues beyond the simple tabulation of frequencies and pairwise dependencies must depend in part on the quality of the data ... any conclusions are in most cases likely to be tentative and in need of independent confirmation”
- “ A large amount of data is in no way synonymous with a large amount of information”

Design and Analysis CD Ch. 1

- choice of material/individuals to study – "units of analysis"
- "For studies of a new phenomenon it will usually be best to examine situations in which the phenomenon is likely to appear in the most striking form, even if this is in some sense artificial"
- statistical analysis needs to take account of the design (even if statistician enters the project at the analysis stage)
- need to be clear at the design stage about broad features of the statistical analysis – more publicly convincing **and** "reduces the possibility that the data cannot be satisfactorily analysed"
- example: Female faculty salary survey
- "it is unrealistic and indeed potentially dangerous to follow an initial plan unswervingly ... it may be a crucial part of the analysis to clarify the research objectives"

Experimental and observational studies CD Ch. 1

- experiment is a study in which all key elements are under the control of the investigator
- in an observational study key elements cannot be manipulated by the investigator.
- “It often, however, aids the interpretation of an observation study to consider the question: what would have been done in a comparable experiment?”
- Example: hormone replacement therapy and heart disease
- observational study – strong and statistically significant reduction in heart disease among women taking hormone replacement therapy
- women’s health study (JAMA, 2002, p.321) – statistically significant increase in risk among women randomized to hormone replacement therapy

Principles of measurement CD Ch. 1

- “construct validity – measurements do actually record the features of concern”
- “record a number of different features sufficient to capture concisely the important aspects”
- reliable – i.e. reasonably reproducible
- “cost of the measurements is commensurate with their importance”
- “measurement process does not appreciably distort the system under study”

Types and phases of analysis CD Ch. 1

- "A general principle, sounding superficial but difficult to implement, is that analyses should be as simple as possible, but no simpler."
- the method of analysis should be transparent
- main phases of analysis
  - data auditing and screening;
  - preliminary analysis;
  - formal analysis;
  - presentation of conclusions

“What are the principles of applied statistics?” CD Ch. 1

- “formulation and clarification of focused research questions of subject-matter importance
- design of individual investigations and sequences of investigations that produce secure answers and open up new possibilities
- production of effective and reliable measurement procedures
- development of simple, and where appropriate, not-so-simple methods of analysis, with suitable software, that address the primary research questions, often through a skilful choice of statistical model, and give some assessment of uncertainty
- effective presentation of conclusions
- structuring of analyses to facilitate their interpretation in subject matter terms and their relationship to the knowledge base of the field.”

Role of randomization

SM Ch. 9

- context: units, treatments
- goal: assign treatments to units at random
- why?
  - avoid confounding (on average) bias, systematic error
  - avoid use of personal judgment in assignment of treatment
  - provide blinding of treatment assignment (placebo effect)
  - provides a basis for analysis
  - provides a basis for causal interpretation

418 9 · *Designed Experiments*

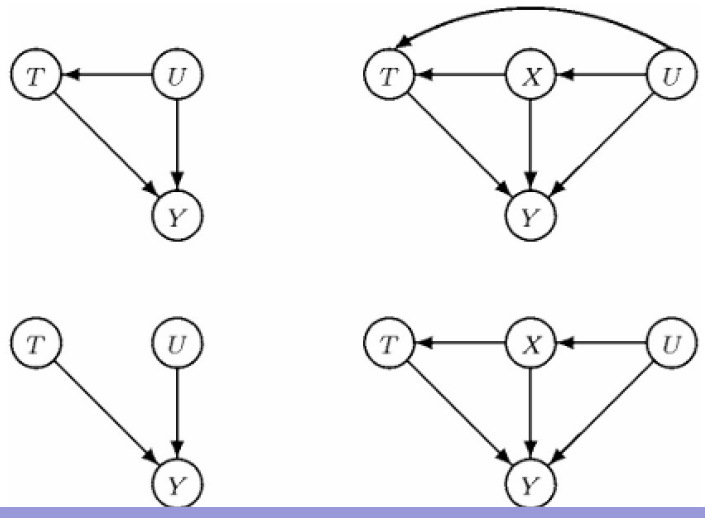


Figure 9.1 Directed acyclic graphs showing consequences of randomization. An arrow from T to Y indicates dependence of Y on T, and so forth. In general both response Y and treatment T may depend on properties U of units (upper left). Randomization (lower left) makes treatments and units independent, so any observed dependence of Y on T cannot be ascribed to joint dependence on U. The upper right graph shows the general dependence of Y, T, and covariates X on U. Randomization makes T

... randomization

- usually restricted in some way, often by blocking or grouping units to be more homogeneous
- analogous to stratified random sampling
- with two groups only, data is paired
- *t*-test based on differences with paired data more powerful than *t*-test based on original data
- two examples in SM: each subject 'partnered' case-control; each subject used as own control
- see also Cox & Snell Example E

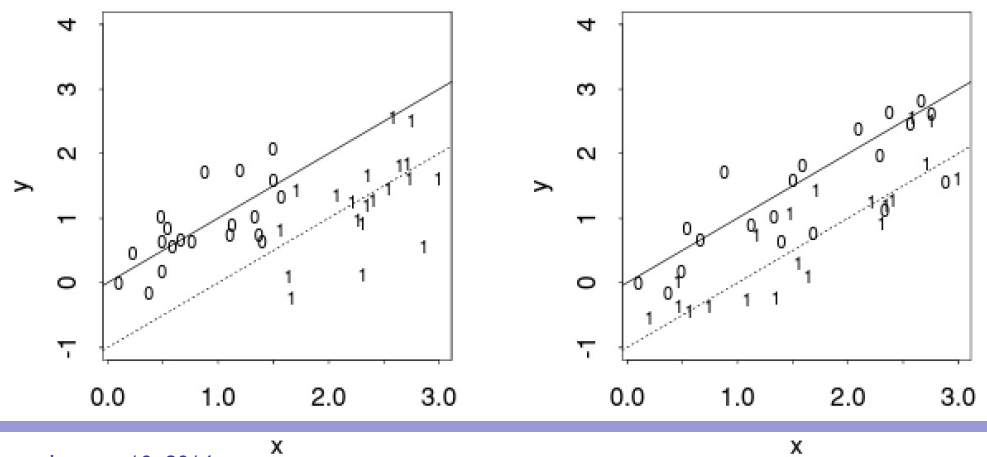


Figure 9.2 Simulated results from experiments to compare the effect of a treatment T on a response Y that varies with a covariate X. The lines show the mean response for $T = 0$ (solid) and $T = 1$ (dots). Left: the effect of T is confounded with dependence on X. Right: the experiment is balanced, with random allocation of T dependent on X.

eft indicates what *could* happen if T and X not independent (based on simulated data); right is ignored until Ex. 9.1 shows how blocking (pairing) using a covariate gives more efficient inference

if covariate not measured until after randomization, could be adjusted for in analysis

... randomization analysis

SM p.421, Ex.9.1

- Response of unit j is $\gamma_j + \delta$ if unit j is assigned T;
  $\gamma_j - \delta$ if unit j is assigned C
- $\gamma_1, \ldots, \gamma_{2m}$ are fixed constants
- theory (2-sample):

$$Y_j = \begin{cases} T_j(\gamma_j + \delta) & T_j = 1 \\ (1 - T_j)(\gamma_j - \delta) & T_j = 0 \end{cases}$$

$$T_j = \text{Bernoulli}(1/2), \quad \mathrm{E}\, T_j = \frac{1}{2}, \mathrm{E}(T_j T_k) = \frac{m-1}{2(2m-1)}$$

- Facts:

$$\begin{aligned} \mathrm{E}_R(\bar{Y}_1 - \bar{Y}_0) &= 2\delta \\ \mathrm{var}_R(\bar{Y}_1 - \bar{Y}_0) &= 2\sum_{j=1}^{2m}(\gamma_j - \bar{\gamma})^2/\{m(2m-1)\} \\ \mathrm{E}_R(\frac{2s^2}{m}) &= \mathrm{var}_R(\bar{Y}_1 - \bar{Y}_0) \end{aligned}$$

... randomization analysis

SM p.421, Ex.9.1

- Result:

$$\frac{\bar{Y}_1 - \bar{Y}_0 - 2\delta}{(2s^2)/m} \xrightarrow{\mathcal{L}} N(0,1),$$

- under the randomization distribution
- $$s^2 = \frac{\sum(Y_j T_j - \bar{Y}_1)^2 + \sum(Y_j(1-T_j) - \bar{Y}_0)^2\}}{2(m-1)}$$

  pooled estimate of variance
- this analysis on p.421 is for 2-sample test: $2m$ units; m allocated to T and m to C

- example 9.1 is for a paired test, based on differences
- no normal approximated needed; distribution of all possible 2^{10} configurations can be computed directly

... randomization analysis

- can be applied to more complex randomization schemes: RCB, LS, GLS, etc.
- "normal theory inference can be justifies as an approximation to the randomization distribution
- requires assumption of a non-testable assumption
- unit-treatment additivity
- stable unit-treatment-value assumption SUTVA
- There exists constants $\gamma_1, \ldots, \gamma_n$ and $\alpha_1, \ldots, \alpha_T$, s.t.
- $$Y_j = \gamma_j + \alpha_t, \quad \text{if unit } j \text{ assigned to treatment } t$$
- independently of treatment allocation to other units

Causality

- in what sense does randomization strength the argument for causality?
- physical: if we change x we will observe a related change in y
- observational: the effect of x on y cannot be explained by dependence on another variable
- experimental: randomization and unit-treatment additivity ensure that the sample mean difference estimates the theoretical difference, baed on counterfactuals

- exercise: Bradford-Hill's criteria for causality

Boy	Material A	Material B	Difference d
1	13.2 (L)	14.0 (R)	0.8
2	8.2 (L)	8.8 (R)	0.6
3	10.9 (R)	11.2 (L)	0.3
4	14.3 (L)	14.2 (R)	–0.1
5	10.7 (R)	11.8 (L)	1.1
6	6.6 (L)	6.4 (R)	–0.2
7	9.5 (L)	9.8 (R)	0.3
8	10.8 (L)	11.3 (R)	0.5
9	8.8 (R)	9.3 (L)	0.5
10	13.3 (L)	13.6 (R)	0.3

Table 9.1 Shoe wear data (Box *et al.*, 1978, p. 100). The table shows the amount of shoe wear in an paired comparison experiment in which two materials A and B were randomly assigned to the soles of the left (L) or right (R) shoe of each of ten boys.

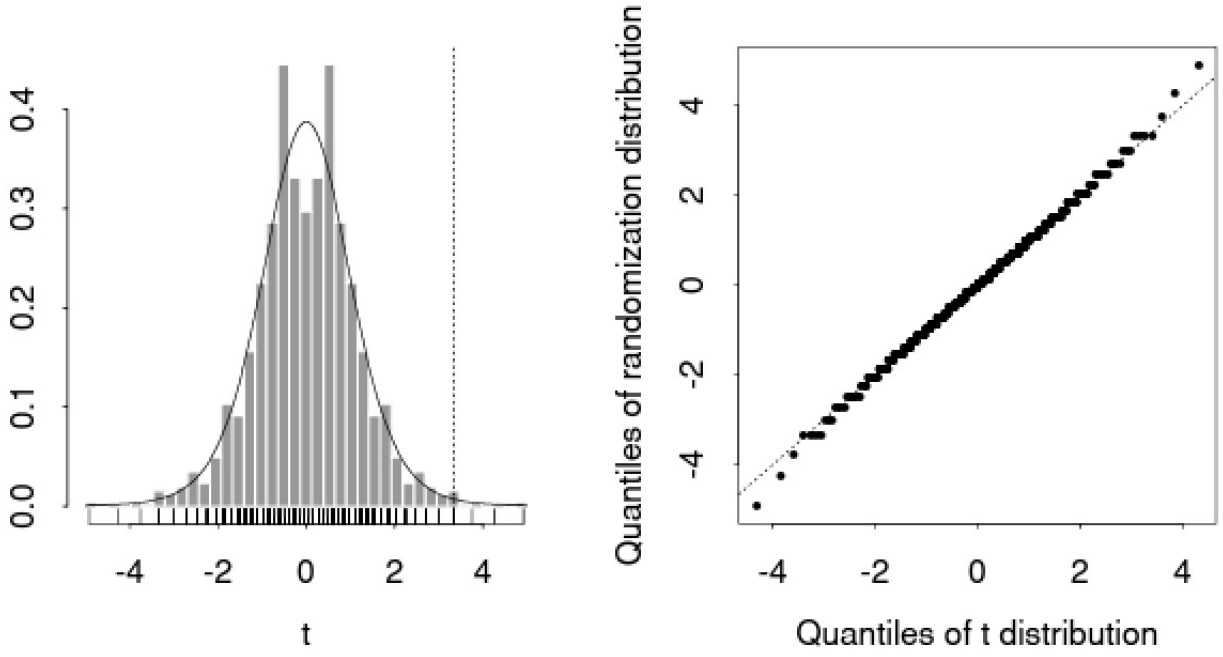


Figure 9.3 Randomization distribution of the t statistic for the shoes data, together with its approximating t_9 distribution. The left panel shows a histogram and rug for the randomized values of Z, with the t_9 density overlaid; the observed value is given by the vertical dotted line. The right panel shows a probability plot of the randomization distribution against t_9 quantiles.

Specialized designs

- completely randomized design (one-way layout)
- randomized block design (two-way layout) – note block differences not of interest
- incomplete block design – not enough units in each block to include each treatment
- Latin Square design – two blocking factors, rows and columns; each treatment appears exactly once in each row and column
- Graeco-Latin Square design – three blocking factors, or two treatment factors

$A\alpha$	$B\beta$	$C\gamma$	$D\delta$	$E\epsilon$
$B\gamma$	$C\delta$	$D\epsilon$	$E\alpha$	$A\beta$
$C\epsilon$	$D\alpha$	$E\beta$	$A\gamma$	$B\delta$
$D\beta$	$E\gamma$	$A\delta$	$B\epsilon$	$C\alpha$
$E\delta$	$A\epsilon$	$B\alpha$	$C\beta$	$D\gamma$

Table 9.10 Analysis of variance table for a Latin square.

Term	df	Sum of squares
Rows	$q-1$	$\sum_{r,c}(\overline{y}_{r\cdot}-\overline{y}_{\cdot\cdot})^2$
Columns	$q-1$	$\sum_{r,c}(\overline{y}_{\cdot c}-\overline{y}_{\cdot\cdot})^2$
Treatments	$q-1$	$\sum_{r,c}(\overline{y}_{t(r,c)}-\overline{y}_{\cdot\cdot})^2$
Residual	$(q-1)(q-2)$	$\sum_{r,c}(y_{rc}-\overline{y}_{r\cdot}-\overline{y}_{\cdot c}-\overline{y}_{t(r,c)}+2\overline{y}_{\cdot\cdot})^2$

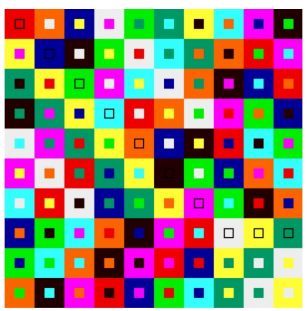
Table 9.11 Field concrete mixer data. Latin square experiment, showing application of treatments — speed in miles per hour (left) — and observed responses — machine efficiency (%) (right) — for 16 combinations of day and run. Below are average efficiencies for day, run, and speed.

	Run					Run			
Day	1	2	3	4	Day	1	2	3	4
1	8	16	4	12	1	64.2	59.8	66.2	63.6
2	16	12	8	4	2	47.5	57.3	67.7	58.6
3	4	8	12	16	3	54.2	59.9	57.1	54.1
4	12	4	16	8	4	60.1	68.4	58.7	63.7

Day	Average	Run	Average	Speed	Average
1	63.45	1	56.50	4	61.85
2	57.78	2	61.35	8	63.88
3	56.33	3	62.43	12	59.53
4	62.73	4	60.00	16	55.03

Aside: fun facts about GL squares

- if $n = p^m$, then there are $n - 1$ mutually orthogonal $n \times n$ Latin squares
- if $n = p_1^{m_1} p_2^{m_2} \ldots$, there are at least $\min(p_1^{m_1}, p_2^{m_2}, \ldots)$ mutually orthogonal $n \times n$ Latin squares
- if $n = 6$ there are none
- conjectured in 1782 that this is true for $n = 2 \mod 4$
- a pair of 10×10 orthogonal Latin squares was discovered in 1960
- it is unknown if there are more than 2
- there is no formula for the number of LS of size n, but it is large!: 2, 12, 576, 161280, 812851200



Picture

Factorial treatment structures

- complete factorial – e.g. 3×4; twelve tmt combinations $A_1B_1, \ldots, A_4B_4$
- two-level factorial – each tmt has high and low level; screening designs
- three-level factorial – response surface designs
- fractional factorials – e.g. $1/2$ fraction of a 2^5 factorial has 16 runs
- fractional factorials confounded in blocks – e.g. $1/4$ fraction of 2^8 run in 8 days, 8 runs/day
- and many more
- used in industrial experimentation, error-correcting codes, numerical integration

Table 9.10 Analysis of variance table for a Latin square.

Term	df	Sum of squares
Rows	$q-1$	$\sum_{r,c}(\overline{y}_{r\cdot}-\overline{y}_{\cdot\cdot})^2$
Columns	$q-1$	$\sum_{r,c}(\overline{y}_{\cdot c}-\overline{y}_{\cdot\cdot})^2$
Treatments	$q-1$	$\sum_{r,c}(\overline{y}_{t(r,c)}-\overline{y}_{\cdot\cdot})^2$
Residual	$(q-1)(q-2)$	$\sum_{r,c}(y_{rc}-\overline{y}_{r\cdot}-\overline{y}_{\cdot c}-\overline{y}_{t(r,c)}+2\overline{y}_{\cdot\cdot})^2$

Table 9.11 Field concrete mixer data. Latin square experiment, showing application of treatments — speed in miles per hour (left) — and observed responses — machine efficiency (%) (right) — for 16 combinations of day and run. Below are average efficiencies for day, run, and speed.

	Run					Run			
Day	1	2	3	4	Day	1	2	3	4
1	8	16	4	12	1	64.2	59.8	66.2	63.6
2	16	12	8	4	2	47.5	57.3	67.7	58.6
3	4	8	12	16	3	54.2	59.9	57.1	54.1
4	12	4	16	8	4	60.1	68.4	58.7	63.7

Day	Average	Run	Average	Speed	Average
1	63.45	1	56.50	4	61.85
2	57.78	2	61.35	8	63.88
3	56.33	3	62.43	12	59.53
4	62.73	4	60.00	16	55.03

Analysis

- balanced designs can be analyzed directly, with suitable constraints
- see Ex. 9.9, 9.6, §9.2.3, p. 430, etc.
- set of linearly estimable parameters are contrasts
- individual tests of factorial effects are independent, in the usual normal theory linear model
- total sums of squares can be uniquely partitioned into components identified with each contrast
- follow independent χ^2 distributions

- can all be fit using the linear model $y = X\beta + \epsilon$
- columns of X are mutually orthogonal
- as soon as this is not the case, individual effect estimates depend on the inclusion of other terms

Analysis: incomplete block design

SM Ex.9.4

Embryo	Treat	y	Treat	y	Embryo	Treat	y	Treat	y
1	—	2.51	His–	2.15	9	His–	2.32	Lys–	2.53
2	—	2.49	Arg–	2.23	10	Arg–	2.15	Thr–	2.23
3	—	2.54	Thr–	2.26	11	Arg–	2.34	Val–	2.15
4	—	2.58	Val–	2.15	12	Arg–	2.30	Lys–	2.49
5	—	2.65	Lys–	2.41	13	Thr–	2.20	Val–	2.18
6	His–	2.11	Arg–	1.90	14	Thr–	2.26	Lys–	2.43
7	His–	2.28	Thr–	2.11	15	Val–	2.28	Lys–	2.56
8	His–	2.15	Val–	1.70					

Table 9.8 Log_{10} dry weight y (μg) of chick bones after cultivation over a nutrient chemical medium, either complete (—), or with single amino acids missing (Cox and Snell, 1981, p. 95). The order of treatment pairs was randomized, but the table shows them systematically.

```
> anova(lm(y ~ Pair + Treat, data = chicks))
Analysis of Variance Table

Response: y
          Df  Sum Sq  Mean Sq F value   Pr(>F)
Pair      14 0.75288 0.053777  8.1728 0.001025 **
Treat      5 0.44620 0.089240 13.5623 0.000347 ***
Residuals 10 0.06580 0.006580
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(y ~ Treat + Pair, data = chicks))
Analysis of Variance Table

Response: y
          Df  Sum Sq  Mean Sq F value    Pr(>F)
Treat      5 0.85788 0.171576 26.0754 1.963e-05 ***
Pair      14 0.34120 0.024371  3.7039   0.02165 *
Residuals 10 0.06580 0.006580
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the news

Washington Post, Jan 6

reading pompeii neuroscience

Web Images Maps News More Search tools

About 8,970,000 results (0.43 seconds)

eScienceCommons: A novel look at how stories may change the brain
esciencecommons.blogspot.com/.../a-novel-look-at-how-stories-may-cha...
Dec 17, 2013 - All of the study subjects **read** the same novel, "**Pompeii**," a 2003 thriller by Robert Harris that is **...** Novelists, **neuroscientists** trade mental notes **...**

Great novels can change your life...and your brain - Telegraph
www.telegraph.co.uk › Science › Science News
3 days ago - **Neuroscientist** Professor Gregory Berns, of Emory University in Atlanta, **...** Researchers enlisted 21 students to **read** the novel **Pompeii** by best **...**

Reading A Novel Could Physically Change Your Brain (But What ...

www.forbes.com/.../**reading**-a-novel-could-physically-change-y...
by Alex Knapp - in 786 Google+ circles
4 days ago - It turns out that **reading** a novel can cause measurable changes in **...** each participant **read** approximately 1/9 of the novel **Pompeii** by **...** This is a prime example of both how much we know about **neuroscience** and how little.

Reading Fiction Improves Brain Connectivity and Function ...
www.psychologytoday.com/.../**reading**-fiction-improves-brain-connectiv...
5 days ago - **Neuroscientists** have discovered that **reading** a novel can improve brain

The Washington Post

Back to previous page

Can reading a novel change your brain? A study of brain scans suggests yes.

By Emory University, Published: January 6

Researchers say that reading a novel can change the biology of your brain

Study Hall presents recent studies as described by researchers and their institutions. This report is from Emory University.

After reading a novel, actual changes linger in the brain, at least for a few days, report researchers at Emory University.

"Stories shape our lives and in some cases help define a person," says neuroscientist Gregory Berns, lead author of the study and the director of Emory University's Center for Neuropolicy. "We want to understand how stories get into your brain, and what they do to it."

Neurobiological research using functional magnetic resonance imaging (fMRI) has begun to identify brain networks associated with reading stories.

The study focused on the lingering neural effects of reading a narrative. Over the course of 19 days, 21 Emory undergraduates read the same novel, "Pompeii," a 2003 thriller based on the real-life eruption of Mount Vesuvius in ancient Italy.

The researchers chose the book due to its page-turning plot.

For the first five days, the participants came in each morning for a base-line fMRI scan of their brains in a resting state. Then they were given nine sections of the novel, about 30 pages each, over a nine-day period. They were asked to read the assigned section in the evening and come in the following morning. After taking a quiz to ensure they had finished the assigned reading, the participants underwent an fMRI scan of their brain in a non-reading, resting state. After completing all nine sections of the novel, the participants returned for five more mornings to undergo additional scans in a resting state.

BRAIN CONNECTIVITY
Volume 3, Number 6, 2013

DOI: 10.1089/brain.2013.0166

Short- and Long-Term Effects of a Novel on Connectivity in the Brain

Gregory S. Berns,[1] Kristina Blaine,[1] Michael J. Prietula,[2] and Brandon E. Pye[1]

Abstract

We sought to determine whether reading a novel causes measurable changes in resting-state connectivity of the brain and how long these changes persist. Incorporating a within-subjects design, participants received resting-state functional magnetic resonance imaging scans on 19 consecutive days. First, baseline resting state data for a "washin" period were taken for each participant for 5 days. For the next 9 days, participants read 1/9th of a novel during the evening and resting-state data were taken the next morning. Finally, resting-state data for a "washout" period were taken for 5 days after the conclusion of the novel. On the days after the reading, significant increases in connectivity were centered on hubs in the left angular/supramarginal gyri and right posterior temporal gyri. These hubs corresponded to regions previously associated with perspective taking and story comprehension, and the changes exhibited a timecourse that decayed rapidly after the completion of the novel. Long-term changes in connectivity, which persisted for several days after the reading, were observed in bilateral somatosensory cortex, suggesting a potential mechanism for "embodied semantics."

Key words: connectivity; fMRI; reading; resting state

A great book should leave you with many experiences, and slightly exhausted at the end. You live several lives while reading. – William Styron, *Conversations with William Styron.*

guistic and literary theories describe what constitutes a story, neurobiological research has just begun to elucidate brain networks that are active when processing stories. To date, these studies have focused on the immediate response

Wired Blog Post

Brain Watch

Reading a Novel Alters Your Brain Connectivity — So What?

BY CHRISTIAN JARRETT 01.06.14 5:30 AM