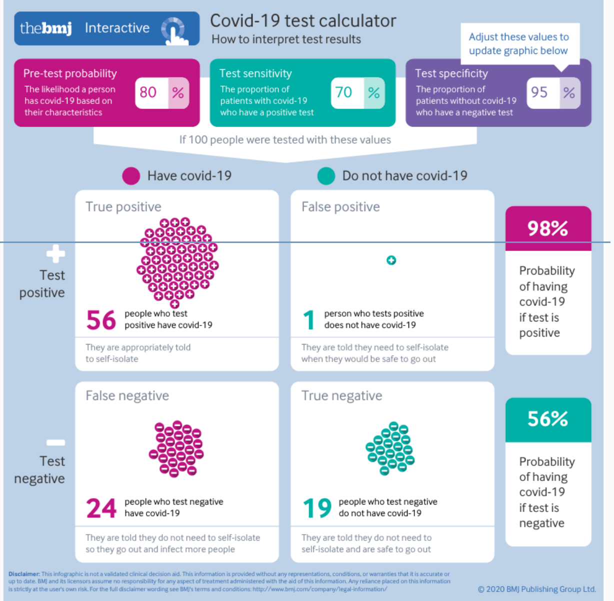
Mathematical Statistics II

STA2212H S LEC9101

Week 5

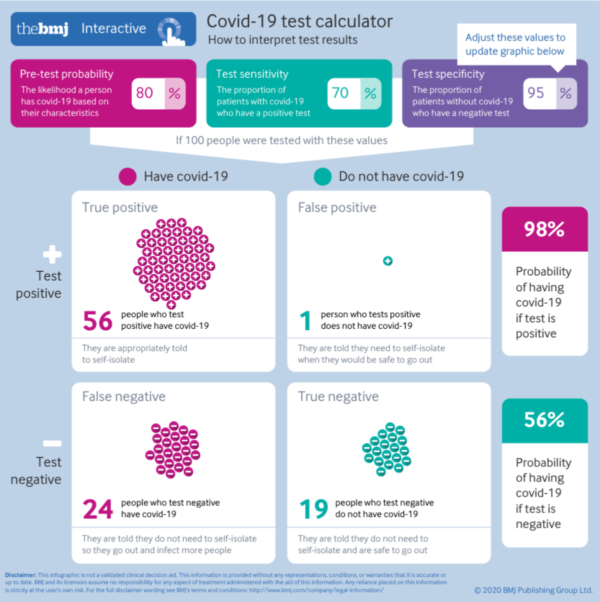
February 10 2021

Start recording!



Online calculator

Link to Calculator



- Confidence intervals – approximate and exact; relationship to testing; optimal confidence intervals; connection to size and power

length

t-test

se N r.v.

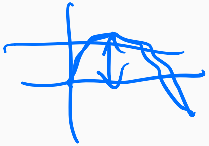
Wald test

$\hat{\theta} \pm \hat{se} \cdot z_{1-\alpha/2}$

LRT

- Likelihood-based confidence intervals and regions
- pure significance tests; simple and composite H_0

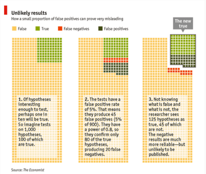
$T \sim H_0$



- goodness-of-fit tests
- empirical cumulative distribution function $\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i \leq t\}$

$E\hat{F}_n(t) = F(t)$

- introduction to multiple testing



Today

Start Recording

1. Friday Feb 12
2. hypothesis vs significance testing
3. diagnostic testing
4. Benjamini-Hochberg method

HW 4 updated

- February 25-26 Workshop

- Feb 22 3.00 pm EST Joshua Speagle
  "Mapping the Milky Way in the Age of Gaia" Link
  Data Science ARES

- Feb 25 1.00 pm EST Dylan Small
  CANSSI National Seminar Series (Journal Club; Slack)

info@canssi.ca

Start Recording

1. ~~Friday Feb 12~~ Reading week: no office hours NR on Zoom 10-12
2. hypothesis vs significance testing
3. diagnostic testing
4. Benjamini-Hochberg method

HW 5 posted Fri
due Feb 25

- February 25-26 Workshop
- Feb 22 3.00 pm EST Joshua Speagle
  "Mapping the Milky Way in the Age of Gaia" Link
  Data Science ARES
- Feb 25 1.00 pm EST Dylan Small
  CANSSI National Seminar Series (Journal Club; Slack)

Toronto Data Workshop on Reproducibility

A two-day workshop focusing on reproducibility in data-centric analysis. Thursday and Friday 25-26 February 2021. Free and hosted via Zoom. All welcome! Register here.

Hypothesis tests and significance tests

- Hypothesis tests typically means:
  - H_0, H_1
  - critical/rejection region $R \subset \mathcal{X}$, sample space
  - level α, power $1-\beta$ #1 error ; #2 error
  - conclusion: "reject H_0 at level α" or "do not reject H_0 at level α" X
  - planning: maximize power for some relevant alternative $\Longleftrightarrow$ n ? minimize type II error

critical region $= \{x ; x \in R\}$

often $\rightarrow \{ t(x) \in I \}$

Hypothesis tests and significance tests

- Hypothesis tests typically means: "reject / not"
  - 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: "p-value is . . ."
  - 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 some 'away' from Ho alternative

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

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0"

] sig → hyp. t.

"reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0

] hyp t → sig. test

Definition 10.11 in AoS

$p^{obs} = 0.03$

H.t. @ $\alpha = 0.03$

at edge of R

in R if $\alpha = 0.05$

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0"

Rice: α-level

"reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0

Definition 10.11 in AoS

Rice, Exercise 9.11.5

Mini-quiz – True or False?

at

1. The significance level of a statistical test is equal to the probability the the null hypothesis is true

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0" "reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0 Definition 10.11 in AoS

Mini-quiz – True or False? Rice, Exercise 9.11.5

1. The significance level of a statistical test is equal to the probability the the null hypothesis is true
2. If the significance level of a test is decreased, the power would be expected to increase ? F

$\alpha \downarrow$ $\beta \uparrow$ $1-\beta \downarrow$

$pr(rej.\ H_0 ; H_0)$

1.96
.05 - level test
T^{old}

2.38
.01 - level test
T^{new}

$pr(rej\ H_0 ; H_1)$

n : power = 80%, say

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0" — "reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0 — Definition 10.11 in AoS

Mini-quiz – True or False? — Rice, Exercise 9.11.5

1. The significance level of a statistical test is equal to the probability the the null hypothesis is true
2. If the significance level of a test is decreased, the power would be expected to increase
3. If the test is rejected at level α, the probability that the null hypothesis is true equals α.

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0" "reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0 Definition 10.11 in AoS

Mini-quiz – True or False? Rice, Exercise 9.11.5

1. The significance level of a statistical test is equal to the probability the the null hypothesis is true
2. If the significance level of a test is decreased, the power would be expected to increase
3. If the test is rejected at level α, the probability that the null hypothesis is true equals α.
4. The probability that the null hypothesis is falsely rejected is equal to the power of the test

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0" "reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0 Definition 10.11 in AoS

Mini-quiz – True or False? Rice, Exercise 9.11.5

1. The significance level of a statistical test is equal to the probability the the null hypothesis is true
2. If the significance level of a test is decreased, the power would be expected to increase
3. If the test is rejected at level α, the probability that the null hypothesis is true equals α.
4. The probability that the null hypothesis is falsely rejected is equal to the power of the test
5. A type I error occurs when the test statistic falls in the rejection region of the test

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0" "reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0 Definition 10.11 in AoS

Mini-quiz – True or False? Rice, Exercise 9.11.5

1. The significance level of a statistical test is equal to the probability the the null hypothesis is true
2. If the significance level of a test is decreased, the power would be expected to increase
3. If the test is rejected at level α, the probability that the null hypothesis is true equals α.
4. The probability that the null hypothesis is falsely rejected is equal to the power of the test
5. A type I error occurs when the test statistic falls in the rejection region of the test .05 α
6. A type II error is more serious than a type I error convention is to for α small(ish)
   & β hence a bit larger
   .20 → power 0.8 (.9)

Hypothesis tests and significance tests

- overlap: sometimes (not recommended)
  $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0" "reject H_0"

- overlap: p^{obs} is the smallest α-level
  at which the corresponding hypothesis test would reject H_0 Definition 10.11 in AoS

Mini-quiz – True or False? Rice, Exercise 9.11.5

1. The significance level of a statistical test is equal to the probability the the null hypothesis is true
2. If the significance level of a test is decreased, the power would be expected to increase
3. If the test is rejected at level α, the probability that the null hypothesis is true equals α.
4. The probability that the null hypothesis is falsely rejected is equal to the power of the test
5. A type I error occurs when the test statistic falls in the rejection region of the test
6. A type II error is more serious than a type I error
7. The power of a test is determined by the null distribution of the test statistic ✓ False

1. Hypothesis testing

1 test

		H_0 not rejected	H_0 rejected	
truth	H_0 true	$1-\alpha$ ✓	type 1 error α	1
	H_1 true	type 2 error β	$1-\beta$ ✓	1

← textbook

2. Diagnostic testing

many tests

link

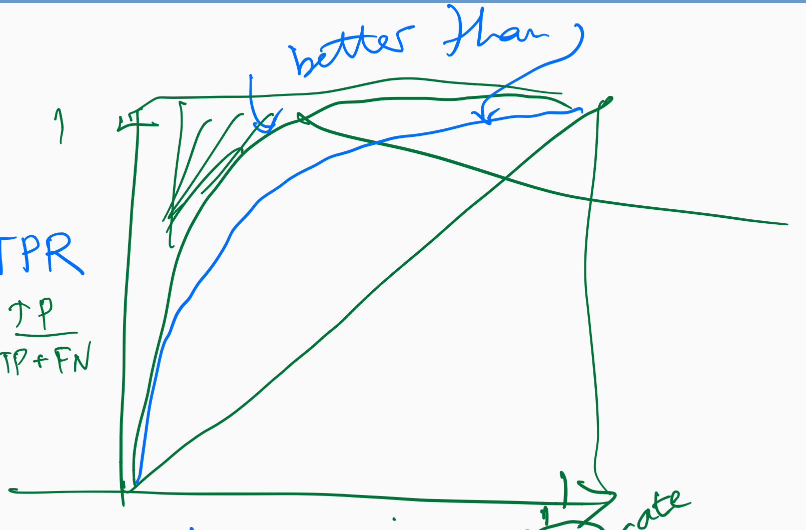
		test negative	test positive	
truth	C19 neg	TN ✓	α FP ×	N
	C19 pos	FN β	TP ✓	P
				N+P

$? \; \frac{TN}{TN+FP} = \frac{TN}{N}$ = true-ve rate = specificity

$? \; \frac{TP}{TP+FN} = \frac{TP}{P}$

true +ve RATE
= sensitivity

Diagnostic testing and ROC



TPR $= \frac{TP}{P} = \frac{TP}{TP+FN}$

receiver operating characteristic

(size / power)

truth		test −	test +ve	
	c19 −	TN	FP	N
	c19 +	FN	TP	P

FPR rate $= 1 - TNR$

$= \frac{FP}{FP+TN}$

$= \frac{FP}{N}$

	do not rej H_0	rej H_0
H_0	$1-\alpha$	α
H_1	β	$1-\beta$

2. Diagnostic testing

link

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

✓

$\hat{=} 0$ if $R=0$

$\frac{V}{R}$ False disc. proportion

$E\left(\frac{V}{R}\right) = \text{FD Rate}$

3. Multiple testing

		H_0 not rejected	H_0 rejected	
truth	H_0 true	U	small V ✗	m_0 ?
	H_1 true	T ✗	S large	m_1 ?
		$m - R$	✓ R	m

Lots of H_{0i} $i=1,\dots,m$

p_i $i=1,\dots,m$

Diagnostic testing

→ it's possible to have FDR $\leq q \simeq 0.1$

when FWER needs a cutoff of 10^{-3}, 10^{-4} (m)

reported p-value is $\frac{p^{obs}}{m}$

α - level to control $\Pr\{$ never reject any false $H_{0i}\}$ $\leq \alpha$ need $\frac{\alpha}{m}$ critical value

FWER

$z_{\alpha/2} = 1.96$

$\boxed{m = 10}$ $z_{.005/2}$ to reject ~ 3.1 ??

Benjamini-Hochberg

- order the p-values $p_{(1)}, \ldots, p_{(m)}$

Benjamini-Hochberg

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

$$p_{(i)} \leq \frac{i}{m} q \qquad p_{(i)} \leq \frac{i(.1)}{m}$$

Benjamini-Hochberg

- 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

Benjamini-Hochberg

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

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

	H_0 not rej.	H_0 rej.	
H_0 true		V	m_0
H_0 not		S	m_1
		R	m

$E\left(\frac{V}{R}\right) \leq q$ ✓

- Let BH_q be the rule that rejects H_{0i} for $i \leq i_{max}$, not rejecting otherwise
- Theorem: If the p-values corresponding to valid null hypotheses are independent of each other, then

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

π_0 unknown but close to 1

proportion of true nulls

e.g. .1

Benjamini-Hochberg

- 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_{0i} for $i \leq i_{max}$, not rejecting otherwise
- Theorem: If the p-values corresponding to valid null hypotheses are independent of each other, then

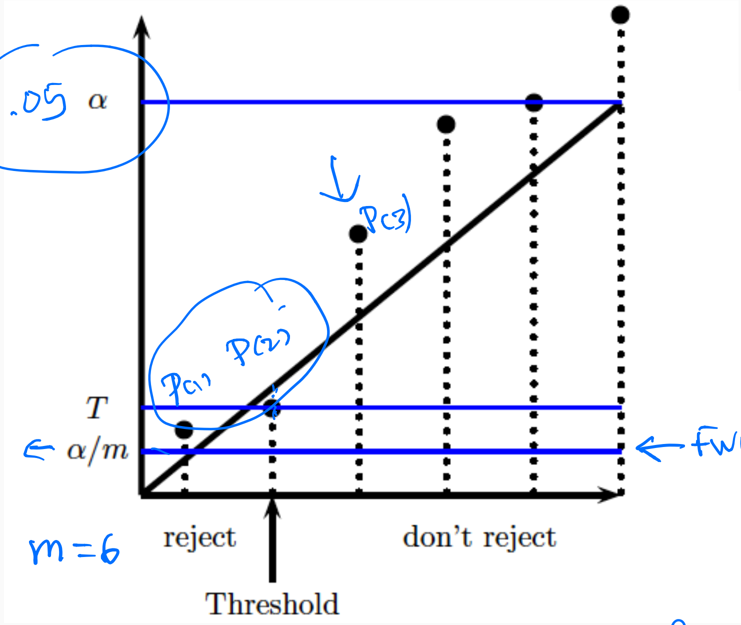
$$FDR(BH_q) = \pi_0 q \leq q, \qquad \text{where } \pi_0 = m_0/m$$

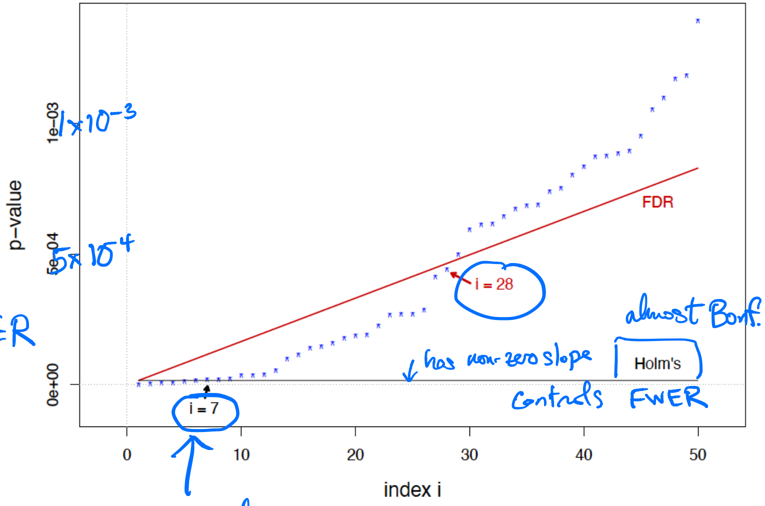
π_0 unknown but close to 1

- change the bound under dependence

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

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





biggest i s.t. $p_{(i)}$ below line

7 discoveries ≈ Bonferroni ~7000 tests

i										
index	1.00000	2.00000	3.00000	4.00000	5.00000	6.00000	7.0000	8.0000	9.0000	10.0000
$P_{(i)}$ 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~~	~~0.01464~~	~~0.02929~~	~~0.04393~~	~~0.05858~~	~~0.07322~~	~~0.08787~~	~~0.1025~~	~~0.1172~~	~~0.1318~~	~~0.1464~~

provided

⟶ Chrome: Google search; papers; Spiegelhalter; Gelman; Genovese; Ferreira & Zwinderman

5 p-vals
5 H_0's
of interest

$$\frac{iq}{m} \quad \frac{i \cdot .05}{10} = .005 \quad i=1, \quad .010 \quad i=2$$

$$\left(\frac{i\alpha}{m}\right) \qquad \frac{iq}{m C_m} \qquad C_m = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{10} \stackrel{?}{=} 2.7$$

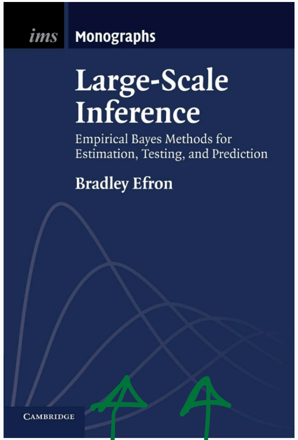
Theorem: If the p-values corresponding to valid null hypotheses are independent of each other, then

$X \sim U(0,1)$

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

$\Pr(X \leq s \mid X \leq t)$

$U(0,t)$



H_0 H_0 aV m_0

$\bar{H}_0$	$\bar{H}_0$	b S	m_1
		R $R(t)$	m

$P_{(1)} \le P_{(2)} \le \dots \le P_{(m)}$

Gelman's blog

link