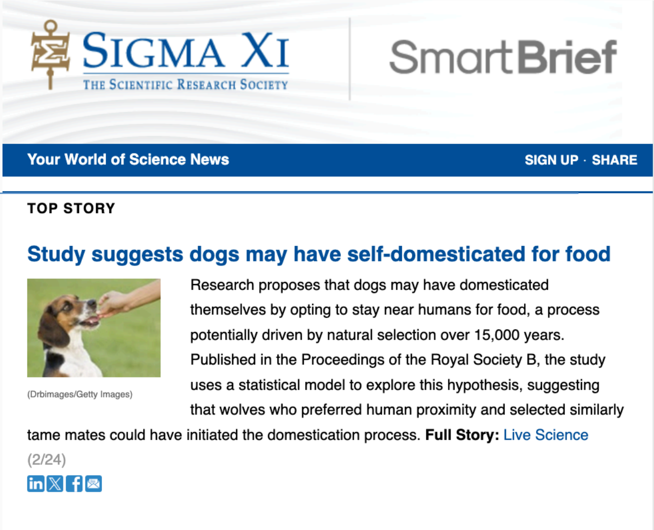
Mathematical Statistics II

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)

PROCEEDINGS
OF THE ROYAL SOCIETY B

BIOLOGICAL SCIENCES

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

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

MS 7.1,2; AoS 6.3.2

- $p = 1$: $(1 - \alpha)$ confidence interval, confidence bound

- $p > 1$: $(1 - \alpha)$ confidence region

  see next slides

- $p = 1$: $(1 - \alpha)$ credible interval, credible bound

  prior $\pi(\theta)$

- $p > 1$: $(1 - \alpha)$ 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

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

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

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

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

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

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

- approximate normality of posterior

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

Recap: Confidence and credible regions

- Bayesian: HPD region C for $\boldsymbol{\theta}$:

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

$$\begin{aligned} &(1) \qquad \int_C \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = 1 - \alpha \\ &(2) \quad \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \geq \pi(\boldsymbol{\theta}^* \mid \boldsymbol{x}), \quad \text{any } \theta \in C \text{ and } \theta^* \notin C \end{aligned}$$

- approximate pivotal quantities

$$(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T I_n(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$

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

- approximate confidence regions

$$R_1(\widehat{\boldsymbol{\theta}}) = \{\boldsymbol{\theta} : (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T I_n(\widehat{\boldsymbol{\theta}})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \chi^2_{p,1-\alpha}\},$$

$$R_2(\widehat{\boldsymbol{\theta}}) = \{\boldsymbol{\theta} : w(\boldsymbol{\theta}) \leq \chi^2_{p,1-\alpha}\}$$

regions in Θ, depend on $\widehat{\boldsymbol{\theta}}$

Confidence intervals for sub-parameters

MS 7.4

- model $Y \sim f(y;\psi,\lambda), \quad \psi \in \mathbb{R}^d, \lambda \in \mathbb{R}^{p-d}, \quad \theta = (\psi,\lambda)$ $y = (y_1,\dots,y_n)$
- profile log-likelihood function $\ell_p(\psi) = \ell(\psi,\hat\lambda_\psi)$ maximize over λ
- approximate pivotal quantities MS Thm 7.4,5

$$(\widehat{\psi}-\psi)^{\mathrm{T}} j_{\mathrm{p}}(\widehat{\psi})(\widehat{\psi}-\psi)$$

$$2\{\ell_p(\hat\psi)-\ell_p(\psi)\}$$

- approximate confidence regions

$$\{\psi : (\widehat{\psi}-\psi)^{\mathrm{T}} j_{\mathrm{p}}(\widehat{\psi})(\widehat{\psi}-\psi) \le \chi^2_{d,1-\alpha}\}$$

$$\{\psi : 2\{\ell_p(\hat\psi)-\ell_p(\psi)\} \le \chi^2_{d,1-\alpha}\}$$

Aside: Likelihood ratio test

- Textbook version

$$\Lambda = \frac{\sup_{\theta \in \Theta} f(\boldsymbol{X};\theta)}{\sup_{\theta \in \Theta_0} f(\boldsymbol{X};\theta)} = \frac{\sup_{\theta \in \Theta} L(\theta)}{\sup_{\theta \in \Theta_0} L(\theta)}$$

- Theorem 7.5

$$2 \log \Lambda_n \xrightarrow{d} V \sim \chi^2_d$$

- constraint $\theta \in \Theta_0$ has d restrictions on $\theta = (\theta_1, \ldots, \theta_p)$

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

Introduction to Formal Hypothesis testing

MS 7.3, AoS Ch. 10

$X_1, \ldots, X_n \sim f(\boldsymbol{x}; \theta)$

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

- Null and alternative hypothesis

- Test function
- Rejection region

- Type I and Type II error

- Test statistic and critical value

… Hypothesis testing: intro

MS 7.3, AoS AoS Ch. 10

$X_1, \dots, X_n \sim f(\boldsymbol{x}; \theta)$ AoS range $\boldsymbol{x} \in \mathcal{X}$

- 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: $\phi : \mathcal{X} \to \{0, 1\}$ MS 7.3

  $\phi(\boldsymbol{X}) = 1$ decide $\theta \in \Theta_1$, else decide $\theta \in \Theta_0$
- Rejection region: $R \subset \mathcal{X}$; if $\boldsymbol{x} \in R$ "reject" H_0 AoS 10.0

  $R = \{\boldsymbol{x} : \phi(\boldsymbol{x}) = 1\}$
- Type I and Type II error: $\Pr\{\boldsymbol{X} \in R\}$, $\theta \in \Theta_0$, $\Pr\{\boldsymbol{X} \notin R\}$, $\theta \in \Theta_1$
- Test statistic and critical value: $R = \{\boldsymbol{x} \in \mathcal{X} : t(\boldsymbol{x}) > c\}$ c to be chosen

… Hypothesis testing: intro

MS 7.3; AoS Ch. 10

- 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

size α

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

- subject to this constraint, minimize

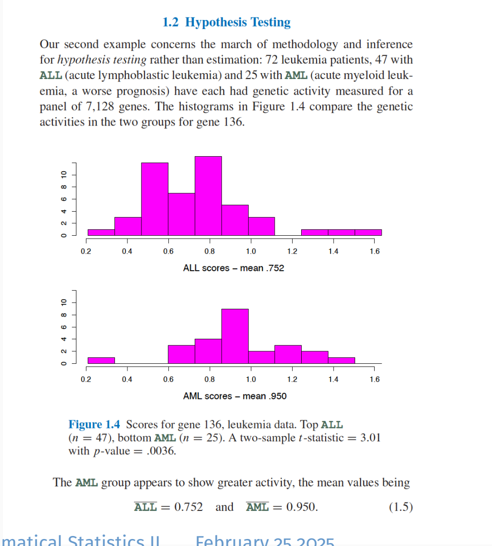
$$E_\theta\{\phi(\boldsymbol{X})\}, \quad \theta \in \Theta_1 \quad (2)$$

- find a test function $\phi(\boldsymbol{X})$ to satisfy (1) and (2)
- find a test statistic, $T = t(\boldsymbol{X})$, and define $\phi(\boldsymbol{x}) = 1\{t(\boldsymbol{x}) \geq c\}$

c to be chosen

Example: Two-sample t-test

EH §1.2



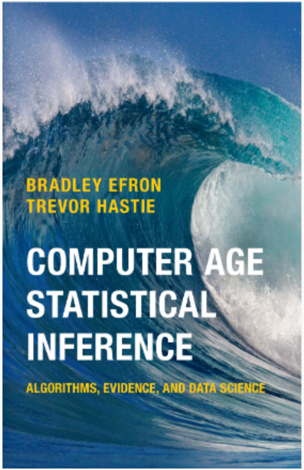
1.2 Hypothesis Testing

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

$$\overline{\mathbf{ALL}} = 0.752 \quad \text{and} \quad \overline{\mathbf{AML}} = 0.950. \tag{1.5}$$



… *t*-test

```
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
```

… *t*-test

```
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
t.test(all136,aml136, var.equal = TRUE)
##
Two Sample t-test

data:  all136 and aml136
t = -3.014, df = 70, p-value = 0.003589
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.32817995 -0.06680742
sample estimates:
```

… t-test

AoS Ex.10.8

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

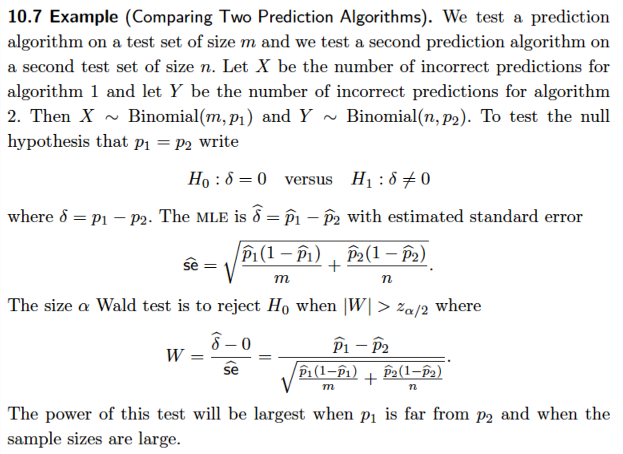
A word on the *t*-test

```
## 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
```

Example: comparing two proportions

AoS Ex.10.7

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

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

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

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

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

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

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

Example: comparing two proportions

AoS Ex.10.7

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

two prediction algorithms

Likelihood ratio statistic

MS 7.4; AoS 10.6

- composite null hypothesis $H_0 : \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}$ $r < p$
- definition $\Lambda_n =$ AoS Def 10.21 λ

Likelihood ratio statistic

MS 7.4; AoS 10.6

- composite null hypothesis $H_0 : \theta_1 = \theta_{10}, \ldots, \theta_r = \theta_{r0}$ $r < p$
- definition $\Lambda_n =$ AoS Def 10.21 λ
- Theorem MS 7.5, AoS 10.22

Neyman-Pearson Lemma

MS 7.3; AoS 10.10.1

- 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

Neyman-Pearson Lemma

MS 7.3; AoS 10.10.1

- 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 $\mathrm{pr}(\text{reject } H_0 \mid H_0) \leq \alpha$, maximize $\mathrm{pr}(\text{reject } H_0 \mid H_1)$

Neyman-Pearson Lemma

- 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 $\mathrm{pr}(\text{reject } H_0 \mid H_0) \leq \alpha$, maximize $\mathrm{pr}(\text{reject } H_0 \mid H_1)$

- Neyman-Pearson Lemma MS Thm 7.2

Neyman-Pearson Lemma

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

The test with test function

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

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

$$\alpha = \mathrm{E}_0\{\phi(\boldsymbol{X})\}.$$

Neyman-Pearson Lemma

MS Thm 7.2; AoS 10.10.1

p-values

- 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

p-values

MS 7.5; AoS 10.2; SM 7.3.1

- 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: $p(\boldsymbol{x}) = \inf\{\alpha : \phi_\alpha(\boldsymbol{x}) = 1\}$ 7.5
- AoS definition: p-value $= \inf\{\alpha : T(X_n) \in R_\alpha\}$ Def 10.11
- SM definition $p_{obs} = \Pr_{H_0}\{T(X_n) \geq t_{obs}\}$

Example: two-sample t-test

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

$H_0 : \mu_1 = \mu_2$

LRT, Wald, score, exact

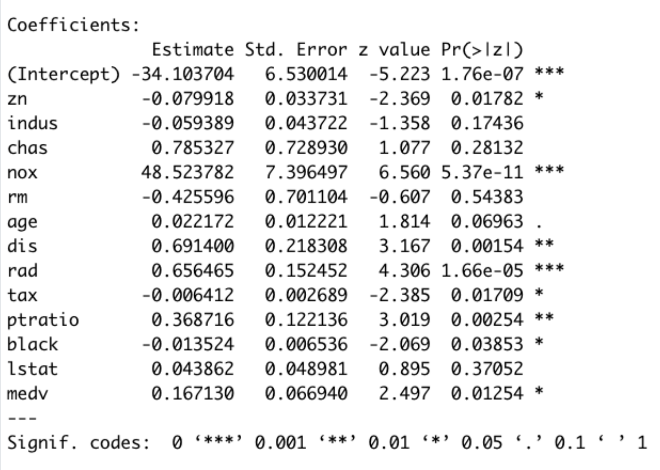
$p(t) = \mathrm{pr}_{H_0}(|T| > t)$

Example: Poisson

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

$H_0 : \lambda = \lambda_0$

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
```

Sign test

SM Ex.7.26

- $X_1, \ldots, X_n$ 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

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

- under H_0,

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

- p-value

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

Power of the sign test

SM Ex.7.30

- $H_0 : F^{-1}(1/2) = \mu_0 \qquad H_1 : F^{-1}(1/2) > \mu_0$
- 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 μ and for F

Power of the sign test

SM Ex.7.30

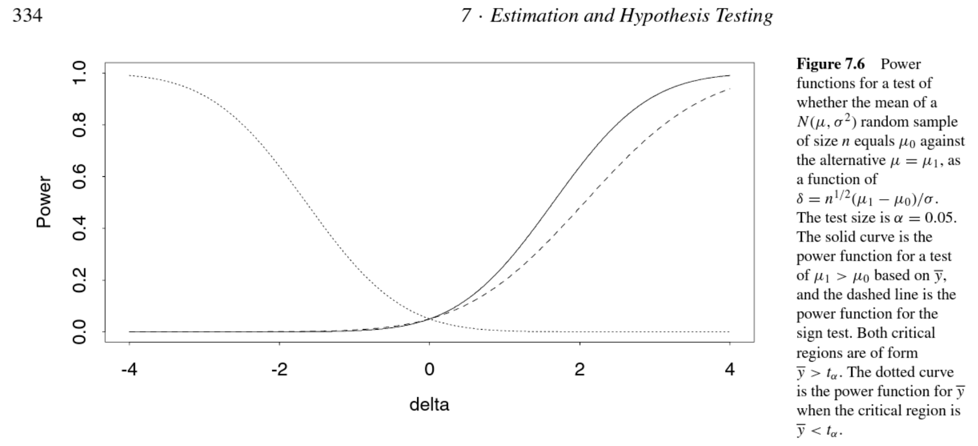
- $H_0 : F^{-1}(1/2) = \mu_0 \qquad H_1 : F^{-1}(1/2) > \mu_0$
- Test statistic $T = \sum_{i=1}^{n} 1\{X_i > \mu_0\}$
- $\mathrm{pr}_{H_0}(\text{reject } H_0) = \mathrm{pr}(T \geq c_\alpha \mid H_0) = \alpha \Rightarrow c_\alpha \approx n/2 - n^{1/2} z_\alpha/2$
- $\mathrm{pr}_{H_1}(\text{reject } H_0)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

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

$$\doteq \Phi\{z_\alpha + \delta(2/\pi)^{1/2}\}$$

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

… power of sign test



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 $\overline{y}$, and the dashed line is the power function for the sign test. Both critical regions are of form $\overline{y} > t_\alpha$. The dotted curve is the power function for $\overline{y}$ when the critical region is $\overline{y} < t_\alpha$.

Other test statistics

$X_1, \ldots, X_k \sim Mult(n; p_1, \ldots, p_k)$

Permutation test

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
```

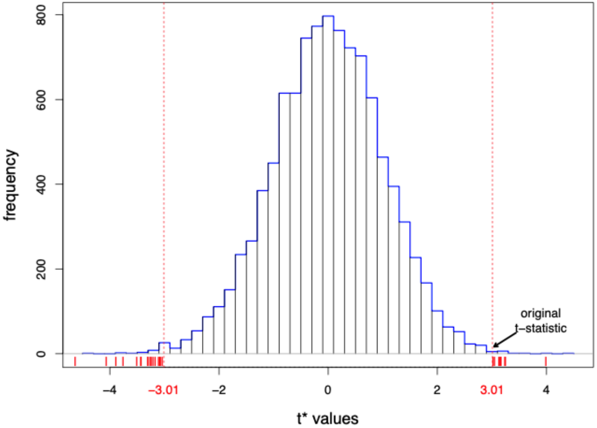


Figure 4.3 10,000 permutation t^*-values for testing **`ALL`** vs **`AML`**, for gene 136 in the **`leukemia`** data of Figure 1.3. Of these, 26 t^*-values (red ticks) exceeded in absolute value the observed t-statistic 3.01, giving permutation significance level 0.0026.

AoS 10.20 uses median

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