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
- Likelihood-based confidence intervals and regions
- pure significance tests; simple and composite Ho
- goodness-of-fit tests
- empirical cumulative distribution function
- introduction to multiple testing



Today

HW 4 updated

- 1. Friday Feb 12
- 2. hypothesis vs significance testing
- 3. diagnostic testing
- 4. Benjamini-Hochberg method
- 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 <u>hore</u>.





1. Friday Feb 12

Reading week: no office hours

- 2. hypothesis vs significance testing
- 3. diagnostic testing
- 4. Benjamini-Hochberg method
- 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 <u>here</u>.



- Hypothesis tests typically means:
 - H₀, H₁
 - critical/rejection region $R \subset \mathcal{X}$,
 - + level $\alpha {\rm , \ power \ 1} \beta$
 - 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 typically means:
 - H₀, H₁
 - critical/rejection region $R \subset \mathcal{X}$,
 - level α , power 1 β
 - 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*_o,
 - test statistic T
 - observed value t^{obs},
 - p-value $p^{obs} = Pr(T \ge t^{obs}; H_0)$
 - alternative hypothesis often only implicit

large T points to alternative

- Hypothesis tests typically means:
 - H₀, H₁
 - critical/rejection region $R \subset \mathcal{X}$,
 - level α , power 1 β
 - 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*_o,
 - test statistic T
 - observed value t^{obs},
 - p-value $p^{obs} = Pr(T \ge t^{obs}; H_0)$
 - alternative hypothesis often only implicit

large T points to alternative

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_0

"reject Ho"

Definition 10.11 in AoS

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_0

Mini-quiz – True or False?

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

"reject Ho"

Rice. Exercise 9.11.5

Definition 1011 in AoS

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_0

Mini-quiz – True or False?

"reject H_0 "

Definition 10.11 in AoS

- 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

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_0

Mini-quiz – True or False?

Definition 10.11 in AoS

"reject Ho"

- 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 .

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_0

Mini-quiz – True or False?

Definition 10 11 in AoS

"reject Ho"

- 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

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_{0} Definition 1011 in AoS

Mini-guiz – True or False?

"reject Ho"

- 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

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_0

Mini-quiz – True or False?

Definition 10.11 in AoS

"reject Ho"

- 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

- overlap: sometimes (not recommended) $p^{obs} < 0.05 \longrightarrow$ "evidence against H_0 "
- overlap: p^{obs} is the smallest α -level at which the corresponding hypothesis test would reject H_0

Mini-quiz – True or False?

"reject Ho"

Definition 10.11 in AoS

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 Mathematical Statistics II February 10 2021

1. Hypothesis testing

		H _o not rejected	$H_{\rm o}$ rejected
	H _o true		type 1 error
truth			
	H₁ true	type 2 error	

2. Diagnostic testing

test negativetest positiveC19 negTNFPNtruthFNTPPC19 posFNTPP

link

Diagnostic testing and ROC

2. Diagnostic testing

		test negative	test positive	
	C19 neg	TN	FP	Ν
truth				
	C19 pos	FN	TP	Р

3. Multiple testing

		H _o not rejected	$H_{\rm o}$ rejected	
	H _o true	U	V	mo
truth				
	H₁ true	Т	S	m_1
		m-R	R	m

link

Diagnostic testing

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

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

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

- 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

$$\mathsf{FDR}(\mathsf{BH}_q) = \pi_{\mathsf{o}} q \leq q, \qquad ext{where } \pi_{\mathsf{o}} = m_{\mathsf{o}}/m$$

 $\pi_{\rm O}$ unknown but close to 1

- 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

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

 $\pi_{\rm O}$ unknown but close to 1

· change the bound under dependence

$$p_{(i)}\leq rac{i}{mC_m}q,$$

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



index 1.0000 2.0000 3.0000 4.0000 5.0000 6.0000 7.000 8.000 9.000 10.000 pval 0.00017 0.00448 0.00671 0.00907 0.0122 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

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

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

 $FDR(BH_q) = \pi_o q \leq q$, where $\pi_o = m_o/m$



Gelman's blog

link