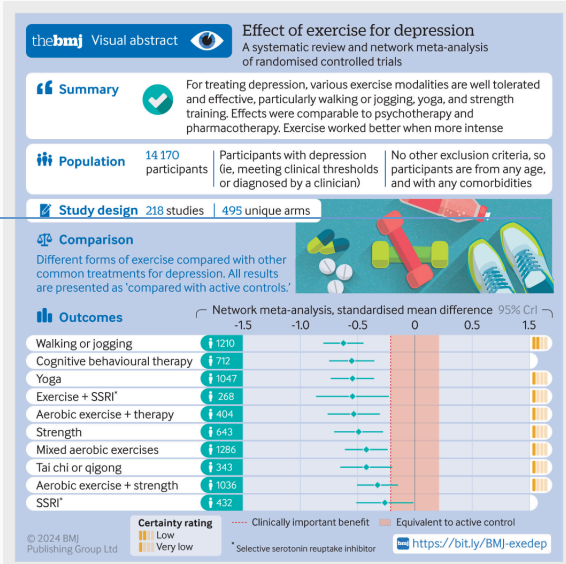
Mathematical Statistics II

STA2212H S LEC9101

Week 8

March 4 2025



thebmj Visual abstract

Effect of exercise for depression

A systematic review and network meta-analysis of randomised controlled trials

Summary

For treating depression, various exercise modalities are well tolerated and effective, particularly walking or jogging, yoga, and strength training. Effects were comparable to psychotherapy and pharmacotherapy. Exercise worked better when more intense

Population

14 170 participants | Participants with depression (ie, meeting clinical thresholds or diagnosed by a clinician) | No other exclusion criteria, so participants are from any age, and with any comorbidities

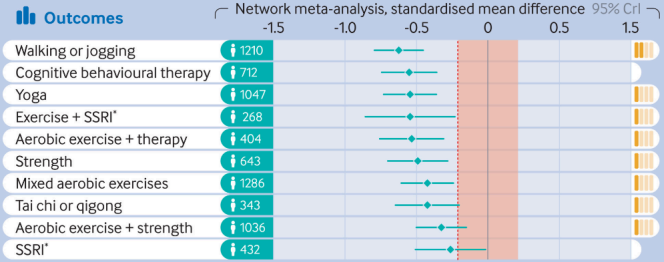
Study design

218 studies | 495 unique arms

Comparison

Different forms of exercise compared with other common treatments for depression. All results are presented as 'compared with active controls.'

Outcomes



Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials

Michael Noetel,[1] Taren Sanders,[2] Daniel Gallardo-Gómez,[3] Paul Taylor,[4] Borja del Pozo Cruz,[5,6] Daniel van den Hoek,[7] Jordan J Smith,[8] John Mahoney,[9] Jemima Spathis,[9] Mark Moresi,[4] Rebecca Pagano,[10] Lisa Pagano,[11] Roberta Vasconcellos,[2] Hugh Arnott,[2] Benjamin Varley,[12] Philip Parker,[13] Stuart Biddle,[14,15] Chris Lonsdale[13]

For numbered affiliations see end of the article

Correspondence to: M Noetel
m.noetel@uq.edu.au
(or @mnoetel on Twitter;
ORCID 0000-0002-6563-8203)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2024;384:e075847
http://dx.doi.org/10.1136/bmj-2023-075847

Accepted: 15 January 2024

ABSTRACT

OBJECTIVE

To identify the optimal dose and modality of exercise for treating major depressive disorder, compared with psychotherapy, antidepressants, and control conditions.

DESIGN

Systematic review and network meta-analysis.

METHODS

Screening, data extraction, coding, and risk of bias assessment were performed independently and in duplicate. Bayesian arm based, multilevel network meta-analyses were performed for the primary

g −0.42, −0.65 to −0.21). The effects of exercise were proportional to the intensity prescribed. Strength training and yoga appeared to be the most acceptable modalities. Results appeared robust to publication bias, but only one study met the Cochrane criteria for low risk of bias. As a result, confidence in accordance with CINeMA was low for walking or jogging and very low for other treatments.

CONCLUSIONS

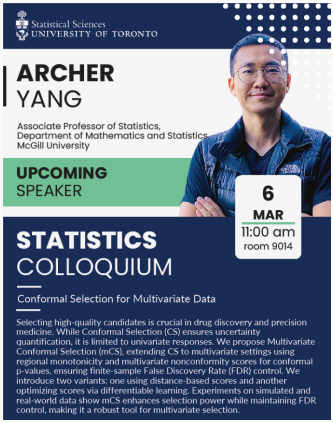
Exercise is an effective treatment for depression, with walking or jogging, yoga, and strength training more effective than other exercises, particularly when intense. Yoga and strength training were well

Today

1. Recap Feb 25 Formal testing, NP Lemma, size and power, *p*-values
2. Significance testing, nonparametric tests
3. Diagnostic testing
4. Multiple testing
5. Project Selections and Guidelines, HW 7

Upcoming seminar

Department Seminar Thursday March 6 11.00 – 12.00
Hydro Building, Room 9014 Conformal selection
Archer Yang, McGill University



Project Guidelines

link

Project Guidelines STA 2212S: Mathematical Statistics II 2025

Presentation on April 1, 2025.
Report submission due April 16, 2025.

Part 1: Presentation [10 points]

On the last day of class (April 1), you will present your final project. This includes:

- Emailing a .pdf version of your team's slide deck pdf to `nancym.reid@utoronto.ca` by 09.00 April 1. You are responsible for the slides corresponding to your sections of the write-up. Please email **one** complete version for each team.
- Presenting the slides in no more than 10 minutes; each team member to present for no more than 5 minutes.

Recap: hypothesis testing

$X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta), \theta \in \Theta \subset \mathbb{R}^p$

- Null and alternative hypotheses
- Size and power
- Test statistic $T = t(\boldsymbol{X})$ testing function
- Rejection region $\{\boldsymbol{x} : T \geq c_\alpha\}$
- P-value $\mathrm{pr}_{H_0}(T \geq t^{obs})$

Recap: Neyman-Pearson lemma

MS Thm 7.2; AoS Thm 10.30

- for testing simple H_0 against simple H_1
- test statistic

$$T = \frac{L(\theta_1; \boldsymbol{x})}{L(\theta_0; \boldsymbol{s})} = \frac{f(\boldsymbol{x}; \theta_1)}{f(\boldsymbol{x}; \theta_0)}$$

- critical region

$$\{\boldsymbol{x} : t(\boldsymbol{x}) \geq k\}$$

- Choose $k = k_\alpha$ to satisfy

$$\mathrm{pr}_{H_0}(T \geq k_\alpha) = \alpha$$

- This test is a most powerful test of H_0 against H_1 at level α.

A neatly-typed proof (from MS)

Let $\phi(\boldsymbol{x})$ be the test function for the test based on T.
Let $\psi(\boldsymbol{x})$ be any other function that maps $\boldsymbol{x}$ to $[0, 1]$.
If

$$\mathrm{E}_{H_0}\{\psi(\boldsymbol{X})\} \leq \mathrm{E}_{H_0}\{\phi(\boldsymbol{X})\} = \alpha$$

then it must follow that

$$\mathrm{E}_{H_1}\{\psi(\boldsymbol{X})\} \leq \mathrm{E}_{H_1}\{\phi(\boldsymbol{X})\}$$

Proof: $\forall\, \boldsymbol{x}$,

$$\psi(\boldsymbol{x})\{f_1(\boldsymbol{x}) - kf_0(\boldsymbol{x})\} \leq \phi(\boldsymbol{x})\{f_1(\boldsymbol{x}) - kf_0(\boldsymbol{x})\}$$

Integrate and re-arrange terms to get the result

A neatly-typed proof (from SM 7.3)

Let R be the rejection region for the test based on $T = f_1(\boldsymbol{x})/f_0(\boldsymbol{s})$

$$R = \{\boldsymbol{x} : T(\boldsymbol{x}) \geq k_\alpha\}$$

Let R' be some other rejection region also of size α $\leq \alpha$

$$\alpha = \int_R f_0(\boldsymbol{x})d\boldsymbol{x} = \int_{R'} f_0(\boldsymbol{x})d\boldsymbol{x}$$

$$\int_{R-R'} f_0(\boldsymbol{x})d\boldsymbol{x} = \int_{R'-R} f_0(\boldsymbol{x})d\boldsymbol{x}$$

On LHS $f_1(\boldsymbol{x}) \geq k_\alpha f_0(\boldsymbol{x})$. $R - R' \subset R$

On RHS $f_1(\boldsymbol{x}) < k_\alpha f_0(\boldsymbol{x})$. $R' - R \subset R^c$

$$\int_{R-R'} f_1(\boldsymbol{x})d\boldsymbol{x} \geq \int_{R'-R} f_1(\boldsymbol{x})d\boldsymbol{x}$$

Add integral over intersection $R \cap R'$

Choosing test statistics

1. Optimal choice – Neyman-Pearson lemma — Might be UMP (HW 7)
2. Pragmatic choice – likelihood-based test statistics
3. Pragmatic choice – nonparametric test statistics

(a) Need to know distribution of test statistic under H_0

(b) Test statistic should be large when H_0 is not true — in probability

(c) Test statistic should have maximum power to detect departures from H_0

Choosing test statistics

1. Optimal choice – Neyman-Pearson lemma

Might be UMP (HW 7)

2. Pragmatic choice – likelihood-based test statistics

3. Pragmatic choice – nonparametric test statistics

Choosing test statistics

1. Optimal choice – Neyman-Pearson lemma

Might be UMP (HW 7)

2. Pragmatic choice – likelihood-based test statistics

3. Pragmatic choice – nonparametric test statistics

Example: Sign test

SM Ex.7.26

- $X_1, \ldots, X_n$ i.i.d. $F(\cdot)$
- $H_0 : \mu = \mu_0$, $\mu = F^{-1}(1/2)$ median of distribution
- $H_1 : \mu > \mu_0$

both H composite

- test statistic

$$T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$$

- under H_0,

$$T \sim Binom(n, 1/2)$$

- p-value

$$p_{obs} = \mathrm{pr}_{H_0}(T \geq t_{obs}) = \sum_{r=t_{obs}}^{n} \binom{n}{r} \frac{1}{2^n} \doteq 1 - \Phi\left\{\frac{2(t_{obs} - n/2)}{n^{1/2}}\right\}.$$

Power of the sign test

SM Ex.7.30

- $H_0 : \mu = \mu_0 \qquad H_1 : \mu > \mu_0$ — $\mu = F^{-1}(1/2)$
- Test statistic $T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$
- Rejection region $R = \{T \geq c_\alpha\}$
- $c_\alpha \approx n/2 - n^{1/2} z_\alpha / 2$ — Normal approx

- Power = $\mathrm{pr}_{H_1}(\text{reject } H_0) = \mathrm{pr}_{H_1}(T \geq c_\alpha)$ — Need distribution of T under H_1
- to calculate power we need values for μ and for F

Power of the sign test

SM Ex.7.30

- $H_0 : \mu = \mu_0 \qquad H_1 : \mu > \mu_0$ $\quad \mu = F^{-1}(1/2)$
- Test statistic $T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$
- Rejection region $R = \{T \geq c_\alpha\}$
- $c_\alpha \approx n/2 - n^{1/2} z_\alpha / 2$ Normal approx

- Power = $\mathrm{pr}_{H_1}(\text{reject } H_0) = \mathrm{pr}_{H_1}(T \geq c_\alpha)$ Need distribution of T under H_1
- to calculate power we need values for μ and for F

- SM assumes F is $N(\mu, \sigma^2)$, so $\quad \delta = n^{1/2}(\mu_1 - \mu_0)/\sigma$

$$\mathrm{pr}_{\mu_1}(T \geq c_\alpha) = \mathrm{pr}_{\mu_1}(T \geq n/2 - n^{1/2} z_\alpha/2) \doteq \Phi\left\{ \frac{n\Phi(n^{-1/2}\delta) - n/2 + n^{1/2} z_\alpha}{[n\Phi(n^{-1/2}\delta)\{1 - \Phi(n^{-1/2}\}]} \right\}$$
$$\doteq \Phi\{z_\alpha + \delta(2/\pi)^{1/2}\}$$

- test based on $\bar{X}$ has power $\Phi(z_\alpha + \delta)$

... power of sign test

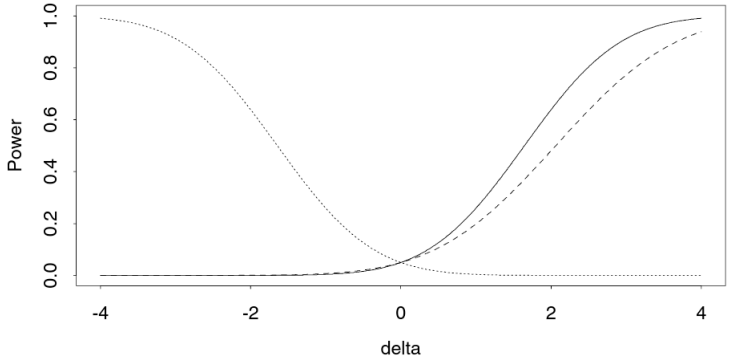


Figure 7.6 Power functions for a test of whether the mean of a $N(\mu, \sigma^2)$ random sample of size n equals μ_0 against the alternative $\mu = \mu_1$, as a function of $\delta = n^{1/2}(\mu_1 - \mu_0)/\sigma$. The test size is $\alpha = 0.05$. The solid curve is the power function for a test of $\mu_1 > \mu_0$ based on $\overline{y}$, and the dashed line is the power function for the sign test. Both critical regions are of form $\overline{y} > t_\alpha$. The dotted curve is the power function for $\overline{y}$ when the critical region is $\overline{y} < t_\alpha$.

Permutation test

leukemia data (EH): $X_1, \ldots, X_{47}$; $Y_1, \ldots, Y_{25}$

AoS Ex. 10.20

```
oneline
          ALL    ALL.1     ALL.2     ALL.3     ALL.4     ALL.5     ALL.6     ALL.7
136 0.9186952 1.634002 0.4595867 0.6379664 0.3440379 0.8614784 0.5132176 0.9790902
        ALL.8     ALL.9    ALL.10    ALL.11  ALL.12    ALL.13    ALL.14    ALL.15   ALL.16
136 0.2105782 0.8016072 0.6006949 0.3614374 1.04632 0.9697635 0.4873159 0.4976364 1.101717
       ALL.17   ALL.18   ALL.19       AML     AML.1     AML.2     AML.3     AML.4     AML.5
136 0.8563937 0.661415 0.817711 0.7671718 0.9793741 1.425479 1.074389 0.9839282 0.9859271
        AML.6     AML.7    AML.8     AML.9    AML.10    AML.11    AML.12   AML.13    AML.20
136 0.3247027 0.7110302 1.09625 0.9675151 0.975123 0.7775957 0.9472205 1.261352 0.5679544
       ALL.21    ALL.22    ALL.23    ALL.24    ALL.25    ALL.26   ALL.27    ALL.28
136 0.8462901 0.8838616 0.7239931 0.7327029 0.7823618 0.5435396 0.832537 0.5527333
       ALL.29    ALL.30    ALL.31    ALL.32   ALL.33    ALL.34    ALL.35    ALL.36
136 0.7327029 0.5510955 0.8214005 0.6418498 0.720798 0.5830999 0.7657568 0.5262976
      ALL.37    ALL.38    ALL.39   ALL.40    ALL.41    ALL.42    ALL.43    ALL.44
136 1.466999 0.5445589 0.5725049 1.362768 0.8533535 0.8132982 0.8538596 0.5689876
       ALL.45   ALL.46    AML.14    AML.15   AML.16   AML.17    AML.18    AML.19   AML.20
136 0.6930355 1.067526 0.9677959 0.9338141 1.138926 1.161753 0.6242354 0.6590103 1.215186
       AML.21   AML.22   AML.23    AML.24
136 0.9340861 1.310376 0.771426 0.7556606
```

$H_0 : F_X = F_Y \qquad H_1 \qquad T = T(\boldsymbol{X}, \boldsymbol{Y}) =$

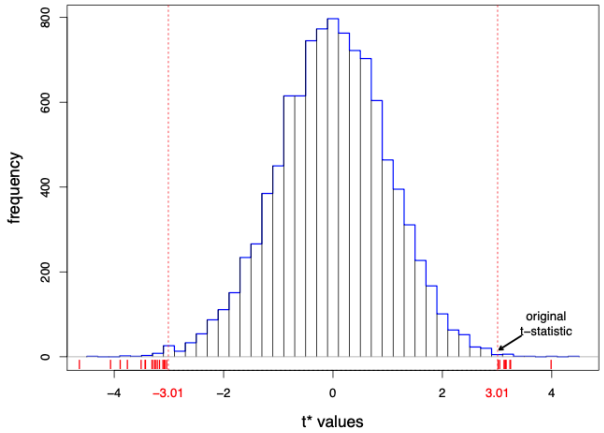


Figure 4.3 10,000 permutation t^*-values for testing **ALL** vs **AML**, for gene 136 in the **leukemia** data of Figure 1.3. Of these, 26 t^*-values (red ticks) exceeded in absolute value the observed t-statistic 3.01, giving permutation significance level 0.0026.

Hypothesis tests and significance tests

- Hypothesis tests typically means:
  - H_0, H_1
  - critical/rejection region $R \subset \mathcal{X}$,
  - level α, power $1 - \beta$
  - conclusion: "reject H_0 at level α" or "do not reject H_0 at level α"
  - planning: maximize power for some relevant alternative minimize type II error

Hypothesis tests and significance tests

- Hypothesis tests typically means:
  - H_0, H_1
  - critical/rejection region $R \subset \mathcal{X}$,
  - level α, power $1 - \beta$
  - conclusion: "reject H_0 at level α" or "do not reject H_0 at level α"
  - planning: maximize power for some relevant alternative — minimize type II error

- Significance tests typically means:
  - H_0,
  - test statistic T
  - observed value t^{obs},
  - p-value $p^{obs} = \Pr(T \geq t^{obs}; H_0)$
  - alternative hypothesis often only implicit — large T points to alternative

Diagnostic testing

Wikipedia

1. Hypothesis testing

AoS Table 10.1

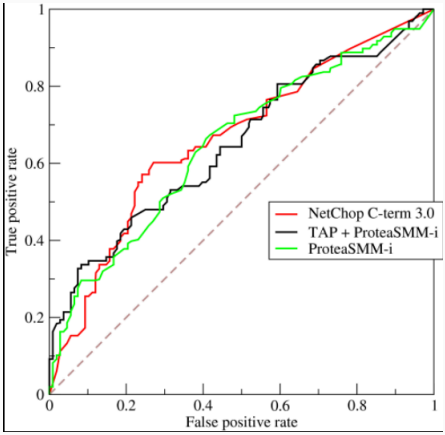
truth	H_0 not rejected	H_0 rejected
H_0 true		type 1 error
H_1 true	type 2 error	

2. Diagnostic testing

link

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

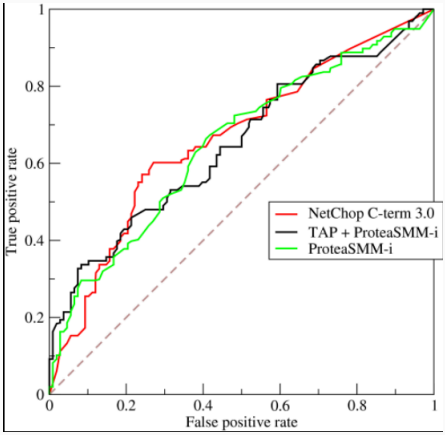
Diagnostic testing and ROC



True positive rate =
sensitivity =
TP/P

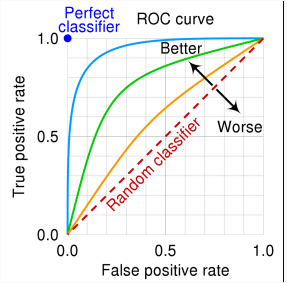
False positive rate =
1− specificity =
$1 - TN/N$

Diagnostic testing and ROC



True positive rate =
sensitivity =
TP/P

False positive rate =
1− specificity =
$1 - TN/N$



Rapid flow test, care home link

truth	test negative	test positive	
C19 neg	114,993	101	115,094
C19 pos	371	128	499

Sensitivity = TP/P = 128/499 = 0.257
Specificity = TN/N = 114,993/115094 =0.999

Cochrane review meta-analysis

“consistently high specificities”

“sensitivity varied widely: average sensitivities by brand ranged from 34.3% to 91.3% ”

1. Hypothesis testing

AoS Table 10.1

		H_0 not rejected	H_0 rejected
truth	H_0 true		type 1 error
	H_1 true	type 2 error	

3. Multiple testing

AoS Table 10.2

		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

FDP, FDR

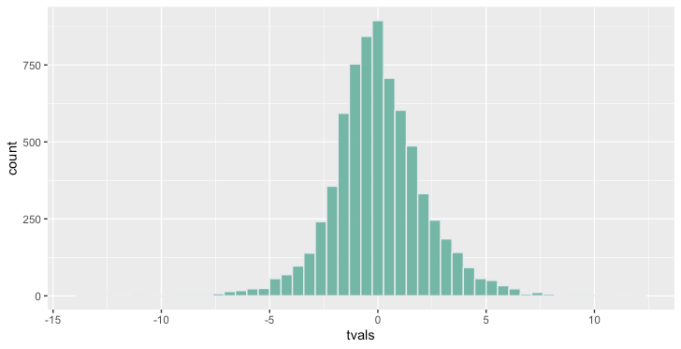
```
leukemia_big <- read.csv
  ("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")
  dim(leukemia_big)
  [1] 7128   72
```

- each row is a different gene; 47 AML responses and 25 ALL responses
- we could compute 7128 t-statistics for the mean difference between AML and ALL

```
tvals <- rep(0,7128)
for (i in 1:7128){
  leukemia_big[i,] %>% select(starts_with("ALL")) %>% as.numeric() -> x
  leukemia_big[i,] %>% select(starts_with("AML")) %>% as.numeric() -> y
  tvals[i] <- t.test(x,y,var.equal=T)$statistic
  }
```

Multiple testing

EH 1.2, 15.2




```
summary(tvals)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-13.52611  -1.20672  -0.08406   0.02308   1.20886  12.26065
```

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

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

$FDR = \mathrm{E}(FDP)$

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

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

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

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

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

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

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

$$FDR(BH_q) = \pi_o q \leq q, \qquad \text{where } \pi_o = m_o/m$$

π_o unknown but close to 1

- change the bound under dependence

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

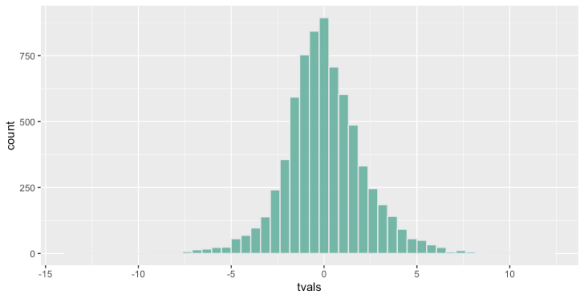
Example

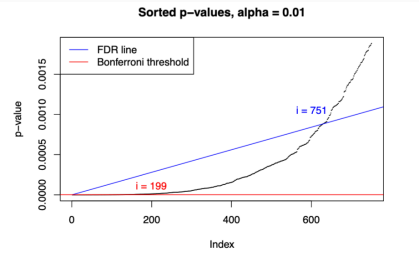
AoS Ex.10.28

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

Multiple testing

EH 1.2, 15.2

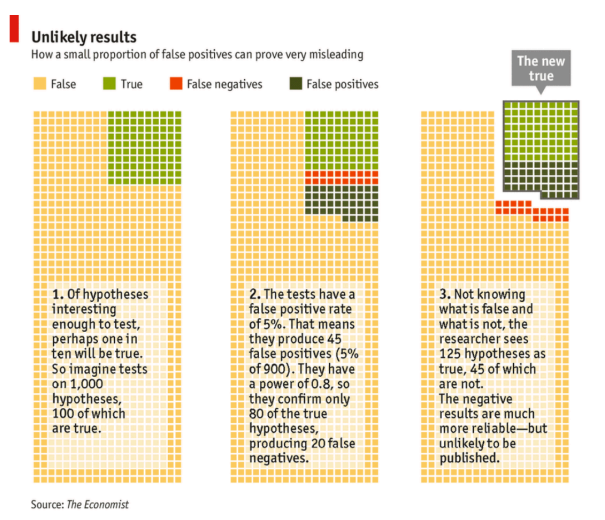




The figure above shows sorted p-values of the $N = 7128$ t-tests. The red line corresponds to the threshold α/N from the Bonferroni method, and the blue line is the FDR line $(i/N)\alpha$. The

```
> summary(ttest)
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-13.52611  -1.20672  -0.08406   0.02308   1.20886  12.26065
```

Multiple testing



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

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

The Annals of Statistics
2006, Vol. 34, No. 4, 1827–1849
DOI: 10.1214/009053606000000425

ON THE BENJAMINI–HOCHBERG METHOD

BY J. A. FERREIRA[1] AND A. H. ZWINDERMAN

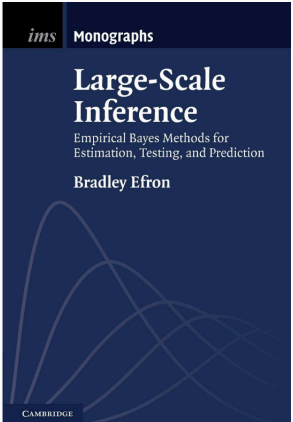
University of Amsterdam

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 $X = \{X_1, X_2, \ldots, X_m\}$ 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 $X_1, X_2, \ldots, X_{m_0}$ has distribution function (d.f.) F and $X_{m_0+1}, \ldots, X_m$ 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 $S_m = \#(\mathcal{R} \cap \{X_1, \ldots, X_{m_0}\})$, 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



- $X_1, \ldots, X_n$ i.i.d.
- $H_0 : X_i \sim f(x; \theta)$; $H_1 : X_i$ arbitrary distribution
- Define k sets $A_1, \ldots, A_k$ s.t.

$$\mathrm{pr}(X_i \in \cup_{j=1}^{k} A_j\} = 1$$

- Define

$$Y_j = \sum_{i=1}^{n} 1\{X_i \in A_j\}$$

number of obs in category j

Multinomial goodness of fit tests

MS 9.2; AoS 10.8

- $X_1, \ldots, X_n$ i.i.d.
- $H_0 : X_i \sim f(x; \theta)$; $H_1 : X_i$ arbitrary distribution
- Define k sets $A_1, \ldots, A_k$ s.t.

$$\mathrm{pr}(X_i \in \cup_{j=1}^{k} A_j\} = 1$$

- Define

$$Y_j = \sum_{i=1}^{n} 1\{X_i \in A_j\}$$

number of obs in category j

- $Y = (Y_1, \ldots, Y_k) \sim Mult_k(n; p)$
- $\mathrm{pr}(Y_1 = y_1, \ldots, Y_k = y_k; p) =$
- $H_0 : p = p(\theta)$; $H_1 : p$ arbitrary

Multinomial goodness of fit tests

MS 9.2, AoS 10.8

- log-likelihood function

- generalized likelihood ratio test

- log-likelihood function

- generalized likelihood ratio test

- Theorem 9.1 (MS): Under H_0 $p = \dim(\theta)$

$$W = 2\sum_{j=1}^{k} Y_j \log\left(\frac{Y_j}{np_j(\tilde{\theta})}\right) \xrightarrow{d} \chi^2_{k-1-p}$$

- log-likelihood function

- generalized likelihood ratio test

- Theorem 9.1 (MS): Under H_0 $p = \dim(\theta)$

$$W = 2\sum_{j=1}^{k} Y_j \log\left(\frac{Y_j}{np_j(\tilde{\theta})}\right) \xrightarrow{d} \chi^2_{k-1-p}$$

- Theorem 92. (MS): Under H_0

$$Q = \sum_{j=1}^{k} \frac{\{Y_j - np_j(\hat{\theta})\}^2}{np_j(\hat{\theta})} \xrightarrow{d} \chi^2_{k-1-p}$$

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

$$p_0(\lambda) = 1 - \sum_{j=0}^{4} p_j(\lambda); \quad p_j(\lambda) = e^{-\lambda}\lambda^j/j!, \quad \tilde{\lambda} = 1.3118$$

$$Q = 11.09; \quad W = 10.87; \quad \mathrm{pr}(\chi^2_4 > [11.09, 10.87]) = [0.026, 0.028]$$

		Antigen 'B'		
		Absent	Present	Total
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.

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

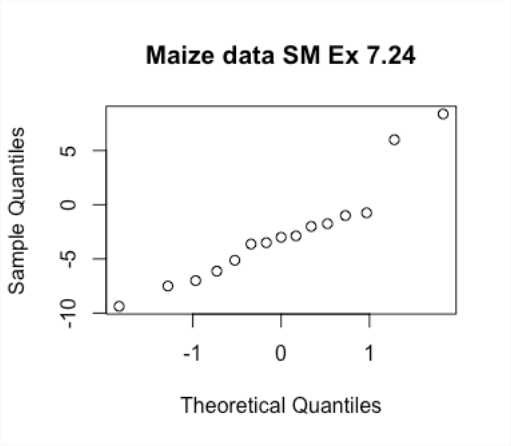
$p < 10^{-5}$

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

$p = 0.09; 0.07$

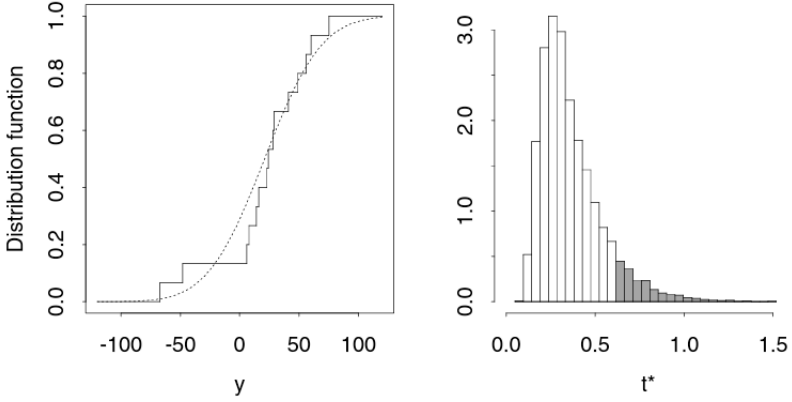
Smooth goodness-of-fit tests

MS 9, SM p.327-9




```
library(SMPracticals)
data(darwin)
cross <- seq(1,30,by=2)
self <- cross+1
diffs <- darwin[self,4]-darwin[cross,4]
qqnorm(diffs)
```

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

Smooth goodness-of-fit tests

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

- three test statistics:
  1. $\sup_t |\widehat{F}_n(t) - F_0(t)|$
  2. $\int \{\widehat{F}_n(t) - F_0(t)\}^2 dF_0(t)$
  3. $\int \frac{\{\widehat{F}_n(t) - F_0(t)\}^2}{F_0(t)\{1 - F_0(t)\}} dF_0(t)$

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

Smooth goodness-of-fit tests

- Special case $H_0 : F(t) = F_0(t) = t$ $\quad X_i \sim U(0,1)$
- Recall

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

- What about distribution of

$\sup_t |\widehat{F}_n(t) - t|$ $\qquad$ $\int \{\widehat{F}_n(t) - t\}^2 dt$ $\qquad$ $\int \dfrac{\{\widehat{F}_n(t) - t\}^2}{F_0(t)\{1-t\}} dt$

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

- Special case $H_0 : F(t) = F_0(t) = t$ $X_i \sim U(0,1)$
- Recall

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

- What about distribution of

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

- need joint density of $\widehat{F}_n(t)\ \forall\ t$
- define stochastic process $B_n(t) = \sqrt{n}(\widehat{F}_n(t) - t)$
- vector $(B_n(t_1), \ldots, B_n(t_k)) \xrightarrow{d} N_k(0, C), \quad C_{ij} = \min(t_i, t_j) - t_i t_j$ MS 9.3
- a Brownian bridge is a continuous function on $(0,1)$
  with all finite-dimensional distributions as above

Smooth goodness-of-fit tests

- 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

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

Smooth goodness-of-fit tests

- Kolmogorov-Smirnov test

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

- 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

$$K_n \xrightarrow{d} K, \qquad W_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2\pi^2}, \qquad A_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)}$$

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

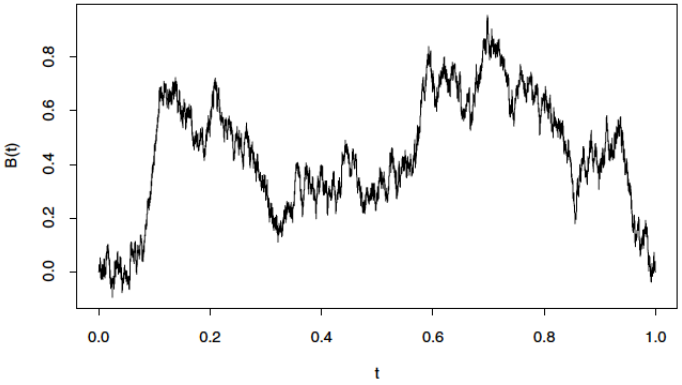


Figure 9.1 *A simulated realization of a Brownian bridge process.*