2004-10-27

STATS 3N03/3J04

18-1

TERM MARKS ARE POSTED ON THE WEB. (CHECK YOURS!)

MAKE-UP TESTS

FRI NOV 5 9:30AM REF-102

FRI NOV 26 9:30 AM MDCL-1102

TEST PREPARATION

- PROBLEMS IN TEXT
- A02, PAST A02, PAST T02
- TRIVIA, DEFINITIONS
- SEE "WHERE ARE WE" PAGE FOR SECTIONS OMITTED

WHAT ARE WE DOING NOW???

MEASUREMENT MODEL

MAKE n INDEPENDENT REPETITIONS OF THE SAME MEASUREMENT, EACH SUBJECT TO RANDOM ERROR.

TRUE VALUE IS μ.

EACH REPETITION HAS THE SAME DISTRIBUTION OF ERROR ABOUT μ, EXPRESSED BY A STANDARD DEVIATION σ.

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ IS OUR ESTIMATE OF μ.

BUT HOW PRECISE IS IT?

THEORY SHOWS THAT, AT LEAST WHEN n IS LARGE, IT IS USUALLY SAFE TO ASSUME THAT

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

IN ORDER TO MAKE INFERENCES ABOUT μ.

~~eg~~ $\bar{X}$ WILL BE WITHIN $\pm 1.96 \frac{\sigma}{\sqrt{n}}$ OF μ WITH PROBABILITY = 95%.

RESTATED:

μ WILL BE WITHIN $\pm 1.96 \frac{\sigma}{\sqrt{n}}$ OF $\bar{X}$ WITH ~~PROBABILITY~~ CONFIDENCE = 95%

WHY DO WE STUDY THE THEORY ??

- BY SEEING WHERE THE ASSUPTIONS GO IN, WE CAN UNDERSTAND HOW VALID THE RESULT CAN BE WHEN THE ASSUMPTIONS AREN'T EXACTLY TRUE.

DATA: $X_1, \dots, X_n$

MEASUREMENT MODEL:

$$E(X_i) = \mu \qquad Var(X_i) = \sigma^2 \quad \forall i$$

$$Cov(X_i, X_j) = 0 \quad \forall\, i \neq j$$

$$\bar{X} = \frac{1}{n} X_1 + \cdots + \frac{1}{n} X_n$$

RECALL: $Y = AX$

$$E(Y) = A\,E(X)$$

$$Var(Y) = A\,Var(X)\,A'$$

HERE: $\bar{X} = \left(\frac{1}{n}, \cdots, \frac{1}{n}\right)\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$

$$E(X) = \begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} \qquad Var(X) = \sigma^2 I_n$$

$$E(\bar{X}) = \left(\frac{1}{n}, \cdots, \frac{1}{n}\right)\begin{pmatrix} \mu \\ \vdots \\ \mu \end{pmatrix} = \frac{n\mu}{n} = \mu$$

$$Var(\bar{X}) = \sigma^2 \left(\frac{1}{n}, \cdots, \frac{1}{n}\right) I_n \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix}$$

$$= \sigma^2 \frac{n}{n^2} = \frac{\sigma^2}{n}$$

REPRODUCTIVE PROPERTY OF THE NORMAL:

"A LINEAR COMBINATION OF NORMAL RANDOM VARIABLES IS NORMAL"

$X_1, \dots, X_n \sim IN(\mu, \sigma^2)$

$\Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$\Rightarrow \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

EXACT RESULTS

CENTRAL LIMIT THEOREM:

"THE DISTRIBUTION OF A LINEAR COMBINATION OF RANDOM VARIABLES APPROACHS THE NORMAL DISTRIBUTION AS THE NUMBER OF VARIABLES $\to \infty$"

(SOME REGULARITY CONDITIONS APPLY)

$$X_1, \ldots, X_n \sim I*(\mu, \sigma^2)$$

$$\Rightarrow \bar{X} \sim AN\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim AN(0,1)$$

APPROXIMATE NORMALITY

WHAT IF DATA ARE CORRELATED ??

EX WASHER & PIN

EX BAG OF APPLES