# Mathematical Statistics II

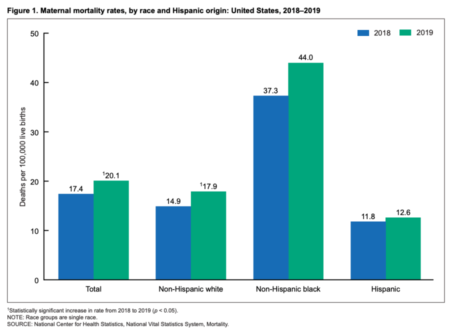
## STA2212H S LEC9101

Week 12

April 7 2021

Start recording!

CDC Apr 4



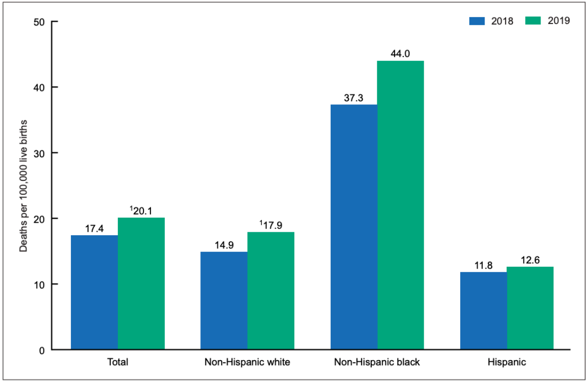
# In the news

B Jarosz
@DataGeekB

Story that has flown under the radar (and should not)
#MaternalMortality rate rose sharply in the US
AND rose most for Black women--who already su
that are unacceptably high.
Source: cdc.gov/nchs/data/hest…
pic.twitter.com/k7IldBwjCh
2021-04-02, 10:18 PM

CDC Apr 4

**Figure 1. Maternal mortality rates, by race and Hispanic origin: United States, 2018–2019**




[1]Statistically significant increase in rate from 2018 to 2019 ($p < 0.05$).
NOTE: Race groups are single race.
SOURCE: National Center for Health Statistics, National Vital Statistics System, Mortality.

# … In the news

**Table. Number of maternal deaths and maternal mortality rates, by race and Hispanic origin and age: United States, 2018 and 2019**

| Race and Hispanic origin and age | 2018 Number of deaths | 2018 Maternal mortality rate[1] | 2019 Number of deaths | 2019 Maternal mortality rate[1] |
|---|---|---|---|---|
| Total[2] | 658 | 17.4 | 754 | 20.1 |
| Under 25 | 96 | 10.6 | 111 | 12.6 |
| 25–39 | 458 | 16.6 | 544 | 19.9 |
| 40 and over | 104 | 81.9 | 98 | 75.5 |
| Non-Hispanic white[3] | 291 | 14.9 | 343 | 17.9 |
| Under 25 | 41 | 10.5 | 49 | 13.1 |
| 25–39 | 207 | 13.8 | 248 | 16.8 |
| 40 and over | 43 | 72.0 | 46 | 75.2 |
| Non-Hispanic black[3] | 206 | 37.3 | 241 | 44.0 |
| Under 25 | 27 | 15.3 | 32 | 18.8 |
| 25–39 | 137 | 38.2 | 179 | 49.7 |
| 40 and over | 42 | 239.9 | 30 | 166.5 |
| Hispanic | 105 | 11.8 | 112 | 12.6 |
| Under 25 | 21 | 7.6 | 23 | 8.5 |
| 25–39 | 72 | 12.4 | 71 | 12.2 |
| 40 and over | 12 | * | 18 | * |

* Rate does not meet National Center for Health Statistics standards of reliability.
[1]Maternal mortality rates are deaths per 100,000 live births.
[2]Total includes deaths for race and origin groups not shown separately, including deaths among multiple-race women and deaths with origin not stated. Race groups are single race.
[3]Race groups are single race.
NOTES: Maternal causes are those assigned to categories A34, O00–O95, and O98–O99 of the *International Classification of Diseases, 10th Revision*, 1992. Maternal deaths occur while pregnant or within 42 days of being pregnant.

# In PNAS

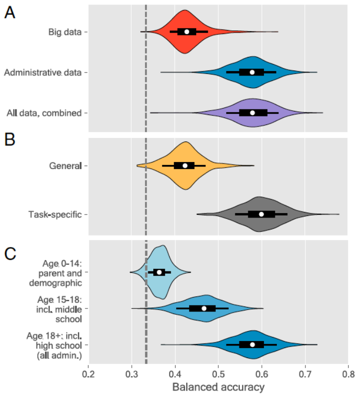



**Fig. 1.** Balanced accuracy of model out of sample on test data when using various feature sets. (*A*) Big data vs. administrative data, (*B*) task-related vs. general information, and (*C*) comparison of feature sets gathered over the lifespan of the student. Models are estimated using logistic regression with $L_2$ regularization and using feature selection; see *Materials and Methods* for details. Each violin represents the distribution of weighted accuracy from 1,000 resamples. Inside the violins, the thick bar represents the bottom and top quartiles, and the thin lines represent the bottom and top deciles. The dashed, black line indicates the performance of a baseline random guessing model.

Bjerre-Nielsen, et al. "Task-specific information outperforms surveillance-style big data in predictive analytics"

# In PNAS

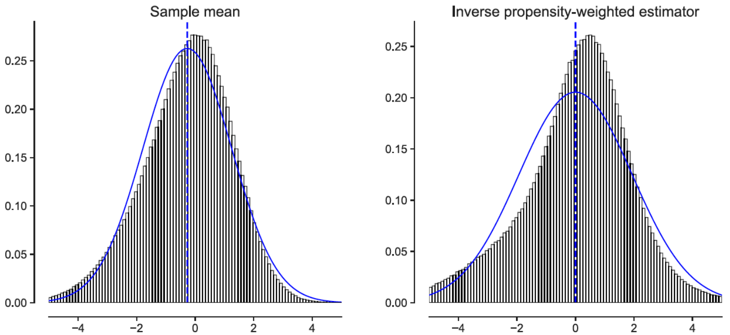



**Fig. 1.** Distribution of the estimates $\widehat{Q}^{AVG}(1)$ and $\widehat{Q}^{IPW}(1)$ described in the introduction. The plots depict the distribution of the estimators for $T = 10^6$, scaled by a factor $\sqrt{T}$ for visualization. The distributions are overlaid with the normal curve that matches the first two moments of the distribution, along with a dashed line that denotes the mean. All numbers are aggregated over 1 million replications.

Hadad , et al. “Confidence intervals for policy evaluation in adaptive experiments”

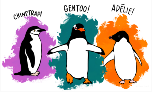
# Today

Start Recording

1. HW 11 due Friday April 9
   Take-home posted April 9 due April 19 — April 8 if I can
2. Friday April 9: Inference for Visualization, Chenghui Zheng;
   Data science at CIBC, Manuel Blain
3. Week of April 12: No classes. Usual Office hours. — Monday 7pm; Thursday, Friday 11am

4. Course evaluations available until April 12
5. Sketch of some ideas in Causal Inference
6. Review of course

- Apr 12 3.00 – 4.00 pm EDT, Data Science ARES, Alison Hill, RStudio, “Crafting kind tools”



# Recap

- multivariate distributions: normal, multinomial — AoS Ch 14
- testing independence: correlation $\rho_{jk}$, $\chi^2$ tests — AoS Ch 15
- classification – parametric and nonparametric — AoS 22

# Recap

- multivariate distributions: normal, multinomial — AoS Ch 14
- testing independence: correlation $\rho_{jk}$, $\chi^2$ tests — AoS Ch 15
- classification – parametric and nonparametric — AoS 22

- Markov random fields — SM 6.2
- Directed acyclic graphs — SM 6.2.2
- Graphical Gaussian models — SM 6.3.3

# Recap

- multivariate distributions: normal, multinomial — AoS Ch 14
- testing independence: correlation $\rho_{jk}$, $\chi^2$ tests — AoS Ch 15
- classification – parametric and nonparametric — AoS 22

- Markov random fields — SM 6.2
- Directed acyclic graphs — SM 6.2.2
- Graphical Gaussian models — SM 6.3.3

methods for studying relationships between variables

a word on modelling

# A word about models

# Causal Inference

- $X$ – binary treatment indicator — "treatment"
- $Y$ – binary outcome — could be continuous
- "X causes Y"
- "X is associated with Y"

- $X$ – binary treatment indicator — “treatment”
- $Y$ – binary outcome — could be continuous
- “X causes Y”
- “X is associated with Y”

- potential outcomes $C_0, C_1$
- consistency equation — two expressions

| $X$ | $Y$ | $C_0$ | $C_1$ |
|---|---|---|---|
| 0 | 4 | 4 | * |
| 0 | 7 | 7 | * |
| 0 | 2 | 2 | * |
| 0 | 8 | 8 | * |
| 1 | 3 | * | 3 |
| 1 | 5 | * | 5 |
| 1 | 8 | * | 8 |
| 1 | 9 | * | 9 |

| Type | $C_0$ | $C_1$ |
|---|---|---|
| Survivors | 1 | 1 |
| Responders | 0 | 1 |
| Anti-responders | 1 | 0 |
| Doomed | 0 | 0 |

# Causal treatment effect



$\theta =$

$\alpha =$

If $X$ is randomly assigned, then

# Example 16.2

| $X$ | $Y$ | $C_0$ | $C_1$ |
|---|---|---|---|
| 0 | 0 | 0 | 0* |
| 0 | 0 | 0 | 0* |
| 0 | 0 | 0 | 0* |
| 0 | 0 | 0 | 0* |
| 1 | 1 | 1* | 1 |
| 1 | 1 | 1* | 1 |
| 1 | 1 | 1* | 1 |
| 1 | 1 | 1* | 1 |

| $X$ | $Y$ | $C_0$ | $C_1$ |
|---|---|---|---|
| 0 | 0 | 0 | 0* |
| 1 | 0 | 0 | 0* |
| 1 | 0 | 0 | 0* |
| 1 | 0 | 0 | 0* |
| 1 | 1 | 1* | 1 |
| 1 | 1 | 1* | 1 |
| 1 | 1 | 1* | 1 |
| 1 | 1 | 1* | 1 |

# Definitions of causality

# Causal regression function

# No unmeasured confounding



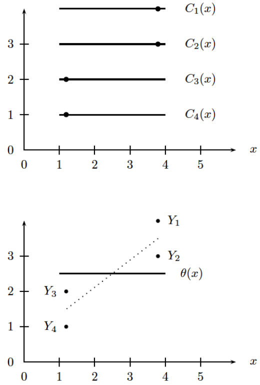



FIGURE 16.2. The top plot shows the counterfactual function $C(x)$ for four subjects. The dots represent their $X$ values. Since $C_i(x)$ is constant over $x$ for all $i$, there is no causal effect. Changing the dose will not change anyone's outcome. The lower plot shows the causal regression function $\theta(x) = (C_1(x) + C_2(x) + C_3(x) + C_4(x))/4$. The four dots represent the observed data points $Y_1 = C_1(X_1)$, $Y_2 = C_2(X_2)$, $Y_3 = C_3(X_3)$, $Y_4 = C_4(X_4)$. The dotted line represents the regression $r(x) = \mathbb{E}(Y|X = x)$. There is no causal effect since $C_i(x)$ is constant for all $i$. But there is an association since the regression curve $r(x)$ is not constant.

# No unmeasured confounding



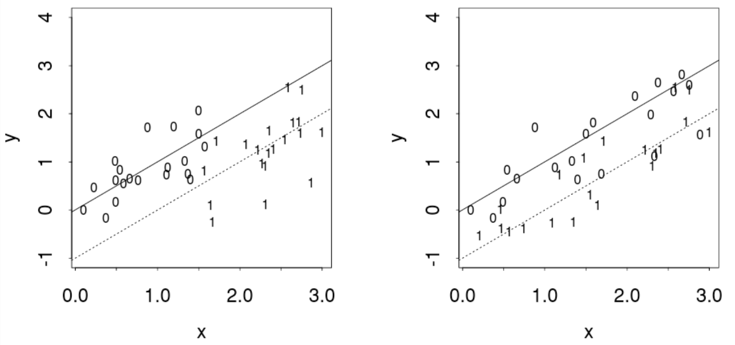
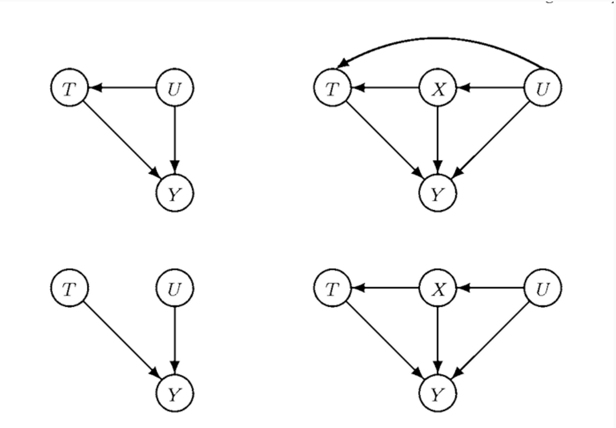



**Figure 9.2** Simulated results from experiments to compare the effect of a treatment $T$ on a response $Y$ that varies with a covariate $X$. The lines show the mean response for $T = 0$ (solid) and $T = 1$ (dots). Left: the effect of $T$ is confounded with dependence on $X$. Right: the experiment is balanced, with random allocation of $T$ dependent on $X$.

# No unmeasured confounding





# Criteria for causality

2101 Oct 22 slide 6

# Summary

- likelihood function, MLE, asymptotic normality of MLE, LR, delta method, sufficient statistics, exponential family, MLEs are "BAN", Newton-Raphson, EM algorithm, quasi-Newton
- hypothesis testing, null hypothesis, alternative hypothesis, type I and type II error, size, power, test statistic, critical regions, optimal tests, NP Lemma, tests based on likelihood function (Wald, LRT, GLRT)
- significance testing, $p$-values, $p$-hacking, goodness-of-fit tests, diagnostic testing, multiple testing, FWER, Bonferroni, Benjamini-Hochberg

# Summary

- Bayes' theorem, Bayesian inference, priors, likelihood, credible intervals, posterior modes, posterior distribution, HPD intervals, approximations to posterior, two-sided $p$-values,
- types of priors – conjugate, subjective, objective, noninformative, flat, matching, Jeffreys', Bayesian computation – Laplace, importance sampling, MCMC, Bayesian philosophy, empirical and epistemic probability,
- posterior predictive distributions, hierarchical models, empirical Bayes
- decision theory, loss functions, risk function, Bayes risk, admissibility, Stein's paradox, minimax risk, Bayes rules

# Summary

- multivariate normal distribution, inference for correlation, maximum likelihood estimates, partial correlation, gaussian graphical models, conditional independence
- multinomial distribution, $\chi^2$ testing, independence testing, classification and regression
- directed acyclic graphs, gaussian graphical models, Markov property, confounding, causality, potential outcomes,