Statistical Theory for Data Science

STA2212H S LEC9101

Week 5

February 3 2026

Today

1. Recap: Formal theory of testing
2. Some examples
3. Goodness-of-fit tests
4. Diagnostic testing

- null and alternative hypotheses: H_0 and H_1
- type 1 and type 2 error: probability of a wrong conclusion

  reject H_0 when true; do not reject H_0 when false
- rejection region, critical region, define a subset of the sample space
- either through an indicator (test) function or more usually as $\{\boldsymbol{x} : t(\boldsymbol{x}) > c\}$

  $t(\cdot)$, c to be determined
- for parametric model, with $H_0 : \theta \in \Theta_0, H_1 : \theta \in \Theta_1$ power function $\beta(\theta) = \mathrm{pr}_\theta(\boldsymbol{X} \in R)$
- size of the test $\alpha = \sup_{\Theta_0} \beta(\theta)$
- p-value = $\inf\{\alpha : T(\boldsymbol{X}) \in R\} = \sup_{\Theta_0} \mathrm{pr}_\theta\{T(\boldsymbol{X}) \geq T(\boldsymbol{x}^0)\}$ obs'd 'just reject'
- simple H_0 specifies distribution of $T(\boldsymbol{X})$; composite H_0 does not

dist^n $T(\underline{X})$ is known

$\to H_0: \theta_1 = \theta_{10}$, θ_2 unspecified

$H_0: \theta \leq \theta_0$

Recap: Neyman-Pearson lemma

- for testing simple H_0 against simple H_1
- test statistic

$$T = \frac{L(\theta_1; \boldsymbol{x})}{L(\theta_0; \boldsymbol{x})} = \frac{f(\boldsymbol{x}; \theta_1)}{f(\boldsymbol{x}; \theta_0)}$$

$\theta \in \mathbb{R}^k$

- critical region

$$\{\boldsymbol{x} : t(\boldsymbol{x}) \geq k\}$$

- Choose $k = k_\alpha$ to satisfy

$$\mathrm{pr}_{H_0}(T \geq k_\alpha) = \alpha$$

- This test is a most powerful test of H_0 against H_1 at level α

A neatly-typed proof (from SM 7.3)

Let R be the rejection region for the test based on $T = f_1(\boldsymbol{x})/f_0(\boldsymbol{x})$

$$R = \{\boldsymbol{x} : T(\boldsymbol{x}) \geq k_\alpha\}$$

$\underline{\underline{\phi(\underset{\sim}{x})}} = 1\{\underset{\sim}{x} \in R\}$

Let R' be some other rejection region also of size α $\leq \alpha$

$$\alpha = \int_R f_0(\boldsymbol{x})d\boldsymbol{x} = \int_{R'} f_0(\boldsymbol{x})d\boldsymbol{x}$$

$$\int_{R-R'} f_0(\boldsymbol{x})d\boldsymbol{x} = \int_{R'-R} f_0(\boldsymbol{x})d\boldsymbol{x}$$

On LHS $f_1(\boldsymbol{x}) \geq k_\alpha f_0(\boldsymbol{x})$. $R - R' \subset R$

On RHS $f_1(\boldsymbol{x}) < k_\alpha f_0(\boldsymbol{x})$. $R' - R \subset R^c$

$$\int_{R-R'} f_1(\boldsymbol{x})d\boldsymbol{x} \geq \int_{R'-R} f_1(\boldsymbol{x})d\boldsymbol{x}$$

Add integral over intersection $R \cap R'$

A neatly-typed proof (from MS)

Let $\phi(\boldsymbol{x})$ be the test function for the test based on T.
Let $\psi(\boldsymbol{x})$ be any other function that maps $\boldsymbol{x}$ to $[0, 1]$.
If

$$\mathrm{E}_{H_0}\{\psi(\boldsymbol{X})\} \leq \mathrm{E}_{H_0}\{\phi(\boldsymbol{X})\} = \alpha$$

then it must follow that

$$\mathrm{E}_{H_1}\{\psi(\boldsymbol{X})\} \leq \mathrm{E}_{H_1}\{\phi(\boldsymbol{X})\}$$

Proof: $\forall$ $\boldsymbol{x}$,

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

Integrate and re-arrange terms to get the result

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, H_1 implicit, not well-specified
  - 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

Choosing test statistics

1. Optimal choice – Neyman-Pearson lemma — Might be UMP
2. Pragmatic choice – pivotal quantity $g(\underset{\sim}{x};\theta)$ with known dist^{n} — exact or approximate
3. Pragmatic choice – nonparametric test statistics

(a) Need to know distribution of test statistic under H_0

(b) Test statistic should be large when H_0 is not true — in probability

(c) Test statistic should have maximum/good power to detect departures from H_0

Example 1: N-P lemma

$X_1, \dots, X_n$ iid expl family

$\theta \in \{\theta_0, \theta_1\}$

- $f(\mathbf{x};\theta) = \prod_{i=1}^n \left[h(x_i)\exp\{\theta s(x_i) - A(\theta)\}\right]$ $\quad s, \theta \in \mathbb{R}$, $\quad H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1 > \theta_0$

$H_1' : \theta > \theta_0$

$\theta_1 < \theta_0$

- MP test of H_0 vs H_1 has critical region

$$R = \{\mathbf{x} : \frac{f(\mathbf{x};\theta_1)}{f(\mathbf{x};\theta_0)} \geq k\}$$

$$\left\{ \underset{\sim}{x} : \frac{\cancel{\{\prod_{i=1}^n h(x_i)\}} e^{\theta_1 \Sigma s(x_i) - nA(\theta_1)}}{\cancel{\{\prod h(x_i)\}} e^{\theta_0 \Sigma s(x_i) - nA(\theta_0)}} \geq k \right.$$

dist^{n} is cont^{s}

$$\iff (\theta_1 - \theta_0)s - n\{A(\theta_1) - A(\theta_0)\} \geq k'$$

- UMP against H_1':

$$\iff (\theta_1 - \theta_0)s \geq k''$$

$\theta_1 > \theta_0$

$$\iff \boxed{s \geq c} - \text{free of } \theta_1$$

$$\boxed{R = \{s > c_\alpha\}}$$

$s < c$

$H_1'' : \theta \neq \theta_0$ no UMP test

$$\int^c f(s;\theta_0)\,ds = 1 - \alpha$$

UMP for H_1'

Example 2: approximate pivotal quantities

$= \int_{-\infty}^{c} e^{\theta_0 s - nA(\theta_0)} \tilde{h}(s)\,ds$

$\theta \in \mathbb{R}$

- Wald test:

  $|(\hat{\theta} - \theta_0) I^{1/2}(\hat{\theta})| > z_{\alpha/2}$

  AoS Def 10.3

  $|(\hat{\theta} - \theta_0) J^{1/2}(\hat{\theta})| > z_{\alpha/2}$

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

  $H_0: \theta = \theta_0$ vs $(H_1: \theta \neq \theta_0)$

- Score test:

  $|\ell'(\theta_0) I^{-1/2}(\hat{\theta})| > z_{\alpha/2}$

  $J^{-1/2}(\hat{\theta})$

  $\boxed{|\ell'(\theta_0) I^{-1/2}(\theta_0)| > z_{\alpha/2}}$

- Likelihood ratio test:

  AoS §10.6

  $2\{\ell(\hat{\theta}) - \ell(\theta_0)\} \geq \chi^2_{1,(1-\alpha)} (= 3.84)$ @ $\alpha = .05$

$\underline{H_0: \theta = \theta_0 \text{ simple}}$

Example 2: approximate pivotal quantities

$\theta = (\psi, \lambda) \quad (\theta_1, \theta_2)$

interest

nuisance

$H_0: \psi = \psi_0$

- Wald test: profile log likelihood

AoS Def 10.3

$$|(\hat\psi - \psi_0) J_p^{1/2}(\hat\psi)| > z_{\alpha/2}$$

$$\ell_p(\psi) = \ell(\psi, \hat\lambda_\psi)$$

- Score test:

$$|\ell_p'(\hat\psi) J_p^{-1/2}(\psi_0)| > z_{\alpha/2}$$

- Likelihood ratio test:

$$2\{\ell_p(\hat\psi) - \ell_p(\psi_0)\} > \chi^2_{q,1-\alpha}$$

AoS §10.6

$$-2 \log \frac{\sup_{\Theta} L(\theta; x)}{\sup_{\Theta_0} L(\theta; x)} \;\dot\sim\; \chi^2_q$$

p-value functions

Wald $\quad (\hat{\psi} - \psi) J_p^{1/2}(\hat{\psi}) \sim N(0,1)$ / under $H_0: \psi = \psi$

$\Phi\{ \qquad \} \triangleq$ p-value fn of ψ $\quad$ no prior

sort of like a Bayes posterior

$Pr_\pi\{\Psi \geq \psi \mid \underset{\sim}{x}\}$ $\quad \leftarrow$ prior

$\theta = (\mu_1, \dots, \mu_k)$ $\quad \boxed{\psi = \sum \mu_i^2}$

$\pi(\theta) d\theta = d\theta$

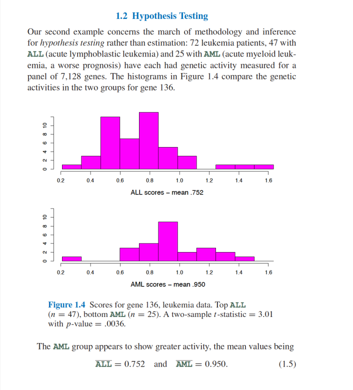
Example 2: two-sample t-test

- model $X_1, \dots, X_{n_1}$ iid $N(\mu_1, \sigma^2)$ $Y_1, \dots, Y_{n_2}$ iid $N(\mu_2, \sigma^2)$
- null and alternative hypothesis $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
- rejection region
- test statistics and critical value
- type I and type II error

$$R = \left\{ \underbrace{\left| \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right|}_{T(\bar{x}, \bar{y})} \geq t^{\alpha}_{n_1+n_2-2} \right\}$$

p value for H_0: $P_{H_0}\left(T(\bar{X}, \bar{Y}) \geq T(\bar{x}, \bar{y})\right)$

$\sim$ student t, n_1+n_2-2 d.f.



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

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

```
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:
```

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
```

$$\frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

est σ^2

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}}$$

$\sigma_x^2 \overset{\text{est.}}{=} \hat{\sigma}_y$

Satterthwaite's approx$^{\text{n}}$

ALL   AML

leukemia data (EH): $X_1, \dots, X_{47}$; $Y_1, \dots, Y_{25}$ AoS Ex. 10.20

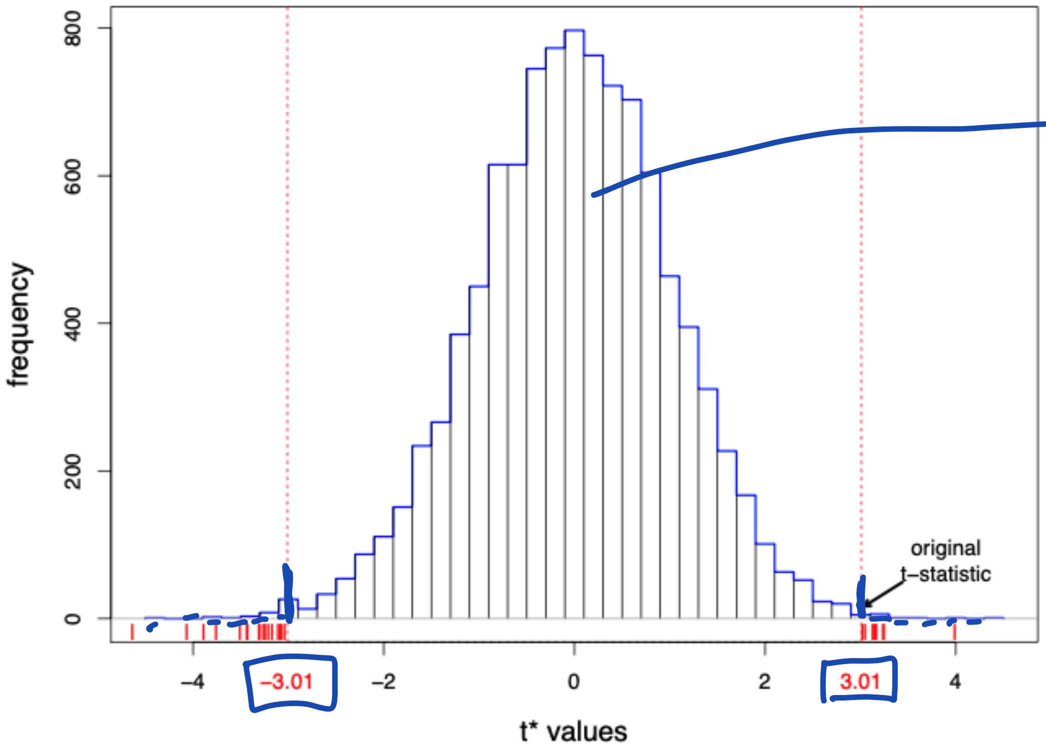
n_1 n_2

```
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
```

all $\binom{72}{47}$ possible allocations of "sp" to the obs.

AoS uses $\text{med}(\tilde{X}) - \text{med}(\tilde{Y})$

$H_0 : F_X = F_Y$ H_1 $F_X \neq F_Y$ $T = T(\boldsymbol{X}, \boldsymbol{Y}) = \dfrac{|\bar{X}^* - \bar{Y}^*|}{\sqrt{S_p^{2*}\left(\frac{1}{47} + \frac{1}{25}\right)}}$ ← test-statistic



T*

sub-sample of 10,000

(perm.test)

p-value is sum of frequencies for the extreme t statistics

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.

.0035

← I think uses var.eq = F for t^*

AoS 10.20 uses median

- $X_1, \ldots, X_n$ i.i.d.
- $\boxed{H_0 : X_i \sim f(x; \theta);}$ $H_1 : X_i$ has an arbitrary distribution
- Define k sets $\underline{I_1, \ldots, I_k}$ s.t.

could be intervals

$$\underline{\mathrm{pr}(X_i \in \cup_{j=1}^{k} I_j\} = 1}$$

- Define

$$\underline{Y_j = \sum_{i=1}^{n} 1\{X_i \in I_j\}}$$

number of obs in category j

$L(\theta; \underline{x}) = \prod_{i=1}^{n} f(x_i; \theta)$

- $X_1, \ldots, X_n$ i.i.d.
- $H_0 : X_i \sim f(x; \theta)$; $H_1 : X_i$ has an arbitrary distribution
- Define k sets $I_1, \ldots, I_k$ s.t.

could be intervals

$$\mathrm{pr}(X_i \in \cup_{j=1}^{k} I_j\} = 1$$

- Define

$$Y_j = \sum_{i=1}^{n} 1\{X_i \in I_j\}$$

number of obs in category j

- $Y = (Y_1, \ldots, Y_k) \sim Mult_k(n; p)$
- $\mathrm{pr}(Y_1 = y_1, \ldots, Y_k = y_k; p) =$ $\prod_{j=1}^{k} p_j^{y_j}$ $\quad \Sigma y_j = n$; $\Sigma p_j = 1$
- $\boxed{H_0 : p = p(\theta);}$ $H_1 : p$ arbitrary

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

$\boxed{\theta = (\psi, \lambda)}$

- log-likelihood function

$$L(p; y) = \prod p_j^{y_j}$$

$$\underline{\ell(p; y) = \sum_{j=1}^{k} y_j \log p_j}$$

$\sum p_j = 1$

$\sum y_j = n$

- generalized likelihood ratio test

$$w \triangleq 2\{\ell(\hat{p}; y) - \ell(p(\hat{\theta}); y)\}$$

$\hat{p}$ ↑ mult.

$p(\hat{\theta})$ ↑ mult w/ p's specified

$\underline{H_0: p_j} = p_j(\theta)$

$j = 1, \ldots, n$

$$\hat{\theta} \text{ max's}_{\theta} \sum_{j=1}^{k} y_j \log p_j(\theta)$$

$$\hat{p} \text{ max's} \sum_{j=1}^{k} y_j \log p_j \quad > \sum p_j = 1$$

$$\frac{\partial \ell}{\partial p_j} = \frac{y_j}{p_j} \quad ; \; j = 1, \ldots, k \quad = 0$$

Multinomial goodness of fit tests

$\ldots \quad \hat{p}_j = y_j/n$

- log-likelihood function

- generalized likelihood ratio test

$$2\left\{\sum y_j \log(y_j/n) - \sum y_j \log p_j(\hat{\theta})\right\}$$

$$= 2 \sum y_j \log\left[y_j / n p_j(\hat{\theta})\right]$$

- Theorem 10.22: Under H_0

$p = \dim(\theta)$

$$W = 2\sum_{j=1}^{k} Y_j \log\left(\frac{Y_j}{np_j(\tilde{\theta})}\right) \xrightarrow{d} \chi^2_{k-1-p}$$

$y_j/np_j(\hat{\theta})$

$$\sum_{j=1}^{k} O_j \log \frac{O_j}{E_j}$$

$\# p's = k-1$

$\# \theta's = p \quad (\theta \in \mathbb{R}^p)$

- log-likelihood function

$$Q = \sum_{j=1}^{k} \frac{\{Y_j - np_j(\hat\theta)\}^2}{np_j(\hat\theta)} \qquad \left(\sum \frac{(O-E)^2}{E}\right)$$

- generalized likelihood ratio test

$$Y_j \log\left\{\frac{Y_j}{np_j(\hat\theta)} + 1 - 1\right\} = Y_j \log\left(1 + \underbrace{\frac{Y_j - np_j(\hat\theta)}{np_j(\hat\theta)}}_{\varepsilon}\right)$$

- Theorem 10.22: Under H_0

$p = \dim(\theta)$

$$W = 2\sum_{j=1}^{k} Y_j \log\left(\frac{Y_j}{np_j(\tilde\theta)}\right) \xrightarrow{d} \chi^2_{k-1-p} \leftarrow Y_j(\log(1+\varepsilon))$$

$$\simeq \varepsilon - \frac{1}{2}\varepsilon^2$$

- Theorem 10.29 Under H_0

$$Q = \sum_{j=1}^{k} \frac{\{Y_j - np_j(\tilde\theta)\}^2}{np_j(\tilde\theta)} \xrightarrow{d} \chi^2_{k-1-p}$$

Pearson's χ^2 stat.

Table 9.1 *Frequency of goals in First Division matches and "expected" frequency under Poisson model in Example 9.2*

O →

Goals	0	1	2	3	4	≥ 5
Frequency	252	344	180	104	28	16
Expected	248.9	326.5	214.1	93.6	30.7	10.2

$$p_0(\lambda) = 1 - \sum_{j=0}^{4} p_j(\lambda); \quad p_j(\lambda) = e^{-\lambda}\lambda^j/j!, \quad \boxed{\tilde{\lambda} = 1.3118}$$

$$Q = 11.09; \quad W = 10.87; \quad \mathrm{pr}(\chi_4^2 > [11.09, 10.87]) = [0.026, 0.028]$$

evidence against H_0

Multinomial goodness-of-fit tests

SM Ex 4.38

		Antigen 'B'		
		Absent	Present	Total
Antigen 'A'	Absent	'O': 202	'B': 35	237
	Present	'A': 179	'AB': 6	185
Total		381	41	422

	Two-locus model		One-locus model	
Group	Genotype	Probability	Genotype	Probability
'A'	$(AA; bb)$, $(Aa; bb)$	$\alpha(1-\beta)$	(AA), (AO)	$\lambda_A^2 + 2\lambda_A\lambda_O$
'B'	$(aa; BB)$, $(aa; Bb)$	$(1-\alpha)\beta$	(BB), (BO)	$\lambda_B^2 + 2\lambda_B\lambda_O$
'AB'	$(AA; BB)$, $(Aa; BB)$, $(AA; Bb)$, $(Aa; Bb)$	$\alpha\beta$	(AB)	$2\lambda_A\lambda_B$
'O'	$(aa; bb)$	$(1-\alpha)(1-\beta)$	(OO)	λ_O^2

χ_1^2 ??

? 3 par. ?

Table 4.3 Blood groups in England (Taylor and Prior, 1938). The upper part of the table shows a cross-classification of 422 persons by presence or absence of antigens 'A' and 'B', giving the groups 'A', 'B', 'AB', 'O' of the human blood group system. The lower part shows genotypes and corresponding probabilities under one- and two-locus models. See Example 4.38 for details.

χ_k^2

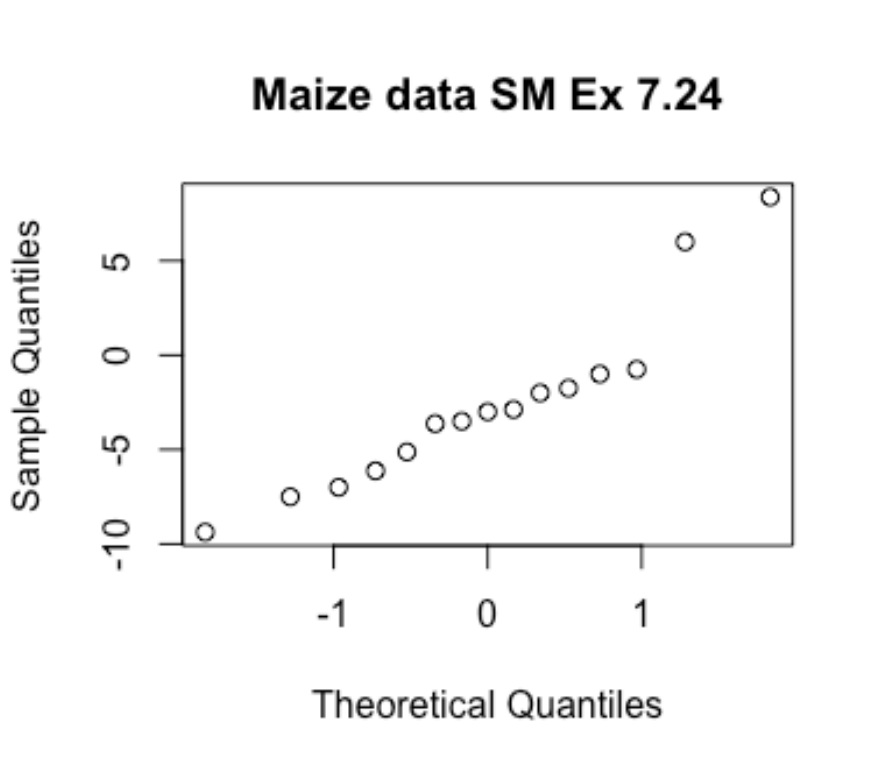
$Q = 15.73$; $W = 17.66$ (two-locus)

$p < 10^{-5}$

$Q = 2.82$; $W = 3.17$ (single locus)

$p = 0.09; 0.07$

Smooth goodness-of-fit tests



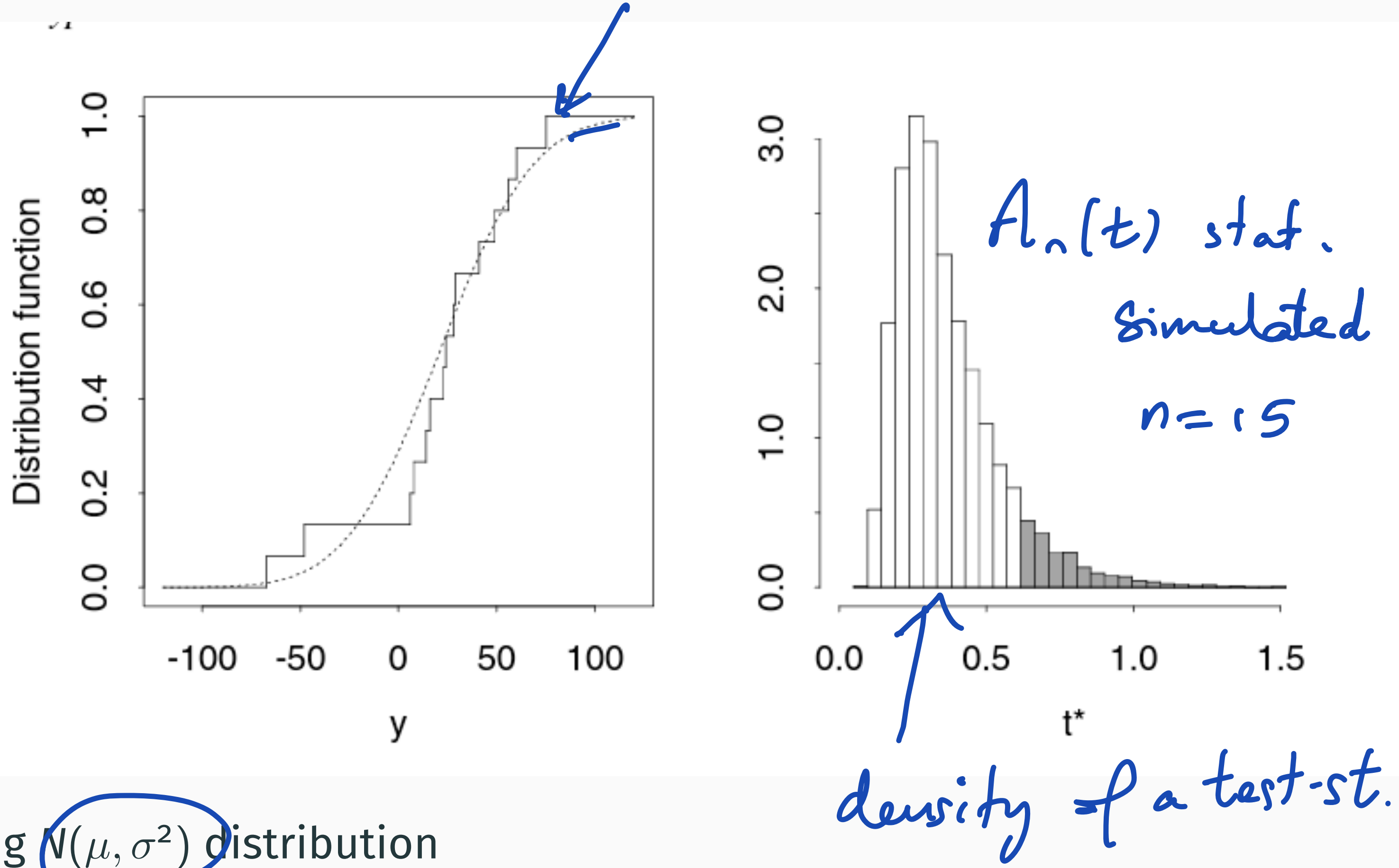

```
library(SMPracticals)
data(darwin)
cross <- seq(1,30,by=2)
self <- cross+1
diffs <- darwin[self,4]-darwin[cross,4]
qqnorm(diffs)
```

H_0: $X_i - Y_i = Z_i \sim N(\mu, \sigma^2)$

H_1: not

Smooth goodness-of-fit tests

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}.



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

- $X_1, \ldots, X_n$ i.i.d. $F(\cdot)$; $H_0 : F = F_0$ simple null cumulative d.f.
- $\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n 1\{X_i \le t\}$
- three test statistics:
  1. $\sup_t |\widehat{F}_n(t) - F_0(t)|$ ← Kolmogorov-Smirnov stat.
  2. $\int \{\widehat{F}_n(t) - F_0(t)\}^2 dF_0(t)$ ← Cramer-von Mises
  3. $\int \frac{\{\widehat{F}_n(t) - F_0(t)\}^2}{F_0(t)\{1 - F_0(t)\}} dF_0(t)$ ← Anderson-Darling
- SM Example 7.24 testing $N(\mu, \sigma^2)$ distribution
- SM Example 7.23; 6.14 testing $U(0, 1)$ distribution

$\{\hat{F}_n(t), 0 \le t \le 1\}$ (empirical process)

Smooth goodness-of-fit tests

$X_i \sim U(0,1)$

- Special case $H_0 : F(t) = F_0(t) = t$
- Recall

$$E_0\{\widehat{F}_n(t)\} = F_0(t) = t, \quad \text{var}\{\widehat{F}_n(t)\} = t(1-t)/n$$

- What about distribution of

$$\sup_t |\widehat{F}_n(t) - t| \qquad \int \{\widehat{F}_n(t) - t\}^2 dt \qquad \int \frac{\{\widehat{F}_n(t) - t\}^2}{\cancel{F_n}(t)\{1-t\}} dt$$

- need joint density of $\widehat{F}_n(t)\ \forall\ t$

Smooth goodness-of-fit tests

- Special case $H_0 : F(t) = F_0(t) = t$ ← uniform $\quad X_i \sim U(0,1)$
- Recall

$$E_0\{\widehat{F_n}(t)\} = F_0(t) = t, \quad \mathrm{var}\{\widehat{F_n}(t)\} = t(1-t)/n$$

generalize to $F_0(t;\tilde{\theta})$

$H_0: F = F_0(t;\tilde{\theta})$

- What about distribution of

$$\sup_t |\widehat{F_n}(t) - t| \qquad \int \{\widehat{F_n}(t) - t\}^2 dt \qquad \int \frac{\{\widehat{F_n}(t) - t\}^2}{F_0(t)\{1-t\}} dt$$

- need joint density of $\widehat{F_n}(t)\ \forall\ t$

- define stochastic process $B_n(t) = \sqrt{n}(\widehat{F_n}(t) - t)$ $\xrightarrow{d}$ Brownian bridge

- vector $(B_n(t_1), \ldots, B_n(t_k)) \xrightarrow{d} N_k(0, C), \quad C_{ij} = \min(t_i, t_j) - t_i t_j$ $\quad$ MS 9.3

$T_n: \sup_t |B_n(t)| \xrightarrow{d} \sup_t |B(t)|$

- a Brownian bridge is a continuous function on $(0,1)$ with all finite-dimensional distributions as above

Smooth goodness-of-fit tests

- Kolmogorov-Smirnov test

$$K_n = \sup_{0 \le t \le 1} |B_n(t)|$$

- Cramer-vonMises test

$$W_n^2 = \int_0^1 B_n^2(t) dt$$

- Anderson-Darling test

$$A_n^2 = \int_0^1 \frac{B_n^2(t)}{t(1-t)} dt$$

- Kolmogorov-Smirnov test

$$K_n = \sup_{0 \le t \le 1} |B_n(t)|$$

- Cramer-vonMises test

$$W_n^2 = \int_0^1 B_n^2(t)dt$$

- Anderson-Darling test

$$A_n^2 = \int_0^1 \frac{B_n^2(t)}{t(1-t)}dt$$

- limit theorems

$Z_j \sim N(0,1)$

$$K_n \xrightarrow{d} K, \qquad W_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j^2\pi^2}, \qquad A_n^2 \xrightarrow{d} \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)}$$

$\mathrm{pr}(K > x) = 2\sum_{j=1}^{\infty}(-1)^{j+1}\exp(-2j^2x^2)$

θ may need to be est'd $\Rightarrow$ needs re-think

 Wikipedia

1. Hypothesis testing

AoS Table 10.1

action

		H_0 not rejected	H_0 rejected
truth	H_0 true	✓	type 1 error
	H_1 true	type 2 error	✓

2. Diagnostic testing

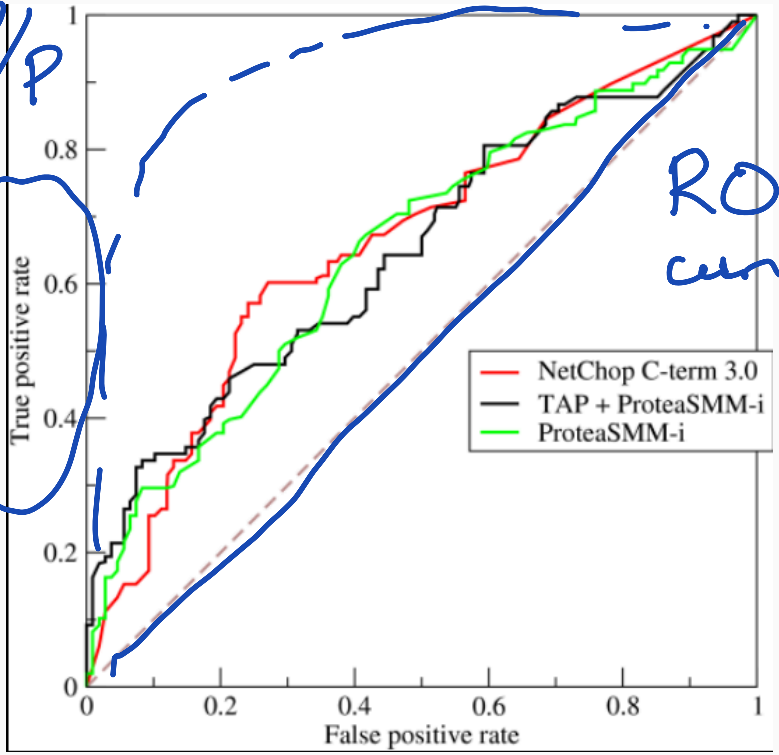
obs ?=

link

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

P ← TP/P (power)

TP/P



1 − TN/N

True positive rate $=$ TP/P

$=$ sensitivity $=$

TP/P

False positive rate $=$ 1 − TN/N

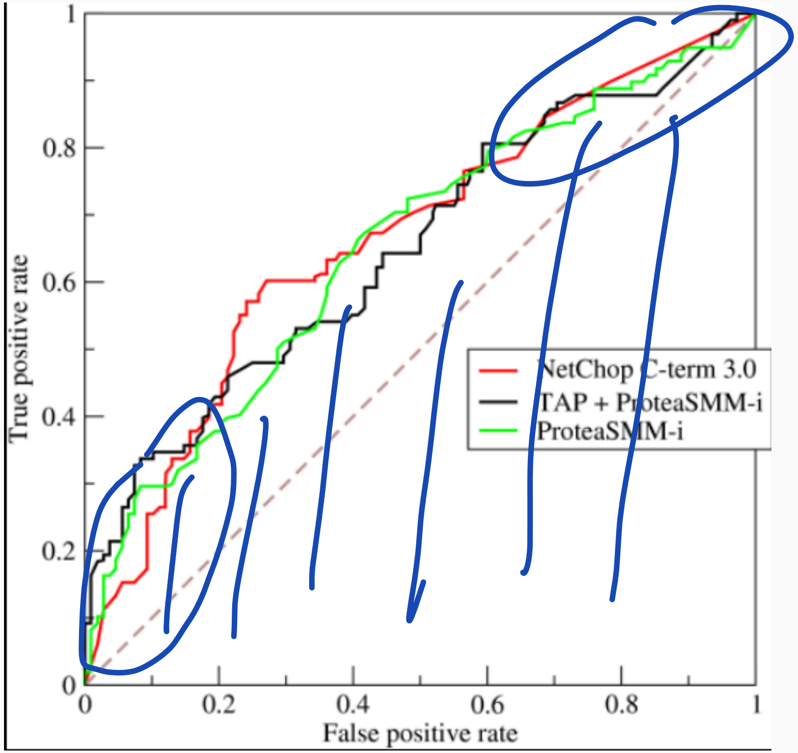
$=$ 1$-$ specificity $=$

$1 - TN/N$

spec. $= \frac{TN}{N}$

PPV positive predictive value

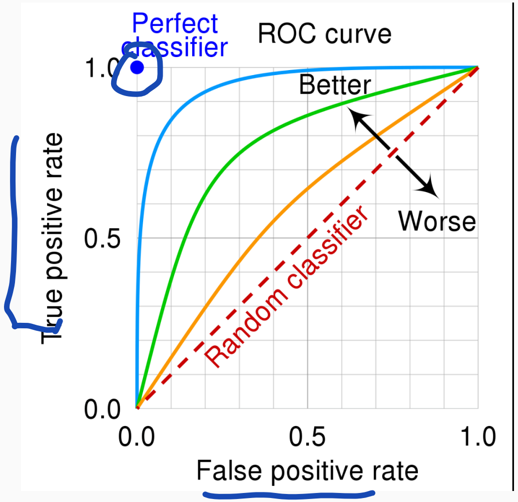
? $f(\frac{TP}{P}, P)$?



True positive rate $=$ sensitivity $=$ TP/P

False positive rate $=$ $1-$ specificity $=$ $1 - TN/N$

AUC = area under curve



Rapid flow test, care home link

		test negative	test positive	
truth	C19 neg	114,993	101	115,094
	C19 pos	371	128	499

Sensitivity = TP/P = 128/499 = 0.257

Specificity = TN/N = 114,993/115094 =0.999

Cochrane review

meta-analysis

"consistently high specificities"

"sensitivity varied widely: average sensitivities by brand ranged from 34.3% to 91.3% "