AI SEMINAR WEEK II

Phil Bording

1) BRIEF REVIEW
2) AGENTS - ACTIONS AND DOMAINS
3) REGRESSION - LINEAR MODELS
4) QUESTIONS

Linear Models I

$$\text{Price} = a \cdot \text{Area} + b$$

$$\begin{bmatrix} P_0 \\ P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} A_0 & 1 \\ A_1 & 1 \\ A_2 & 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix}$$

Three Equations
Two Unknown's

$$\begin{bmatrix} A_0 & A_1 & A_2 \\ 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} P_0 \\ P_1 \\ P_2 \end{bmatrix} = \begin{bmatrix} A_0 & A_1 & A_2 \\ 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} A_0 & 1 \\ A_1 & 1 \\ A_2 & 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix} \Sigma A_i P_i \\ \Sigma P_i \end{bmatrix} = \begin{bmatrix} \Sigma A_i A_i & \Sigma A_i \\ \Sigma A_i & 3 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\begin{bmatrix} \Sigma A_i^2 & \Sigma A_i \\ \Sigma A_i & 3 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \Sigma A_i P_i \\ \Sigma P_i \end{bmatrix}$$

Solve

Two Equations
Two Unknown's

Linear Models II

$$\text{Price} = a\,\text{Area} + b\,\text{Bath} + c$$

$$
\begin{aligned}
P_0 &= aA_0 + bB_0 + c \\
P_1 &= aA_1 + bB_1 + c \\
P_2 &= aA_2 + bB_2 + c
\end{aligned}
$$

$$
\begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{bmatrix} = \begin{bmatrix} A_0 & B_0 & 1 \\ A_1 & B_1 & 1 \\ A_2 & B_2 & 1 \\ A_3 & B_3 & 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix}
$$

$$
\begin{bmatrix} A_0 & A_1 & A_2 & A_3 \\ B_0 & B_1 & B_2 & B_3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{bmatrix} =
$$

$$
\begin{bmatrix} A_0 & A_1 & A_2 & A_3 \\ B_0 & B_1 & B_2 & B_3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} A_0 & B_0 & 1 \\ A_1 & B_1 & 1 \\ A_2 & B_2 & 1 \\ A_3 & B_3 & 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix}
$$

Linear Models

$$\begin{bmatrix} \Sigma A_i P_i \\ \Sigma B_i P_i \\ \Sigma P_i \end{bmatrix} = \begin{bmatrix} \Sigma A_i^2 & \Sigma A_i B_i & \Sigma A_i \\ \Sigma A_i B_i & \Sigma B_i^2 & \Sigma B_i \\ \Sigma A_i & \Sigma B_i & 4 \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

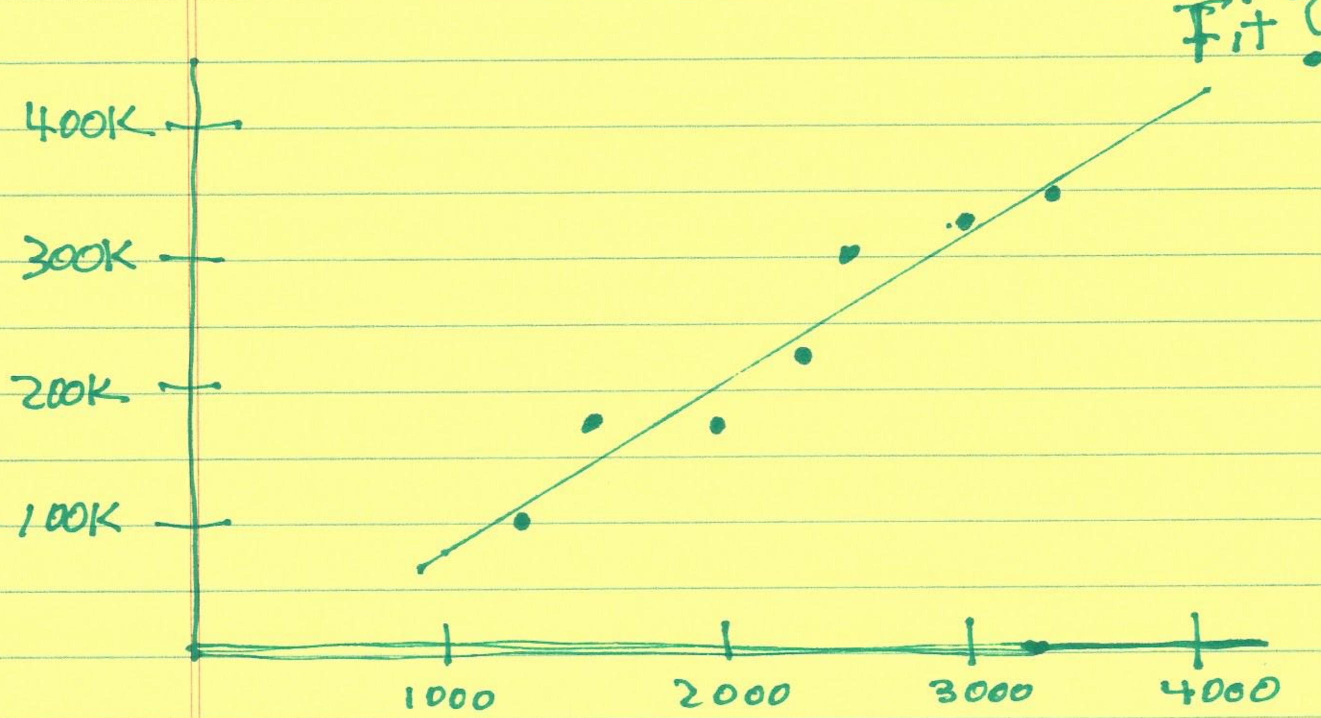
In General

$$\begin{bmatrix} \Sigma A_i^2 & \Sigma A_i B_i & \Sigma A_i \\ \Sigma A_i B_i & \Sigma B_i^2 & \Sigma B_i \\ \Sigma A_i & \Sigma B_i & n \end{bmatrix} \times \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \Sigma A_i P_i \\ \Sigma B_i P_i \\ \Sigma P_i \end{bmatrix}$$

Three Equations - Three Unknowns

Linear Models I

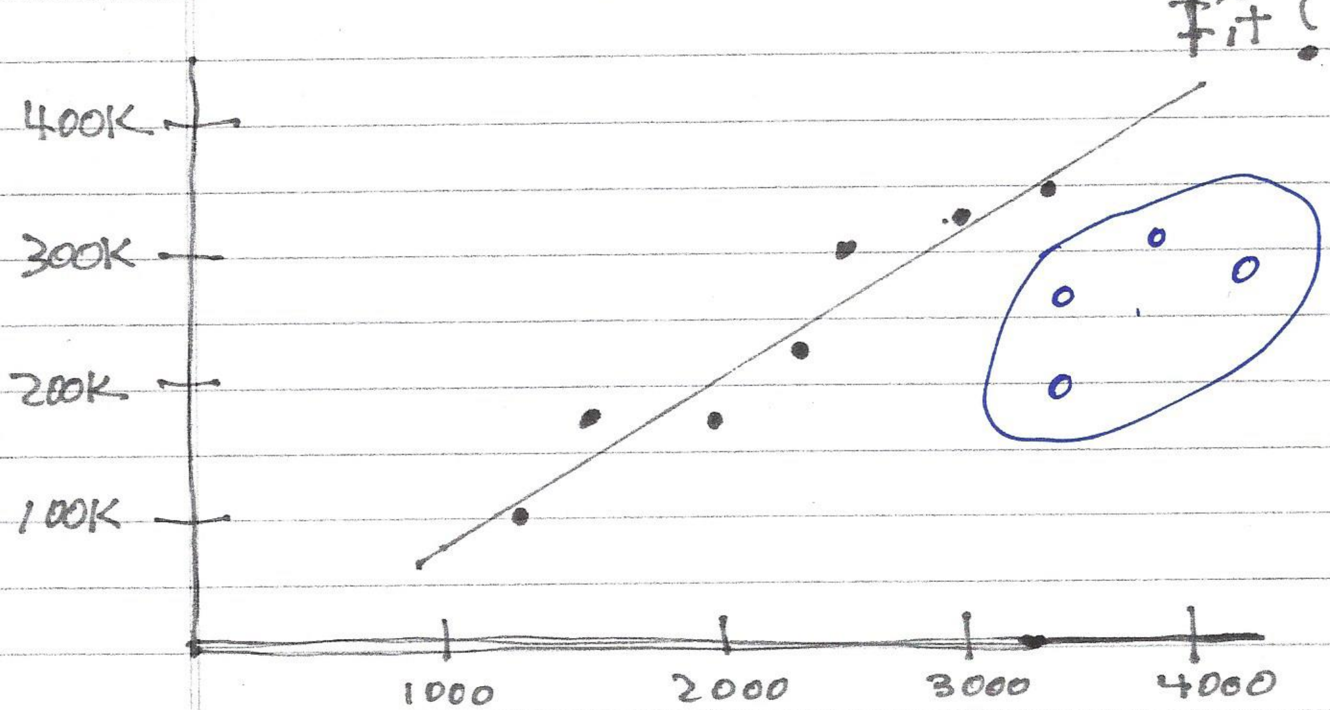
Price = a · Area + b



Price	Area
$100,000	1200
$220,000	2400
$180,000	2000
$320,000	3000
$350,000	3600
$180,000	1500
$300,000	2600

Linear Models I

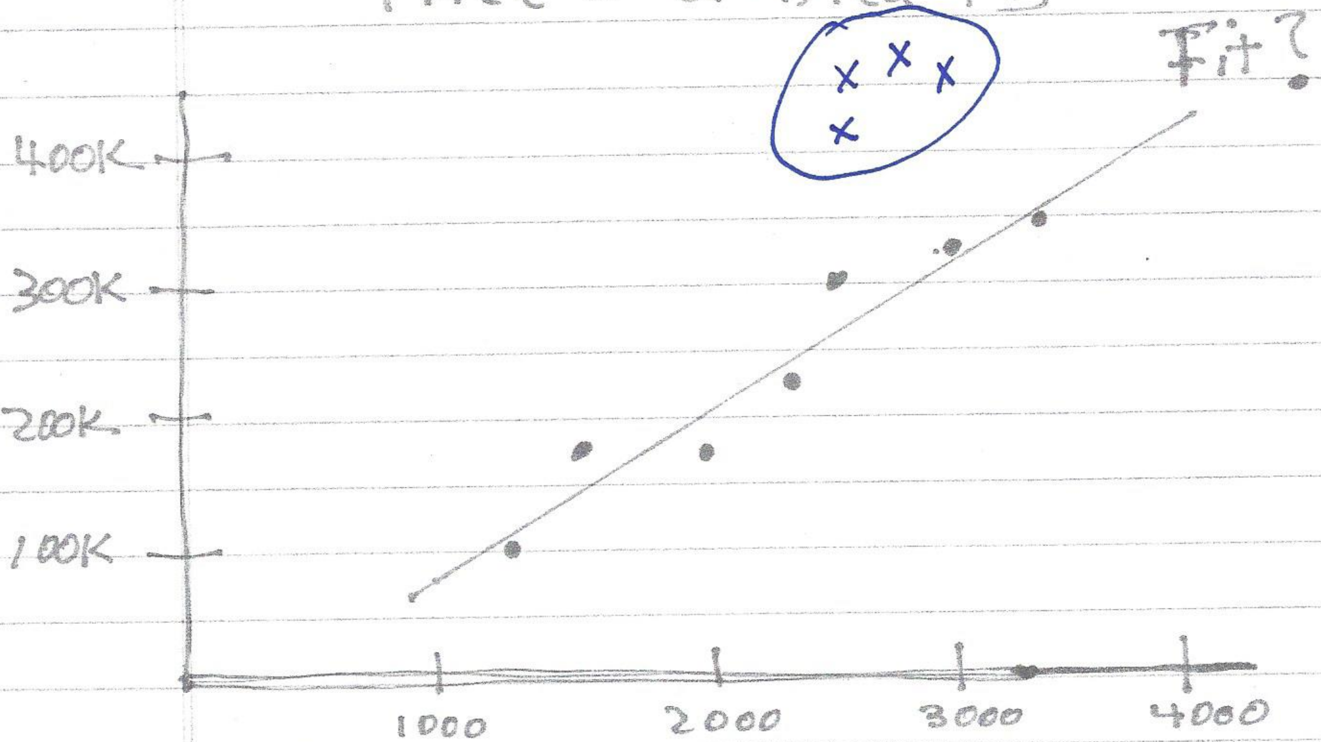
Price = a · Area + b



Price	Area
\$100,000	1200
\$220,000	2400
\$180,000	2000
\$320,000	3000
\$350,000	3600
\$180,000	1500
\$300,000	2600

Linear Models I

Price = a · Area + b



Price	Area
$100,000	1200
$220,000	2400
$180,000	2000
$320,000	3000
$350,000	3600
$180,000	1500
$300,000	2600

Linear Models - Housing Data

$$\text{Price} = \text{Area} \cdot a + \text{Brm} \cdot b + c$$

\$		
232,000	1416	2
330,000	1600	3
369,000	2400	3
400,000	2104	3
540,000	3000	4

$$\text{Price} = 63.84 \cdot \text{Area} + 103436 \cdot \text{Brm} - \$70434.00$$

```
1416.    2.    232000.
1600.    3.    330000.
2400.    3.    369000.
2104.    3.    400000.
3000.    4.    540000.
```

Living Area	Bedrooms	Price,k	$ ft	$ brd
1416.	2.	232000.	163.842	116000.0
1600.	3.	330000.	206.250	110000.0
2400.	3.	369000.	153.750	123000.0
2104.	3.	400000.	190.114	133333.3
3000.	4.	540000.	180.000	135000.0

amT x am

23751872.00	33144.00	10520.00	4203712000.00	4203712000.00
33144.00	47.00	15.00	5921000.00	5921000.00
10520.00	15.00	5.00	1871000.00	1871000.00

weights solution

63.84	103436.05	-70434.60

Data * weights = price

	Model Price $	Asking Price $	~~Variance~~ Difference $
1	226839.71	232000.00	-5160.29
2	342022.94	330000.00	12022.94
3	393097.64	369000.00	24097.64
4	374200.00	400000.00	-25800.00
5	534839.71	540000.00	-5160.29

Sum of ~~Variance~~ Differences 0.000

amT x am

23751872.00	10520.00	4203712000.00	4203712000.00
10520.00	5.00	1871000.00	1871000.00
10520.00	0.32	0.20	1871000.00

weights solution

165.12	26789.88

Data * weights = price

	Model Price $	Asking Price $	~~Variance~~ Difference $
1	260598.21	232000.00	28598.21
2	290980.09	330000.00	-39019.91
3	423075.19	369000.00	54075.19
4	374200.00	400000.00	-25800.00
5	522146.51	540000.00	-17853.49

Sum of ~~Variance~~ Differences 0.000

```
amT x am
       47.00          15.00      5921000.00      5921000.00
       15.00           5.00      1871000.00      1871000.00
    10520.00           0.32            0.20      1871000.00

weights solution

   154000.00      -87800.00


Data * weights = price
                                   Difference
  Model Price     Asking Price     ~~Variance~~
      $               $               $

 1  220200.00     232000.00     -11800.00
 2  374200.00     330000.00      44200.00
 3  374200.00     369000.00       5200.00
 4  374200.00     400000.00     -25800.00
 5  528200.00     540000.00     -11800.00

Sum of ~~Variance~~           0.000
       Differences
```

```
        program house
!
      implicit real*8 (a-h,o-z)
!
!     Linear least squares of house data -- Ng's example - Stanford
!
!     my code - Ralph Phillip Bording
!
        real*8  am(3,5)
        real*8  cm(3,5)
        real*8  dm(3,5)
!
        real*8  xm(3)
        real*8  bm(5)


        nvar   =  3
        nhouse  = 5

        do i=1,nvar
          xm(i) = 0.0
        enddo


!
      open(8,file="house.data",form="formatted")
!
!     read inputs
!
      write(6,*) "        "
     write(6,*)"    Living Area   Bedrooms    Price,k  $ ft     $ brd "
      write(6,*) "        "
!
      do in=1,nhouse
         read(8,*) area,bed,price
         dpft = price/area
         dpbrm = price/bed
         write(6,200) area,bed,price,dpft,dpbrm
!
         am(1,in) = area
         am(2,in) = bed
         am(3,in) = 1.0
         cm(1,in) = area
         cm(2,in) = 1.0
         cm(3,in) = 0.0
         dm(1,in) = bed
         dm(2,in) = 1.0
         dm(3,in) = 0.0

         bm(in) = price

      enddo

       close(8)

 200  format(5x,3f9.0,f9.3,f12.1)
!
!
!      make linear least squares model of house data
!
```

```
!
      call lsq(am,bm,xm,nhouse,nvar)

      nvar = 2

      call lsq(cm,bm,xm,nhouse,nvar)
      call lsq(dm,bm,xm,nhouse,nvar)

      end
      subroutine lsq(ain,bin,xm,nhouse,nvar)

      implicit real*8 (a-h,o-z)

      real*8  ain(3,5)
      real*8  am (3,4)
      real*8  xm(*)
      real*8  bin(*)
      real*8  ym(5)
      real*8  sumvar
      real*8  sum
!
!
        write(6,*) "                     "
        do i=1,nvar
         do k=1,nvar
          am(i,k) = 0.0
          do j=1,nhouse
          am(i,k) =  am(i,k) +  ain(i,j) * ain(k,j)
          enddo
         enddo
        enddo
!
!       form rhs =  xlT * y
!
           do j=1,nvar
           ym(j) = 0.0
            do i=1,nhouse
            ym(j) =  ym(j) +  ain(j,i) * bin(i)
           enddo
           am(j,nvar+1) = ym(j)
         enddo

         write(6,*) "                     "
         write(6,*) "  amT x am                "
         do i=1,3
          write(6,230) (am(i,k),k=1,nvar+1),ym(i)
 230      format(8f16.2)
         enddo
         write(6,*) "                     "
         nm = nvar
         nmd = 3
           call solve(am,xm,nm,nmd,ierr)

         write(6,*) " weights solution  "
         write(6,*) "                     "
          write(6,230) (xm(k),k=1,nvar)
         write(6,*) "                     "

!
!
```

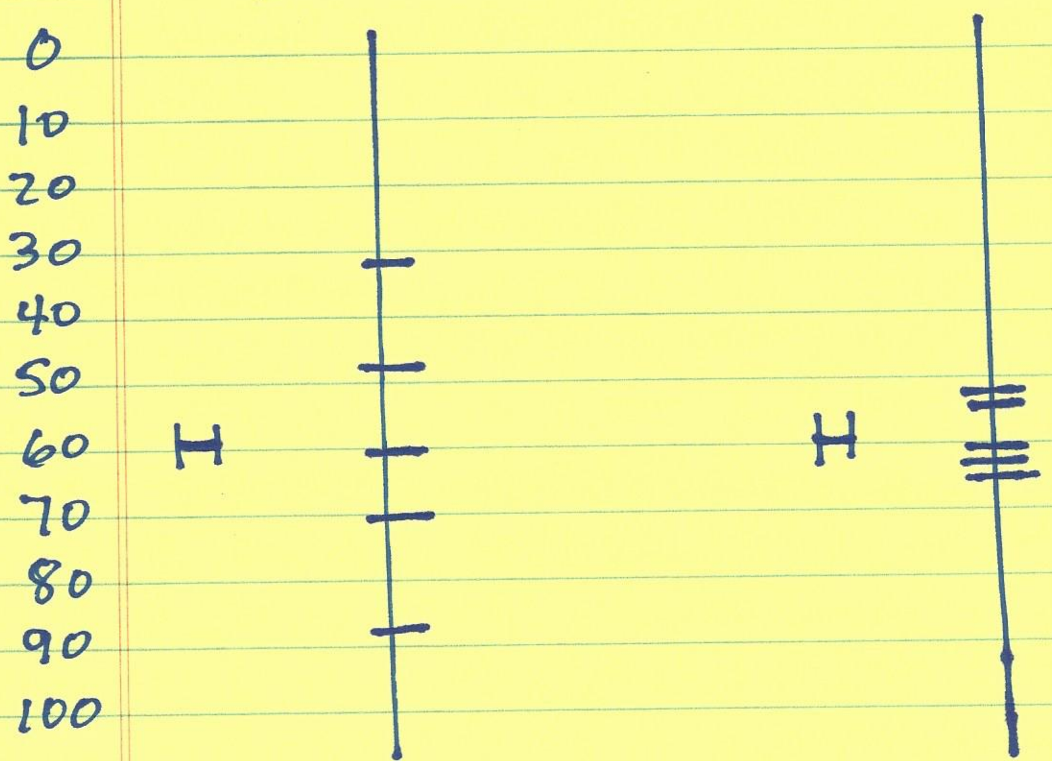
```
!       see how good least squares fit is............
!
!
        write(6,*)  "                          "
        write(6,*)  "   Data * weights = price  "
        write(6,*)  "                          "
                                                         Difference
        write(6,*)  "     Model Price     Asking Price     ~~Variance~~"
        write(6,*)  "          $                $              $    "
        write(6,*)  "                          "
!
        sumvar = 0.0
        do i=1,nhouse
         sum = 0.0
        do j=1,nvar
           sum = sum +   ain(j,i) * xm(j)
         enddo

           sumvar =  sumvar + sum-bin(i)
           write(6,221) i,sum,bin(i),sum-bin(i)
 221       format(i5,f12.2,2f14.2)
        enddo
        write(6,*)  "                          "
        write(6,233)  sumvar
 233    format("  Sum of ~~Variance~~ ",f14.3)
        write(6,*)  "   Differences"

      return
      end

       SUBROUTINE SOLVE(a, x, n, nd, errflag)
!      IMPLICIT NONE
!Declarations
!      IMPLICIT NONE
       implicit real*8 (a-h,o-z)
       INTEGER, INTENT(IN) :: n   !Stores the number of unknowns
       INTEGER, INTENT(IN) :: nd   !Stores the number of unknowns
       INTEGER, INTENT(OUT) :: errflag
       REAL*8, DIMENSION(nd,nd+1) :: a!An n x n+1 matrix which stores the simultaneous
equations
       REAL*8, INTENT(OUT), DIMENSION(nd) :: x !Array to store the solutions
       INTEGER :: i,j,k !Counters for Loops
       INTEGER :: largest
       REAL*8 :: temp, tm, sums
       LOGICAL :: FLAG
       tm = 0.0d00
       !Solve Using Gaussian Elimination
       FLAG = .FALSE.
       x = 0
       DO k = 1, n-1
              DO j = 1, n
                     IF (a(j, 1) /= 0.0d00 ) FLAG = .TRUE.
              END DO
              IF (FLAG .EQV. .FALSE.) THEN
                     PRINT*,"No Unique Solution"
                     errflag = -1
                     x = 0.0d00
                     EXIT
              ELSE
                     largest = k
                     !Find largest coefficient of first unknown
```

```
                    DO j = k, n
                     IF (ABS(a(j, k)) > ABS(a(largest,k))) largest = j
                    END DO
                      !Make the equation with largest first coefficient as the first equation
                      !Largest coefficient is chosen to prevent round-off errors as far as
possible
                      DO j = 1, n + 1
                             temp = a(k, j)
                             a(k,j) = a(largest,j)
                             a(largest,j)=temp
                      END DO
               ENDIF
               !Convert the input matrix to Upper Traingualar form
               DO j = k+1, n
                      tm = a(j,k)/a(k,k)
                      DO i = k+1, n+1
                             a(j,i) = a(j,i) - tm*a(k,i)
                      END DO
               END DO
               !No unique solution exists if the last element in the upper triangular matrix
is zero
               IF (a(n,n) == 0.0d00) THEN
                      PRINT*,"No Unique Solution"
                      errflag = -1
                      x = 0.0d00
                      EXIT
               ELSE
                      !Find xn
                      x(n) = a(n,n+1)/a(n,n)
                      !Find the remaining unknowns by back-substitution
                      DO i = n-1, 1, -1
                             sums = 0.0d00
                             DO j = i+1, n
                                    sums = sums + a(i,j)*x(j)
                             END DO
                             x(i) = (a(i,n+1) - sums)/a(i,i)
                      END DO
               ENDIF
        END DO
        END SUBROUTINE SOLVE
```

KARAOKE SCORES

TEAM MIU		TEAM RISA	
Miu	48	Risa	67
Yuko	32	Asuka	55
Aiko	88	Nana	61
Maya	61	Yuki	63
Marie	71	Rika	54
Average	60	Average	60

PLOT MIU PLOT RISA



MEAN VALUE
VARIANCE

$$\frac{\sum \text{scores}}{N} = \text{Mean Value}$$

48	67
32	55
88	61
61	63
71	54
300 / 5 = 60	300 / 5 = 60

$(48-60)^2$	$(67-60)^2$
$(32-60)^2$	$(55-60)^2$
$(88-60)^2$	$(61-60)^2$
$(61-60)^2$	$(63-60)^2$
$(71-60)^2$	$(54-60)^2$
1834	120

$$\frac{1834}{n-1} = \frac{1834}{4} = 458.8 \quad \text{Variance}$$

$$\frac{120}{n-1} = \frac{120}{4} = 30 \quad \text{Variance}$$

VARIANCE

Variance

CLOSE TO AVERAGE $\Rightarrow$ SMALL

FAR FROM AVERAGE $\Rightarrow$ LARGE

STANDARD DEVIATION

$$S.D. = \sqrt{VAR}$$

TEAM MIU

$$\sqrt{458.8} \doteq 21.4 \equiv S.D._{MIU}$$

TEAM RISA

$$\sqrt{30} = 5.5 \equiv S.D._{RISA}$$

DATA – MEAN

TEAM MIU	TEAM RISA
$D - \overline{m}$	$D - \overline{m}$
−12	7
−28	−5
28	1
+1	3
+11	−6
"0"	"0"
ZERO	ZERO

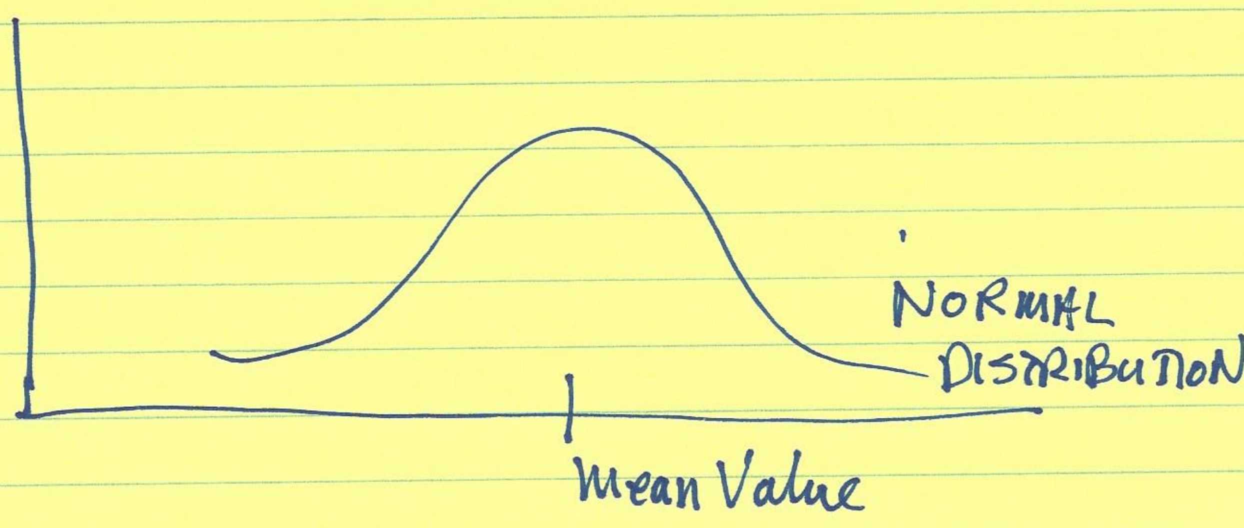
DATA HAS INFORMATION

DATA HAS VARIABILITY

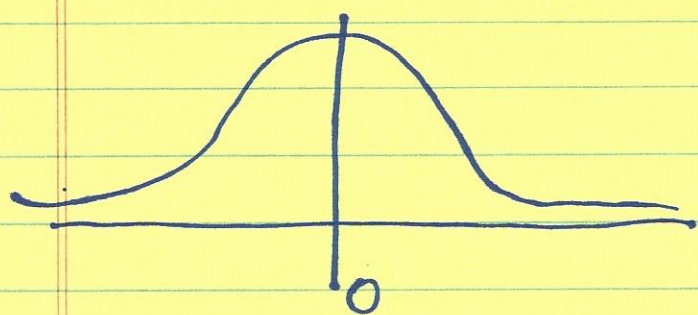
MEASUREMENT HAS
1) ACCURACY

2) ERROR
   a) RANDOM
   b) SYSTEMATIC

Removing Mean and Drift from data is usefull in study of error and accuracy.



NORMAL DISTRIBUTIONS



1) BELL CURVE

2) EQUATION, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

3) $x \to \infty \quad f(x) > 0$
$x \to -\infty \quad f(x) > 0$

4) Probability Density Function

5) Area under the curve $\equiv 100\%$

NEXT WEEK

1) SHORT REVIEW

2) PERCEPTRON'S

3) MINES AND ROCKS DATA

4) GENERALIZATION
   TO
   CONVOLUTIONAL
   NEURAL
   NETS

5) DISCUSSION