p-value

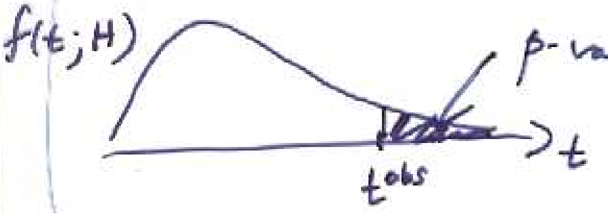
If we have a function of data $t(Y) = T$, say, and a model for the data, hypothesis, then p-value for hypothesis, based on t,

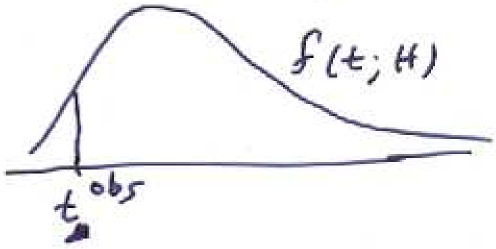
$$\boxed{\text{probability}\,(T \geq t^{obs}\,;\ \text{model})}$$

$t^{obs} = t(y)$

(1-sided p-value)

probability (result as or more extreme than observed) ↳ both or one?





$1 - F_T(t^{obs})$

dist^n f^n for T

we would probably use

$p = pr\,(T \leq t^{obs})$ bec. evidence in other direction

- we need a 1-dim $t(\cdot)$ for this to work ✓ t-stat $\frac{\bar{y}-\bar{x}}{s\sqrt{\cdots}}$
- we need to know its dist^n under model ✓ t_{n+m-2}

Density Estimation §5.6

Sample $y_1, \dots, y_n$ assume i.i.d. $f(y)$

$f(y)$ = density fn for y i.e.

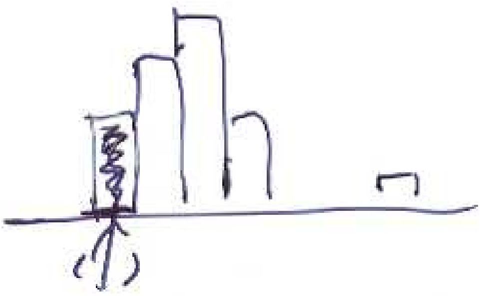
$f(y)dy = pr(Y \in (y, y+dy))$ where Y is a r.v. ~~with same~~

$$f(y) = \frac{d}{dy} pr(Y \le y)$$

How to estimate $f(y)$?

- histogram is simplest   >hist(y) counts

  >hist(y, prob=T) freq.



- not smooth, even if $f(\cdot)$ is

$$\frac{\# \, y_i\text{'s in } (,)}{n}$$

- depends a lot on how you choose bins; # + start pt.

>truehist in MASS library is better; easier to specify bins   Fig 5.8

- kernel density estimator (better)

$$\hat{f}(y) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{y-y_i}{b}\right)$$

$K(\cdot)$ is some function e.g. $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

b is some parameter "bandwith"

- default for $K(\cdot)$ is Gaussian (normal), but there are others (but not usually so crucial)

- choice of b - called bandwith (bw in R) is very crucial

$$E \int \{\hat{f}(y) - f(y)\}^2 \, dy$$

$\hat{f}(\cdot) = \hat{f}(\cdot\,; y_1, \dots, y_n)$
fn. of data

Integrated mean squared error

$$= \frac{1}{nb} \int K^2(y)\,dy + \frac{1}{4} b^4 \int f''(y)^2 \, dy \left[\int y^2 K(y)\,dy\right]^2 + O\left(\frac{1}{nb} + b^4\right)$$

$n \to \infty$, $b \to 0$, $nb \to \infty$

"If we neglect $O(\ )$, then optimal b is

$$b = \left[\frac{\int K^2(y)\,dy}{n \int f''(y)^2 dy \left(\int y^2 K(y)\,dy \right)^2} \right]^{1/5}$$

- now 2 possibilities ① use an est. of $f''(\cdot)$ to find the b e.g. $\hat{f}''(\cdot)$

② solve this eqⁿ with b on both sides

- these are called

bw = "SJ.dpi" ← "direct plug-in" ①

or bw = "SJ.ste" ← "solve the eq'n" ②

default is neither, but bw = "~~[illegible]~~ nrd0" which ~~means~~ is ~~[illegible]~~ a slight modⁿ of

$$b = \underline{0.9 \min\left(\hat{\sigma}, \frac{IQR}{1.34}\right)} n^{-1/5}$$

"ad-hoc" suggestion that's often ok

Default choice may need improving;

next best is bw = SJ.dpi //
or ste

Chapter 7 Generalized Linear Models

Linear model $y_i = x_i^T \beta + \sigma e_i \qquad e_i \sim f_0(\cdot)$

$$f(y_1, \dots, y_n) = \prod_{i=1}^{n} f_0\left(\frac{y_i - x_i^T\beta}{\sigma} \right) \cdot \frac{1}{\sigma} \qquad \begin{matrix} ||| \\ N \end{matrix}$$

~~loc~~ regression - scale

Gen. lin model $\quad f(y_1, \dots, y_n) = \prod_{i=1}^{n} f_1(y_i ; x_i^T\beta, \sigma)$

$f_1(\cdot)$ is a density $f^=$, but not location-scale as above

in exponential family $\Leftarrow$ to be defined

- extend linear inference, or linear regression to more choice of models