Mathematical Statistics II

STA2212H S LEC0101

Week 9

March 14 2023

HEALTH

A New Turn in the Fight Over Masks

A crucial pandemic question is deceptively hard to answer.

By Yasmin Tayag

Masks


HEALTH

A New Turn in the Fight Over Masks

A crucial pandemic question is deceptively hard to answer.

By Yasmin Tayag


OPINION

ZEYNEP TUFEKCI

Here's Why the Science Is Clear That Masks Work

March 10, 2023 8 MIN READ

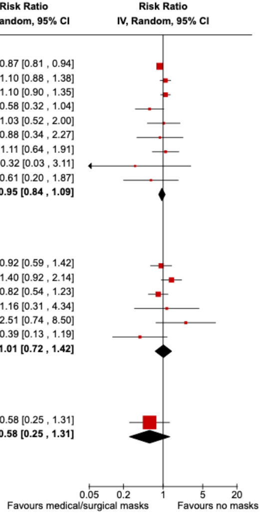
Atlantic (Y. Tayag)

NY Times (Z. Tufekci)

Masks

Cochrane Review

Study or Subgroup	log[RR]	SE	Medical/surgical masks Total	No masks Total	Weight	Risk Ratio IV, Random, 95% CI
1.1.1 Influenza/COVID-like illness						
Abaluck 2022 (1)	-0.135	0.036	111525	155268	41.4%	0.87 [0.81 , 0.94]
Aiello 2012	0.095	0.115	392	370	19.8%	1.10 [0.88 , 1.38]
Alfelali 2020	0.095	0.105	3864	3823	21.9%	1.10 [0.90 , 1.35]
Barasheed 2014	-0.55	0.3	75	89	4.6%	0.58 [0.32 , 1.04]
Canini 2010	0.025	0.342	148	158	3.6%	1.03 [0.52 , 2.00]
Cowling 2008	-0.128	0.483	61	205	1.9%	0.88 [0.34 , 2.27]
MacIntyre 2009	0.1	0.28	186	100	5.2%	1.11 [0.64 , 1.91]
MacIntyre 2016	-1.139	1.16	302	295	0.3%	0.32 [0.03 , 3.11]
Suess 2012	-0.494	0.571	26	30	1.4%	0.61 [0.20 , 1.87]
Subtotal (95% CI)			**116579**	**160338**	**100.0%**	**0.95 [0.84 , 1.09]**
Heterogeneity: Tau² = 0.01; Chi² = 11.44, df = 8 (P = 0.18); I² = 30%						
Test for overall effect: Z = 0.71 (P = 0.48)						
1.1.2 Laboratory-confirmed influenza or SARS-cov-2						
Aiello 2012	-0.083	0.223	392	370	25.9%	0.92 [0.59 , 1.42]
Alfelali 2020	0.34	0.215	3864	3823	26.7%	1.40 [0.92 , 2.14]
Bundgaard 2021 (2)	-0.2	0.208	2392	2470	27.4%	0.82 [0.54 , 1.23]
Cowling 2008	0.148	0.674	61	205	5.8%	1.16 [0.31 , 4.34]
MacIntyre 2009	0.92	0.6225	186	100	6.6%	2.51 [0.74 , 8.50]
Suess 2012	-0.942	0.57	26	30	7.7%	0.39 [0.13 , 1.19]
Subtotal (95% CI)			**6921**	**6998**	**100.0%**	**1.01 [0.72 , 1.42]**
Heterogeneity: Tau² = 0.07; Chi² = 8.52, df = 5 (P = 0.13); I² = 41%						
Test for overall effect: Z = 0.07 (P = 0.95)						
1.1.3 Laboratory-confirmed other respiratory viruses						
Bundgaard 2021	-0.55	0.42	2392	2470	100.0%	0.58 [0.25 , 1.31]
Subtotal (95% CI)			**2392**	**2470**	**100.0%**	**0.58 [0.25 , 1.31]**
Heterogeneity: Not applicable						
Test for overall effect: Z = 1.31 (P = 0.19)						



Footnotes

(1) Covid-like-illness

(2) SARS-cov-2

Today

1. Next lectures
2. Recap
3. Robust estimation
4. Asymptotic theory

Upcoming

- March 20 3.30 – 4.30 online Details
  "Using Data Science to Optimize Business – Opportunities & Challenges"

  Alison Burnham, Digitization Office, RepairSmith Inc

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

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	Robust Estimation Likelihood Asymptotics	MS 8.4, 8.6; SM 8.4 SM 4.4, 4.5
10	Mar 21	Causal Inference	AoS 16, 17
11	Mar 28	Classification	AoS 22
12	Apr 4	Course Summary; Presentations	

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)

Recap

- multiple testing: family-wise error rate (FWER); false discovery rate (FDR); Benjamini-Hochberg method — controls FDR
- goodness-of-fit tests based on empirical cdf $\hat{F}_n(\cdot)$
  - Kolmogorov-Smirnov
  - Cramer-vonMises
  - Anderson-Darling
- Brownian bridge; limit distributions
- goodness-of-fit tests based on multinomial distribution

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

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
- change the bound under dependence

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

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
- change the bound under dependence

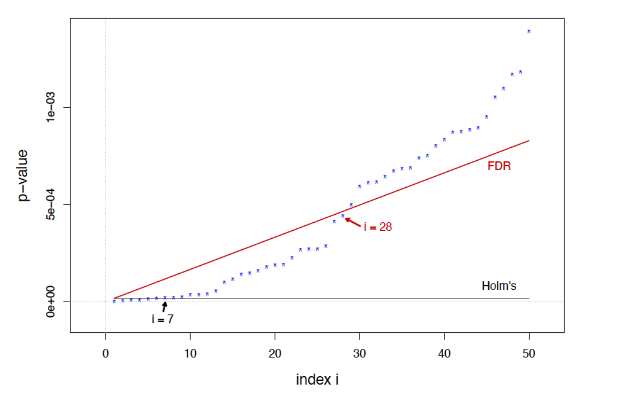
$$p_{(i)} \leq \frac{i}{mC_m} q \qquad C_m = \sum_{i=1}^{m} \frac{1}{i}$$

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

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

π_0 unknown but close to 1

Benjamini-Hochberg



BH-q: reject H_{0i} for $p_{(i)} \leq \frac{i}{m} q$

$p_{(1)} \leq \cdots \leq p_{(m)}$

Proof

BH-q: reject H_{0i} for $p_{(i)} \le \frac{i}{m} q$ $\qquad$ $FDR(BH_q) = \pi_0 q \le q$, $\qquad$ where $\pi_0 = m_0/m$

$0 < t \le 1$:

$$
\begin{aligned}
R(t) &= \#\{p_i \le t\} \\
V(t) &= \#\{p_i \le t, H_{0i} \text{ true }\} \\
FDP(t) &= V(t)/\max(R(t), 1) \\
Q(t) &= mt/\max(R(t), 1) \\
t_q &= \sup_t \{Q(t) \le q\}
\end{aligned}
$$

		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

1.
2.
3.

Proof

BH-q: reject H_{0i} for $p_{(i)} \leq \frac{i}{m} q$ $\quad$ $FDR(BH_q) = \pi_0 q \leq q$, $\quad$ where $\pi_0 = m_0/m$

$0 < t \leq 1$:

$$\begin{aligned} R(t) &= \#\{p_i \leq t\} \\ V(t) &= \#\{p_i \leq t, H_{0i} \text{ true }\} \\ FDP(t) &= V(t)/\max(R(t), 1) \\ Q(t) &= mt/\max(R(t), 1) \\ t_q &= \sup_t \{Q(t) \leq q\} \end{aligned}$$

		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

$R(p_{(i)}) = i \implies Q(p_{(i)}) = mp_{(i)}/i$

1. BH-q $\iff$: reject H_{0i} for $p_{(i)} \leq t_q$
2. $A(t) = V(t)/t, \quad \mathrm{E}\{A(s) \mid A(t)\} = A(t), s \leq t$

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

$\implies \mathrm{E}\{A(t_q)\} = E\{A(1)\} = m_0$

3. $\max\{R(t_q), 1\} = \frac{mt_q}{Q(t_q)} = \frac{mt_q}{q} \implies FDP(t_q) = \frac{q}{m} \frac{V(t_q)}{t_q} = q \frac{m_0}{m}$

2. $E\{A(s) \mid A(t)\} = A(t), \quad s \leq t$

$A(t) = V(t)/t = \#\{p_i \leq t, H_{0i} \text{ true }\}/t$

$X_1, \ldots, X_n$ i.i.d. $\sim U(0,1)$

$$\mathrm{pr}(X \leq s \mid X \leq t) = \cdots = \frac{s}{t}$$

- Linear regression $Y = X\beta + \sigma\epsilon$ — $E(\epsilon) = 0, \text{var}(\epsilon) < \infty$

- Least squares estimator

$$\hat{\beta}_{LS} = (X^TX)^{-1}X^TY$$

- Gauss-Markov theorem: — BLUE; MVUE

$$\text{var}(\hat{\beta}_{LS}) \leq \text{var}(\tilde{\beta}),$$

- fun fact: if $\epsilon_i \sim t_\nu$ are independent,

$$\text{var}(\hat{\beta}_{LS}) = \sigma^2(X^TX)^{-1}\frac{\nu}{\nu-2}; \quad \text{a.var}(\hat{\beta}_{MLE}) = \sigma^2(X^TX)^{-1}\frac{\nu+3}{\nu+1}$$

- $\hat{\beta}_{LS}$ has asymptotic relative efficiency $\dfrac{(\nu-2)(\nu+3)}{\nu(\nu+1)}$ — $\nu = 5, 10, 20; \mathit{eff} = .80, .95, .99$

Robust estimation

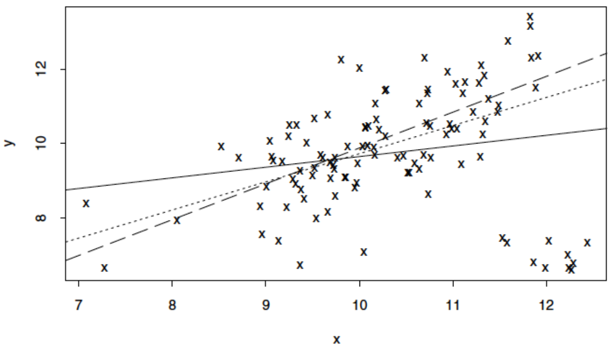


Figure 8.2 *Estimated regression lines; the solid line is the least squares line, the dotted line is the L_1 line and the dashed line is the LMS line. Notice how the least squares line is pulled more towards the 10 "outlying" observations than are the other two lines.*

M-estimation

- Linear regression: $y = X\beta + \sigma\epsilon$

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \longrightarrow \min_{\beta} \sum_{i=1}^{n} \rho(y_i - x_i^T \beta)$$

$\rho(\cdot)$ to be chosen

- or more typically, $$\min_{\beta} \sum_{i=1}^{n} \rho\{(y_i - x_i^T \beta)/\sigma\}$$

- various choices for $\rho(u)$: $u^2/2, \quad |u|, \quad \nu \log(1 + u^2/\nu)/2,$

Huber: $\rho(u) = \begin{cases} u^2, & |u| \leq c \\ c(2|u| - c) & \text{otherwise} \end{cases}$

- MS also considers Least Median Squares estimator

$$\min_{\beta} \sum_{i=1}^{n} \rho\{(y_i - x_i^T \beta)/\sigma\}$$

- Theorem: if $\rho(\cdot)$ is convex, and MS Thm 8.6

$$\psi(t) = \rho'(t),$$

is non-decreasing, then

$$\max_{1 \le i \le n} x_i^T (X^T X)^{-1} x_i \to 0, \quad n \to \infty$$

$$\implies A_n(\hat{\beta} - \beta) \xrightarrow{d} N_p(0, \gamma^2 I), \qquad \gamma^2 = \frac{\mathrm{E}\{\psi^2(\epsilon_i)\}}{\mathrm{E}^2\{\psi'(\epsilon)\}}$$

Aside: ordinary least squares

$$
\begin{aligned}
\hat{\beta}_{n,LS} &= (X_n^T X_n)^{-1} X_n^T Y_n \\
\hat{\beta}_{n,LS} - \beta &= (X_n^T X_n)^{-1} X_n^T \epsilon_n \\
A_n^2 &= X_n^T X_n \\
A_n(\hat{\beta}_{n,LS} - \beta) &= A_n^{-1} X_n^T \epsilon_n
\end{aligned}
$$

$$
\max_{1 \le i \le n} x_i^T (X_n^T X_n)^{-1} x_i \to 0 \implies A_n(\hat{\beta}_{n,LS} - \beta) \xrightarrow{d} N(0, \sigma^2 I)
$$

Note that $x_i^T (X_n^T X_n)^{-1} x_i = h_{ii}$, where $H =$ and $\text{trace}(H) =$

Simple linear regression: $x_i = i; x_i = 2^i$

Estimating equations

- $g(\theta; X)$ is an unbiased estimating equation for θ if

$$\mathrm{E}_\theta\{g(\theta; X)\} = 0, \quad \mathrm{E}_\theta\{g(\theta; X)g^T(\theta; X)\} < \infty$$

- given $X_1, \ldots, X_n$ i.i.d. with density $f(x; \theta)$, define the estimator $\tilde{\theta}_g$ by

$$\sum_{i=1}^{n} g(\tilde{\theta}_g; X_i) = 0$$

- then

$$\sqrt{n}(\tilde{\theta} - \theta) \xrightarrow{d} N\{0, V(\theta)\}$$

-

$$V(\theta) = J^{-1}(\theta)I(\theta)J^{-1}(\theta)$$

-

$$J(\theta) = \mathrm{E}_\theta\{-g'(\theta; X)\}, \quad I(\theta) = \mathrm{E}_\theta\{g(\theta; X)g^T(\theta; X)\}$$

... Estimating equations

Robust regression

- limit theory derived from theory of estimating equations
- more generally, from the theory of model misspecification MS 5.5
- true model $X_1, \ldots, X_n$ i.i.d. $h(\cdot)$, say
- assumed model $X_1, \ldots, X_n$ i.i.d. $f(\cdot;\theta)$

- maximum likelihood estimator SM 4.6

$$\ell'(\hat\theta; \boldsymbol{X}) = 0$$

$$\hat\theta \xrightarrow{p} \arg\min_\theta \mathrm{E}_h \log\left\{\frac{h(x)}{f(x;\theta)}\right\}$$

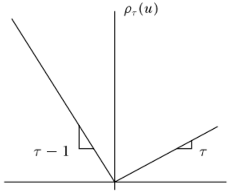
relative entropy; K-L divergence

$$\sqrt{n}(\hat\theta - \theta) \xrightarrow{d} N\{0, J_h^{-1}(\theta_h) I_h(\theta_h) J_h^{-1}(\theta_h)\}$$

- forms the basis for GEE approach to longitudinal data

Quantile regression

Figure 2
Quantile Regression ρ Function



$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i^T\beta)$$

Solution by linear programming; solution has approximately τ/n positive residuals

R package `quantreg`

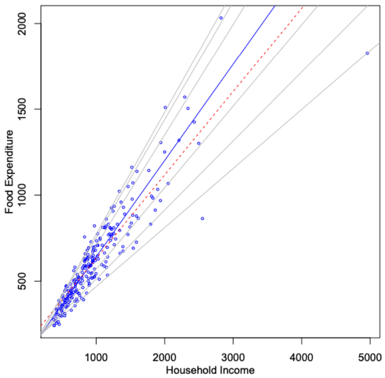


FIGURE 1. Scatterplot and Quantile Regression Fit of the Engel Food Expenditure Data: The plot shows a scatterplot of the Engel data on food expenditure vs household income for a sample of 235 19th century working class Belgian households. Superimposed on the plot are the $\{.05, .1, .25, .75, .90, .95\}$ quantile regression lines in gray, the median fit in solid black, and the least squares estimate of the conditional mean function as the dashed (red) line.

Likelihood asymptotics

```
> cover_prob
                1       5      10      50
Exact CI 0.95005 0.95148 0.95044 0.95101
q        0.94834 0.95741 0.95567 0.95159
s        0.71684 0.86771 0.90188 0.93992
r        0.93352 0.94634 0.94774 0.94939
r*       0.95165 0.95064 0.94940 0.94984
```

Figure 1: Simulation with 100000 times. Sample size $n = 1, 5, 10, 50$.

$X_1, \ldots, X_n$ i.i.d. $f(x;\theta) = \theta \exp(-\theta x)$

$$q = (\hat{\theta} - \theta) j^{1/2}(\hat{\theta}), \quad s = \ell'(\theta) j^{1/2}(\hat{\theta}), \quad r = \pm\sqrt{2\{\ell(\hat{\theta}) - \ell(\theta)\}}, \quad r^* = r + \frac{1}{r}\log\left(\frac{q}{r}\right)$$

First-order theory

- central limit theorem: $Y_1, \ldots, Y_n$ i.i.d. $\mathrm{E}(Y_i) = \mu$, $\mathrm{var}(Y_i) = \sigma^2$,

$$\sqrt{n}(\bar{Y} - \mu)/\sigma \xrightarrow{d} N(0, 1)$$

- normal approximation