

Mathematical Statistics II

STA2212H S LEC0101

Week 8

March 7 2023

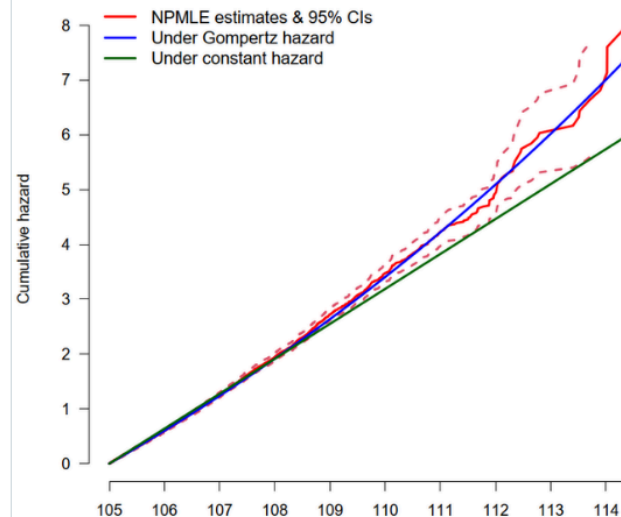
← Tweet



Demographic Research
@DemographicRes

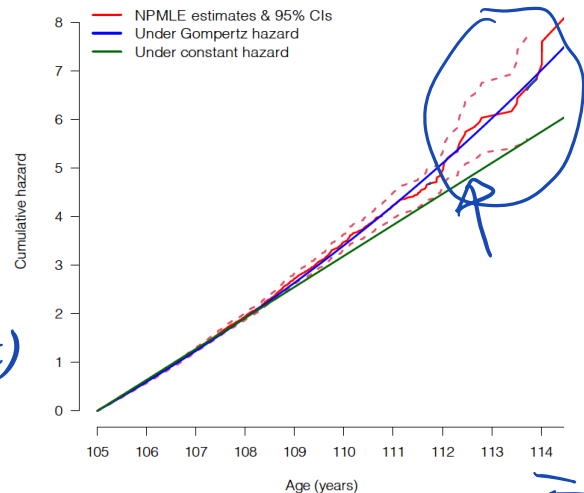
...

Evidence for human mortality plateau was claimed in a 2018 article [@ScienceMagazine](#). Replication on French data shows this finding is not universal: the plateau can't be proven yet. Read the new paper by [@linhhkdang](#) et al. [@InedFr](#) [@Demo_UdeM](#) [@Inserm](#). [demographic-research.org/volumes/vol48/...](#)



Mortality plateau

Figure 1: Estimated cumulative hazard using nonparametric and parametric approaches, French females born 1883–1901, ages 105 and above



Notes: NPMLE is for nonparametric maximum likelihood estimate. This figure is equivalent to Figure 2 in Barbi et al. (2018: p. 1461).
Source: Authors' calculations based on data from the Répertoire national d'identification des personnes physiques (RNIPP).

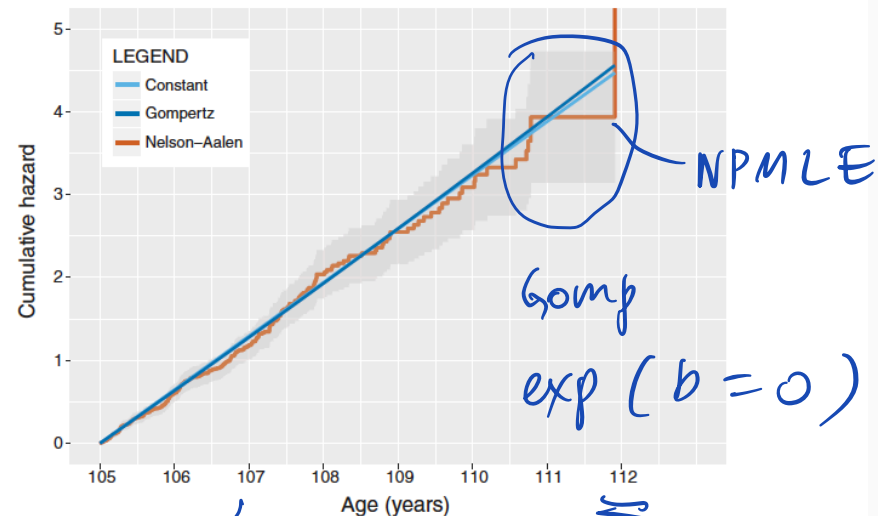


Fig. 2. Cumulative hazard beyond age 105 for the cohort of Italian women born in 1904, as determined by the Nelson-Aalen estimator. Straight lines represent cumulative hazards of the estimated plateau predicted from ISTAT data, under a constant hazard (light blue) and a Gompertz hazard model (darker blue). The shaded area indicates the 95% confidence bands of the Nelson-Aalen estimate.

France 2022 (Dang et al.)

Gompertz

Mathematical Statistics II

March 7 2023

$$h(t; \mathbf{x}) = \frac{f(t; \mathbf{x})}{1 - F(t; \mathbf{x})} = a \exp(bt + \beta_1 x_1 + \beta_2 x_2); \quad H_0 : b = 0$$

Gender ← Cohort (yr. of birth) ⁻²

Italy 2017 (Barbi et al.)



WIKIPEDIA

The Free Encyclopedia

Search Wikipedia

Wiki Loves FOLKLORE

Photograph your folkloric traditions on Wikipedia and

List of women in statistics

Article

Talk

Read

From Wikipedia, the free encyclopedia

This is a list of women who have made noteworthy contributions to or achievements in [statistics](#).^{[1][2]}

◄See TIM.

Contents

[A](#) · [B](#) · [C](#) · [D](#) · [E](#) · [F](#) · [G](#) · [H](#) · [I](#) · [J](#) · [K](#) · [L](#) · [M](#) · [N](#) · [O](#) · [P](#) · [Q](#) · [R](#) · [S](#) · [T](#) · [U](#) · [V](#) · [W](#) · [X](#) · [Y](#) · [Z](#)

[See also](#) · [References](#) · [External links](#)

A [edit]

- Helen Abbey** (1915–2001), American biostatistician known for prolific mentorship of students
- Edith Abbott** (1876–1957), American economist, social worker, educator, and author
- Sarah Abramowitz** (born 1967), American statistics educator and textbook author
- Dorothy Adkins** (1912–1975), psychologist concentrating on psychometrics
- Susan Ahmed** (born 1946), American biostatistician and educational statistician

Mathematical Statistics II

March 7 2023

-1

Today

1. Next lectures
2. Recap
3. Multiple testing
4. Goodness-of-fit tests

Upcoming

- March 9 3.30 – 4.30 DoSS 9014 [Details](#)
“Valid statistical inference with privacy constraints”
Aleksandra Slavković, Penn State
- March 13 3.30 – 4.30 DoSS 9014 & online [Details](#)
“Training Scientists to Perform Robust Bayesian Inference”
Justin Bois, CalTech



Next Lectures

- March 14 10.00 – 13.00
- March 21 11.00 – 13.00
- March 28 10.00 – 12.00
- April 4 Project presentations

Next Lectures

- March 14 10.00 – 13.00
- March 21 11.00 – 13.00
- March 28 10.00 – 12.00
- April 4 Project presentations

STA 2212S: Mathematical Statistics II Syllabus

Spring 2023

Week	Date	Methods	References
1	Jan 10	Likelihood inference: review of ML estimation; mis-specified models; computation; nonparametric mle	MS §§5.1–7, SM Ch 4
2	Jan 17	Bayesian estimation; Bayesian inference	MS §5.8; AoS §§ 11.1–4; SM §§11.1,2
3	Jan 24	Optimality in estimation	MS Ch 6; AoS Ch 12; SM §7.1, 11.5.2
4	Jan 31	Interval estimation; Confidence bands	MS §§7.1,2; AoS Ch 7; SM §7.1.4
5	Feb 7	Hypothesis testing; likelihood ratio tests	MS §§7.1–4 AoS Ch 10.6, SM
6	Feb 14	Significance testing	MS §7.5; AoS §10.2,6; SM Ch 4, §7.3.1
	Feb 21	Break	
7	Feb 28	Significance testing	SM 7.3.1
7	Feb 28	Goodness-of-fit testing	MS Ch 9; AoS §§10.3,4,5,8; SM p.327-8 (hard)
8	Mar 7	Multiple testing and FDR	AoS Ch 10.7, EH Ch 15.1,2
9	Mar 14	Empirical Bayes ✓	EH Ch 6, SM Ch 11.5
10	Mar 21	Multivariate Models	AoS Ch 14; SM Ch 6.3
11	Mar 28	Introduction to Causal Inference	AoS Ch 16, 17
12	Apr 4	Recap	

Next Lectures

- March 14 10.00 – 13.00
- March 21 11.00 – 13.00
- March 28 10.00 – 12.00
- April 4 Project presentations

2	Jan 17	Bayesian estimation; Bayesian inference	MS §§9.8; AoS §§ 11.1–4; SM §§11.1,2
3	Jan 24	Optimality in estimation	MS Ch 6; AoS Ch 12; SM §7.1, 11.5.2
4	Jan 31	Interval estimation; Confidence bands	MS §§7.1,2; AoS Ch 7; SM §7.1.4
5	Feb 7	Hypothesis testing; likelihood ratio tests	MS §§7.1–4 AoS Ch 10.6, SM
6	Feb 14	Significance testing	MS §7.5; AoS §10.2,6; SM Ch 4, §7.3.1
	Feb 21	Break	
7	Feb 28	Significance testing	SM 7.3.1
7	Feb 28	Goodness-of-fit testing	MS Ch 9; AoS §§10.3,4,5,8; SM p.327-8 (hard)
8	Mar 7	Multiple testing and FDR	AoS Ch 10.7, EH Ch 15.1,2
9	Mar 14		
10	Mar 21	Likelihood asymptotics; robust estimation (SM 7.2); causal inference (AoS 16); linear and generalized linear models (MS 8); graphical models (AoS 17,18); nonparametric curve estimation (AoS 20); classification (AoS 22); any of 1-8	
11	Mar 28		
12	Apr 4	Course Summary; Presentations	

Subject to adjustment as the course progresses.

References

MS: *Mathematical Statistics* by K. Knight (Chapman & Hall/CRC).

AoS: *All of Statistics* by L. Wasserman (Springer) If your copy has a Chapter 1. Introduction, then all Chapter numbers increase by 1.

SM: *Statistical Models* by A.C. Davison (Cambridge University Press)

revisit

Recap

- Neyman-Pearson lemma; simple and composite hypotheses; power and size

- p -values: definition, interpretation; diagnostic tests

- sign test; permutation test; intro to multiple testing

$$H_0: f = f_0 \quad H_1: f = f_1$$

"reject H_0 " if

$$\frac{f_1(\underline{n})}{f_0(\underline{n})} > k$$

↑
MP test of size
might be UMP
 $H_1: \eta > \eta_0$

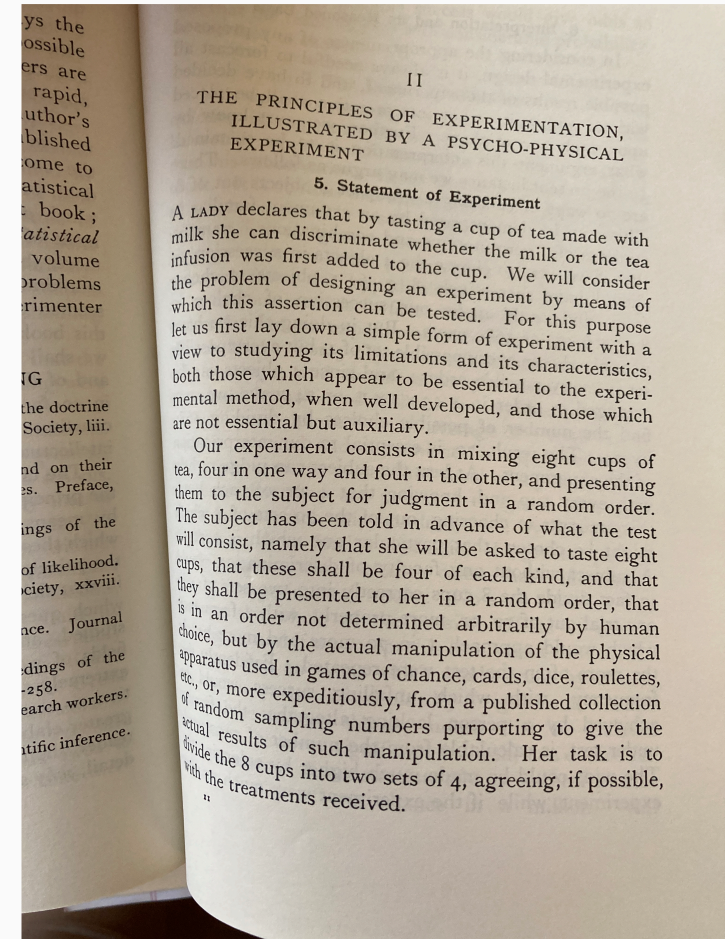
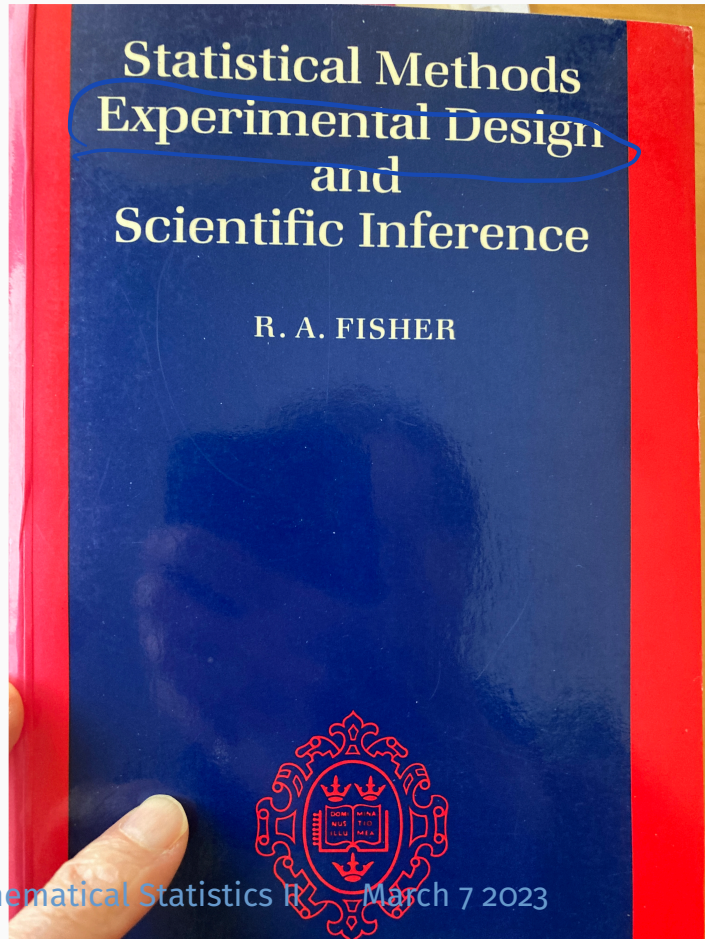
$$p = \text{Pa} \left(T \geq t^{\text{obs}}; H_0 \right)$$

"prob. (of a result as or more extreme than observed")

$|T|$
not T



$\uparrow$
 3.7
 $H_0: \sigma^2 = \sigma_0^2$
 $H_1: \sigma^2 > \sigma_0^2$
 $t \dots$
 Lehmann
 TS H



undoubtedly occur more than its appropriate frequency, however surprised we may be that it should occur to us. In order to assert that a natural phenomenon is experimentally demonstrable we need, not an isolated record, but a reliable method of procedure. In relation to the test of significance, we may say that a phenomenon is experimentally demonstrable when we know how to conduct an experiment which will rarely fail to give us a statistically significant result.

Returning to the possible results of the psychophysical experiment, having decided that if every cup were rightly classified a significant positive result would be recorded, or, in other words, that we should admit that the lady had made good her claim, what should be our conclusion if, for each kind of cup, her judgments are 3 right and 1 wrong? We may take it, in the present discussion, that any error in one set of judgments will be compensated by an error in the other, since it is known to the subject that there are 4 cups of each kind. In enumerating the number of ways of choosing 4 things out of 8, such that 3 are right and 1 wrong, we may note that the 3 right may be chosen, out of the 4 available, in 4 ways and, independently of this choice, that the 1 wrong may be chosen, out of the 4 available, also in 4 ways. So that in all we could make a selection of the kind supposed in 16 different ways. A similar argument shows that, in each kind of judgment, 2 may be right and 2 wrong in 36 ways, 1 right and 3 wrong in 16 ways and none right and 4 wrong in 1 way only. It should be noted that the frequencies of these five possible results of the experiment make up together, as it is obvious they should, the 70 cases out of 70.

It is obvious, too, that 13 successes, 10 failures, although showing a bias, or deviation, in the right

direction, evidence of frequency, that rests on frequency, "3" occurs good including obvious have 1 chance out not seen have

expected sign two sign not with which sign possible some class sign sign name

NULL HYPOTHESIS

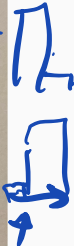
15

direction, could not be judged as statistically significant evidence of a real sensory discrimination. For its frequency of chance occurrence is 16 in 70, or more than 20 per cent. Moreover, it is not the best possible result, and in judging of its significance we must take account not only of its own frequency, but also of the frequency of any better result. In the present instance "3 right and 1 wrong" occurs 16 times, and "4 right" occurs once in 70 trials, making 17 cases out of 70 as good as or better than that observed. The reason for including cases better than that observed becomes obvious on considering what our conclusions would have been had the case of 3 right and 1 wrong only 1 chance, and the case of 4 right 16 chances of occurrence out of 70. The rare case of 3 right and 1 wrong could not be judged significant merely because it was rare, seeing that a higher degree of success would frequently have been scored by mere chance.

8. The Null Hypothesis

Our examination of the possible results of the experiment has therefore led us to a statistical test of significance, by which these results are divided into two classes with opposed interpretations. Tests of significance are of many different kinds, which need to be calculated here. Here we are only concerned with the calculation in permutations of our test of

$$70 = \binom{8}{4}$$



Recap: Choosing test statistics

1. Context
2. Optimal choice – Neyman-Pearson Lemma ✓
3. Pragmatic choice – likelihood-based statistics ✓
4. Pragmatic choice – nonparametric test statistics ✓

and its extensions; MS 7.3

Wald, score, LRT

sign, permutation

"model-free"

"model - weak

1. Hypothesis testing

	H_0 not rejected	H_0 rejected	
H_0 true	✓	<u>type 1 error</u>	$\leq \alpha$
H_1 true	type 2 error		$\downarrow$ $\checkmark 1 - \text{t.2} = \text{power}$

2. Diagnostic testing

	test negative	test positive	
C19 neg	TN ✓	<u>FP</u> ✗	N
C19 pos	<u>FN</u> ✗	TP ✓	P

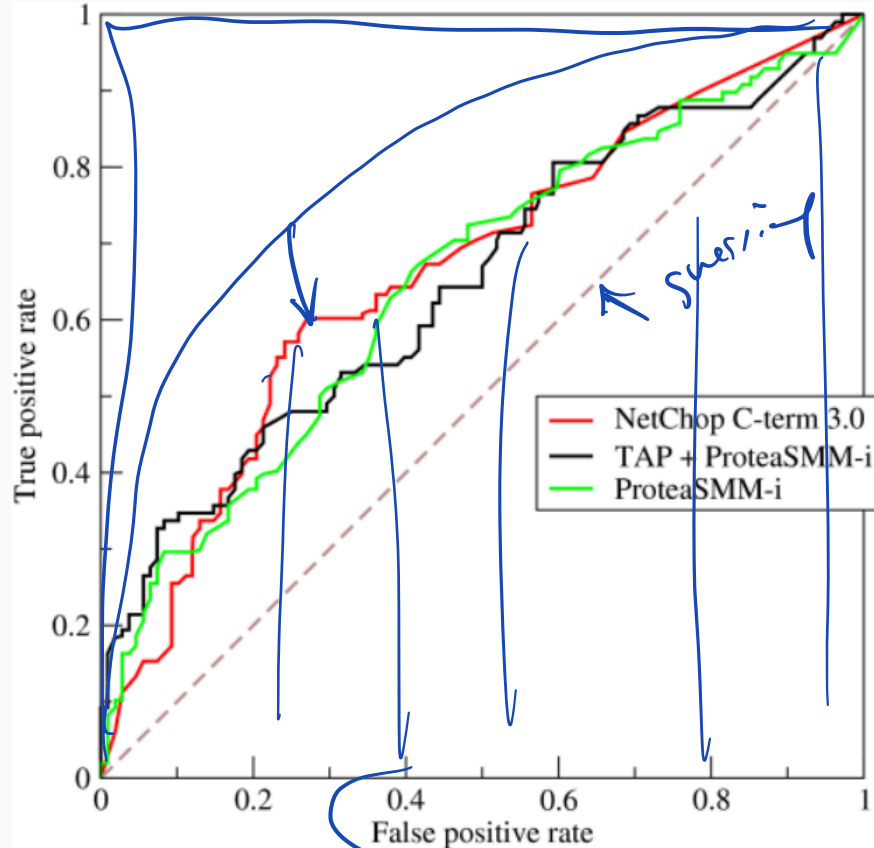
specificity = TN/N

sensitivity = TP/P

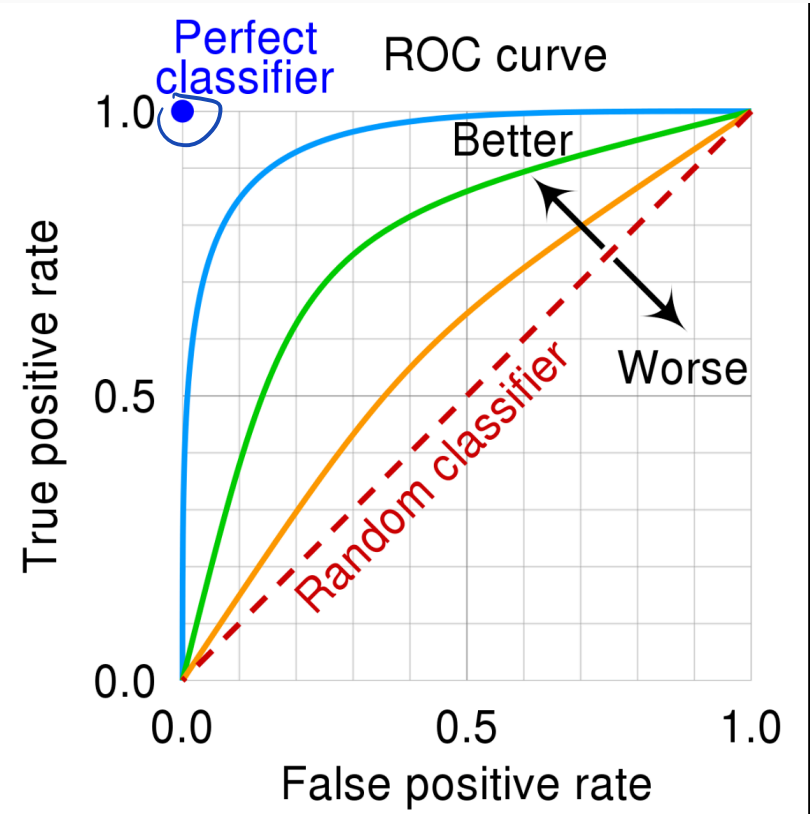
[link](#)

Diagnostic testing and ROC

Sensitivity (TP)



1 - Specificity FP



AUC $\leftarrow$ area under c.

2. Diagnostic testing

	test negative	test positive	
truth			
C19 neg	TN	<u>FP</u>	N
C19 pos	<u>FN</u>	TP	P

p value for H_{0i} is
 $\text{sign. if } < \frac{.05}{7128}$

link
control
FWER

3. Multiple testing

	H_0 not rejected	H_0 rejected	
truth			
H_0 true	U	<u>V</u>	<u>m_0</u>
H_1 true	T	S	m_1
	$m - R$	<u>R</u>	m

$H_{0i} \quad i = 1, \dots, 7128$
 $\mu_{1i} = \mu_{2i} \quad t\text{-test}$

$$\text{FDP} = V/R$$

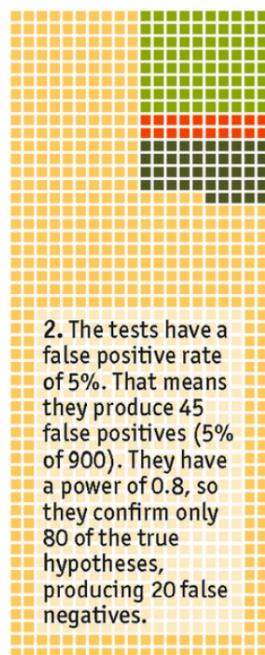
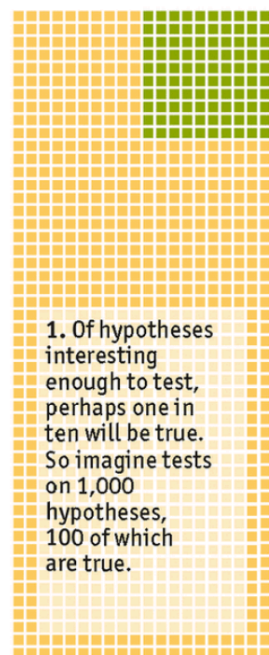
$\equiv 0$ if $R = 0$

$$\text{FDR} = E(V/R)$$

Multiple testing

Unlikely results
How a small proportion of false positives can prove very misleading

False True False negatives False positives



Source: *The Economist*

	H_0 not rejected	H_0 rejected	
H_0 true	U	V	m_0
H_1 true	T	S	m_1
	$m - R$	R	m

	test negative	test positive	
C19 neg	TN	FP	N
C19 pos	FN	TP	P

- H_{0i} versus H_{1i} , $i = 1, \dots, m$

- p -values $p_1, \dots, p_m$

- Bonferroni method: reject H_{0i} if $p_i < \alpha/m$

- $\text{pr}(\text{any } H_0 \text{ falsely rejected}) \leq \alpha$

family-wise error rate

FWER

very conservative

- H_{0i} versus H_{1i} , $i = 1, \dots, m$
- p -values $p_1, \dots, p_m$
- Bonferroni method: reject H_{0i} if $p_i < \alpha/m$
- $\text{pr}(\text{any } H_0 \text{ falsely rejected}) \leq \alpha$

very conservative

- FDR method controls the number of rejections that are false

FDP = V/R

		H_0 not rejected	H_0 rejected	
truth	H_0 true	U	V	m_0
	H_1 true	T	S	m_1
		$m - R$	R	m

- order the p -values $p_{(1)}, \dots, p_{(m)}$
- find $i_{\max}$, the largest index for which

$$p_{(i)} \leq \frac{i}{m} q$$

- Let BH_q be the rule that rejects H_{0i} for $i \leq i_{\max}$, not rejecting otherwise

FDR control
value
 $q \rightarrow \alpha$

- order the p -values $p_{(1)}, \dots, p_{(m)}$
- find $i_{\max}$, the largest index for which

$$p_{(i)} \leq \frac{i}{m}q$$

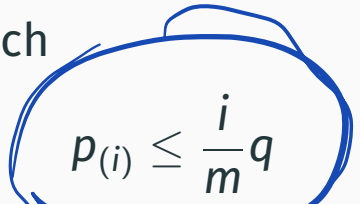
- Let BH_q be the rule that rejects H_{0i} for $i \leq i_{\max}$, not rejecting otherwise
- change the bound under dependence

$$p_{(i)} \leq \frac{i}{m c_m} q$$

$$c_m = \sum_{i=1}^m \frac{1}{i}$$

more stringent

- order the p -values $p_{(1)}, \dots, p_{(m)}$
- find $i_{\max}$, the largest index for which


$$p_{(i)} \leq \frac{i}{m}q$$

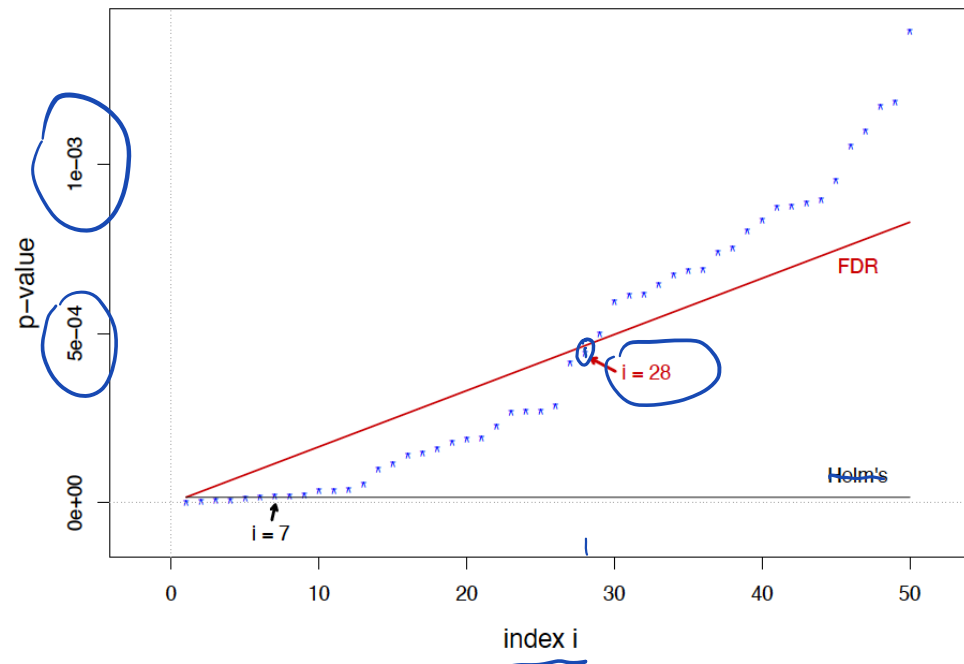
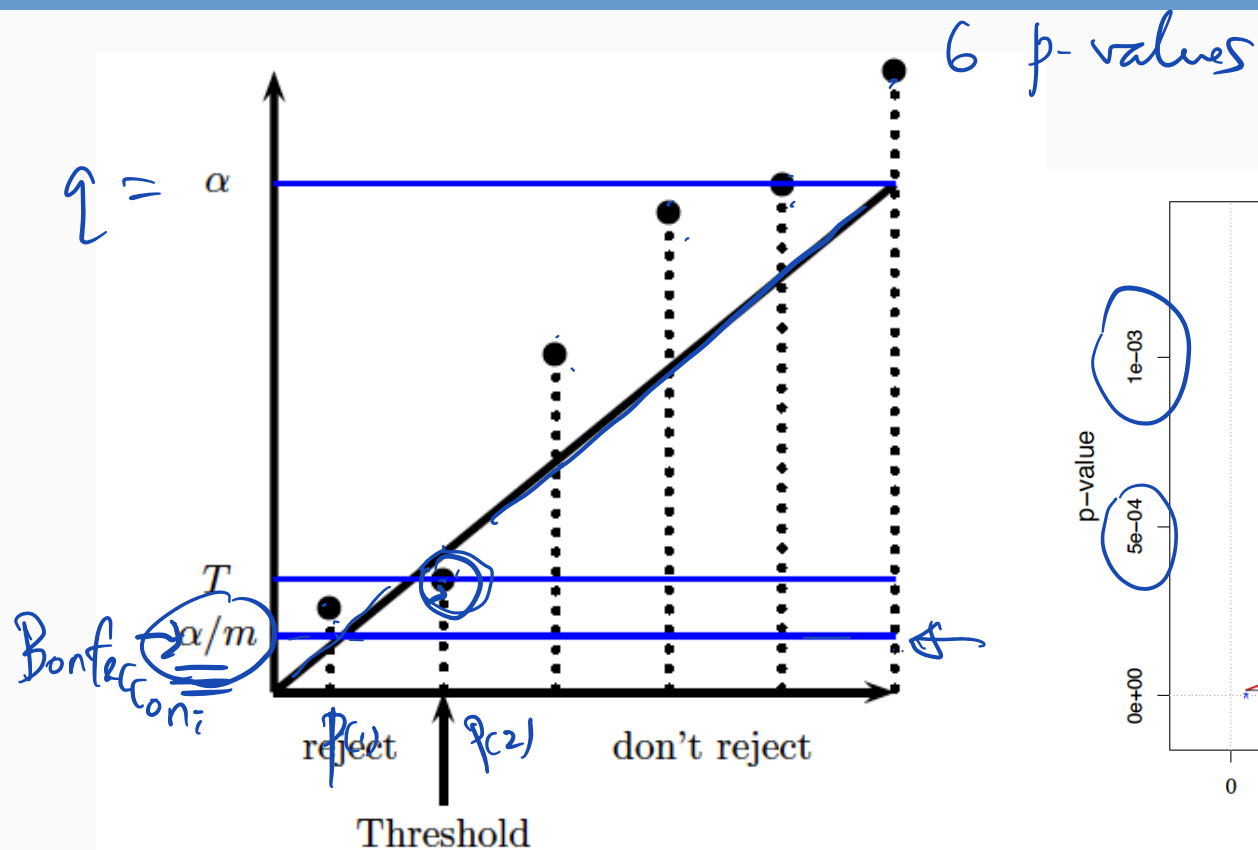
- Let BH_q be the rule that rejects H_{0i} for $i \leq i_{\max}$, not rejecting otherwise
- change the bound under dependence

$$p_{(i)} \leq \frac{i}{mC_m}q \qquad C_m = \sum_{i=1}^m \frac{1}{i}$$

- **Theorem:** If the p -values corresponding to valid null hypotheses are independent of each other, then

$$FDR(BH_q) = \pi_0 q \leq q, \quad \text{where } \pi_0 = m_0/m$$

π_0 unknown but close to 1



reject H_1, H_2

$$p_{(i)} \leq \frac{i}{m} \alpha$$

$$\alpha = 0.1$$

$$p_{(i)} \leq \frac{0.1}{m} i$$

not H_3, H_4, H_5, H_6

index	1	2	3	4	5	6	7	8	9	10
pval	0.00017	0.00448	0.00671	0.00907	0.01220	0.33626	0.3934	0.5388	0.5813	0.9862
cut1	0.00500	0.01000	0.01500	0.02000	0.02500	0.03000	0.0350	0.0400	0.0450	0.0500
cut2	0.00171	0.00341	0.00512	0.00683	0.00854	0.01024	0.0119	0.0137	0.0154	0.0171

C_m

✓

reject 1 H_0

$$q = 0.05$$

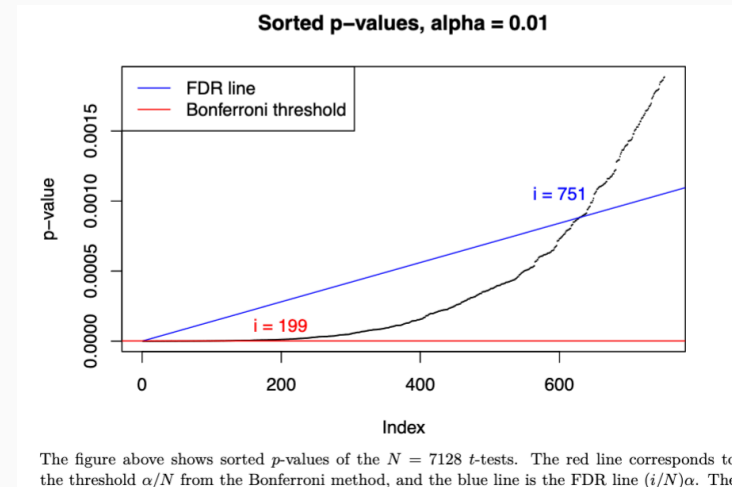
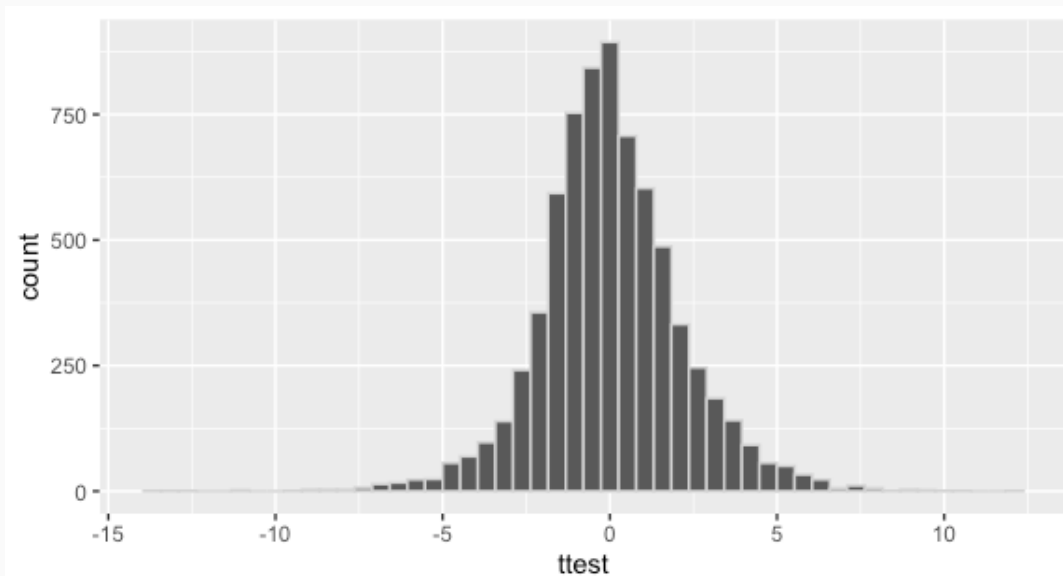
$$q^{i/m} = 10$$

evidence against

5 H_0 's

```
leukemia_big <- read.csv  
  ("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")  
dim(leukemia_big)  
[1] 7128 72  
df <- t(leukemia_big)  
df_all <- df[startsWith(rownames(df), "ALL"), ]  
df_aml <- df[startsWith(rownames(df), "AML"), ]  
n1 <- nrow(df_all); n2 <- nrow(df_aml)  
m1 <- apply(df_all, 2, mean); m2 <- apply(df_aml, 2, mean)  
s1 <- apply(df_all, 2, sd); s2 <- apply(df_aml, 2, sd)  
  
pooled <- sqrt(((n1 - 1) * s1^2 + (n2 - 1) * s2^2) / (n1 + n2 - 2))  
ttest <- (m1 - m2) / pooled / sqrt(1 / n1 + 1 / n2)  
pvalues <- 2 * pt(abs(ttest), df = n1 + n2 - 2, lower.tail = F)
```

← s_{pooled}^2



```
> summary(ttest)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-13.52611	-1.20672	-0.08406	0.02308	1.20886	12.26065

Theorem: If the p -values corresponding to valid null hypotheses are independent of each other, then

$$\underline{FDR(BH_q)} = \pi_0 q \leq q, \quad \text{where } \pi_0 = m_0/m$$

α

The Annals of Statistics
2006, Vol. 34, No. 4, 1827-1849
DOI: 10.1214/000000000000000425
© Institute of Mathematical Statistics, 2006

ON THE BENJAMINI-HOCHBERG METHOD

BY J. A. FERREIRA¹ AND A. H. ZWINDERMAN

University of Amsterdam

We investigate the properties of the Benjamini-Hochberg method for multiple testing and of a variant of Storey's generalization of it, extending and complementing the asymptotic and exact results available in the literature. Results are obtained under two different sets of assumptions and include asymptotic and exact expressions and bounds for the proportion of rejections, the proportion of incorrect rejections out of all rejections and two other proportions used to quantify the efficacy of the method.

1. Introduction. Let $X = \{X_1, X_2, \dots, X_m\}$ be a set of m random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ such that, for some positive integer $m_0 \leq m$, each of $X_1, X_2, \dots, X_{m_0}$ has distribution function (d.f.) F and $X_{m_0+1}, \dots, X_m$ all have d.f.'s different from F , and consider the problem of choosing a set $\mathcal{R} \subseteq X$ in such a way that the random variable (r.v.)

$$\Pi_{1,m} = \frac{S_m}{R_m \vee 1},$$

where $R_m = \#\mathcal{R}$ and $S_m = \#\{\mathcal{R} \cap \{X_1, \dots, X_{m_0}\}\}$, is guaranteed to be small in some probabilistic sense. In more ordinary language, the problem is that of discovering observations in X which do not have d.f. F without incurring a high proportion of incorrect rejections—the proportion $\Pi_{1,m}$ of rejected observations which in fact come from F .

Benjamini and Hochberg [2] have proposed a method of choosing $\mathcal{R}$ specifically aimed at discovering r.v.'s taking values in the interval $[0, 1]$ that tend to be smaller than standard uniform r.v.'s and which, given $\delta > 0$, guarantees that $E(\Pi_{1,m}) \leq \delta$ under certain conditions. The method consists of fixing $q \in [0, 1]$, computing

$$(1.1) \quad R_m = \max \left\{ i : X_{i:m} \leq q \frac{i}{m} \right\},$$

where $0 \leq X_{1:m} \leq \dots \leq X_{m:m} \leq 1$ denote the order statistics of X , and setting $\mathcal{R} = \{X_{1:m}, \dots, X_{R_m:m}\}$. In its simplest form, the Benjamini-Hochberg theorem states that if $\mathcal{R}$ is chosen according to this procedure and $X_1, X_2, \dots, X_{m_0}$

Received February 2005; revised August 2005.

¹Supported in part by "Nederlandse Horstschilling," the Dutch Heart Foundation, through Grant 2001-011.

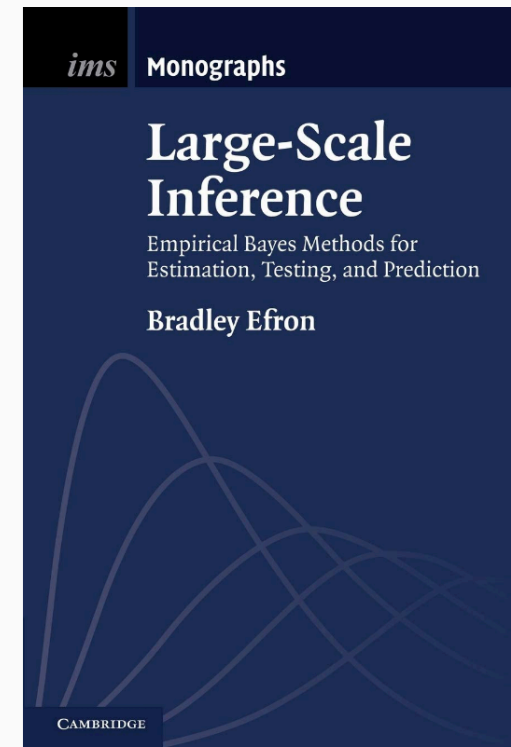
AMS 2000 subject classifications. 62J15, 62G30, 60F05.

Key words and phrases. Multiple testing, goodness of fit, empirical distributions, false discovery rate.

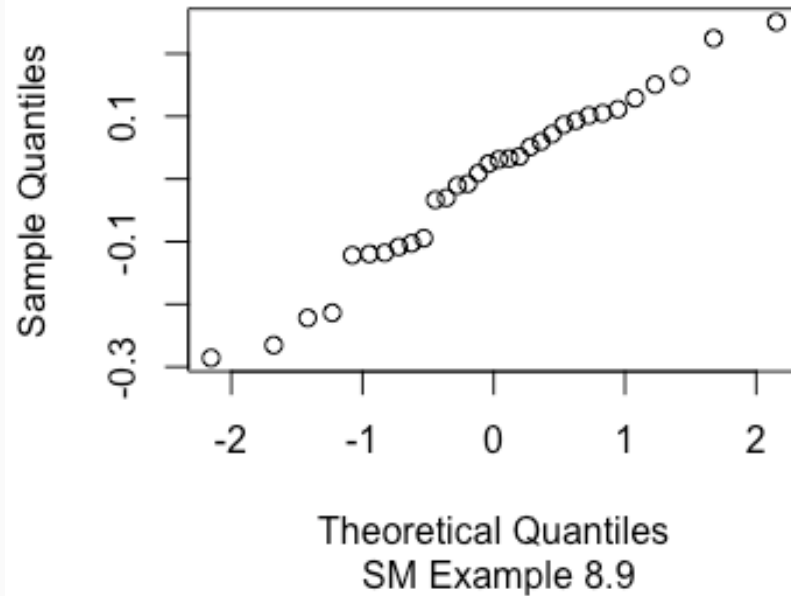
$$p_{(i)} \leq \frac{iq}{m} \quad \left(\text{or } \frac{iq}{mC_m} \right)$$

↑
ind't p-values

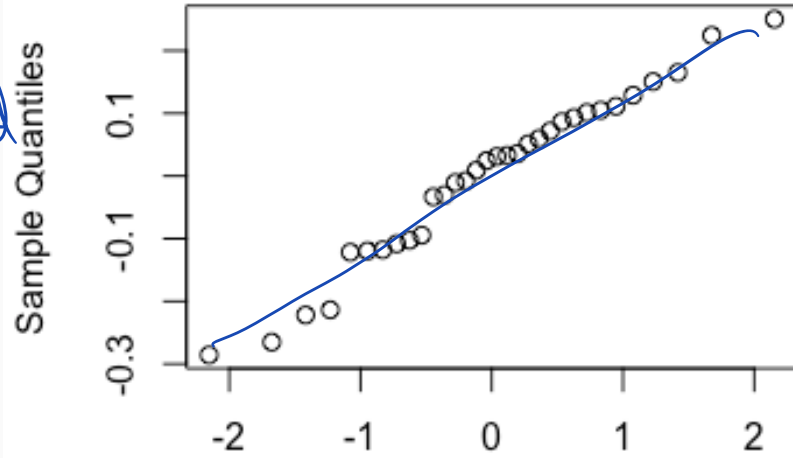
$$q = 0.1$$



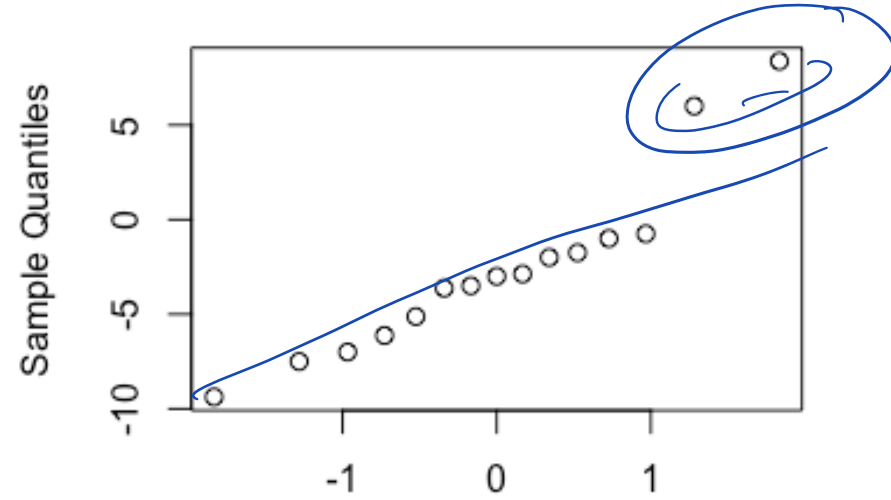
residuals from linear regression



residuals from linear regression



Maize data SM Ex 7.24



Theoretical Quantiles
SM Example 8.9

Theoretical Quantiles

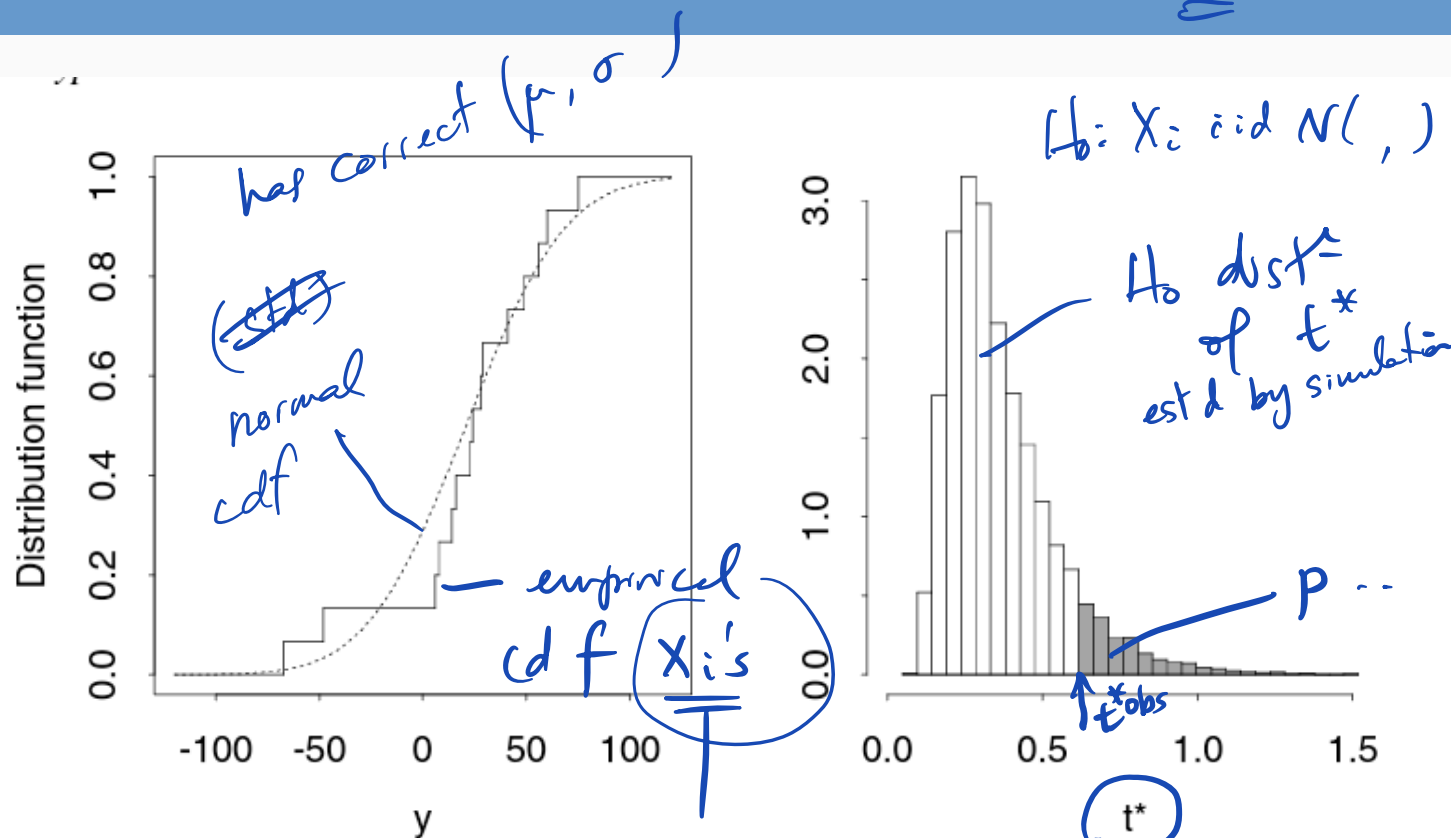
Shapiro-Wilk test $E Z_{(i)} \sim N(0,1)$
combines 3rd & 4th moment $\leftarrow$
Mathematical Statistics II March 7 2023

$$\sum_{i=1}^n \frac{1}{n} \left(\hat{\varepsilon}_{(i)} - \bar{\hat{\varepsilon}}_{(i)} \right)^3$$

$\uparrow$ skewness

(H₀) $\rightarrow E N(0,1)^3 = 0$

Figure 7.5 Analysis of maize data. Left: empirical distribution function for height differences, with fitted normal distribution (dots). Right: null density of Anderson-Darling statistic T for normal samples of size $n = 15$ with location and scale estimated. The shaded part of the histogram shows values of T^* in excess of the observed value t_{obs} .



SM Example 7.24 testing $N(\mu, \sigma^2)$ distribution

$$t_{AD}^* = \frac{\int \{\hat{F}_n(x) - F_0(x)\}^2 dF_0(x)}{F_0(x)\{1 - F_0(x)\}}$$

- $X_1, \dots, X_n$ i.i.d. $F(\cdot)$; $H_0 : F = F_0$
- $\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n 1\{X_i \leq t\}$

H_0 simple
(F_0 has no parameters)
cumulative d.f.

- three test statistics:

1. $\sup_t |\hat{F}_n(t) - F_0(t)| \leftarrow$ Kolmogorov

2. $\int \{\hat{F}_n(t) - F_0(t)\}^2 dF_0(t) \leftarrow$ Cramer-vonMises

3. $\int \frac{\{\hat{F}_n(t) - F_0(t)\}^2}{F_0(t)\{1 - F_0(t)\}} dF_0(t) \leftarrow$ Anderson-Darling

- SM Example 7.24 testing $N(\mu, \sigma^2)$ distribution
- SM Example 7.23; 6.14 testing $U(0, 1)$ distribution

$X \sim F_0$
 $F_0(x) \sim U(0, 1)$

- Special case $H_0 : \underline{F(t) = F_0(t) = t}$
- Recall

$U(0,1)$

$X_i \sim U(0,1)$

$$E_0\{\hat{F}_n(t)\} = F_0(t) = t, \quad \text{var}\{\hat{F}_n(t)\} = t(1-t)/n$$

fixed t

- What about distribution of

$$\sup_{0 \leq t \leq 1} |\hat{F}_n(t) - t|$$

$$\int_0^1 \{\hat{F}_n(t) - t\}^2 dt$$

$$\int_0^1 \frac{\{\hat{F}_n(t) - t\}^2}{t(1-t)} dt$$

t

- need joint density of $\hat{F}_n(t) \forall t$

- Special case $H_0 : F(t) = F_0(t) = t$
- Recall

$$X_i \sim U(0, 1)$$

$$E_0\{\hat{F}_n(t)\} = F_0(t) = t, \quad \text{var}\{\hat{F}_n(t)\} = t(1-t)/n$$

- What about distribution of

$$\sup_t |\hat{F}_n(t) - t| \quad \int \{\hat{F}_n(t) - t\}^2 dt \quad \int \frac{\{\hat{F}_n(t) - t\}^2}{F_0(t)\{1-t\}} dt$$

- need joint density of $\hat{F}_n(t) \forall t$

- define ~~Brownian bridge~~ $B_n(t) = \sqrt{n}(\hat{F}_n(t) - t)$ $\forall t \in [0, 1]$

$B_n(\cdot) \xrightarrow{d}$
Brownian
bridge

- vector $(B_n(t_1), \dots, B_n(t_k)) \xrightarrow{d} N_k(\mathbf{0}, \mathbf{C})$, $C_{ij} = \min(t_i, t_j) - t_i t_j$

MS 9.3

- a **Brownian bridge** is a continuous function on $(0, 1)$

with all finite-dimensional distributions as above

- Kolmogorov-Smirnov test
- Cramer-vonMises test
- Anderson-Darling test

$$K_n = \sup_{0 \leq t \leq 1} |B_n(t)|$$

$$W_n^2 = \int_0^1 B_n^2(t) dt$$

$$A_n^2 = \int_0^1 \frac{B_n^2(t)}{t(1-t)} dt$$



- Kolmogorov-Smirnov test

$$K_n = \sup_{0 \leq t \leq 1} |B_n(t)|$$

- Cramer-vonMises test

$$W_n^2 = \int_0^1 B_n^2(t) dt$$

- Anderson-Darling test

$$A_n^2 = \int_0^1 \frac{B_n^2(t)}{t(1-t)} dt$$

- limit theorems

$$K_n \xrightarrow{d} K,$$

↑
AoS

$$W_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2 \pi^2}$$

$$A_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)}$$

$$\text{pr}(K > x) = 2 \sum_{j=1}^{\infty} (-1)^{j+1} \exp(-2j^2 x^2)$$

$Z_j \sim N(0,1)$
 $Z_j^2 \sim \chi^2_1$

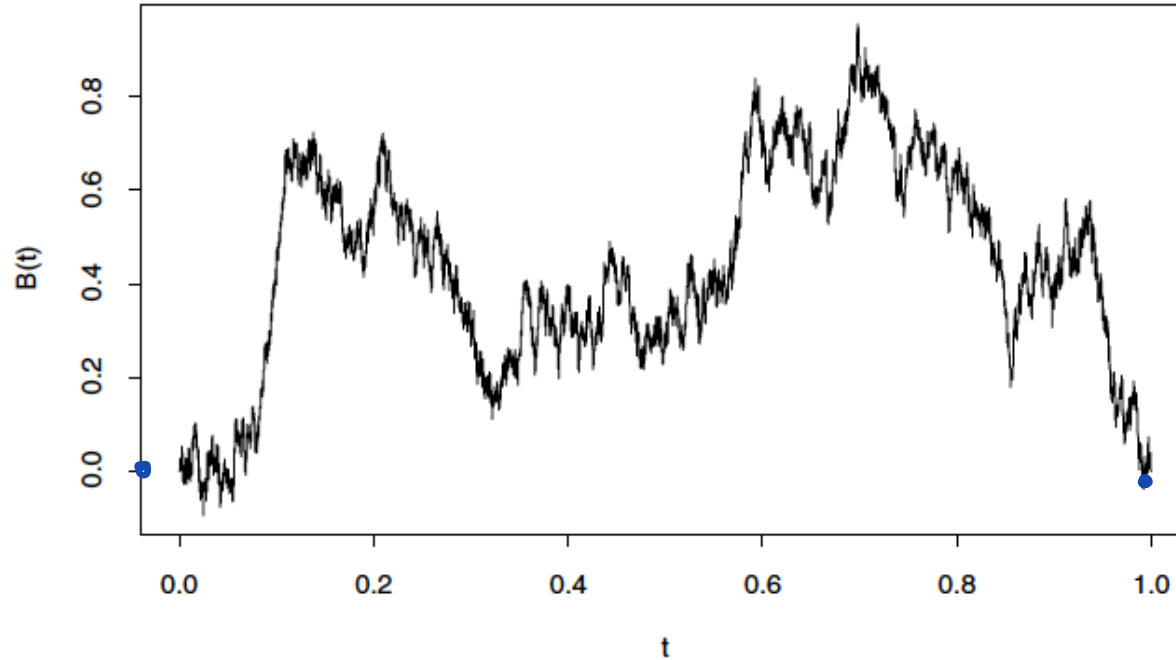


Figure 9.1 *A simulated realization of a Brownian bridge process.*

- $X \sim \text{Mult}_k(n; p)$

X_j = number of obs in category j

- $L(p) = \text{pr}(\underline{X_1 = x_1}, \dots, \underline{X_k = x_k}; p) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}; \quad \underline{\sum p_j = 1}; \quad \underline{\sum x_j = n}$
 $j = 1, \dots, k$

- $E(\underline{X}) = n \underline{p} \quad E(X_j) = np_j \quad \text{var}(X_j) = np_j(1-p_j)$

- $\text{cov}(X) = \underline{\text{Cov}(X_j, X_k) = -np_j p_k} \quad (?) \text{ not bc. } \text{AoS Thm 14.4}$

- $\hat{p} = \frac{X}{n} \quad \hat{p}_j = \frac{X_j}{n} \quad \text{and} \quad \hat{p}_k = 1 - \hat{p}_1 - \hat{p}_2 - \dots - \hat{p}_{k-1}$

- $\text{cov}(\hat{p}) = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \dots & -p_1 p_k \\ -p_1 p_2 & p_2(1-p_2) & \dots & -p_2 p_k \\ \vdots & \vdots & \ddots & \vdots \\ -p_{k-1} p_k & \dots & \dots & p_{k-1}(1-p_{k-1}) \end{pmatrix} \quad X_k = n - X_1 - X_2 - \dots - X_{k-1}$
 $k \times k \text{ matrix rank}(k-1)$

- $X \sim \text{Mult}_k(n; p)$

X_j = number of obs in category j

- $\text{pr}(X_1 = x_1, \dots, X_k = x_k; p) =$

- $E(X) =$

- $\text{cov}(X) =$

AoS Thm 14.4

- $\hat{p} =$

- $\text{cov}(\hat{p}) =$

- log-likelihood function
- Fisher information

- $H_0 : X_1, \dots, X_n$ i.i.d. $f(x; \theta)$, $x \in \mathbb{R}$; $\theta \in \mathbb{R}^s$
- Let $I_1, \dots, I_k$ be disjoint intervals on $\mathbb{R}$
- Define $N_j = \sum_{i=1}^n 1\{X_i \in I_j\}$
- $(N_1, \dots, N_k) \sim \text{Mult}_k(n, \underline{p}(\theta))$



1. χ^2 goodness-of-fit st.: $\sum_{j=1}^k \frac{\{N_j - np_j(\theta)\}^2}{np_j(\theta)} = \chi^2$ under model

if θ is unknown, est. it by max. lik $\rightarrow \tilde{\theta}$

$$\chi^2 = \sum_{j=1}^k \frac{\{N_j - np_j(\tilde{\theta})\}^2}{np_j(\tilde{\theta})}$$

$$\sum_{\text{cells}} \frac{(O - E)^2}{E} \xrightarrow{d} \chi^2_{k-s-1}$$

- $H_0 : X_1, \dots, X_n$ i.i.d. $f(x; \theta)$, $x \in \mathbb{R}$; $\theta \in \mathbb{R}^s$
- Let $I_1, \dots, I_k$ be disjoint intervals on $\mathbb{R}$
- Define $N_j = \sum_{i=1}^n \mathbf{1}\{X_i \in I_j\}$
- $(N_1, \dots, N_k) \sim$

• $L(\theta) = \prod_{j=1}^k p_j(\theta)^{N_j}$; $\ell(\theta) = \sum_{j=1}^k N_j \log\{p_j(\theta)\}$; $\tilde{\theta} = \arg \max_{\theta} \ell(\theta)$

~~$\frac{n!}{x_1! \dots x_k!} p_1(\theta)^{x_1} \dots p_k(\theta)^{x_k} = L(\theta)$~~

Under big space
(no θ)

$\hat{p}_j = \frac{N_j}{n} = \left(\frac{x_j}{n} \right)$

$\sup_{\theta \in \Theta} L(\theta)$ $\leftarrow$ X_i 's i.i.d.
 $\sup_{\theta \in \Theta_0} L(\theta) \leftarrow f(x, \theta)$

$\frac{\partial \ell(\theta)}{\partial \theta} \Big|_{\tilde{\theta}} \approx$

$\frac{\partial \ell(\theta)}{\partial \theta_s} \Big|_{\tilde{\theta}}$

$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow \tilde{\theta} \rightarrow p_j(\tilde{\theta})$

Multinomial goodness-of-fit tests

AoS 10.8; MS 9.2

- $H_0 : X_1, \dots, X_n$ i.i.d. $f(x; \theta)$, $x \in \mathbb{R}$; $\theta \in \mathbb{R}^s$
- Let $I_1, \dots, I_k$ be disjoint intervals on $\mathbb{R}$
- Define $N_j = \sum_{i=1}^n 1\{X_i \in I_j\}$
- $(N_1, \dots, N_k) \sim$

$H_0 : X_i \stackrel{iid}{\sim} f(x_i; \theta)$ Composite
 $H_1 : X_i \stackrel{iid}{\sim}$ something else

$$W_n \equiv LRT = 2 \log \frac{L(\hat{P})}{L(\tilde{\theta})} = \frac{\prod_{j=1}^k \left(\frac{N_j}{n}\right)^{N_j}}{\prod_{j=1}^k p_j(\tilde{\theta})^{N_j}}$$

MS Thm 9.2

- $L(\theta) = \prod_{j=1}^k p_j(\theta)^{N_j}$; $\ell(\theta) = \sum_{j=1}^k N_j \log\{p_j(\theta)\}$; $\tilde{\theta} = \arg \max_{\theta} \ell(\theta)$

- Theorem 10.29 (AoS): Under H_0 ,

Pearson's χ^2

$$Q = \sum_{j=1}^k \frac{\{N_j - np_j(\tilde{\theta})\}^2}{np_j(\tilde{\theta})} \xrightarrow{d} \chi_{k-1-s}^2$$

- Theorem 9.1 (MS): Under H_0

LRT

$$W = 2 \sum_{j=1}^k N_j \log \left(\frac{N_j}{np_j(\tilde{\theta})} \right) \xrightarrow{d} \chi_{k-1-s}^2$$

$k-1$ # par
in $L(\hat{P})$

~~$k-s$~~ # par = $(k-s)$

Table 9.1 Frequency of goals in First Division matches and “expected” frequency under Poisson model in Example 9.2

Goals	0	1	2	3	4	≥ 5
Frequency	252	344	180	104	28	16
Expected	248.9	326.5	214.1	93.6	30.7	10.2

under H_0

$$L(\lambda) = \prod_{j=0}^5 p_j(\lambda)^{N_j}$$

$$p_0(\lambda) = 1 - \sum_{j=1}^5 p_j(\lambda); \quad p_j(\lambda) = e^{-\lambda} \lambda^j / j!; \quad \tilde{\lambda} = 1.3118$$

$p_2(2 \text{ goals})$

$$\frac{e^{-\lambda} \lambda^2}{2}$$

$Q = 11.09$; $W = 10.87$; $\text{pr}(\chi_4^2 > [11.09, 10.87]) = [0.026, 0.028]$

$$6 - 1 - 1 = 4$$

136

4 · Likelihood

		Antigen 'B'		Total
		Absent	Present	
Antigen 'A'	Absent	'O': 202	'B': 35	237
	Present	'A': 179	'AB': 6	185
Total		381	41	422

Table 4.3 Blood groups in England (Taylor and Prior, 1938). The upper part of the table shows a cross-classification of 422 persons by presence or absence of antigens 'A' and 'B', giving the groups 'A', 'B', 'AB', 'O' of the human blood group system. The lower part shows genotypes and corresponding probabilities under one- and two-locus models. See Example 4.38 for details.

Group	Two-locus model		One-locus model	
	Genotype	Probability	Genotype	Probability
'A'	(AA; bb), (Aa; bb)	$\alpha(1 - \beta)$	(AA), (AO)	$\frac{\lambda_A^2}{\lambda_A^2 + 2\lambda_A\lambda_O}$
'B'	(aa; BB), (aa; Bb)	$(1 - \alpha)\beta$	(BB), (BO)	$\frac{\lambda_B^2}{\lambda_B^2 + 2\lambda_B\lambda_O}$
'AB'	(AA; BB), (Aa; BB), (AA; Bb), (Aa; Bb)	$\alpha\beta$	(AB)	$2\lambda_A\lambda_B$
'O'	(aa; bb)	$(1 - \alpha)(1 - \beta)$	(OO)	λ_O^2

$Q = 15.73; W = 17.66$ (two-locus)

$p < 10^{-5}$

$Q = 2.82; W = 3.17$ (single locus)

$p = 0.09; 0.07$

$$\sum_{j=1}^k (n_{j.} - np_{j.}(\theta))^2 / np_{j.}(\theta)$$

LRT

! better for than the old one