

Mathematical Statistics II

STA2212H S LEC9101

Week 7

February 25 2025

**SIGMA XI**
THE SCIENTIFIC RESEARCH SOCIETY

SmartBrief

Your World of Science News

SIGN UP · SHARE

TOP STORY

Study suggests dogs may have self-domesticated for food



(Drbimages/Getty Images)

Research proposes that dogs may have domesticated themselves by opting to stay near humans for food, a process potentially driven by natural selection over 15,000 years. Published in the Proceedings of the Royal Society B, the study uses a statistical model to explore this hypothesis, suggesting that wolves who preferred human proximity and selected similarly tame mates could have initiated the domestication process. **Full Story:** [Live Science](#)

(2/24)

[in](#) [X](#) [f](#) [e](#)

PROCEEDINGS OF THE ROYAL SOCIETY B

BIOLOGICAL SCIENCES

 Open Access

 Check for updates

View PDF

 Tools  Share

Cite this article 

Section

Abstract

Research articles

Rapid evolution of prehistoric dogs from wolves by natural and sexual selection emerges from an agent-based model

David C. Elzinga, Ryan Kulwicki, Samuel Iselin, Lee Spence and Alex Capaldi 

Published: 12 February 2025 | <https://doi.org/10.1098/rspb.2024.2646>

 **Review history**

Abstract

Today

1. Recap Feb 11 confidence intervals and regions, exact and approximate, approximate posterior
2. Hypothesis testing and significance testing
3. Project: Choice of papers ✓
4. HW6

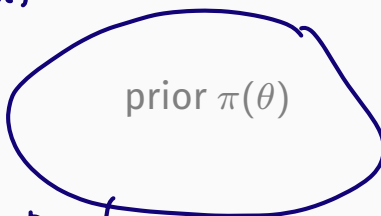
Recap: Confidence intervals

$$\underline{x} \sim f(\underline{x}; \underline{\theta}) \quad \underline{\theta} \in \mathbb{R}^p \quad \text{MS 7.1,2; AoS 6.3.2}$$

$\underline{x} \in \mathbb{R}^n$ $p < n$

- $p = 1$: $(1 - \alpha)$ confidence interval, confidence bound
- $p > 1$: $(1 - \alpha)$ confidence region
- $p = 1$: $(1 - \alpha)$ credible interval, credible bound
- $p > 1$: $(1 - \alpha)$ credible set
- $(1 - \alpha)$ is the coverage of the confidence interval or region or the probability of the credible interval or set
- coverage of confidence regions and intervals can be exact or approximate

} aleatory or empirical prob. "how well method works" see next slides

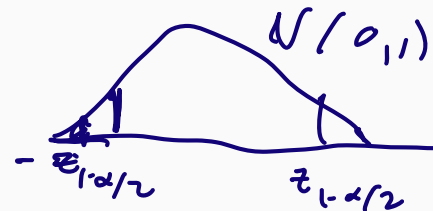
} ← epistemic prob. "how uncertain you are..."


Recap $\mathbf{X} \sim f(\mathbf{x}; \theta), \quad \theta \in \mathbb{R}$

- approximate $(1 - \alpha)$ confidence interval

$$\hat{\theta} \pm z_{1-\alpha/2} \hat{se}(\hat{\theta}) = \hat{\theta} \pm z_{1-\alpha/2} j^{-1/2}(\hat{\theta})$$

$\hat{\theta}$ mle



- approximate pivotal quantity $(\hat{\theta} - \theta) j^{1/2}(\hat{\theta}) \sim Z$

$$-l''(\hat{\theta}, \mathbf{x}) = j(\hat{\theta}) = O_p(n)$$

- approximate $(1 - \alpha)$ credible interval

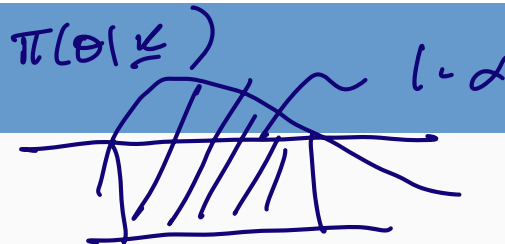
$$\tilde{\theta} \pm z_{1-\alpha/2} \tilde{j}^{-1/2}(\tilde{\theta})$$

$$\tilde{\theta} = \arg \sup_{\theta} \pi(\theta | \mathbf{x})$$

- approximate normality of posterior

$$\pi(\theta | \mathbf{x}) \approx N\{\hat{\theta}, j^{-1/2}(\hat{\theta})\} \quad \text{or} \quad \pi(\theta | \mathbf{x}) \approx N\{\tilde{\theta}, \tilde{j}^{-1/2}(\tilde{\theta})\}$$

Recap: Confidence and credible regions



$\theta \in \mathbb{R}$ or $\mathbb{R}^p$

- Bayesian: HPD region C for θ :

$$\left\{ \begin{array}{l} (1) \quad \int_C \pi(\theta | \mathbf{x}) = 1 - \alpha \\ (2) \quad \pi(\theta | \mathbf{x}) \geq \pi(\theta^* | \mathbf{x}), \quad \text{any } \theta \in C \text{ and } \theta^* \notin C \end{array} \right.$$

motivation



- approximate pivotal quantities

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

$\theta \in \mathbb{R}$

$$(\hat{\theta} - \theta)^T I_n(\hat{\theta})(\hat{\theta} - \theta)$$

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

$$I_n(\theta) = E_{\theta} \left\{ -\ell''(\theta; \underline{x}) \right\}$$

(M.S. not)

χ_p^2

LRT

- approximate confidence regions

$$R_1(\hat{\theta}) = \{ \underline{\theta} : (\hat{\theta} - \theta)^T I_n(\hat{\theta})(\hat{\theta} - \theta) \leq \chi_{p,1-\alpha}^2 \}$$

$$R_2(\hat{\theta}) = \{ \theta : w(\theta) \leq \chi_{p,1-\alpha}^2 \}$$

regions in Θ , depend on $\hat{\theta}$

- model $X \sim f(x; \psi, \lambda)$, $\psi \in \mathbb{R}^d$, $\lambda \in \mathbb{R}^{p-d}$, $\theta = (\psi, \lambda)$ $d=1$ usually

$$x = (x_1, \dots, x_n)$$

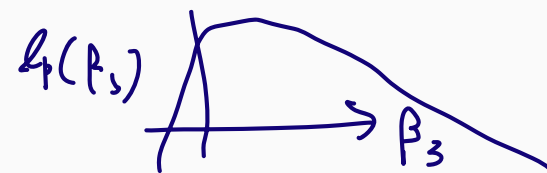
- profile log-likelihood function $\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$ $\psi \in \mathbb{R}^d$
- approximate pivotal quantities

maximize over λ
for each ψ
MS Thm 7.4,5

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

$$(\hat{\psi} - \psi)^T j_p(\hat{\psi})(\hat{\psi} - \psi) \leftarrow 2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\}$$

- approximate confidence regions



$$\left. \begin{aligned} \{\psi : (\hat{\psi} - \psi)^T j_p(\hat{\psi})(\hat{\psi} - \psi) \leq \chi_{d,1-\alpha}^2\} \\ \{\psi : 2\{\ell_p(\hat{\psi}) - \ell_p(\psi)\} \leq \chi_{d,1-\alpha}^2\} \end{aligned} \right\}$$

- Textbook version

$$\Lambda = \frac{\sup_{\theta \in \Theta} f(\mathbf{X}; \theta)}{\sup_{\theta \in \Theta_0} f(\mathbf{X}; \theta)} = \frac{\overset{\text{full max}}{\sup_{\theta \in \Theta} L(\theta)}}{\underbrace{\sup_{\theta \in \Theta_0} L(\theta)}_{\text{constrained}}}$$

$H_0: \theta = \underline{\theta}_0$

- Theorem 7.5

$$\underbrace{2 \log \Lambda_n}_{H_0} \xrightarrow{d} V \sim \chi_d^2$$

- constraint $\underline{\theta} \in \Theta_0$ has d restrictions on $\underline{\theta} = (\theta_1, \dots, \theta_p)$

- version using profile fixes $\psi = \psi(\theta_1, \dots, \theta_d), d < p$

$$\psi = \theta_1$$

$$\theta = (\psi, \lambda)$$

$$X_1, \dots, X_n \sim f(\mathbf{x}; \theta)$$

alt.

AoS range $\mathbf{x} \in \mathcal{X}$

- Null and alternative hypothesis H_0 vs H_1 , $= \{ \mathbf{x} \in \mathcal{R} \}$

- Test function how to decide between them? $\phi(\mathbf{x}) = \{0, 1\}$

- Rejection region if $\mathbf{x} \in \mathcal{R}$, then H_1
 $\mathbf{x} \in \mathcal{R}$ H_0 $\phi(\mathbf{x}) = 1$ we decide H_1
 $= 0$ " " H_0

- Type I and Type II error

$\phi(\mathbf{x}) = 1$ but H_0 is true

- Test statistic and critical value

ϕ or $\mathcal{R}$ to be found

$$P_{H_0} \{ \mathbf{x} \in \mathcal{R} \}$$

prob. type 1 error

$$= E_{H_0} \phi(X) = P_{H_0} \{ \phi(X) = 1 \}$$

$$P_{H_1} \{ \mathbf{x} \notin \mathcal{R} \}$$

prob type 2 error

$$= E_{H_1} \{ 1 - \phi(X) \} = P_{H_1} \{ \phi(X) = 0 \}$$

$$X_1, \dots, X_n \sim f(\mathbf{x}; \theta)$$

AoS range $\mathbf{x} \in \mathcal{X}$

- **Null** and **alternative** hypothesis: $H_0 : \theta \in \Theta_0$; $H_1 : \theta \in \Theta_1$, $\Theta_0 \cup \Theta_1 = \Theta$

- Test (decision) function: $\phi : \mathcal{X} \rightarrow \{0, 1\}$

MS 7.3

$\phi(\mathbf{X}) = 1$ decide $\theta \in \Theta_1$, else decide $\theta \in \Theta_0$

- Rejection region: $R \subset \mathcal{X}$; if $\mathbf{x} \in R$ “reject” H_0

AoS 10.0

$$R = \{\mathbf{x} : \phi(\mathbf{x}) = 1\}$$

- **Type I** and **Type II** error: $\Pr\{\mathbf{X} \in R\}, \theta \in \Theta_0$, $\Pr\{\mathbf{X} \notin R\}, \theta \in \Theta_1$

- Test statistic and critical value: $R = \{\mathbf{x} \in \mathcal{X} : t(\mathbf{x}) > c\}$

c to be chosen

$\uparrow$
evidence against H_0 if $\mathbf{x} \in R$

• goal is to identify R , or $\phi(\cdot)$ with small Type I and Type II errors

• can't reduce both errors at once

see text following Ex. 7.10

• classical solution: require

$$\underline{E_{\theta}\{\phi(\mathbf{X})\}} \leq \alpha, \quad \theta \in \Theta_0 \quad (1)$$

• subject to this constraint, minimize

$$\underline{E_{\theta}\{\phi(\mathbf{X})\}}, \quad \theta \in \Theta_1 \quad (2)$$

• find a **test function** $\phi(\mathbf{X})$ to satisfy (1) and (2)

• find a **test statistic**, $T = t(\mathbf{X})$, and define $\phi(\mathbf{x}) = 1\{t(\mathbf{x}) \geq c\}$

size α

$$P_{\pi}\{\text{type I}\} = P_{\pi_{H_0}}\{\text{reject } H_0\}$$

$$\min_{H_1} P_{\pi}\{\phi(\mathbf{X}) = 0\}$$

$$\max_{H_1} P_{\pi}\{\phi(\mathbf{x}) = 1\}$$

c to be chosen

power = $1 - \text{prob}_{\pi_2}$

1.2 Hypothesis Testing

Our second example concerns the march of methodology and inference for *hypothesis testing* rather than estimation: 72 leukemia patients, 47 with **ALL** (acute lymphoblastic leukemia) and 25 with **AML** (acute myeloid leukemia, a worse prognosis) have each had genetic activity measured for a panel of 7,128 genes. The histograms in Figure 1.4 compare the genetic activities in the two groups for gene 136.

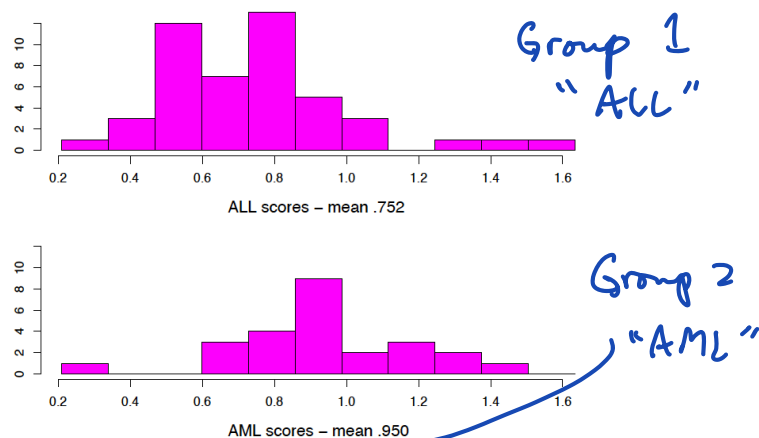
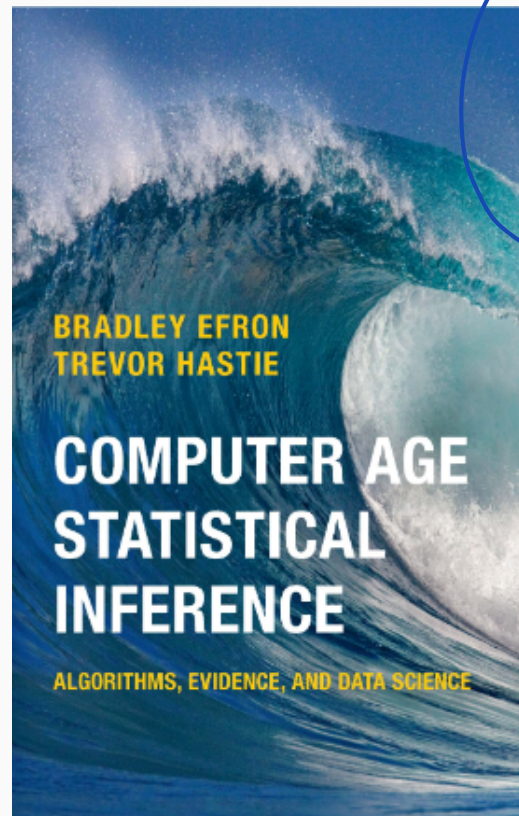


Figure 1.4 Scores for gene 136, leukemia data. Top **ALL** ($n = 47$) bottom **AML** ($n = 25$). A two-sample t -statistic = 3.01 with p -value = .0036.

The **AML** group appears to show greater activity, the mean values being

$$\bar{\text{ALL}} = 0.752 \text{ and } \bar{\text{AML}} = 0.950. \quad (1.5)$$



$$T_1 = \text{Wald}$$

$$(\hat{\theta} - \theta)^T \dots ()$$

$$T_2 = \text{LRT}$$

$$= 2\{\ell_p(\hat{\gamma}) - \ell_p(\gamma)\}$$

```
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+n2-1)  
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] -3.014003
```

```

library("tidyverse")
leukemia_big<- read.csv
  ("http://web.stanford.edu/~hastie/CASI_files/DATA/leukemia_big.csv")
leukemia_big[136,] %>% select(starts_with("ALL")) %>% as.numeric() -> all136 *
leukemia_big[136,] %>% select(starts_with("AML")) %>% as.numeric() -> aml136 y
t.test(all136,aml136, var.equal = TRUE)
##
Two Sample t-test

```

since $p < .05$

data: all136 and aml136

$t = -3.014$, $df = 70$, $p\text{-value} = 0.003589$

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.32817995 -0.06680742

sample estimates:

we reject $H_0: \mu_{ALL} = \mu_{AML}$
 $\mu_1 = \mu_2$

$$X_{i,ALL} \sim N(\mu_1, \sigma^2)$$

$$X_{i,AML} \sim N(\mu_2, \sigma^2)$$

0 not in CI

- model
- null and alternative hypothesis
- rejection region
- test statistics and critical value
- type I and type II error

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

$$t(\underline{x}) \sim \underline{T}_{n-2} \text{ exact}$$

$$= \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

S_p = pooled std. dev.

$$R = \{ \underline{x} : |t(\underline{x})| \geq c \}$$

$$\phi(\underline{x}) = 1 \quad \text{if} \quad |t(\underline{x})| \geq c$$

$$c = .025 \text{ critical value of } T_{70} \text{ d.v. } t$$

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, ...)
```

$$T_2 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}}$$

$$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{m})$$

$$s_x^2 = \frac{1}{m-1} \sum (x_i - \bar{x})^2$$

$$\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n})$$

`t.test(all136, aml136, var.eq = T)`

```
> 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
```

`t.test(all136, aml136)`

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

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

moment-matching
argument

0.0049
.000
.05
7000
↑

model, null, alternative, rejection reg, test stat, ...

10.7 Example (Comparing Two Prediction Algorithms). We test a prediction algorithm on a test set of size m and we test a second prediction algorithm on a second test set of size n . Let X be the number of incorrect predictions for algorithm 1 and let Y be the number of incorrect predictions for algorithm 2. Then $X \sim \text{Binomial}(m, p_1)$ and $Y \sim \text{Binomial}(n, p_2)$. To test the null hypothesis that $p_1 = p_2$ write

$$H_0 : \delta = 0 \quad \text{versus} \quad H_1 : \delta \neq 0$$

$$\hat{p}_1 = \frac{X}{m} \quad \hat{p}_2 = \frac{Y}{n}$$

where $\delta = p_1 - p_2$. The MLE is $\hat{\delta} = \hat{p}_1 - \hat{p}_2$ with estimated standard error

$$\widehat{\text{se}} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}$$

The size α Wald test is to reject H_0 when $|W| > z_{\alpha/2}$ where

$$W = \frac{\hat{\delta} - 0}{\widehat{\text{se}}} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}}$$

The power of this test will be largest when p_1 is far from p_2 and when the sample sizes are large.

$$t = \hat{p}_1 - \hat{p}_2$$

$$H_0: p_1 = p_2$$

- $X \sim \text{Binom}(m, p_1), Y \sim \text{Binom}(n, p_2)$
- $\delta = p_1 - p_2; H_0 : \delta = 0$
- maximum likelihood estimate of δ
- estimated standard error

two prediction algorithms

- composite null hypothesis $H_0 : \theta_1 = \theta_{10}, \dots, \theta_r = \theta_{r0}$

$$r < p$$

- definition $\Lambda_n =$

AoS Def 10.21 λ

$$\underline{X} \sim f(\underline{x}; \underline{\theta}) \quad \underline{\theta} \in \mathbb{R}^p$$

- composite null hypothesis $H_0 : \theta_1 = \theta_{10}, \dots, \theta_r = \theta_{r0}$

$$r < p$$

- definition $\Lambda_n = \frac{\sup_{\theta \in \Theta_0} L(\theta; \underline{x})}{\sup_{\theta \in \Theta} L(\theta; \underline{x})}$

AoS Def 10.21 λ

- Theorem MS 7.5, AoS 10.22

$$\underline{x} = (x_1, \dots, x_n)$$

$$2 \log \Lambda_n \xrightarrow{d} \chi_d^2 \quad ; \quad d = \begin{array}{l} \# \text{ of restrictions} \\ \text{in } H_0 \end{array}$$

approximate test
statistic~~loss~~
C Reg

$$(\hat{\theta} - \theta_0)^T I_n(\hat{\theta})(\hat{\theta} - \theta_0) \\ 2 \{ \ell(\hat{\theta}_0) - \ell(\theta_0) \}$$

or

$$(\hat{\psi} - \psi_0)^T J_p(\hat{\psi})(\hat{\psi} - \psi_0) \\ 2 \{ \ell_p(\hat{\psi}) - \ell_p(\psi_0) \}$$

$$\xrightarrow{d} \chi_d^2$$

under H_0 :

- Neyman-Pearson lemma gives the “best” test statistic
- among all tests with type I error $\leq \alpha$
- find the test with smallest type II error $\equiv$ largest power

$$\phi(\underline{x}) \rightarrow \{0, 1\} \quad \text{s.t.} \quad E_{H_0}\{\phi(\underline{x})\} \leq \alpha$$
$$E_{H_1}\{\phi(\underline{x})\} \text{ is maxim'd}$$

$$H_0: \underline{X} \sim f_0(\underline{x}) \quad H_1: \underline{X} \sim f_1(\underline{x}) \quad \text{simple vs. simple}$$

- Neyman-Pearson lemma gives the “best” test statistic
- among all tests with type I error $\leq \alpha$
- find the test with smallest type II error $\equiv$ largest power

$H_0: \underline{x} \sim f_0(\cdot) \quad H_1: \underline{x} \sim f_1(\cdot)$

? $f(x; \theta) \quad f(x; \theta)$
 ? $\theta \in \Theta_0 \quad \theta \in \Theta_1$

- fix $\text{pr}(\text{reject } H_0 \mid H_0) \leq \alpha$,

maximize $\text{pr}(\text{reject } H_0 \mid H_1)$

$\phi(\underline{x}) = \begin{cases} 1 \\ 0 \end{cases}$

if $\frac{f_1(\underline{x})}{f_0(\underline{x})} \geq k$
 $\frac{f_1(\underline{x})}{f_0(\underline{x})} < k$

solⁿ

- Neyman-Pearson lemma gives the “best” test statistic
- among all tests with type I error $\leq \alpha$
- find the test with smallest type II error $\equiv$ largest power
- fix $\text{pr}(\text{reject } H_0 \mid H_0) \leq \alpha$, maximize $\text{pr}(\text{reject } H_0 \mid H_1)$
- Neyman-Pearson Lemma

MS Thm 7.2

Suppose $\mathbf{X} = (X_1, \dots, X_n) \sim f(\mathbf{x})$. Under H_0 , $f(\mathbf{X}) = f_0(\mathbf{x})$, and under H_1 , $f(\mathbf{X}) = f_1(\mathbf{x})$.

The test with test function

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } f_1(\mathbf{x}) > k f_0(\mathbf{x}), \\ 0 & \text{otherwise} \end{cases}$$

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

$f_0(\cdot), f_1(\cdot)$
cont^s

(for some $0 < k < \infty$) is a most powerful test of H_0 vs H_1 at level

$$\underline{\underline{\alpha = E_0\{\phi(\mathbf{X})\}}}.$$

Proof: Suppose $\psi(\underline{x}) \rightarrow [0, 1]$ $0 \leq \psi(\underline{x}) \leq 1$
 $E_0 \psi(\underline{x}) \leq \alpha = E_0 \phi(\underline{x})$

$$\int \psi(x) \{f_1(x) - k f_0(x)\} dx \leq \int \phi(x) \{f_1(x) - k f_0(x)\} dx \quad E_k \phi = \alpha$$

$$\frac{E_{f_1} \psi(X) - E_{f_0} \psi(X)}{\phi(x) = 1} \leq k \{E_{f_1} \psi(X) - E_{f_0} \psi(X)\} \leq 0 \quad E_{f_0} \psi \leq \alpha$$

$$\phi(x) = 1 \Rightarrow f_1(x) - k f_0(x) \geq 0$$

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

1) $\phi(x) = 1$ true

1) $\psi(x) \{ \dots \} dx \leq \dots \} \checkmark \psi(x) \leq 1$

2) $\phi(x) = 0$ $\psi(x) \{ -ve \} \leq 0 \checkmark \psi(x) \geq 0$

- The formal theory of testing imagines a decision to “reject H_0 ” or not, according as $X \in R$ or $X \notin R$, for some defined region $R \subset \mathcal{X}$ e.g. $|Z| > 1.96$
- This is useful for deriving the form of optimal tests, but not useful in practice.
- Doesn't distinguish between $Z = 1.97$ and $Z = 19.7$, for example.
- P -values give more precise information about the null hypothesis

- The formal theory of testing imagines a decision to “reject H_0 ” or not, according as $X \in R$ or $X \notin R$, for some defined region $R \subset \mathcal{X}$ e.g. $|Z| > 1.96$
- This is useful for deriving the form of optimal tests, but not useful in practice.
- Doesn't distinguish between $Z = 1.97$ and $Z = 19.7$, for example.
- P-values give more precise information about the null hypothesis

- MS definition: $p(\mathbf{x}) = \inf\{\alpha : \phi_\alpha(\mathbf{x}) = 1\}$
- AoS definition: p-value $= \inf\{\alpha : T(X_n) \in R_\alpha\}$
- SM definition $p_{obs} = \Pr_{H_0}\{T(X_n) \geq t_{obs}\}$

$P = 0.001$

7.5

$\alpha \downarrow .001$
still reject H_0 .
Def 10.11

Example: two-sample t -test

$$T(\underline{X}_n, \underline{Y}_n) = (\bar{X}_n - \bar{Y}_n) / \dots \quad \text{MS Ex.7.24}$$

$X_1, \dots, X_m$ i.i.d. $N(\mu_1, \sigma^2)$, $Y_1, \dots, Y_n$ i.i.d. $N(\mu_2, \sigma^2)$

$$H_0 : \mu_1 = \mu_2$$

LRT, Wald, score, exact

$$p(t) = \text{pr}_{H_0}(|T| > t)$$

Example: Poisson

$X_1, \dots, X_n$ i.i.d. $Po(\lambda)$

$H_0: \lambda = \lambda_0$

$H_1: \lambda = \lambda_1 > \lambda_0$ $f_1(\underline{x}) = \lambda_1^{\sum x_i} e^{-n\lambda_1}$ $f_0(\underline{x}) = \lambda_0^{\sum x_i} e^{-n\lambda_0}$

Simple vs simple

$$t(\underline{x}) = \frac{f_1(\underline{x})}{f_0(\underline{x})} \geq k \iff \left(\frac{\lambda_1}{\lambda_0}\right)^s e^{-n(\lambda_1 - \lambda_0)} \geq k$$

$$s = \sum x_i$$

$$\underbrace{s(\log \lambda_1 - \log \lambda_0)}_{>0} - \underbrace{n(\lambda_1 - \lambda_0)}_{>0} \geq k$$

$$\iff s \geq c$$

$$S \sim_{\lambda_0} Po(n\lambda_0)$$

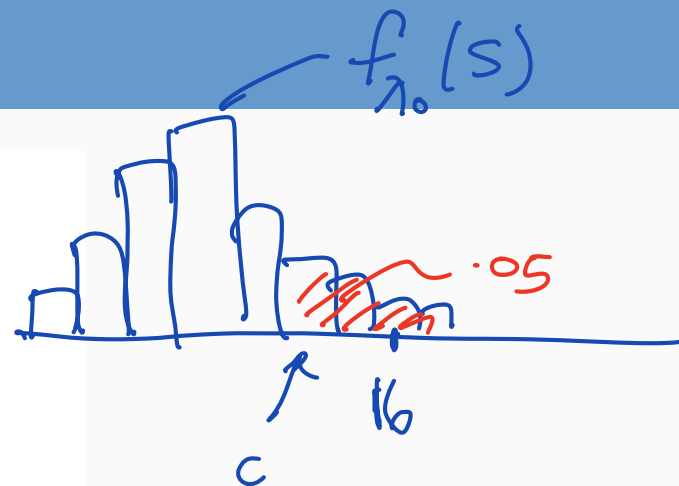
$$P_{\lambda_0} \{S \geq c_\alpha\} \leq \alpha \quad \text{determines } c$$

Example: logistic regression

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



- $X_1, \dots, X_n$ i.i.d. $F(\cdot)$
- $H_0 : \mu = \mu_0, \mu = F^{-1}(1/2)$ median of distribution
- $H_1 : \mu > \mu_0$
- test statistic

both H composite

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

- under H_0 ,

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

- p -value

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

- $H_0 : F^{-1}(1/2) = \mu_0 \quad H_1 : F^{-1}(1/2) > \mu_0$
- Test statistic $T = \sum_{i=1}^n 1\{X_i > \mu_0\}$
- $\text{pr}_{H_0}(\text{reject } H_0) = \text{pr}(T \geq c_\alpha \mid H_0) = \alpha \Rightarrow c_\alpha \approx n/2 - n^{1/2}z_\alpha/2$
- $\text{pr}_{H_1}(\text{reject } H_0) = \text{pr}(T \geq c_\alpha \mid H_1)$ Need distribution of T under H_1
- to calculate power we need values for μ and for F

- $H_0 : F^{-1}(1/2) = \mu_0 \quad H_1 : F^{-1}(1/2) > \mu_0$
- Test statistic $T = \sum_{i=1}^n 1\{X_i > \mu_0\}$
- $\text{pr}_{H_0}(\text{reject } H_0) = \text{pr}(T \geq c_\alpha \mid H_0) = \alpha \Rightarrow c_\alpha \approx n/2 - n^{1/2}z_\alpha/2$
- $\text{pr}_{H_1}(\text{reject } H_0) = \text{pr}(T \geq c_\alpha \mid H_1)$ Need distribution of T under H_1
- to calculate power we need values for μ and for F
- e.g. change to $H_1 : F^{-1}(1/2) = \mu_1$ $\text{pr}_{F_{\mu_1}}(X > \mu_0)$
- SM assumes F is $N(\mu, \sigma^2)$, so $\delta = n^{1/2}(\mu_1 - \mu_0)/\sigma$

$$\begin{aligned} \text{pr}_{\mu_1}(T \geq c_\alpha) &= \text{pr}_{\mu_1}(T \geq n/2 - n^{1/2}z_\alpha/2) \doteq \Phi \left\{ \frac{n\Phi(n^{-1/2}\delta) - n/2 + n^{1/2}z_\alpha}{[n\Phi(n^{-1/2}\delta)\{1 - \Phi(n^{-1/2})\}]} \right\} \\ &\doteq \Phi\{z_\alpha + \delta(2/\pi)^{1/2}\} \end{aligned}$$

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

334

7 · Estimation and Hypothesis Testing

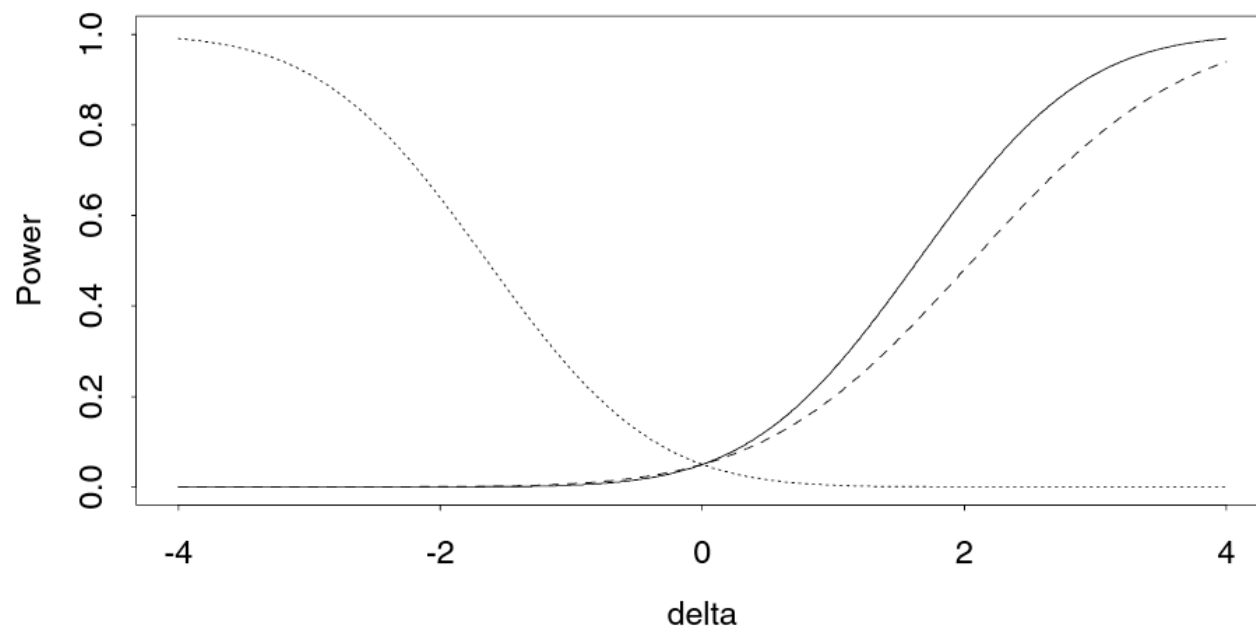


Figure 7.6 Power functions for a test of whether the mean of a $N(\mu, \sigma^2)$ random sample of size n equals μ_0 against the alternative $\mu = \mu_1$, as a function of $\delta = n^{1/2}(\mu_1 - \mu_0)/\sigma$. The test size is $\alpha = 0.05$. The solid curve is the power function for a test of $\mu_1 > \mu_0$ 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 $\bar{y} < t_\alpha$.

$$X_1, \dots, X_k \sim \text{Mult}(n; p_1, \dots, p_k)$$

leukemia data (EH): $X_1, \dots, X_{47}; Y_1, \dots, Y_{25}$

AoS Ex. 10.20

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					

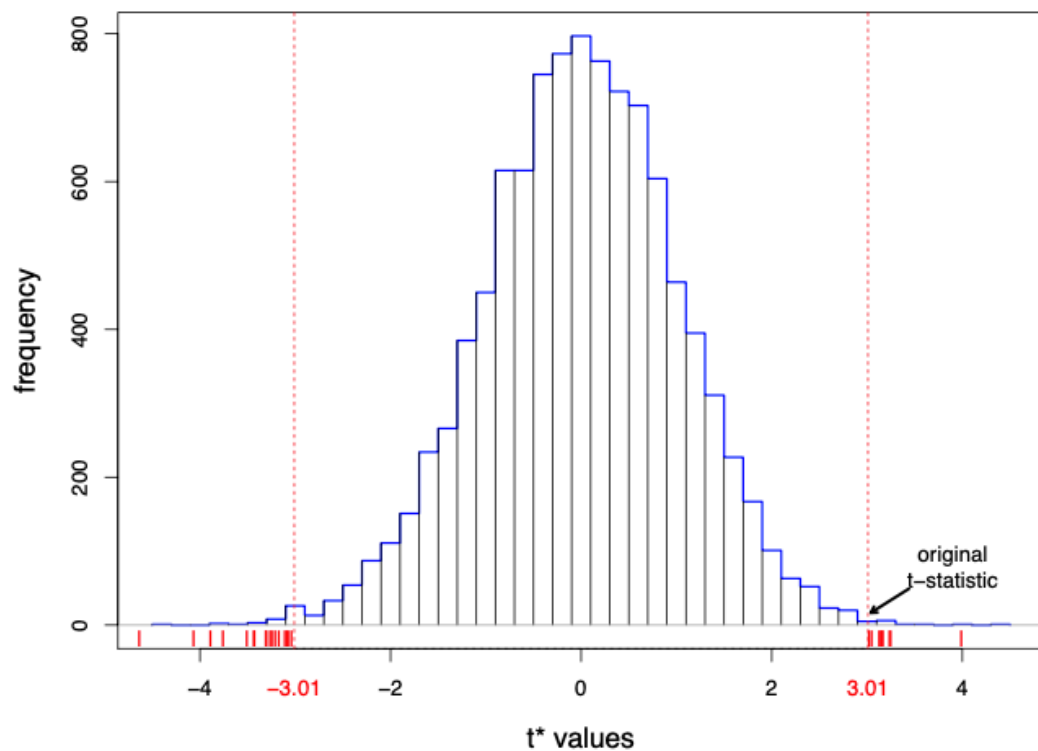


Figure 4.3 10,000 permutation t^* -values for testing **ALL** vs **AML**, for gene 136 in the **leukemia** data of Figure 1.3. Of these, 26 t^* -values (red ticks) exceeded in absolute value the observed t -statistic 3.01, giving permutation significance level 0.0026.

Choosing test statistics

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

Hypothesis tests and significance tests

- **Hypothesis tests** typically means:

- H_0, H_1
- critical/rejection region $R \subset \mathcal{X}$,
- level α , 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 and significance tests

- **Hypothesis tests** typically means:

- H_0, H_1
- critical/rejection region $R \subset \mathcal{X}$,
- level α , 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

- **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 α -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 α -level
at which the corresponding hypothesis test would reject H_0

Definition 10.11 in AoS