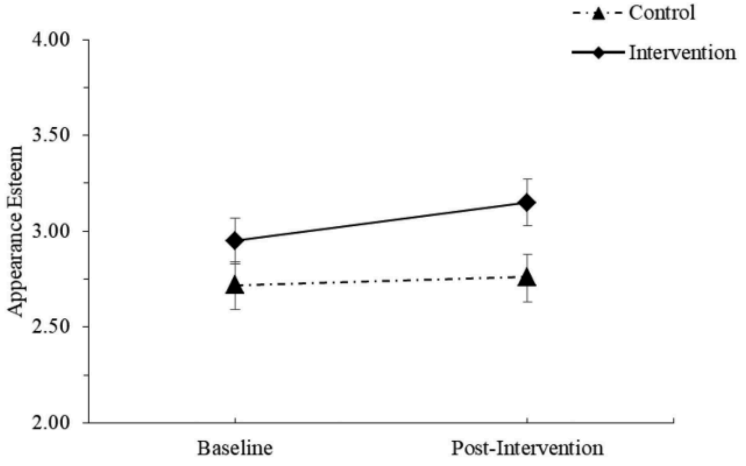
# Mathematical Statistics II

STA2212H S LEC0101

Week 7

February 28 2023

**Figure 2**
*Effect of Reducing Social Media Use on Levels of Appearance Esteem by Condition*




*Note.* Error bars represent standard errors.

**Table 2**

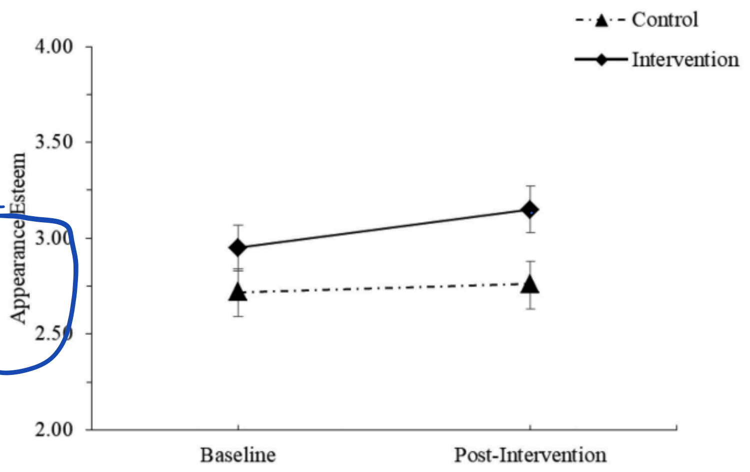
*ANOVA Main Effects and Interaction Effects on Appearance and Weight Esteem*

| Source | *MS* | *F* | *p* | *d* |
|---|---|---|---|---|
| Appearance esteem | | | | |
| Condition | 2.18 | 1.03 | .311 | 0.02 |
| Time | **0.81** | **5.40** | **.021** | **0.29** |
| Gender | **17.78** | **8.42** | **.004** | **0.37** |
| Condition × Time | **0.80** | **5.33** | **.022** | **0.28** |
| Condition × Gender | 1.33 | 0.63 | .43 | 0.00 |
| Time × Gender | 0.00 | 0.03 | .868 | 0.00 |
| Condition × Time × Gender | 0.52 | 3.50 | .063 | 0.22 |
| Error (within) | 0.15 | | | |
| Error (between groups) | 2.11 | | | |
| Weight esteem | | | | |
| Condition | **12.11** | **8.34** | **.004** | **0.37** |
| Time | **1.18** | **8.35** | **.004** | **0.37** |
| Gender | 2.66 | 1.83 | .178 | 0.12 |
| Condition × Time | **0.71** | **5.04** | **.026** | **0.27** |
| Condition × Gender | 1.69 | 1.16 | .282 | 0.06 |
| Time × Gender | 0.03 | 0.18 | .676 | 0.00 |
| Condition × Time × Gender | 0.27 | 1.91 | .168 | 0.13 |
| Error (within) | 0.14 | | | |
| Error (between groups) | 1.45 | | | |

*Note*. Boldface indicates statistical significance ($p < .05$); $N = 218$.

**Figure 2**

*Effect of Reducing Social Media Use on Levels of Appearance Esteem by Condition*




*Note*. Error bars represent standard errors.

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychology of Popular Media


ISSN: 2689-6567

https://doi.org/10.1037/ppm0000460

BRIEF REPORT

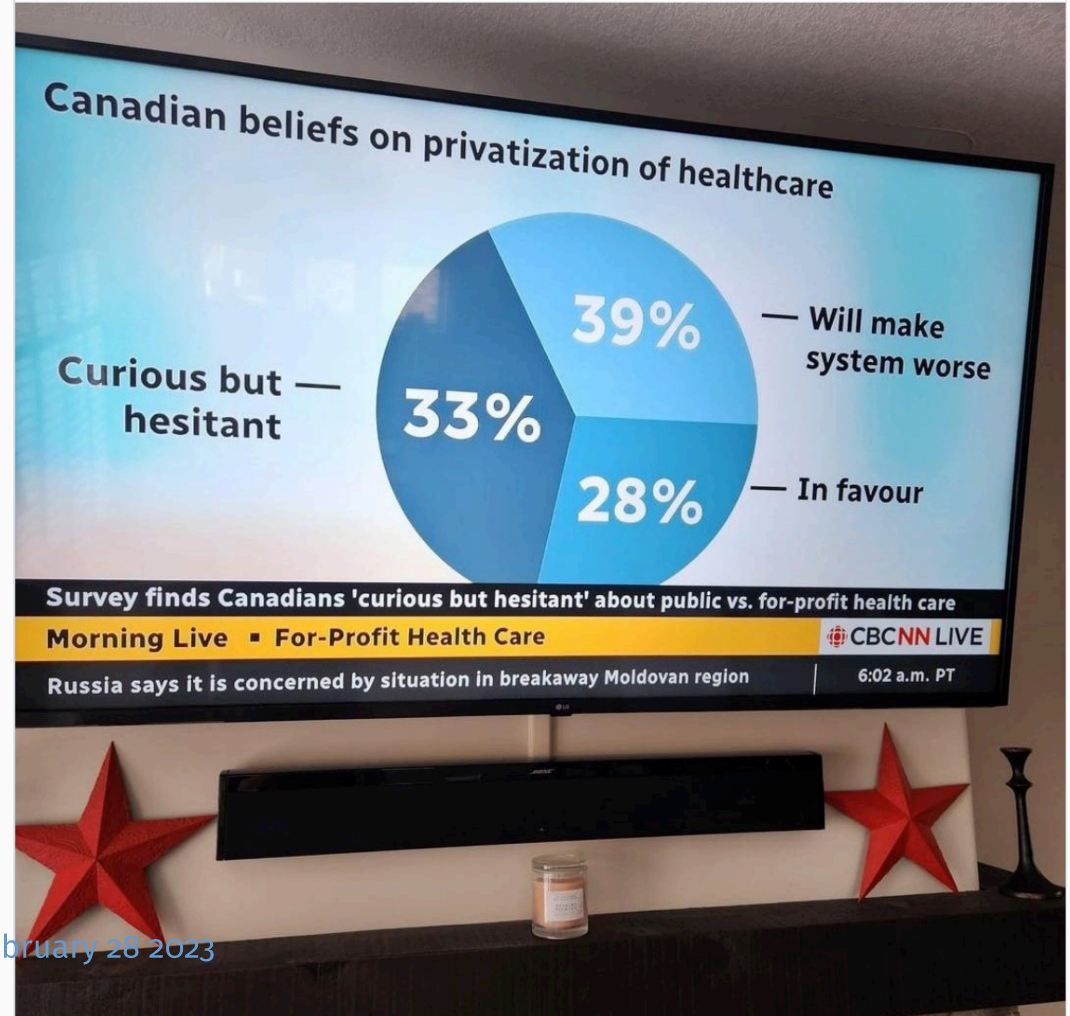
# Reducing Social Media Use Improves Appearance and Weight Esteem in Youth With Emotional Distress

Helen Thai[1], Christopher G. Davis[2], Wardah Mahboob[2, 3], Sabrina Perry[2, 3], Alex Adams[2, 3], and Gary S. Goldfield[2, 3, 4]

[1] Department of Psychology, McGill University
[2] Department of Psychology, Carleton University
[3] Healthy Active Living and Obesity (HALO) Research Group, Children's Hospital of Eastern Ontario
[4] Department of Pediatrics, University of Ottawa

# This just in

Twitter



# Today

1. Project guidelines
2. Recap
3. Hypothesis testing
4. Significance testing
5. Multiple testing

Upcoming

- March 6 3.30 – 4.30 (Zoom) Details

  “Private hypothesis testing over sensitive groups”
  Rina Friedberg, Senior ML Engineer, LinkedIn

  Data Science Applied Research and Education Seminar

link

**Project Guidelines** STA 2212S: Mathematical Statistics II 2023

Presentation on April 4, 2023.
Report submission due April 14, 2023. and slides

**Part 1: Write-up [30 points]**

Your write-up should be: (1): no more than 8 pages, 12 point font, 1.5 vertical spacing; (2) Contain the four sections below, each partner to complete two sections; (3) Include a title page with the title and authors of the paper, the first and last names of the report authors and which section they wrote. (4) Include a list of references.
*The title page and references, and any figures, do not count towards the 10 page limit.*

The sections to include and the questions to answer in each section are:

1. Introduction and Motivation
    (a) What is the problem being addressed?
    (b) What previous work exists?
    (c) Why is the previous work insufficient to solve the problem?

# Recap

- formal structure for hypothesis testing: $H_0$, $H_1$, test function, critical region, Type I error, Type II error, size, power

null alt. $\phi(\underset{\sim}{x}) \in \{0, 1\}$

$+1$ $1-$ $t\,2$

"do not reject $H_0$" (wrong = t. II)

"reject $H_0$" (wrong = type I)

- simple and composite hypotheses

$H_0$ / $H_1$ : dist$^n$ of $\underset{\sim}{X}$ completely known

$\mathcal{R} \subset \mathcal{X}$ : critical region

if $\underset{\sim}{x} \in \mathcal{R}$ then "reject $H_0$"

o.w. not

$\bar{X} \sim N(\mu, \frac{1}{n})$ $H_0: \mu = 0$ $\quad H_1: \mu > 0 \leftarrow$ composite

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $H_0: \mu = 0$ $\quad H_1: \mu > 0 \leftarrow$ composite

# Recap

- formal structure for hypothesis testing: $H_0$, $H_1$, test function, critical region, Type I error, Type II error, size, power

Wald test

$H_0$ simple

LRT

- simple and composite hypotheses

$H_0: \theta = \theta_0$

- three likelihood-based test statistics

(MS §7.3)

1. $\hat{\theta} - \theta_0 \dot{\sim} N_k(0, I_n^{-1}(\theta_0))$

$\hat{\theta}(\underset{\sim}{x})$

exp'd F. info.

$\hat{\theta} - \theta_0 \dot{\sim} N(0, j^{-1}(\hat{\theta}))$

mle

obs'd F. info

2. $w(\theta_0) = 2\{\ell(\hat{\theta}) - \ell(\theta_0)\}$

$\dot{\sim} \chi^2_{\cancel{\dim}(\theta)=k}$ $H_0: \theta = \theta_0$

3. $\ell'(\theta_0) \dot{\sim} N(0, I_n(\theta_0))$

(score test)

rejection region using N distn

# Recap

- formal structure for hypothesis testing: $H_0$, $H_1$, test function, critical region, Type I error, Type II error, size, power

- simple and composite hypotheses

- three likelihood-based test statistics

  if $\theta = (\psi, \lambda)$ $H_0: \psi = \psi_0$

  $\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$

  1., 2., 3. $\sim N, \chi^2$ under $H_0$

- Neyman-Pearson lemma

# Three likelihood-based test statistics

# Neyman-Pearson Lemma

Suppose $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f(\boldsymbol{x})$ and we wish to test the null hypothesis $H_0 : f(\boldsymbol{X}) = f_0(\boldsymbol{x})$, against the alternative hypothesis $H_1 : f(\boldsymbol{X}) = f_1(\boldsymbol{x})$.

The test with test function

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } f_1(\boldsymbol{x}) > kf_0(\boldsymbol{x}), \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{f_1(\underset{\sim}{x})}{f_0(\underset{\sim}{x})} > k$$

(for some $0 < k < \infty$) is a most powerful test of $H_0$ vs $H_1$ at level $\alpha = \mathrm{E}_0\{\phi(\boldsymbol{X})\}$.

$\mathrm{pr}_{H_0}\{\phi(\underset{\sim}{X}) = 1\}$

$\exists\ \psi(\underset{\sim}{x}) : \cancel{\mathcal{X}}^n \to [0,1]$ $\quad 0 \le \psi(\underset{\sim}{x}) \le 1 \quad \forall\, \underset{\sim}{x}$

has size $\alpha$ : $\quad E_0\, \psi(\underset{\sim}{X}) \le \alpha$

show $\underline{E_1\, \psi(\underset{\sim}{X})} \le E_1\, \phi(\underset{\sim}{X})$ : less powerful

$= \mathrm{Pr}_{H_1}\{\phi(\underset{\sim}{X}) = 1\} = \mathrm{Pr}_{H_1}\{\text{reject } H_0\}$

(max'd this)

- Suppose there is another function $0 \leq \psi(\mathbf{x}) \leq 1$ with $\mathrm{E}_0\{\psi(\mathbf{x})\} \leq \mathrm{E}_0\{\phi(\mathbf{x})\}$

LHS $\int \psi(x)\{f_1(x) - k f_0(x)\} dx \leq \int \phi(x)\{f_1(x) - k f_0(x)\} dx$ RHS

$\leq f_1(x) - k f_0(x)$

bec $\psi \leq 1$

$\leq 0$ b.c. $\psi(x) \geq 0$, $\{f_1 - k f_0\} < 0$

$\hookrightarrow$ if $= 1$

term $\leq f_1 - k f_0$

$\hookrightarrow$ if $= 0$ RHS $= 0$

$$\int \{\psi(x) - \phi(\underline{x})\} f_1(\underline{x})\, d\underline{x} \leq k \int \{\psi(x) - \phi(x)\} f_0(x)\, d\underline{x}$$

$$E_{H_1} \psi(\underline{X}) - E_{H_1} \phi(\underline{X}) \leq 0$$

bec. size - size

$k\, E_0 \psi(\underline{X}) - E_0 \phi(\underline{X})$

- Suppose there is another function $0 \le \psi(\boldsymbol{x}) \le 1$ with $\mathrm{E}_0\{\psi(\boldsymbol{x})\} \le \mathrm{E}_0\{\phi(\boldsymbol{x})\}$
-

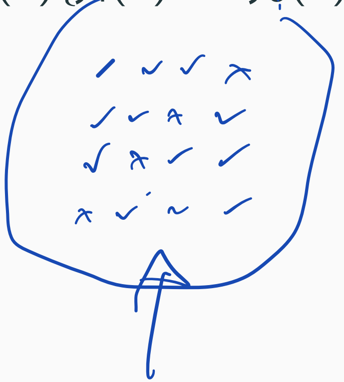
$$\psi(\boldsymbol{x})\{f_1(\boldsymbol{x}) - kf_0(\boldsymbol{x})\} \le$$

- 

$$\int \psi(\boldsymbol{x})\{f_1(\boldsymbol{x}) - kf_0(\boldsymbol{x})\}dx \le$$

$\dfrac{\binom{4}{3} \cdot 4}{70}$

$\dfrac{8 \times 7 \times 6 \times 5}{4 \times 3 \times 2}$ 2



8 cups 4 C

4 P

$Pr_{H_0}(4 \text{ correct}) = \dfrac{1}{\binom{8}{4}} = \dfrac{1}{70}$ ≈ $< .05$

$H_0$: can't dist. them

$Pr_{H_0}\{3 \text{ correct } \& \ 1 \times\}$ $\boxed{0.014}$

$= \dfrac{16}{70}$ (?) $\boxed{0.243}$

# Comments on NP-Lemma

$\frac{16+1}{70} = \frac{17}{70} > .05$

$T(\underset{\sim}{x})$ $\|$

T.S.H. UMP Lehmann
UMPU
UMPI
⋮

- both $H_0$ and $H_1$ must be simple
- the critical region is $R = \{ \underset{\sim}{x} : \boxed{\frac{f_1(\underset{\sim}{x})}{f_0(\underset{\sim}{x})}} > k \}$ ← $k$ to be det'd
- if the distribution of $T(\boldsymbol{X})$ is continuous, size $\alpha$ always exist Ex. 7.11; HW 6
- if the distribution of $T(\boldsymbol{X})$ is discrete, " " " don't always exist Ex 7.12
  size $\le \alpha$ always

Example 7.12

- if $H_0$ and/or $H_1$ are composite

sp. case MP test is UMP for some $\theta_1 > \theta_0$
(lucky)
one-sided only

$$\phi(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \frac{f_1}{f_0} > k \\ c & \text{if } \frac{f_1}{f_0} = k \\ 0 & \text{if } \frac{f_1}{f_0} < k \end{cases}$$

(randomized test)

# A word on the $t$-test

```
## Default S3 method:
t.test(x, y = NULL,
       alternative = c("two.sided", "less", "greater"),
       mu = 0, paired = FALSE, var.equal = FALSE,
       conf.level = 0.95, ...)

  > t.test(x= oneline[1,one], y= oneline[1,two], var.equal=T)
  t = -3.014, df = 70, p-value = 0.003589

  > t.test(x= oneline[1,one], y= oneline[1,two])
  t = -3.1323, df = 54.667, p-value = 0.002786

  > pt(-3.1323, df=54.667) #[1] 0.001392839
  > pt(-3.014, df=70)  # [1] 0.001794297
```

AoS

$$S_p^2 = \frac{\Sigma(x_i-\bar{x})^2+\Sigma(y_i-\bar{y})^2}{n_1+n_2-2}$$

$72-2 = n_1+n_2-2$

$$t = \frac{\bar{x}-\bar{y}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$t_{AoS} = \frac{\bar{x}-\bar{y}}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$$

$\bar{X} \sim N(\mu, \sigma^2)$

$\bar{Y} \sim N(\mu, \sigma_2^2)$

7128 × 72

```
leukemia_big <- read.csv
  ("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")
oneline <- leukemia_big[136,]
one <- c(1:20, 35:61) #  I had to extract these manually,
two <- c(21:34, 62:72) #  couldn't figure out the data frame
n1 <- length(one); n2 <- length(two)
mean_one <- sum(oneline[1,one])/n1.  ##[1] 0.7524794
mean_two <- sum(oneline[1,two])/n2.  ##[1] 0.9499731
var_one <-  sum((oneline[1,one]-mean_one)^2)/(n1-1)
var_two <- sum((oneline[1,two]-mean_two)^2)/(n2-1)
pooled <- ((n1-1)*var_one + (n2-1)*var_two)/(n1+n2-2)
taos <- (mean_one-mean_two)/sqrt((var_one/n1)+(var_two/n2))
##[1] -3.132304
tbe <- (mean_one-mean_two)/sqrt(pooled*((1/n1)+(1/n2)))
##[1] -3.085455
```

gene # 136

$t_{stat} = -3.08 < t_{.025, 70}$

- MS definition: $p(\boldsymbol{x}) = \inf\{\alpha : \phi_\alpha(\boldsymbol{x}) = 1\}$ 7.5

p-value 0.0018 × 2 = .0036

ident.

- AoS definition: p-value $= \inf\{\alpha : T(\boldsymbol{x}) \in R_\alpha\}$ Def 10.11

- SM definition $p_{obs} = \Pr_{H_0}\{T(\boldsymbol{X}) \geq t_{obs}\}$

$T(X) = |t\text{-stat}|$

- "Probability of a result as or more extreme than that observed" ??

more away from null

'implied' $H_1$

$\phi(x) = 1\{x : |t| > 1.99\}$ at level .05 evidence against $H_0$

$H_0: \mu_x = \mu_y$ $H_1: \mu_x \neq \mu_y$

# A non-parametric test

$X_i$ = (after − before)$_i$ patient $i = 1, \dots, n$

- $X_1, \dots, X_n$ 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

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

$\sim N\left(\frac{n}{2}, \frac{n}{4}\right)$

- under $H_0$,

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

normal approx

- $p$-value

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

some # out of $n$

# Power of the sign test

power > α

334 7 · Estimation and Hypothesis Testing

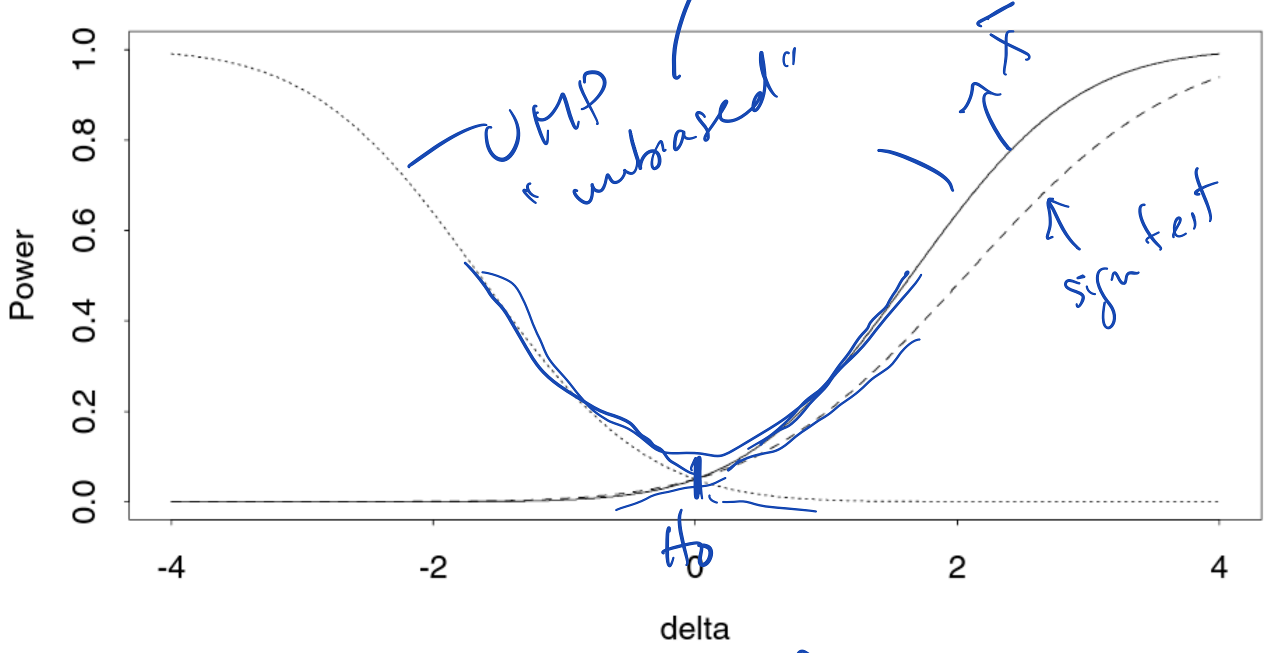



**Figure 7.6** Power functions for a test of whether the mean of a $N(\mu, \sigma^2)$ random sample of size $n$ equals $\mu_0$ against the alternative $\mu = \mu_1$, as a function of $\delta = n^{1/2}(\mu_1 - \mu_0)/\sigma$. The test size is $\alpha = 0.05$. The solid curve is the power function for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

if $X_i \sim N(\mu, \sigma^2)$

- $H_0 : F^{-1}(1/2) = \mu_0$ $H_1 : F^{-1}(1/2) > \mu_0$
- Test statistic $T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$
- $\mathrm{pr}_{H_0}(\text{reject } H_0) = \mathrm{pr}(T \geq c_\alpha \mid H_0) = \alpha \Rightarrow c_\alpha \approx n/2 - n^{1/2} z_\alpha/2$

N approx

$$\mathrm{pr}_{H_1}(T \geq c_\alpha) = \mathrm{pr}_{H_1}\left\{ T \geq \frac{n}{2} - \frac{\sqrt{n}}{2} z_\alpha \right\}$$

$H_1 : F^{-1}(\tfrac{1}{2}) = \mu_1 > \mu_0$

$$= \mathrm{pr}_{H_1}\{ \ ? \ \}$$

exercise

$$E_1\left\{ \sum_{i=1}^{n} 1\{X_i > \mu_0\} \right\} = \mathrm{pr}_{\mu_1}(X_i > \mu_0) = \mathrm{pr}_{\mu_1}(X_i - \mu_1 > \mu_0 - \mu_1)$$

$$= F_1(\mu_0 - \mu_1)$$

# Power of the sign test



- $H_0 : F^{-1}(1/2) = \mu_0 \qquad H_1 : F^{-1}(1/2) > \mu_0$
- Test statistic $T = \sum_{i=1}^n 1\{X_i > \mu_0\}$
- $\mathrm{pr}_{H_0}(\text{reject } H_0) = \mathrm{pr}(T \geq c_\alpha \mid H_0) = \alpha \Rightarrow c_\alpha \approx n/2 - n^{1/2} z_\alpha/2$
- $\mathrm{pr}_{H_1}(\text{reject } H_0) = \mathrm{pr}(T \geq c_\alpha \mid H_1)$

Need distribution of $T$ under $H_1$

- to calculate power we need values for $\mu$ and for $F$
- e.g. change to $H_1 : F^{-1}(1/2) = \mu_1$

$\mathrm{pr}_{F_{\mu_1}}(X > \mu_0)$

- SM assumes $F$ is $N(\mu, \sigma^2)$, and uses normal approximation to dist'n of $T$

sign test

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

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

$\delta = n^{1/2}\underline{(\mu_1 - \mu_0)}/\sigma$

UMP test (HW6)

normal cdf

leukemia data (EH): $\underbrace{X_1, \ldots, X_{47}}_{\text{ALL}}$ ; $\underbrace{Y_1, \ldots, Y_{25}}_{\text{AML}}$ AoS Ex. 10.20

gene 136

```
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    ALL.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
```

$t = -3.01$

$p \approx .003$

$X_i \sim N$

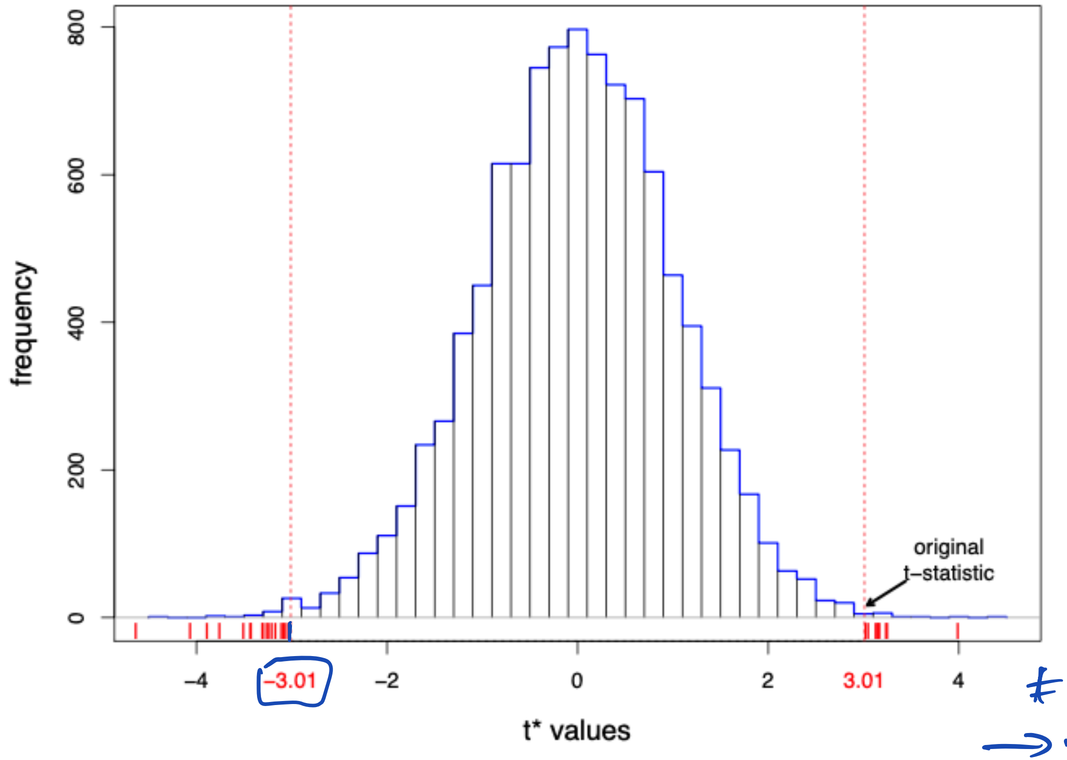
$Y_i \sim N$

$H_0 : F_X = F_Y$

nonparametric

permute the row; relabel
47 ALL
25 AML

$\binom{72}{47}$




# perm
→ |t| > 3.01

**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~~.

.0036 from t dist

AoS 10.20 uses median

# Choosing test statistics

1. Context
2. Optimal choice – Neyman-Pearson Lemma
3. Pragmatic choice – likelihood-based statistics
4. Pragmatic choice – nonparametric test statistics

Need

1. $T = T(\underline{X})$ : test st.

~~need~~ 2. know $P_{H_0}\{T(\underline{X}) > c\}$

null dist

3. Assess it's properties

# Hypothesis tests and significance tests

- Hypothesis tests typically means:
  - $H_0$, $H_1$
  - critical/rejection region $R \subset \mathcal{X}$,
  - level $\alpha$, power $1 - \beta$
  - conclusion: “reject $H_0$ at level $\alpha$” or “do not reject $H_0$ at level $\alpha$”
  - 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 $\alpha$, power $1 - \beta$
  - conclusion: "reject $H_0$ at level $\alpha$" or "do not reject $H_0$ at level $\alpha$"
  - 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

more extreme

large $T$ points to alternative

# Hypothesis tests and significance tests

- Hypothesis tests typically means:
  - $H_0$, $H_1$
  - critical/rejection region $R \subset \mathcal{X}$,
  - level $\alpha$, power $1 - \beta$
  - conclusion: "reject $H_0$ at level $\alpha$" or "do not reject $H_0$ at level $\alpha$"
  - 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 $H_0$" "reject $H_0$"

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

# 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 $\alpha$-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 X
2. If the significance level of a test is decreased, the power would be expected to ~~increase~~ +1 X
3. If the test is rejected at level $\alpha$, the probability that the null hypothesis is true equals $\alpha$. X
4. The probability that the null hypothesis is falsely rejected is equal to the power of the test X

$P_{H_0}\{T > t_\alpha\}$

$P_{H_1}\{T > t_\alpha\}$

# Hypothesis tests and significance tests

$\alpha = 0.05$

power = 80%

type 2 error = .2 or .1 90%

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

- overlap: $p^{obs}$ is the smallest $\alpha$-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 $\alpha$, the probability that the null hypothesis is true equals $\alpha$.
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 mostly
7. The power of a test is determined by the null distribution of the test statistic ✗

### 1. Hypothesis testing

AoS Table 10.1

| | | $H_0$ not rejected | $H_0$ rejected |
|---|---|---|---|
| truth | $H_0$ true | ✓ | type 1 error ← $\phi$ |
| | $H_1$ true | type 2 error | ✓ |

### 2. Diagnostic testing

link

| | | test negative | test positive | |
|---|---|---|---|---|
| truth | $H_0$ C19 neg | TN | FP | N |
| | C19 pos | FN | TP | P |

TP / P sensitivity

TN / N specificity

# An underpowered study

Hernandez et al, JAMA 2019

| | Died | Lived | |
|---|---|---|---|
| New | 74 | 138 | 212 |
| Old | 92 | 120 | 212 |
| Total | 166 | 258 | 424 |

septic shock treat

RCT

cannot reject $H_0$
@ level .05

mortality difference 92/212 = 43%; 74/212 = 34%; difference = -8%

clinically significant

*p*-value (comparing two binomial proportions) 0.07

95% confidence interval for difference $(-18.2\%, 1.2\%)$

0 is in here

"we planned to enrol 420 patients. We calculated that with this sample size the study would have 90% power to detect a reduction in 28-day mortality from 45% to 30%, at an $\alpha$-level of 0.05"

$\Pr_{H_1}\{T > c_\alpha\}$

15% red.

# Diagnostic testing and ROC

## 2. Diagnostic testing

link

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

## 3. Multiple testing

| | | $H_0$ not rejected | $H_0$ rejected | |
|---|---|---|---|---|
| truth | $H_0$ true | $U$ | $V$ | $m_0$ |
| | $H_1$ true | $T$ | $S$ | $m_1$ |
| | | $m - R$ | $R$ | m |

false discovery rate

# Multiple testing

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

gene 136

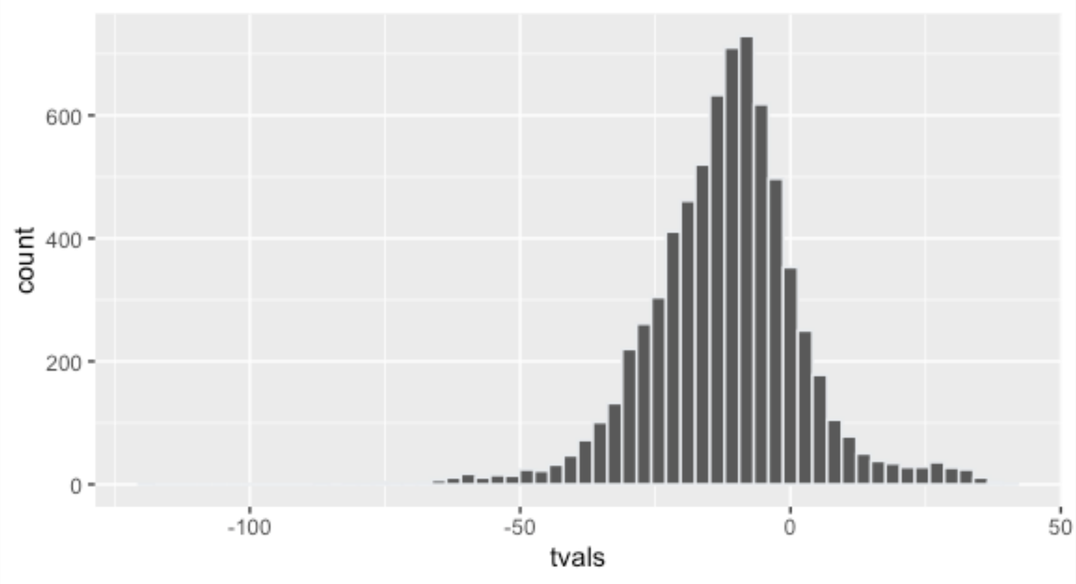
new cutoff $\alpha'$ $\dfrac{.05}{7128}$ .00004 ?

- 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

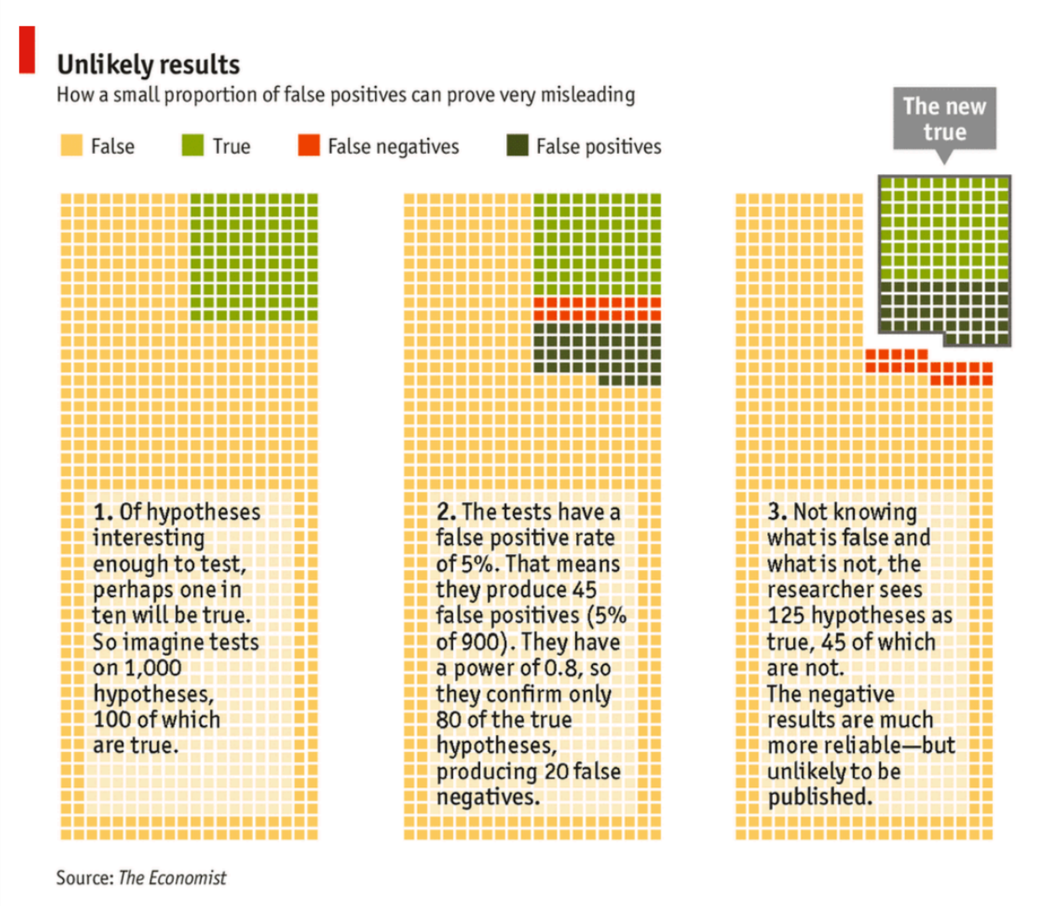





7,128 2-sample
t-stat.

```
> summary(tvals)
Min.    1st Qu.    Median      Mean   3rd Qu.      Max.
-118.793  -19.926  -11.231  -12.019   -4.218   41.015
```

# Multiple testing



# Multiple testing

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

# Multiple testing



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

very conservative

- FDR method controls the number of rejections that are false

FDP = V/R

prop.

| | | $H_0$ not rejected | $H_0$ rejected | |
|---|---|---|---|---|
| truth | $H_0$ true | $U$ | $V$ | $m_0$ |
| | $H_1$ true | $T$ | $S$ | $m_1$ |
| | | $m - R$ | $R$ | m |

multiple testing

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

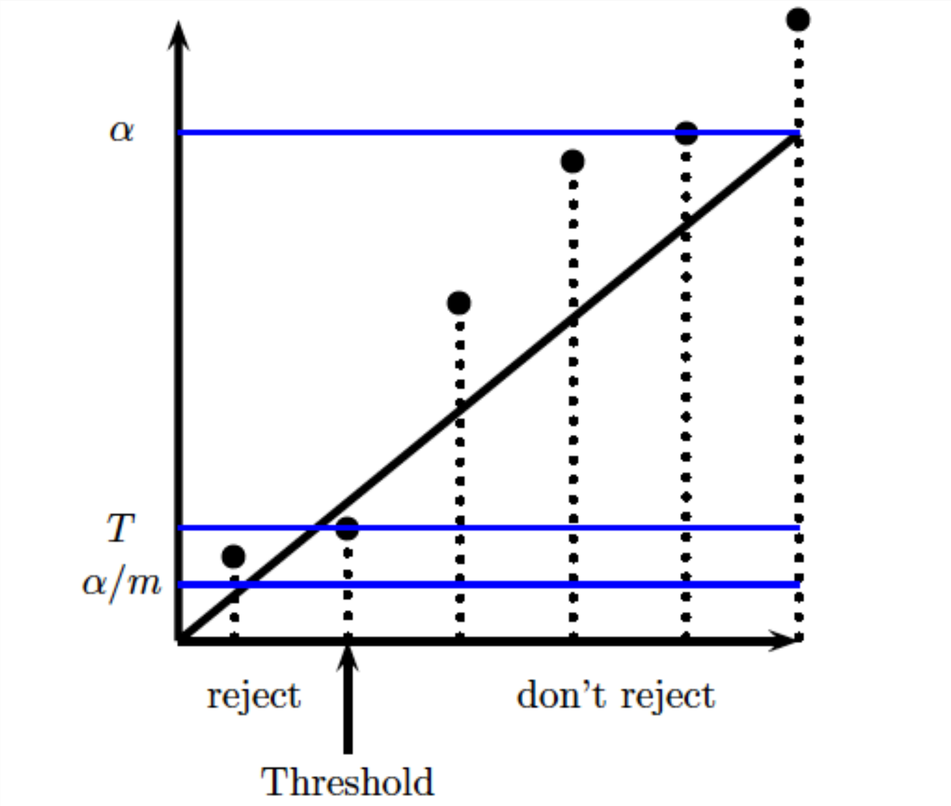
$$FDR(BH_q) = \pi_0 q \leq q, \qquad \text{where } \pi_0 = m_0/m$$

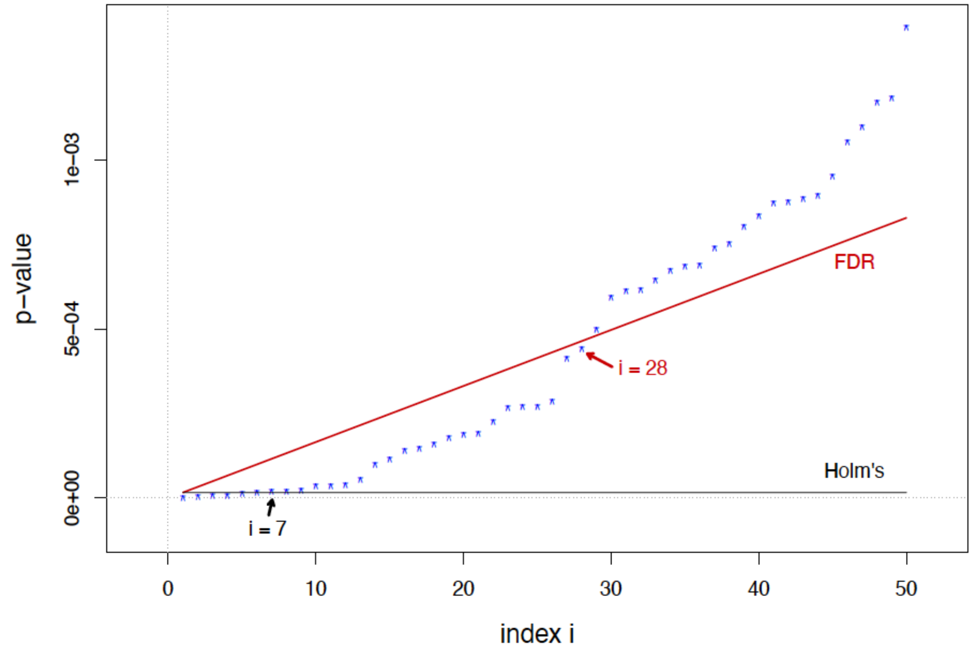
$\pi_0$ unknown but close to 1

- 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



# Example



| 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 |

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

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