Mathematical Statistics II

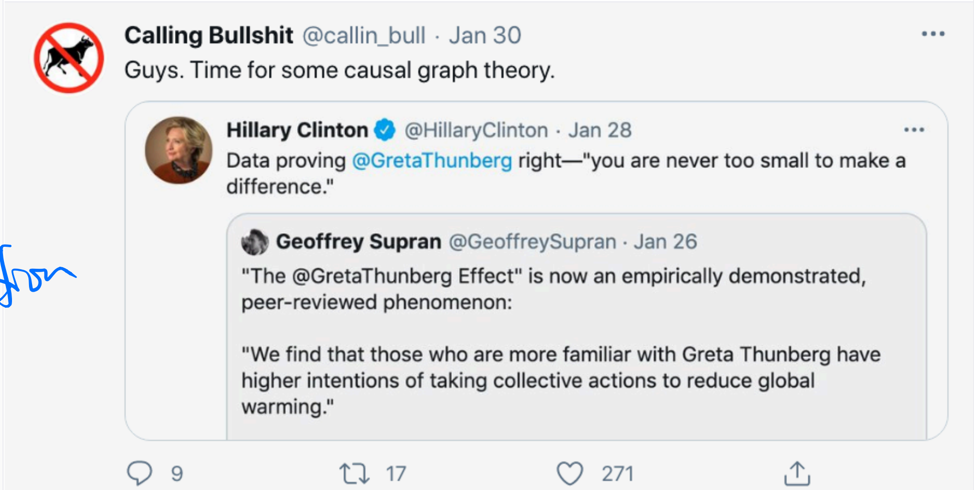
STA2212H S LEC9101

Week 4

February 3 2021

Start recording!

correlation $\neq$ causation



Received: 5 December 2020 | Accepted: 18 December 2020
DOI: 10.1111/jasp.12737

ORIGINAL ARTICLE

Journal of Applied Social Psychology WILEY

The Greta Thunberg Effect: Familiarity with Greta Thunberg predicts intentions to engage in climate activism in the United States

Anandita Sabherwal[1] | Matthew T. Ballew[2] | Sander van der Linden[1] | Abel Gustafson[3] | Matthew H. Goldberg[4] | Edward W. Maibach[5] | John E. Kotcher[5] | Janet K. Swim[6] | Seth A. Rosenthal[4] | Anthony Leiserowitz[4]

Calling Bullshit @callin_bull · Jan 30

Guys. Time for some causal graph theory.

> **Hillary Clinton** @HillaryClinton · Jan 28
>
> Data proving @GretaThunberg right—"you are never too small to make a difference."
>
> > **Geoffrey Supran** @GeoffreySupran · Jan 26
> >
> > "The @GretaThunberg Effect" is now an empirically demonstrated, peer-reviewed phenomenon:
> >
> > "We find that those who are more familiar with Greta Thunberg have higher intentions of taking collective actions to reduce global warming."

9 17 271

Received: 5 December 2020 | Accepted: 18 December 2020
DOI: 10.1111/jasp.12737

ORIGINAL ARTICLE

Journal of Applied Social Psychology WILEY

The Greta Thunberg Effect: Familiarity with Greta Thunberg predicts intentions to engage in climate activism in the United States

Anandita Sabherwal[1] | Matthew T. Ballew[2] | Sander van der Linden[1] | Abel Gustafson[3] | Matthew H. Goldberg[4] | Edward W. Maibach[5] | John E. Kotcher[5] | Janet K. Swim[6] | Seth A. Rosenthal[4] | Anthony Leiserowitz[4]

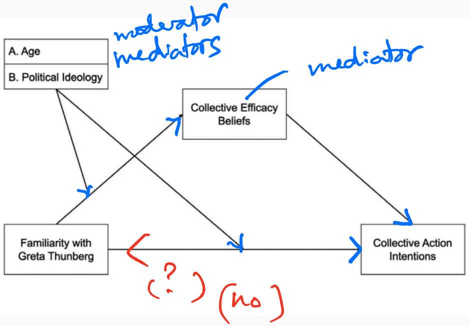


FIGURE 1 Conceptual model for hypotheses. Model tests the effect of familiarity with Greta Thunberg on intentions to take collective action through collective efficacy beliefs, as a simple mediation (Hypothesis 1), moderated by age (A; Hypotheses 2a and 2b), and moderated by political ideology (B; Hypotheses 3a and 3b), respectively

Calling Bullshit @callin_bull · Jan 30

Guys. Time for some causal graph theory.

> **Hillary Clinton** @HillaryClinton · Jan 28
>
> Data proving @GretaThunberg right—"you are never too small to make a difference."
>
> > **Geoffrey Supran** @GeoffreySupran · Jan 26
> >
> > "The @GretaThunberg Effect" is now an empirically demonstrated, peer-reviewed phenomenon:
> >
> > "We find that those who are more familiar with Greta Thunberg have higher intentions of taking collective actions to reduce global warming."

9 17 271

LRT Wald t-test

NP Lemma + we guesstimate $T = t(X)$

- Choosing test statistics – optimality, likelihood, convenience
- Simple and composite hypotheses $\quad H_0: \underbrace{\theta = \theta_0}^{S}; \quad \underbrace{\theta \in \Theta_0}^{C}$
- Generalized likelihood ratio tests $\quad$ C. $W(\theta_0) = 2\{\ell(\hat{\theta}) - \ell(\tilde{\theta}_0)\}$ $\quad \tilde{\theta}_0 = \arg\sup_{\Theta_0} L(\theta)$
- Wald tests $\quad (\hat{\theta} - \theta_0)/\widehat{se} \quad \xrightarrow{\cdot} N(0,1)$
- significance tests $\quad$ p-value: $\Pr(T > t^{obs}; H_0)$

≠ hypoth. "

$p < .05$ *

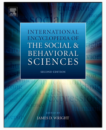
$< .01$ **

↑ report value

? est'd effect, se?

$p \neq pr(H_0 \text{ true})$

$= pr(\text{this data or more; if } H_0)$



interp. β_j

$\cancel{\log\left\{\frac{p(x^T\beta)}{1-p(x^T\beta)}\right\} =}$

$$\log\left\{\frac{pr(y=1 \mid x)}{pr(Y=0 \mid x)}\right\} = x^T\beta$$

N(,)

$\hat\beta$

$\boxed{\hat\beta_j \pm s.\hat{e}_j \approx 95\%}$ CI

odds ratio $e^{\hat\beta_j} = \hat\theta_j$

for $\uparrow$ in x_j

$\left(e^{\hat\beta_L}, e^{\hat\beta_U}\right)$

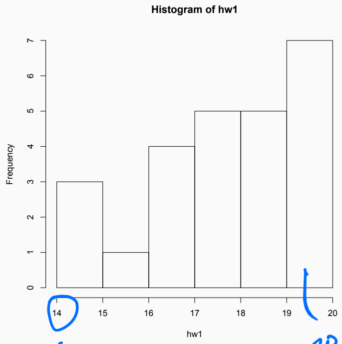
$\xrightarrow{d} N(,)$

$\sim$ 95% CI OR

Today

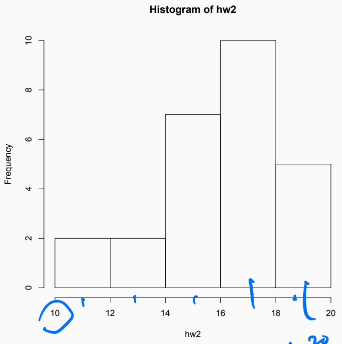
Start Recording

1. Homework 2 (Friday)
2. confidence intervals
3. goodness-of-fit tests Sig. testing
4. multiple testing ← hyp. testing



14 20

, cex = 2



10 19, 20

- February 8 3.00 – 4.00 Paul McNicholas
- “Selected Problems in Classification” Link

Data Science and Applied Research Series

McMaster

Confidence intervals

p_j^{obs}, $\hat\beta_j \pm 2\hat{se}_j$ (95% CI)

AoS Theorem 10.10; SM §7.3.4

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -34.103704   6.530014  -5.223 1.76e-07 ***
zn           -0.079918   0.033731  -2.369  0.01782 *
indus        -0.059389   0.043722  -1.358  0.17436
chas          0.785327   0.728930   1.077  0.28132
nox          48.523782   7.396497   6.560 5.37e-11 ***
rm           -0.425596   0.701104  -0.607  0.54383
age           0.022172   0.012221   1.814  0.06963 .
dis           0.691400   0.218308   3.167  0.00154 **
rad           0.656465   0.152452   4.306 1.66e-05 ***
tax          -0.006412   0.002689  -2.385  0.01709 *
ptratio       0.368716   0.122136   3.019  0.00254 **
black        -0.013524   0.006536  -2.069  0.03853 *
lstat         0.043862   0.048981   0.895  0.37052
medv          0.167130   0.066940   2.497  0.01254 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$\hat\beta$ $\hat{se}$ $\frac{\hat\beta}{\hat{se}}$

$\mathrm{pr}\left\{|N(0,1)| \geq \frac{\hat\beta^{obs}}{se}\right\}$

95% C.I. for β_{age}

$0.022 \pm 1.96 \cdot 0.122$

$= (-0.0?,\ .03?)$

↑ 0 is included

bec. $p > 0.05$

⟷

99% CI include 0

$\Longleftrightarrow p > .01$

$(-0.01,\ .2)$

$p \simeq .06$

logistic regression

Confidence intervals

nuclear ex.

$\boxed{y_i \sim N(x_i^T\beta, \sigma^2)}$

$\dfrac{\hat\beta_j}{\hat{se}} \sim t_{df=21}$ exact

```
> summary(lm1)
Call:
lm(formula = log(cost) ~ date + log(t1) + log(t2) + log(cap) +
    pr + ne + ct + bw + log(cum.n) + pt, data = nuclear)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -14.24198    4.22880  -3.368  0.00291 **
date          0.20922    0.06526   3.206  0.00425 **
log(t1)       0.09187    0.24396   0.377  0.71025
log(t2)       0.28553    0.27289   1.046  0.30731
log(cap)      0.69373    0.13605   5.099 4.75e-05 ***
pr           -0.09237    0.07730  -1.195  0.24542
ne            0.25807    0.07693   3.355  0.00300 **
ct            0.12040    0.06632   1.815  0.08376 .
bw            0.03303    0.10112   0.327  0.74715
log(cum.n)   -0.08020    0.04596  -1.745  0.09562 .
pt           -0.22429    0.12246  -1.832  0.08125 .
---
Residual standard error: 0.1645 on 21 degrees of freedom
```

$H_0: \beta_j = 0$

$\boxed{.258 \pm t_{21,\alpha/2} \cdot 0.08}$

$1-\alpha$ CI based on t

$1-\alpha < .997$

$1-\alpha$ CI not include 0

Confidence intervals

composite ≠

$H_0: \underset{\sim}{\beta} = \underset{\sim}{0}$

(except β_0)

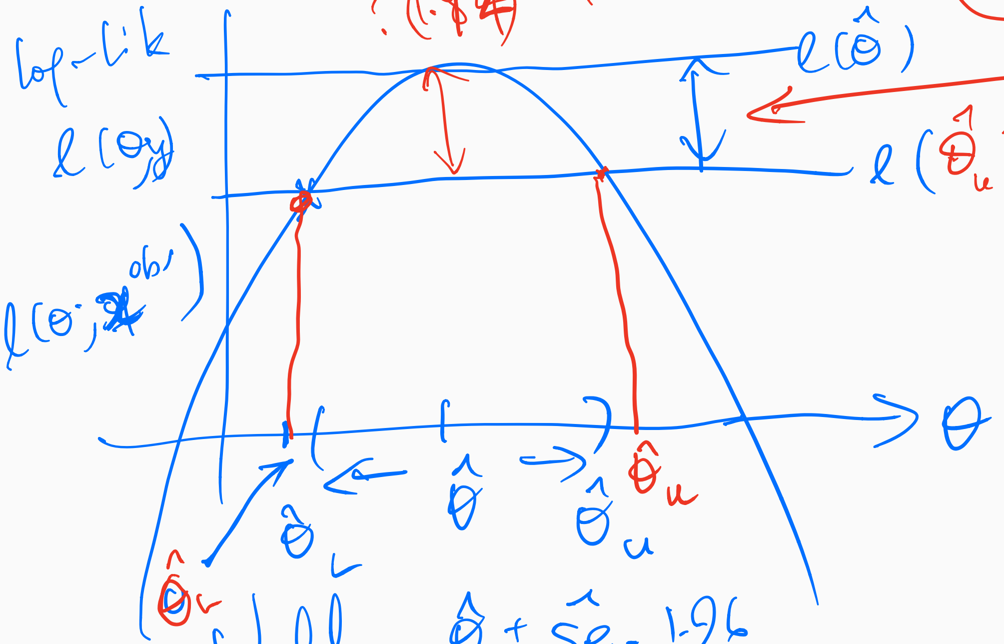
```
log(cap)       0.69373    0.13605   5.099 4.75e-05 ***
pr            -0.09237    0.07730  -1.195  0.24542
ne             0.25807    0.07693   3.355  0.00300 **
ct             0.12040    0.06632   1.815  0.08376 .
bw             0.03303    0.10112   0.327  0.74715
log(cum.n)    -0.08020    0.04596  -1.745  0.09562 .
pt            -0.22429    0.12246  -1.832  0.08125 .
---
Residual standard error: 0.1645 on 21 degrees of freedom
Multiple R-squared:  0.8717,Adjusted R-squared:  0.8106
F-statistic: 14.27 on 10 and 21 DF,  p-value: 3.081e-07
```

$$\frac{SSR/10}{SSE/21} \sim F_{10,21}$$

$W = 2\{\ell(\hat{\theta}) - \ell(\tilde{\theta}_0)\}$ 1-1 fn of

?(1.92) if $1-\alpha = 0.95$ 3.84 — 95% pt (??)

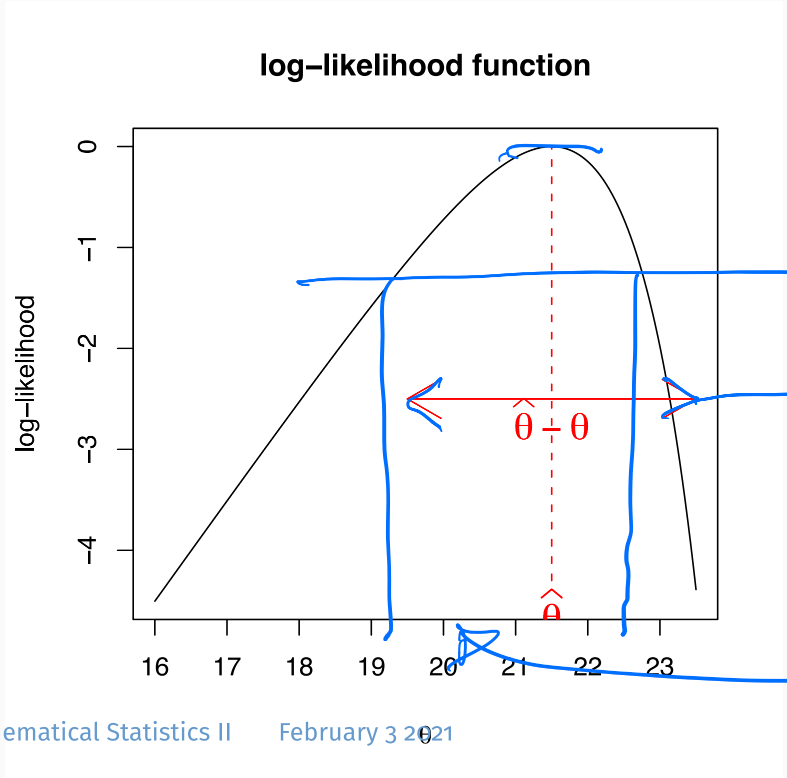
$\left(\frac{1}{2}\chi^2_{1,\alpha}\right)$



Wald $\hat\theta \pm \hat{se} \cdot 1.96$

$$2\{\ell(\hat\theta) - \ell(\theta)\} \overset{\cdot}{\sim} \chi^2_1$$

$H: \theta = \theta_0$



V. 1

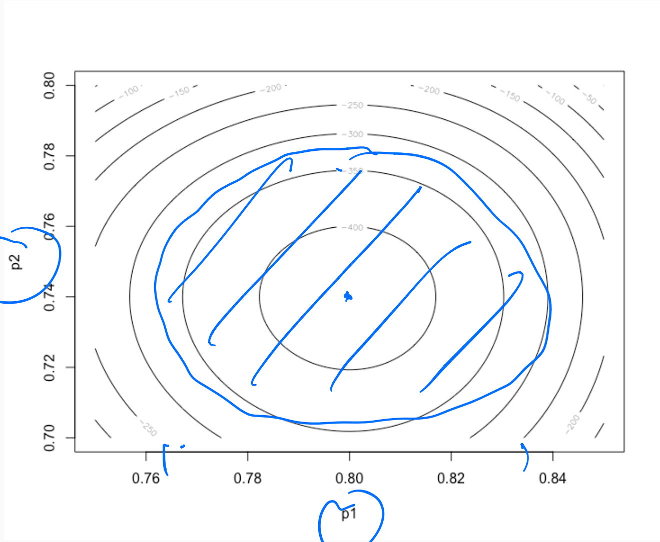
V. 2

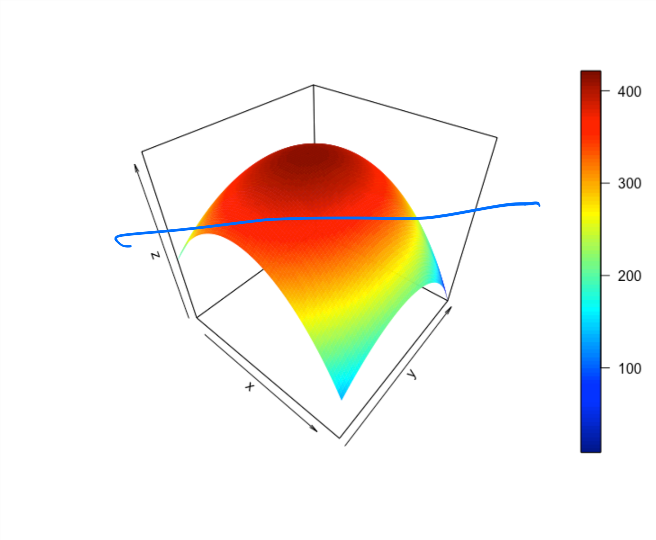
Wald (in R output)

LRT ✓

fit several models

θ changing





χ^2_2 perc. pt.

$H_0: \beta = 0$

single
H_0 $\beta \in \mathbb{R}^2$

these C.I.'s are based on tests

$1-\alpha$ CI $\longleftrightarrow$ α size test

power of a test β type 2 error

$\longleftrightarrow$ length of CI. $1-\beta$ power

(size) $Pr_{H_1}(\text{reject } H_0)$

UMP level α tests high test is "powerful"

lead to smallest upper bound

on 1-sided CI

- interpretation of confidence interval or confidence bound — random endpoint(s)

$\bar{X} \pm 2 \cdot \frac{s}{\sqrt{n}}$ ← s^2 est σ^2

- size of test ↔ confidence level

ahead of time (roughly)

- power of test ↔ width of confidence interval

ahead of time

often length w

$\hat{\theta} \pm 2 \cdot \hat{se}$

5 units effect

$4\hat{se} < 5$

Length $= (4\hat{se})$

$\frac{\hat{\sigma}}{\sqrt{n}}$ small

n large

2-sided ??

- data $x_1, \ldots x_n$ independent, identically distributed
- test statistic $t = t(\boldsymbol{x})$, observed value t^{obs}
-

$$p^{obs} = \Pr(T \geq t^{obs}; H_0)$$

obs'd p-value

H_0

(no H_1)

- Example: sign test

dist^n of T under H_0:

$X_1, \ldots, X_n$ iid $F(\cdot)$ cont^s $\quad F(x) = \Pr(X \leq x)$

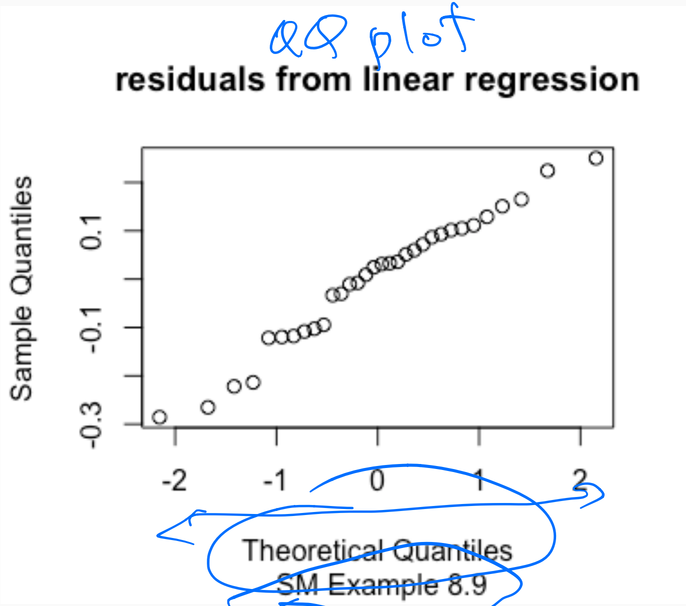
H_0: median$(F) = \mu_0$ $\quad F(\mu_0) = \frac{1}{2}$

✓ $T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$ $\quad P(T \geq t^{obs}; H_0)$ ✓

Bernoulli / H_0

$= \Pr\{Bin(n, \tfrac{1}{2}) \geq t^{obs}\}$

QQ plot



nuclear

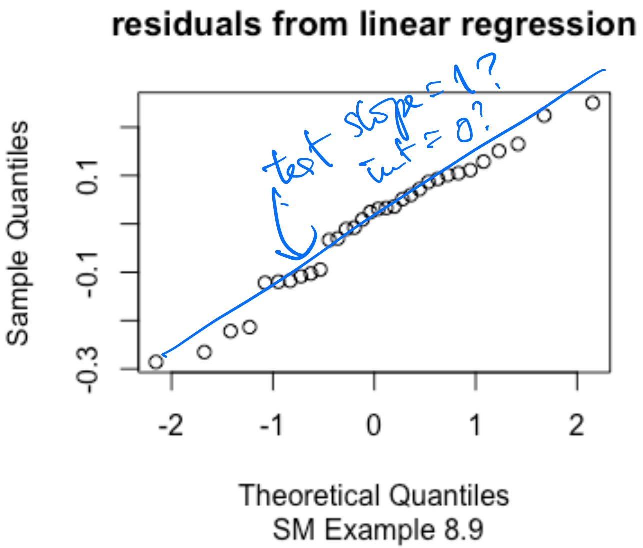
N ass^n is ok?

Test H_0: $\varepsilon_i \sim N(,)$

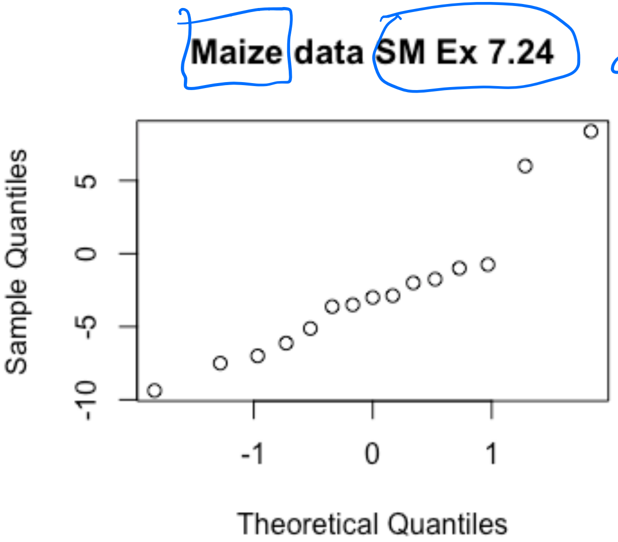
$y_i = x_i^T \beta + \varepsilon_i$ $(x_i = z_i^T\theta + \varepsilon_i)$

$\boxed{\varepsilon_i \sim N(0, \sigma^2)}$

Goodness-of-fit tests



test slope = 1?
int = 0?

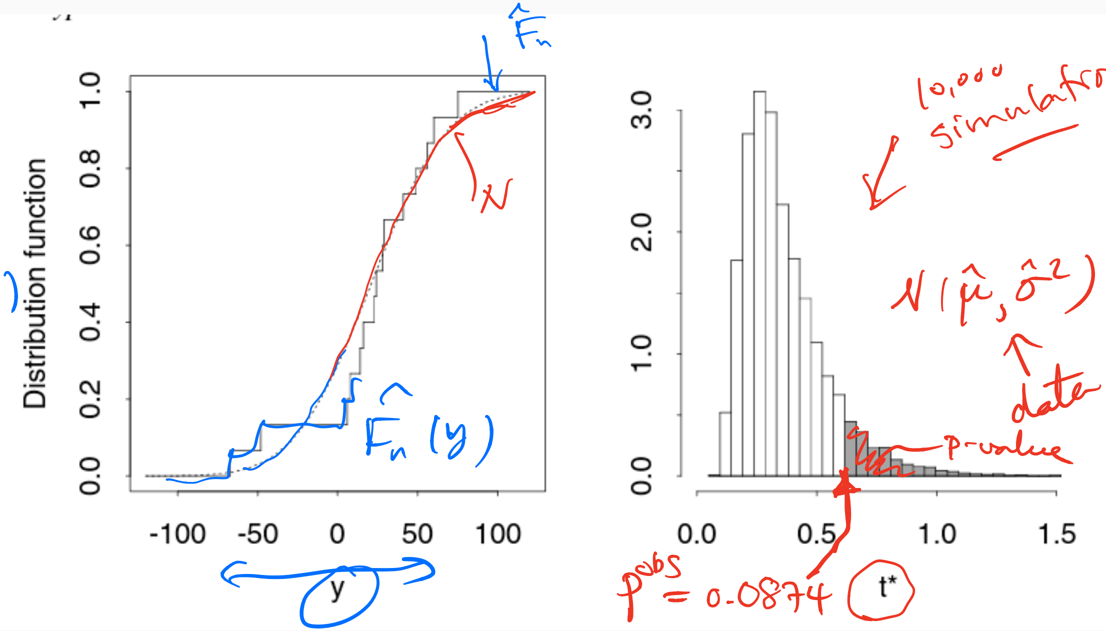


data(darwin)

set of differences in a paired experiment

Figure 7.5 Analysis of maize data. Left: empirical distribution function for height differences, with fitted normal distribution (dots). Right: null density of Anderson–Darling statistic T for normal samples of size $n = 15$ with location and scale estimated. The shaded part of the histogram shows values of T^* in excess of the observed value t_{obs}.

$$T_3 = \int \frac{\{\hat{F}_n(t) - F_0(t)\}^2}{\hat{F}_n(t)\{1-\hat{F}_n(t)\}}\,dt$$



SM Example 7.24 testing $N(\mu, \sigma^2)$ distribution

SM Example 7.23; 6.14 testing $U(0, 1)$ distribution

More interesting example of pure sig test

$X_1, \ldots, X_n$ iid $F(\cdot)$ H_0: $F(\cdot) = F_0(\cdot)$ $H_1: F \neq F_0$

$t(\underline{X}) =$ something need $\mu_{H_0}\{T \geq t^{obs}\}$

ex. $F_0 = U(0,1)$ H_0: simple

$F_0 = N(\underline{\mu, \sigma^2})$ H_0: composite

$$\boxed{\hat{F}_n(t) \triangleq \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \leq t\} \qquad t \in \mathbb{R} \qquad \in [0,1]}$$

Potential Test statistics: "measure dist. from $\hat{F}_n$ to F_0"

(i) $\sup_t |\hat{F}_n(t) - F_0(t)| = T_1$ Kolmogorov-Smirnov test

$pr_{H_0}(T_1 \geq t_1^{obs})$

(ii) $\int \{\hat{F}_n(t) - F_0(t)\}^2 \, dt = T_2$ Cramer-von Mises

(iii) $\int \frac{\{\hat{F}_n(t) - F_0(t)\}^2}{\hat{F}_n(t)\{1 - \hat{F}_n(t)\}} \, dt = T_3$ Anderson-Darling

- $X_1, \ldots, X_n$ *i.i.d.* $F(\cdot)$; $H_0 : F = F_0$ — cumulative d.f.
- $\boxed{\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i \le t\}}$
- test statistic:
  1. $\sup_t |\widehat{F}_n(t) - \tilde{F}_0(t)|$ - KS
  2. $\int \{\widehat{F}_n(t) - \tilde{F}_0(t)\}^2 dt$ - CvM
  3. $\int \frac{\{\widehat{F}_n(t) - \tilde{F}_0(t)\}^2}{\widehat{F}_n(t)\{1 - \widehat{F}_n(t)\}} dt$ - AD
  4. χ^2 tests ← coming.
- SM Example 7.24 testing $N(\mu, \sigma^2)$ distribution
- SM Example 7.23; 6.14 testing $U(0,1)$ distribution

If F_0 has unk. parameters, then replace by m.l.e.'s

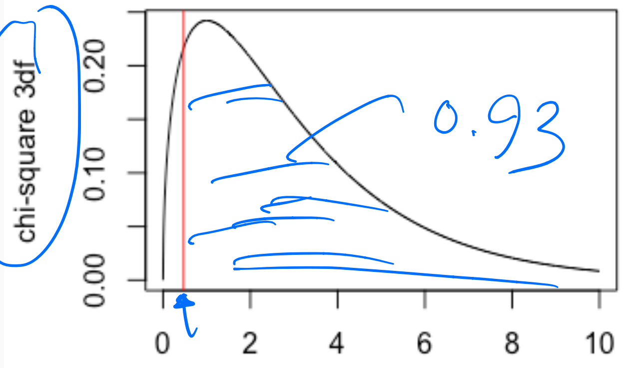
$F_0 : N(\mu, \sigma^2)$ $\quad \hat{\mu} = \bar{X}$ $\quad \hat{\sigma} = \cancel{s^2}$ or

$\tilde{F}_0(t) = \Phi\left(\frac{t - \hat{\mu}}{\hat{\sigma}}\right)$ — AoS 10.4

$\frac{\sum(x_i - \bar{x})^2}{n-1}$ ← unbiased or m.l.e. $\frac{\sum(x_i - \bar{x})^2}{n}$

AoS Example 10.18

$T_4 \sim \chi^2_{k-1-s}$ s: dim θ



don't use mle $\hat{\theta}$

use $\tilde{\theta}$

4. χ^2: test

H_0: $X_1, \ldots, X_n$ iid $f(x; \theta)$

θ nuisance parameter

$$T_4 = \chi^2 := \sum_{j=1}^{k} \frac{\{N_j - p_j(\hat{\theta})\}^2}{n p_j(\hat{\theta})}$$

N_j = # X's in jth interval

$$p_j(\theta) = \int_{I_j} f(x; \theta)\,dx$$

$N_1 \quad N_2 \quad \ldots \quad N_k$

~~Pause~~

Example 10.23 in AoS has a computation re multinomial dist^{n}. Mendel

$\cancel{X} \sim M(n; \underline{p})$ $\quad \underline{p} = (p_1, p_2, p_3, p_4) \quad \sum_{j=1}^{4} p_j = 1$

(simple)

$H_0: \underline{p}_0 = \underline{\left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)}$ various types of pea plants cross-breed

$n = 556 \quad X = (\underline{315, 101, 108, 32})$

$\chi^2_3 = \sum \frac{(N_j - np_j(\hat{\theta}_0))^2}{np_j(\hat{\theta}_0)} = \underline{0.47}$

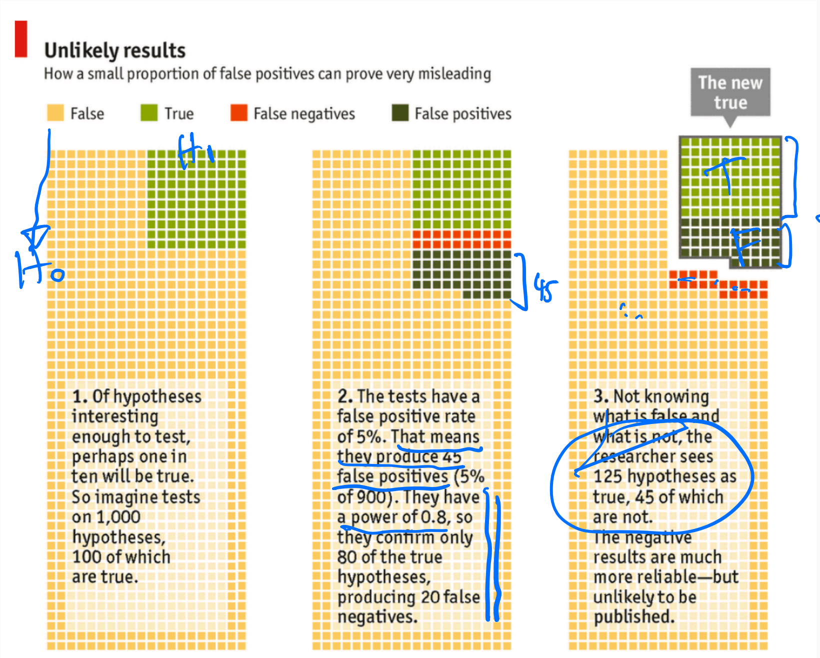
critical value $\alpha = .05$
$= 7.8$

p-value $= 0.93$

Multiple testing

many H_0 tests m H_{0i} p_i $i=1,\dots,m$

	retain H_0	rej. H_0
H_0		α (P_r) .05 say
H_1	β_1 t2	$1-\beta$ (power)



Source: *The Economist*

test 100 H_0's @ .05 (no ~~true~~ diff)

~5 rej H_0 it's false

	ret H_0	rej H_0	
H_0		45	900
H_1		80	100
		125	1000

125 "reject H_0"

80 "true rejections"

$\frac{45}{125} \approx .3$

~~Bonferroni correction~~ FDP

	retain H_0	rej. H_0	
H_0 true	U	V	m_0 ?
H_1 true	T	S	m_1
	$m-R$	R	m

$U+V = m_0$

$T+S = m_1$

V false rejections

R false + true rejections

$\frac{V}{R} \triangleq$ "false discovery proportion"

if $R=0$ $\frac{V}{R} = 0$

Bonferroni

Carry ~~of~~ out m tests H_{0i}, $i=1,\dots,m$

$$\Pr\{\text{reject any } H_{0i} \mid H_{0i};\ i=1,\dots,m\} \le \alpha$$

FWER family-wise error rate

each test at α/m level. $\alpha=.05$ $m=100$

(.0005 e.g.)

$$=\Pr\left[\bigcup_{i=1}^{m}\left\{p_i^{obs} \le \frac{\alpha}{m}\right\};\ H_{0i}\ i\dots\right]$$

e.g. $m=5000$

$$\le \sum_{i=1}^{m}\Pr\left(p_i \le \frac{\alpha}{m}\right) = m\frac{\alpha}{m} = \alpha$$

$\alpha \downarrow \dots$

$p = 0.03$ (~200 tests)

Benjamini-Hochberg

particle physics

$p = 3 \times 10^{-7}$ $\quad 5\sigma$

"claim a discovery"

"5-sigma" $P_1\{N(0,1) \geq 5\}$

adj. "look elsewhere"

Benjamini-Hochberg