UNITED ARAB REPUBLIC

التخطيط القومي

THE INSTITUTE OF NATIONAL PLANNING

Memo No 556

Notes On Statistical Methods

By

Dr. Abd El-Azim Anis

Operations Research Center

March 1965.

Preface

This is a short-period course on statistical methods designed for mathematicians and engineers whose work needs a thorough knowledge of statistical methods.

Because of considerations of time. It was necessary to concentrate on the theory of probability and distributions, leaving the applied side to another course given by Dr. M. W. Mahmoud.

By using the theory of sets, and the matrix notation, it is hoped that the course would make a more sound approach to the theory of probability, and give shorter proofs to a good number of theoriems. It is hoped also that this would help in the field of applications.

The course also includes an introduction to stochastic processes and Random - Walk problems which should be useful to those interested in applications in this field, among engineers and research students in economics, biology and other related fields.

I take this opportunity to thank Mrs. Mary Naguib for the generous help she gave in preparing this course, correcting the proofs, and organizing the publication of this memorandum.

A. A. Anis

9/3/1965.

1. SETS The possible outcomes of an experiment are called (random) events, which may be simple or compound, the latter being aggregates of simple events. It is convenient to represent simple events as points in a space of appropriate dimension, called the observation space. Compound events are then represented by sets of points.

2. Notation: $x \in A$ means x is a member of the set A

$\{x : K(x)\}$	the set of all x's having the property $K(x)$
$A = B$	the sets A, B consist of the same elements- ie. if $x \in A$ then $x \in B$ and conversely
$A \subset B$ or $B \supset A$	A is a subset of B; If $x \in A$ then $x \in B$
Union : $A \cup B$,	$\{x : x \in$ at least one of the sets $A, B\}$
Intersection: AB	$\{x : x \in A$ and also $x \in B\}$
Complement: $\bar{A}$	If S is the whole space, the complement of A (with respect to S) is $\bar{A} = \{x : x \in S$ and $x \notin A\}$
Empty set: 0	denotes the set which has no members

Disjoint sets: A and B are disjoint if $AB = 0$

3. The following properties hold

Commutative laws	:	$A \cup B = B \cup A$, $AB = BA$
Associative laws	:	$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$
		$(AB)C = A(BC) = ABC$
Distributive laws	:	$A(B \cup C) = AB \cup AC$, $A \cup (BC) = (A \cup B)(A \cup C)$
Idempotence	:	$A \cup A = AA = A$
Zero and unit	:	$A \cup 0 = A$, $A0 = 0$, $AS = A$ whenever $S \supset A$
Complementation	:	$\overline{A \cup B} = \bar{A}\,\bar{B}$, $\bar{A} \cup \bar{B} = \overline{AB}$

4. Logical dictionary.

Using the representation of (1), we have the following correspondence

the set A ... the event A	Complement $\bar{A}$ ... Negation, not - A
$A \cup B$ ... disjunction, A or B	AB ... conjunction, both A & B
$A \subset B$ ... implication, A implies B	AB = 0 ... A,B mutually exclusive

(2) AXIOMS OF PROBABILITY (Kolmogorov)

We have a basic set E (corresponding to the observation space) whose members are the simple events. $\mathfrak{F}$ is a set of subsets of E. Then

1. $\mathfrak{F}$ is a field of sets
2. $\mathfrak{F} \supset E$
3. To each set A of $\mathfrak{F}$ is assigned a non-negative real number P(A), the prob. of A.
4. P(E) = 1
5. If AB=0, $P(A \cup B) = P(A) + P(B)$
6. If $A_1 \supset A_2 \supset A_3 \ldots \supset A_n \ldots$ and $A_1 A_2 A_3 \ldots = 0$ then
$$\lim_{n \to \infty} P(A_n) = 0$$
(where "complete additivity" if the A_i are disjoint, $P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$)

2. Basic probability laws.

(i) $P(0)=0$, $0 \leq P(A) \leq 1$, (ii) $P(\bar{A}) = 1 - P(A)$,
(ii) $P(A \cup B) = P(A) + P(B) - P(AB)$
(iv) $P(A \cup B \cup C) = \sum P(A) - P(AB) + P(ABC)$; (v) $P(\cup A_i) \leq \sum P(A_i)$
(vi) If $A \subset B$, $P(A) \leq P(B)$.

3. Independence of experiments

An experiment A_r corresponds to a decomposition of the observation space E into disjoint subsets $A_{r1}, A_{r2}, \ldots A_{rK_r}$. let r = 1,2, ... n. The decompositions are mutually independent provided

$$P(A_{r_1 1}\ A_{r_2 2} \dots A_{r_n n}) = P(A_{r_1 1})\, P(A_{r_2 2}) \dots P(A_{r_n n})$$

for any $r_1, r_2, \dots r_n$

If $A_1, \dots A_n$ are mutually independent, then any m of them ($m < n$) are also independent

4. Independence of events

The n events $A_1, A_2, \dots A_n$ are mutually independent if the decompositions $E = A_k \cup \bar{A}_k$, ($k=1,2,\dots n$) are independent. Hence the N & S conditions for the mutual independence of the events $A_1, A_2, \dots A_n$ are the following $2^n - n - 1$ relations

$$P(A_{r_1}\ A_{r_2} \dots A_{r_m}) = P(A_{r_1}) \dots P(A_{r_m}),\ m = 1, 2, \dots n,$$

$$1 \leq r_1 < r_2 < \dots < r_m \leq n.$$

Note: the independence of events in pairs does not necessarily imply their mutual independence, ie we can have $P(AB) = P(A).P(B), P(BC) = P(B).P(C)$, $P(AC) = P(A).P(C)$, but $P(ABC) \neq P(A).\ P(B)\ P(C)$.
In particular, A and B are independent if and olny if $P(AB) = P(A).\ P(B)$

5. Conditional probability

DEF. $\quad P(B|A) = P(AB)/P(A)$

where $\quad P(AB) = P(A).P(B|A) = P(B).P(A|B)$

Conditional probabilities behave like probabilities :ie

$P(B|A) \geq 0$, $P(E|A) = 1$, $P(B \cup C|A) = P(B|A) + P(C|A)$ provided $BC = 0$

$$P(A|A) = 1$$

If and only if A and B are independent, $P(A|B) = P(A)$, $P(B|A) = P(B)$.

6. Random variables, or variates : A random variable is a single valued real function X (u) defined for all $u \in E$ (where E is the observation space) and for which $\{u: X(u) \leq a\} \subset \mathcal{F}$ for all real a.

We often write X for X(u).

(3) DISTRIBUTION FUNCTIONS

1. <u>Univariate case</u>. Take E to be the real axis, $\mathcal{F}$ the aggregate of all countable unions and intersection of subsets of E; then the non-negative completely additive set function P(A) may be defined by its values for the special intervals $(-\infty, x)$

$$P(-\infty, x) = P(X \leq x) = F(x) \quad = \text{the (cumulative) distrib. fn.,}$$

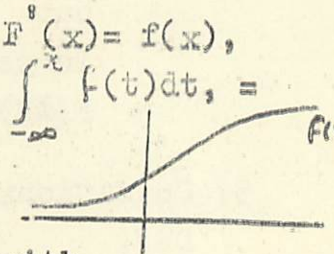
where $F(-\infty) = 0$, $F(+\infty) = 1$, F(x) is a bounded monot. non-decr. fn. since $P(a < X \leq b) = F(b) - F(a) \geq 0$ if $b > a$

At discontinuities we define $F(x) = F(x+0)$

Clearly $P(X=x) = F(x) - F(x-0)$

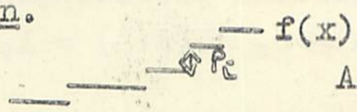
1a. <u>Continuous type</u>: F(x) differentiable, $F'(x) = f(x)$, the <u>prob</u>.(density) <u>fn</u>. $f(x) \geq 0$, $F(x) = \int_{-\infty}^{x} f(t)dt$, $= Pr(X \leq x)$

$Pr(X \in dx) = Pr(x < X \leq x+dx) = f(x)dx$



1b. <u>Discrete type</u>: F(x) a step-function, with jumps of magnitudes p_i at x_i $(i=1,2,\ldots)$, $\sum p_i = 1$

$Pr(X=x_i) = p_i$, the <u>prob</u>. <u>fn</u>.



$$F(x) = \sum_{i(x)} p_i$$

where $i(x) = \{i : x_i \leq x\}$

Alternative notation
$p_i = f(x_i)$

2. Bivariate case.

where $F(x,y) \geq 0$, $F(-\infty,-\infty) = F(-\infty,y) = F(x,-\infty) = F(-\infty,+\infty) = F(\infty,-\infty) = 0$

$F(\infty,\infty) = 1$

$F_{22} - F_{12} - F_{21} + F_{11} \geq 0$ where $F_{ij} = F(x_i, y_j)$ and $x_2 > x_1, y_2 > y_1$

$$\boxed{p(x_1 < X \leq x_2,\ y_1 < Y \leq y_2) = F_{22} - F_{12} - F_{21} + F_{11}}$$

F is monotonic non-decreasing in each variable separately.
At discontinuities $F(x,y) = F(x+0,y) = F(x,y+0)$

2a. Continuous type: $\partial^2 F/\partial x \partial y = f(x,y)$ = Prob. denst. fun. of X, Y

$P(X \in dx, Y \in dy) = f(x,y)\,dx\,dy$, $F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\,du\,d$

2b Discrete type: There is an enumerable set of points (x_r, y_s) and positive numbers p_{rs}; $\sum_{r,s} p_{rs} = 1$; s.t. $F(x,y) = \sum_{rs(xy)} p_{r}$

where $r\,s(xy) = \{ r,s : x_r \leq x,\ y_s \leq y \}$ Then $P(X = x_r, Y = y_s) = p_{rs}$

3. Marginal distributions: Let (X,Y) have the d.f. $F(x,y)$, as in

The marginal d.f. of X is $F_1(x) = F(x,\infty) = P(X \leq x)$, $(= P(X \leq x, Y \leq \infty))$

of Y $\quad F_2(y) = F(\infty,y) = P(Y \leq y)$

Then $F_1(x)$, $F_2(y)$ are univariate d.f.'s.

3a. Continuous type: If $F(x,y)$ is continuous, we define the marginal prob. fn. of X to be $f_1(x) = F_1'(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$

of Y $\quad f_2(y) = F_2'(y)$

3b. Discrete type: If (X,Y) has discrete pr.fn. p_{rs}, the marginal prob. fn. of X is $p_1(r) = \sum_{s} p_{rs}$; of Y, $p_2(S) = \sum_{r}$

$$f_1(x_r) = \sum_{s} f(x_r, y_s)$$

4. Conditional distribution functions From (2), §5, we have

$P(X \leq x \mid Y \leq y) = P(X \leq x, Y \leq y) \;/\; P(Y \leq y)$

$= F(x, y) \;/\; F_2(y)$ by §3

We define $P(X \leq x \mid Y = y)$ to be $\lim_{h \to 0} P(X \leq x \mid y < Y \leq y+h)$

$= \frac{\partial F}{\partial y} \;/\; f_2(y)$ in the continuous case. The conditional prob. fn. of X, given Y=y, is then defined to be $f_{1|2}(x|y)$,

$$f_{1|2}(x|y)\,dx = p(x < X \leq x+dx \mid Y=y)$$

whence

$$f_{1|2}(x|y) = \frac{f(x,y)}{f_2(y)} \quad \text{where} \quad f(x,y) = \frac{\partial^2 F(x,y)}{\partial x\, \partial y}, \quad f_2(y) = \frac{d}{dy}F(\infty, y)$$

(4) INTEGRATION

We write $\int \phi(x)\, dF(x)$ to denote the Steiltjes integral of $\phi(x)$ with respect to $F(x)$.

In the continuous case, $\int \phi(x)\, dF(x) = \int \phi(x) f(x)\,dx$, the ordinary Riemann $\int$ ($f(x) = F'(x)$)

In the discrete case where $F(x)$ is a step function with jumps $f(x_i)$ at x_i, $\int \phi(x)\,dF(x) = \sum \phi(x_i) \cdot f(x_i)$

the R. integral, or sum, being taken over the appropriate range.

Note that if $F(x)$ is the d.f. of the variate X, then

$$\boxed{P(X \in A) = \int_A dF(x)}$$

(5) EXPECTATION

1. If X has d.f. $F(x)$, the expectation of any function of $\psi(X)$ of X is

$$\boxed{\mathcal{E}\psi(X) = \int \psi(x)\, dF(x)}$$ (integral taken over all possible values of X)

Similerly for bivariate distributions, using the natural generalization of § (4)

$$\mathcal{E}\psi(X,Y) = \int \psi(x,y)\, dF(x,y) \quad = \iint \psi(x,y)\, f(x,y)\, dxdy$$
$$= \sum\sum \psi(x_r,y_s)\, p_{rs}$$

Additivity: $\mathcal{E}(X+Y) = \int (x+y)\, dF(x,y)$
$= \int x dF(x,y) + \int y\, dF(x,y) = \mathcal{E}X + \mathcal{E}Y$

We then have : If $X \geqslant 0$, $\mathcal{E}X \geqslant 0$,

$\mathcal{E}(aX+b) = a\,\mathcal{E}X + b$

$\mathcal{E}(X+Y) = \mathcal{E}X + \mathcal{E}Y$

$\mathcal{E}(\ )$ is therefore linear operation

In particular

$$\mathcal{E}(\sum \lambda_i X_i) = \sum \lambda_i \mathcal{E}X_i$$

2. __Moments__ If X has d.f. F(x) the r^{th} moment (about the origin) is $\mu'_r = \mathcal{E}X^r = \int x^r\, dF(x)$

the r^{th} central moment is

$$\mu_r = \mathcal{E}(X-\mu)^r = \int (x-\mu)^r dF(x), \text{ where } \mu \equiv \mu'_1$$

Conditional moments are moments of the appropriate conditional distribution

3. __Variance__ The dispersion of a distribution may be measured by the "standard deviation", whose square is the variance, defined by (var X=) $V(X) = \mathcal{E}(X-\mu)^2 = \mathcal{E}X^2 - \mu^2$ where $\mu = \mathcal{E}(X)$

then we have: $V(X) \geq 0$: in fact $V(X) > 0$ unless $X \equiv$ const.

$V(aX+b) = a^2\, V(X)$

4. __Covariance:__ $\mathcal{C}(X,Y) = \mathcal{E}(XY) - (\mathcal{E}X)(\mathcal{E}Y) = \mathcal{E}(X-\mathcal{E}X)(Y-\mathcal{E}Y) =$
$\mathcal{C}(Y,X) = \text{cov}(X,Y)$

whence

$\mathcal{C}(aX+b, cY+d) = ac\,\mathcal{C}(X,Y)$; $\mathcal{C}(X,a) = 0$,

$\mathcal{C}(X+U,\ Y+V) = \mathcal{C}(X,Y) + \mathcal{C}(U,Y) + \mathcal{C}(X,V) + \mathcal{C}(U,V)$

Hence

$$\boxed{V(aX+bY) = a^2V(X) + 2ab\,\mathcal{C}(X,Y) + b^2V(Y)}$$

If $v(X) = \sigma_1^{\ 2}$, $v(Y) = \sigma_2^{\ 2}$ we define the correlation coefficient as

$$\rho(X,Y) = \frac{\mathcal{C}(X,Y)}{\sigma_1 \sigma_2}$$

whence

$$-1 \leq \rho(X,Y) \leq +1 \qquad \rho(aX+b,\ cX+d) = \rho(X,Y)$$

X,Y are uncorrelated if $\mathcal{C}(X,Y) = 0$

5. Independent variates. X, Y are stochastecally independent if $F(x,y) = F_1(x)\,F_2(y)$, $f(x,y) = f_1(x)\,f_2(y)$

F_1, F_2 are then necessarily the marginal df's (up to a const. multiplier) and f_1, f_2 the marginal probability functions. We then have $\mathcal{E}(XY) = \mathcal{E}X\ \mathcal{E}Y$ as a consequence of the def. So that independence implies uncorrelation(but not conversely).

6. MARKOFF'S INEQUALITY

If $X \geq 0$, and $\mathcal{E}X = \mu$ is finite then for any $k > 0$,

$$P\{X \geqslant k\mu\} \leq 1/k.$$

Proof: Let $\left.\begin{array}{ll} Y = 0 & \text{when } X < k\mu \\ Y = k\mu & \quad ,, \quad X \geqslant k\mu \end{array}\right\}$ then $Y \leq X$ so $\mathcal{E}Y \leq \mu$

But $\mathcal{E}Y = 0 . P(X < k\mu) + k\mu . P(X \geqslant k\mu)$, where the theorem.

7. TCHEBYCHEFF'S INEQUALITY

For any variate X with $\mathcal{E}X = \mu$ and $vX = \sigma^2 \neq 0$,

$$\boxed{P\{\,|X - \mu| \geqslant k\sigma\} \leq \frac{1}{k^2}} \quad \text{for any } k > 0.$$

Proof: In Markoff's inequality replace X by $(X-\mu)/\sigma^2$

Example showing that the = signs are attainable: consider the discrete X for which $P(X=\mu) = 1-1/k^2$, $P(X=\mu \pm k\sigma) = 1/2k^2$

STANDARD DISTRIBUTIONS

The variate is X, (or X_1, X_2... in multivariate cases)

The complete specification of the probability functions listed below is

(in the range quoted, this probability function
(has the value quoted, outside this range the
(probability function is zero.

Name	Probability function	Range
Binomial	$\binom{n}{x} p^x q^{n-x}$, $o < p < 1$, $q = 1-p$,	o,1,2,...,n
Poisson	$e^{-\mu} \mu^x / x!$, $\mu > o$,	o,1,2,...
Hypergeometric	$\binom{a}{x}\binom{k-a}{n-x} \Big/ \binom{k}{n} = \binom{n}{x} \dfrac{a^{(x)} (k-a)^{(n-x)}}{k^{(n)}}$ $o < a < k$, $o < n < k$,	o,1,2,...,n
Neg.binomial	$\binom{k+x-1}{k-1} p^k q^x$, $k > o$, $o < p < 1$, $q = 1-p$	o,1,2,...
Multinomial	$\dfrac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}$, $o < p_i < 1$ $\sum p_i = 1$ $\sum x_i = n$	o,1,2,...n for each x_i
Rectangular	1	$(-\frac{1}{2}, +\frac{1}{2})$
Triangular	$1 - \lvert x \rvert$	(-1,+1)
Exponential	e^{-x}	$x \geqslant o$
Double exponential	$\frac{1}{2} e^{-\lvert x \rvert}$	all real values
Beta	$\dfrac{1}{B(p,q)} x^{p-1} (1-x)^{q-1}$, $p > o$, $q > o$	(o,1)

Standard normal $e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$,, ,, ,,

Bivariate normal $\dfrac{1}{2\pi\,\sigma_1\,\sigma_2\,\sqrt{1-\rho^2}}\; e^{-\frac{1}{2}Q/(1-\rho^2)}$,, ,, ,,

$\sigma_1 > 0$

$\sigma_2 > 0$

$\rho^2 < 1$

$$Q = \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2$$

Standard multinormal $(2\pi)^{-\frac{1}{2}k}\,|\underline{V}|^{-\frac{1}{2}}\,e^{-\frac{1}{2}\underline{x}'V^{-1}\underline{x}}$, all real values. for each x_i

$\underline{V}$ pos.def.

<u>Notes on the Standard distributions</u>

1. <u>Binomial</u> x= number of successes in n indep. trials, where probability of a success, at any trial, is p = const.

 $\mathcal{E}X = n\,p$, $\mathcal{V}X = n\,p\,q$, $\mu_3 = n\,p\,q\,(q-p)$,

 $F(x) = I_x\,(n-x, x+1)$, (incomplete beta function ratio)

 If X binomial (n,p) and Y binomial (m,p), indep.; X+Y binomial (n+m,p)

2. <u>Poisson</u> (a) x = no. of occurrences of a given event in time t, where probability of a single occurrency during $\delta t = \lambda\delta t + O(\delta t)$, probability of >1 occurrences during $\delta t = O(\delta t)$, and no. of occurrences during non-overlapping time intervals are indep. of each other.

 Then x is Poisson, with parameter $\mu = \lambda t$.

(b) $\varphi(x)$ = limit of binomial prob. fn. when $n \to \infty$,

$p \to 0$, $np = \mu$

If X_1, X_2 are indep. Poisson variates with parameters μ_1, μ_2, then $X_1 + X_2$ is Poisson $(\mu_1 + \mu_2)$

$\mathcal{E}X = \mu$, $\mathcal{V}X = \mu$; $F(x) = 1 - I_\mu(x+1)$, (incomplete gamma fn ratio)

3. <u>Hypergeometric</u> x = no. of A's in a sample of n, taken without replacement from a set of k item of which $\underline{\underline{a}}$ were A's

$$\mathcal{E}X = \frac{a}{k}n \qquad \mathcal{E}X^{(r)} = a^{(r)} \; n^{(r)}/k^{(r)}.$$

4. <u>Negative binomial</u> (a) k + X = no. of binomial trials (of prob. p) required to achieve k successes $\mathcal{E}X = kq/p$, $\mathcal{V}X = kq/p^2$

If we put $q' = 1/p$, $p' = q/p$, so that $q' - p' = 1$, the pr. fn. becomes coef. of p'^x in expansion of $(q' - p')^{-k}$, and in this terminology we have $\mathcal{E}X = kp'$, $\mathcal{V}X = kp'q'$.

(b) x may also be regarded as a Poisson variate with varying Poisson parameter. In fact if $P(x,m) = m^x e^{-m}/x!$ while $P(m) = \alpha^\lambda m^{\lambda-1} e^{-\alpha m}/\Gamma(\lambda)$ $m > o$ (a gamma distribution) then $P(x) = \left(\frac{\alpha}{1+\alpha}\right)^\lambda \frac{\Gamma(\lambda+x)}{x!\,\Gamma(\lambda)}(1+\alpha)^{-x}$ which is of one standard neg. binomial form with $p = \alpha/(1+\alpha)$, $q = 1/(1+\alpha)$.

5. <u>Multinomial</u> x_i = no. of occurrences of event A_i $(i=1,2,\ldots n)$ in k indep. trials, where at each trial $p(A_i) = p_i$ = const.
$\mathcal{E} X_i = np_i$, $\mathcal{V} X_i = np_i\,(1-p_i)$, $\mathcal{C}(X_i,X_j) = -n\,p_i\,p_j$

If $X_1, X_2, \ldots X_k$ are indep. Poisson variates with parameters $\mu_1, \ldots \mu_k$, then the conditional joint probability function of the X_i, given $\sum X_i = x$, is multinomial:

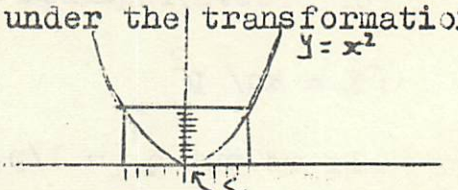
$$P(x_1,\ldots x_k \mid \textstyle\sum x_i = x) = \frac{x!}{\prod x_i!} \prod \left(\frac{\mu_i}{\mu}\right)^{x_i}, \quad (\mu = \textstyle\sum \mu_i)$$

<u>Transformation of variate</u> (Univariate case)

1. Given a variate X, d.f. $F(x)$, pr.fn. $f(x)$, and a single valued function $\theta(x)$ to find the distribution $G(y)$, $g(y)$ of

$$Y = \theta(X)$$

Let S_x be the set of all values of x which are mapped into a specified set S_y of values of y, under the transformation $y = \theta(x)$.



then $p(Y \in S_y) = p(X \in S_x)$

(eg $Y=X^2$. take $S_y = (0,y)$. Then $S_x = (-\sqrt{y}, +\sqrt{y})$
$G(y) = p(o < Y \leq y) = p(-\sqrt{y} < X \leq +\sqrt{y})$
$= F(\sqrt{y}) - F(-\sqrt{y})$)

2. Special case where the transformation is continuous, 1-1.
$G(y) = F(\theta^{-1}(y))$ if f(x) is monotonic increasing
$= 1 - F(\theta^{-1}(y))$ " " " decreasing