THE EFFECT OF GENOTYPE X ENVIRONMENT INTERACTION ON THE ESTIMATE OF GENETIC VARIANCES. A COMPUTER SIMULATION STUDY

Djoko Prajitno *)

RINGKASAN

Pengaruh interaksi antara faktor-faktor genotipe x lingkungan terhadap estimasi varians genetik dipelajari melalui simulasi dengan menggunakan komputer IBM 370 model 135. Program komputer yang ditulis dalam bahasa FORTRAN IV dibuat untuk sistem perkawinan design I dari Comstock dan Robinson (1948). Teknik simulasi dijalankan mengikuti langkah-langkah sebagai berikut :

1. Spesifikasi dalil-dalil untuk penyusunan data phenotipe yang terdiri atas tiga komponen dasar :
   a. Spesifikasi nilai genotipik yang mengikuti dalil-dalil genetika.
   b. Spesifikasi pengaruh faktor-faktor lingkungan.
   c. Spesifikasi pengaruh interaksi antara faktor-faktor genotipe x lingkungan.
2. Estimasi varians genetik.

Hasil penelitian menunjukkan bahwa interaksi antara faktor-faktor genotipe x lingkungan hanyalah dipengaruhi oleh faktor lingkungannya saja. Model-model genetik serta jumlah dari gen tidak menunjukkan pengaruh yang nyata.

Simulasi interaksi genotipe x lingkungan terlihat tidak cukup berpengaruh terhadap estimasi varians genetik yang bersifat aditif. Namun demikian model genetik dan jumlah gen rupa-rupanya menunjukkan pengaruh yang nyata, yaitu : (1) timbulnya gejala bias yang bersifat lebih tinggi akibat pengaruh epistasis dan (2) varians genetik aditif pada model 10 lokus ternyata lebih besar dari model 2 lokus.

Variasi yang besar yang terlihat pada estimasi varians genetik yang bersifat dominan lebih membuktikan bahwa Design I bukan merupakan alat yang baik untuk estimasi varians genetik dominan.

Interaksi antar faktor-faktor genotipe x lingkungan terlihat jelas mengurangi heritabilitas individuil. Heritabilitas individuil antar model-model genetik maupun jumlah gen tidak menunjukkan perbedaan yang nyata.

UNIVERSITAS GADJAHMADA MILIK FAKULTAS PERTANIAN PERPUSTAKAAN YOGYAKARTA

INTRODUCTION

One of the way where the genotype x environment interaction influences the selection procedure is through its effect on the estimate of genetic variances i.e. a large interaction will result to a corresponding reduction in the variances among genotypes. Consequently the effectiveness of most selection procedures will also be reduced.

The measurement of genotype x environment interaction is not easy. While field trials of different genotype over environments have been the usual procedure, it is a very tedious, time consuming and very expensive method. Many genotypes have to be tested in many seasons and locations.

Computer has provided an alternative procedure to the field test. Through computer simulation, the tedious task of actual field trials can be avoided. In this study the use of computer simulation technique is proposed to define the significance of genotype x environment interaction on the estimate of genetic variances.

*) Department of Agronomy, Faculty of Agriculture, Gadjah Mada University, Yogyakarta, Indonesia.

Scanned with CamScanner

MATERIALS AND METHODS

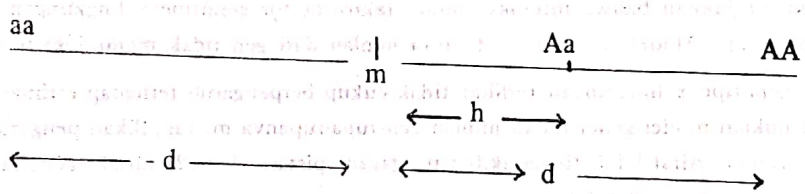
Simulation is a technique for conducting experiments on a model that approximates the important features of the actual system being studied. The simulation procedure used in this study is carried out through the following :

1. Specification of the phenotype. These rules consist of three major components. First, is specification of the genotypic value which essentially follows the accepted genetic theory. Second, is specification of environmental effect and third, is interaction of genotype by environment. These components would be discussed in detail in the succeeding section.
2. Estimation of variances. After (1) was specified, the components of the phenotypic variance were estimated.

The Genetic Components

For simplicity, only two alleles per locus were assumed, that each locus contributed equally to genotypic value, and that the only non-zero epistasis was the first order interaction among loci.

Additive and Dominance Components. The genotypic value of each individual was generated with the use of the following parameters. Consider a locus with (m) is the mid-point between homozygous genotypes, (d) is the deviation of homozygous genotype from (m), and (h) is the deviation of the heterozygous from the mid-point. The model is illustrated below :



The value of (h) depends on the degree of dominance such that h = 0, if there is no dominance, and non-zero otherwise. If dominance is complete, (h) = (d). Thus the degree of dominance is (h/d).

The genotype is expressed as an array of binary numbers in which the dominant gene is represented by (0) and the recessive gene represented by (1). Each two consecutive binary numbers represent one locus. Thus the genotype aABBcCDDeeffGGHhIIjj is represented as 10001000111100010011. The above example represents a quantitative character which is controlled by 10 pairs of genes (10 loci model). Note that the procedure can take care of any number of genes.

If the above example is specified, the values of m, d and h, then for N loci model the genotypic value can be determined as :

$$G = \bar{m} + \sum_{i=1}^{0} k_i d_i + \sum_{j=1}^{E} t_j h_j \qquad (1)$$

where:
G is the genotypic value
0 is the number of homozygous loci
E is the number of heterozygous loci, where:
0 + E = N is the number of loci affecting the character.
d_i and h_j are as defined before.
k_i = 1 for homozygous dominant locus
= −1 for homozygous recessive locus
= 0 for heterozygous locus.
t_j = 1 for heterozygous locus
= 0 for homozygous locus.

$$\bar{m} = \left(\sum_{i=1}^{N} m_i \right) / N$$

Scanned with CamScanner

Non-Allelic Interactions (Epistasis). The model allows for the first order non-allelic interactions. Consider two different genes each with two alleles A-a and B-b, nine genotypes are possible in diploid organism and eight parameters must be used to give a complete description of the differences among phenotypes. Four of these are the d's and h's appropriate to the two genes i.e. d_a, d_b, h_a and h_b. The other four may then be derived conveniently to correspond to the interaction between non-allelic genes. The parameters fall into three classes i.e. :

(1) i_{ab} is the homozygote x homozygote interaction

(2) j_{ab} and j_{ba} are the homozygote x heterozygote interaction, and

(3) l_{ab} is the heterozygote x heterozygote interaction.

Thus the components which compose the genotype AABB is $d_a + d_b + i_{ab}$; AABb is $d_a + h_b + j_{ab}$ and so on. The expression of the phenotypic classes could be done by using an assignment as follows:

Complete dominance : $d_a = h_a$; $d_b = h_b$; $i_{ab} = j_{ab} = j_{ba} = l_{ab}$

Complementary genes : $d_a = d_b = h_a = h_b = i_{ab} = j_{ab} = j_{ba} = l_{ab}$

Recessive epistasis : $d_a \neq d_b$; $d_a = h_a = i_{ab} = j_{ba}$;

$d_b = h_b = j_{ab} = l_{ab}$

The above classification of interaction can be extended to cover interactions between three or more genes. However, in polygenic systems, high order interaction can be expected to become less and less importance. In our model the non-allelic interactions is restricted to the all possible first order interactions (digenic interaction) only. The genetic effects which were tested in this experiment are shown in table 1.

Table 1. List of genetic parameters to be tested in simulation study

Gene number N	Genetic effect				
	d	h	i	j	l
Additive, 2 loci: 4	5	0	0	0	0
Additive, complete dominant, 2 loci: 4	5	5	0	0	0
Additive, 60% dominant, 20% epistasis, 2 loci: 4	5	3	1	1	1
Additive, 10 loci: 20	5	0	0	0	0
Additive, complete dominant, 10 loci: 20	5	5	0	0	0
Additive, 60% dominant, 20% epistasis, 10 loci: 20	5	3	1	1	1

Scanned with CamScanner

Thus the model in equation (2) can be rewritten as :

$$Y_{ijk} = u + G_i + (1 + b_i) S_j + e_{ijk} \quad (4)$$

where e_{ijk} is the pooling error, an accumulation of D_{ij} of equation (3) and e_{ijk} in equation (2).

Because Y_{ijk} follows the normal distribution, e_{ijk} is also normally distributed with mean equal to zero and variance $\sigma^2 e$. In this simulation study the error effect e_{ijk} is assigned by generating a normal independent distributed random number with mean = 0, and standard deviation = σe. The value of σe is assigned from an equation :

$$\sigma e = \text{C.V.} \times \bar{X} \quad (5)$$

where C.V. is coefficient of variation and $\bar{X}$ is the plot mean. In this study C.V. = 10% was used.

Since the genetic correlation among full-sib individuals was equal to one and half-sib was equal to 0.5, the b_i was generated from a uniform distributed random number with the range of $-1 \leqslant b \leqslant 1$ and restrictions in the following manner :

full-sib individuals ———————— same b_i
half-sib individuals ———————— b_i with the same sign
unrelated individuals ———————— no restriction.

The condition of the environmental effects to be tested are shown in table 2.

Variance Estimates

By simulating the mating design I of Comstock and Robinson (1948), the effect of genotype x environment interaction on the estimate of additive and dominance genetic variance were evaluated. One hundred random male individuals, each was mated to three randomly chosen female parents. Five progenies were generated from each mating. The simulated field test involved a single location with three environmental conditions. Figure 3 shows a flow chart of Design I simulation. The analysis of variance is as follows :

Source of Variation	df	E (MS)
Environment (S)	s − 1	$\sigma_e^2 + o\sigma_{SF}^2 + of\sigma_{SM}^2 + ofm\sigma_S^2$
Males (M)	m − 1	$\sigma_e^2 + o\sigma_{SF}^2 + of\sigma_{SM}^2 + s\sigma_O^2 + so\sigma_F^2 + sof\sigma_M^2$
Females (F)/M	(f−1)m	$\sigma_e^2 + o\sigma_{SF}^2 + s\sigma_O^2 + so\sigma_F^2$
Offspring (O)/F/M	(o−1)fm	$\sigma_e^2 + s\sigma_O^2$
S x M	(s−1)(m−1)	$\sigma_e^2 + o\sigma_{SF}^2 + of\sigma_{SM}^2$
S x F/M	(s−1)(f−1)m	$\sigma_e^2 + o\sigma_{SF}^2$
S x O/F/M	(s−1)(o−1)fm	σ_e^2
Total	sofm − 1	

Scanned with CamScanner

where :

σ^2_e = error variance

σ^2_F = variance due to difference among females mated to the same male.
= covariance full-sib – covariance half-sib.
= $\frac{1}{4}\sigma^2 A + \frac{1}{4}\sigma^2 D$

σ^2_M = variance due to differences among males
= covariance half-sib.
= $\frac{1}{4}\sigma^2 A$

σ^2_O = variance due to differences among offsprings from the same matings.

σ^2_{SF} = variance due to interaction between environment x females

σ^2_{SM} = variance due to interaction between environment x males

s is the number of environments = 3

m is the number of males = 100

f is the number of females in each male = 3

o is the number of offsprings in each mating = 5

The additive and dominance variance were estimated through the following formulas :

Additive genetic variance : $\sigma^2_A = 4\ \sigma^2_M$

Dominance genetic variance : $\sigma^2_D = 4\ (\sigma^2_F - \sigma^2_M)$

The actual additive and dominance genetic variance were calculated through the formulas :

$$\sigma^2_A = \sum_{i} 2p_i q_i [d + (q_i - p_i)\ h]^2 \qquad (6)$$

and

$$\sigma^2_D = \sum_{i} (2p_i q_i h)^2 \qquad (7)$$

where p_i and q_i are the frequency of dominant and recessive genes for each locus in the whole population, i.e. both were 0.5 in our simulation study. N, d and h were as mentioned before.

Table 2. List of environmental parameters to be tested in simulation study

Case No.	Seasonal effect		
	Wet	Dry	Summer
1	0	0	0
2	−5	0	5
3	2.5	2.5	−5
4	10	−10	0

Scanned with CamScanner

RESULTS AND DISCUSSION

From the definition of the phenotypic value given in equation :

$$Y_{ijk} = u + G_i + (1 + b_i)S_j + e_{ijk}$$

the magnitude of the genotype x environment (G x E) interaction is given by the product of b_iS_j. Thus the value of G x E interaction is zero when S_j is zero and is expected to be larger as the variability among the S_j becomes larger. Since variance among S_j is the indication of the difference among environment then it can be expected that the G x E interaction will be large when the environmental variation is large.

Other factor affecting the magnitude of G x E interaction is the b_i. In the same manner as the S_j the larger the variation in b_i the larger the G x E interaction. For this expectation, it was assumed b_i to have a uniform distribution range from -1 to 1. This assumption was based on the previous studies that most stability coefficient of yields in soybean varieties fall within the range of 0 to 2 (Buajarern, 1978). This means that the range of $(1 + b_i)$ is 0 to 2 or $-1 \leqslant b_i \leqslant 1$. Thus based on the above assumption, the maximum value of b_iS_j was equal to S_j. This means that the maximum value of G x E interaction effect was equal to its environmental effect.

Another factor influencing the phenotypic value are the genetic model and the number of genes.

Estimation of Genotype x Environment Interaction Variances

In this section, the magnitude of G x E interaction variances was described as affected by different genetic model, number of genes and environmental effect. The G x E interaction variances were expressed through the estimate of σ^2SF and σ^2SM as shown in table 3.

When there is no environmental effect (case 1) it seems that most of σ^2SF and σ^2SM are not significant. Otherwise, a significant σ^2SF and σ^2SM occured in cases 2, 3 and 4. This was in agreement with our expectation, that when environmental variation is large, the G x E interaction variances is also large, otherwise when environmental effect is zero, the G x E interaction will also be zero. A deep discussion for some discrepancies in this case, was given by Prajitno (1979).

Using F value 1.5 as an approximation of standard variance ratio for comparing the estimate of G x E interaction variances under different environmental and genetic model, it seems that when there is an environmental effect (case 2, 3 or 4) the G x E interaction is higher than without environmental effect (case 1). Case 2 and case 3 have a similar G x E interaction, while case 4 has the highest effect.

There are no differences for G x E interaction variances among genetic model, and among number of loci.

Scanned with CamScanner

Table 3. Estimate of G x E interaction variances σ^2SM and σ^2SF using simulation of Design I mating system.

Genetic model	Case 1	Case 2	Case 3	Case 4
Estimate of σ^2SM				
Two loci :				
Additive	-0.48^{ns}	4.91**	4.03**	14.72**
Dominant	0.82*	5.11**	2.75**	17.96**
Epistasis	-0.19^{ns}	3.38**	3.54**	17.08**
Ten loci :				
Additive	0.31^{ns}	3.64**	3.23**	17.72**
Dominant	0.10^{ns}	2.24^{ns}	5.80**	16.65**
Epistasis	1.25^{ns}	6.49^{ns}	7.97*	17.10**
Estimate of σ^2SF				
Two loci :				
Additive	2.23**	2.11**	1.87**	8.44**
Dominant	1.87*	2.76**	3.79**	7.52**
Epistasis	2.53**	3.24**	3.69**	7.88**
Ten loci :				
Additive	-1.89^{ns}	2.25^{ns}	3.05*	2.46*
Dominant	-2.93^{ns}	7.69*	-0.08^{ns}	9.58**
Epistasis	5.94^{ns}	-1.09^{ns}	-4.61^{ns}	12.53*

ns : not significant
* : significant at 5% level
** : significant at 1% level

Estimation of Genetic Variances

This section will describe the role of G x E interaction on the estimation of genetic variances. In all of the discussion, different magnitude of G x E interaction will be labelled in term of cases 1, 2, 3 and 4 as described before.

Additive Genetic Variance (σ^2A)

Shown in table 4 was the effect of G x E interaction on additive genetic variance. Using F = 1.75 as an approximation standard variance ratio for comparing the additive genetic variance, it was shown that the additive genetic variance between cases 1, 2, 3, and 4 were similar. This means that the G x E interaction did not affect the estimate of additive genetic variance.

Scanned with CamScanner

Table 4. **The effect of genotype x environment interaction on the estimate of additive genetic variance ($\sigma^2 A$)**

Genetic model	Direct calculation	Case 1	Case 2	Case 3	Case 4
Two loci :					
Additive	25.00	46.04	42.31	42.37	59.63
Dominant	25.00	48.68	39.20	46.31	34.75
Epistasis	25.00	79.95	77.55	77.80	71.98
Ten loci :					
Additive	125.00	107.02	122.70	99.82	106.92
Dominant	125.00	119.74	127.19	126.93	124.75
Epistasis	125.00	356.49	413.99	373.63	472.65

UNIVERSITAS MILI... FAKULT... PERTANIA... PERPUSTAKAAN ...OGJAKART...

The difference was observed between $\sigma^2 A$ obtained from the direct calculation using formula (6) and $\sigma^2 A$ obtained from simulation of Design I (cases 1, 2, 3 and 4). In two loci model it was clear that $\sigma^2 A$ obtained from Design I was higher than $\sigma^2 A$ from direct calculation. In ten loci model, the results are quite similar.

In direct calculation, the environmental effect was not considered. Direct calculation is more theoritical than estimation through Design I. Thus it is true if different results from these twomethods were obtained. As Empig et.al. (1972) stated that estimation through direct calculation could be used only as a crude guide in choosing intra selection scheme.

Between genetic model, it was shown that $\sigma^2 A$ under epistasis was higher than the other two genetic model.

One of the assumption involved in deriving the genetic interpretation of variance components in using mating design is no epistasis, i.e. the effect of variation in genotype at any single locus is not modified by genes at other locus. Comstock and Robinson (1952) pointed out that epistasis probably caused upward bias in the estimate of genetic variances, but the amount of bias might not be large, about 0.1 to 0.25. However they emphasized also, that the matter had not been considered exhaustively and the possibility remains that in some materials epistasis would be responsible for serious over-estimation This was proved in this simulation study. Epistasis increased the variation of genotypic value through interaction among loci. This increment caused a high $\sigma^2 A$.

All of $\sigma^2 A$ under ten loci model were higher than $\sigma^2 A$ in two loci model. This was true in this simulation study since ten loci gave a large genotypic effect.

Dominance Genetic Variance ($\sigma^2 D$)

Table 5 showed the effect of G x E interaction on the estimate of dominance genetic variance. It was shown that most of the estimate of $\sigma^2 D$ in two loci model was negative while in ten loci was positive.

Scanned with CamScanner

It was assumed from these models that in two loci the genotypic effect was small while in ten loci the genotypic effect was high. Suppose an equal error affected both two and ten loci models, it would be apparent that the two loci would be much influenced by the error comparing to the ten loci model. Thus a particular error that affected the two loci model to become negative probably would not bring a negative result in ten loci model. The other factor is, $\sigma^2 D$ is a function of $\sigma^2 M$ and $\sigma^2 F$, and it was obtained from the relation : $\sigma^2 D = 4 (\sigma^2 F - \sigma^2 M)$. Thus as far as the variation among male parents ($\sigma^2 M$) was higher than variation among females mated the same male ($\sigma^2 F$), the result would be a negative $\sigma^2 D$. It should be noted here a statement from Comstock and Robinson (1952) that the Design I did not provide a good estimate of dominance genetic variance.

Using F value 1.75 as standard variance ratio for comparing the dominance genetic variance under different environmental effect, it was shown that in average, only $\sigma^2 D$ under case 4 was smaller than the other. There were no differences between cases 1, 2 and 3. It was known that case 4 gave the highest G x E interaction variances. Assuming the same phenotypic variance, the higher G x E interaction variances, the lower the genetic variances including $\sigma^2 D$. However the G x E interaction in cases 2 and 3 did not strong enough in reducing $\sigma^2 D$.

$\sigma^2 D$ obtained from direct calculation seemed very different to $\sigma^2 D$ obtained from the simulation. The reason was similar to the same case as shown in estimation of $\sigma^2 A$.

In ten loci, the additive genetic model had the smallest $\sigma^2 D$, while epistasis genetic model had the highest. This was true since in additive genetic model no dominance deviation involved in constructing the genotypic value. Thus $\sigma^2 D$ was only a bias caused by environmental and error effects. However in two loci model, epistasis genetic model had the smallest $\sigma^2 D$. Probably this was due to the fact that two loci model was more affected by G x E interaction and error variations.

Table 5. The effect of genotype x environment interaction on the estimate of dominance genetic variance ($\sigma^2 D$)

Genetic model	Direct calculation	Case 1	Case 2	Case 3	Case 4
Two loci :					
Additive	0.00	−17.19	−14.38	−10.57	−32.57
Dominant	12.50	−10.79	1.16	−12.74	− 4.19
Epistasis	4.50	−38.23	−32.99	−35.98	−26.70
Ten loci :					
Additive	0.00	11.18	− 2.43	19.72	14.01
Dominant	62.50	78.96	68.11	83.47	67.12
Epistasis	22.50	356.16	278.56	393.92	161.82

Individual Heritability (H)

Based on Hanson (1963) paper on heritability, individual heritability was calculated in single plant basis as; $H = \sigma^2 A / \sigma^2 p$ where $\sigma^2 p = \sigma^2 e + \sigma^2 SF + \sigma^2 SM + \sigma^2 O + \sigma^2 F + \sigma^2 M + \sigma^2 S$.

Table 6 shows the effect of G x E interaction on the estimate of individual heritability H. The range of H is 0.20 to 0.95.

Scanned with CamScanner

Table 6. The effect of genotype x environment interaction on the estimate of individual heritability (H)

Genetic model	Case 1	Case 2	Case 3	Case 4
Two loci :				
Additive	0.77	0.49	0.52	0.34
Dominant	0.70	0.38	0.51	0.20
Epistasis	0.95	0.68	0.75	0.36
Ten loci :				
Additive	0.72	0.66	0.59	0.41
Dominant	0.50	0.45	0.48	0.33
Epistasis	0.53	0.59	0.54	0.56

Using F value 1.75 as an approximation standard variance ratio for comparing the individual heritability, it seemed that only individual heritability under case 4 was smaller than case 1. This was due to the fact that in this simulation study only case 4 gave higher G x E interaction variances than case 1. Since the G x E interaction variances were component of phenotypic variance, thus such high interaction caused high phenotypic variance. This means a reduction in heritability.

There were no differences among genetic model and number of loci.

In general, it seemed that genetic model did not affect the individual heritability. The G x E interaction reduced the individual heritability if the interaction was strong enough as occured in case 4.

ACKNOWLEDGEMENT

The author wishes to express his sincere appreciation and gratitude to Dr. Arturo A. Gomez professor in Agronomy, UPLB, for his invaluable advice, guidance and assistance in the execution and completion of this study.

His acknowledgement is also extended to the followings :

- Dr. Henry F. Magalit, professor in Statistics UPLB, for his guidance in computer programming and system analysis.
- Dr. N.T. Ison, former Director of the Agricultural Resource Center Inc. UPLB, for his approval to conduct the study using IBM system 370 model 135 computer facilities at ARC.

REFERENCES

Buajarern, W. (1978) Selection criteria for soybeans under varying cropping systems and environments. Ph. D. thesis UPLB.C.A. (unpublished).

Comstock, R.E. and H.F. Robinson. (1948) The components of genetic variance in a population of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4: 254–266.

——— and ———. (1952) Estimation of average dominance of genes. In J.W. Gowen (ed). Heterosis. Iowa State Coll. pp. 494–516.

Scanned with CamScanner

Empig, L.T., C.O. Gardner and W.A. Compton. (1972) Theoritical gains for different population improvement procedures. Univ. of Nebraska Coll. Agric. Agric. Exp. Sta. MP 26. 22p.

Hanson, W.D. (1963) Heritability. In W.D. Hanson and H.F. Robinson (eds.). Statistical genetics and plant breeding. NAS–NRC Pub. 982. Washington D.C.

Perkins, J.M. and J.L. Jinks. (1968) Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. *Heredity* 23: 339–356.

Prajitno, Djoko. (1979) The effect of genotype x environment interaction on the efficiency of different methods of selection. A Computer simulation study. M.S. Thesis. UPLB. C.A. 103p. (unpublished).

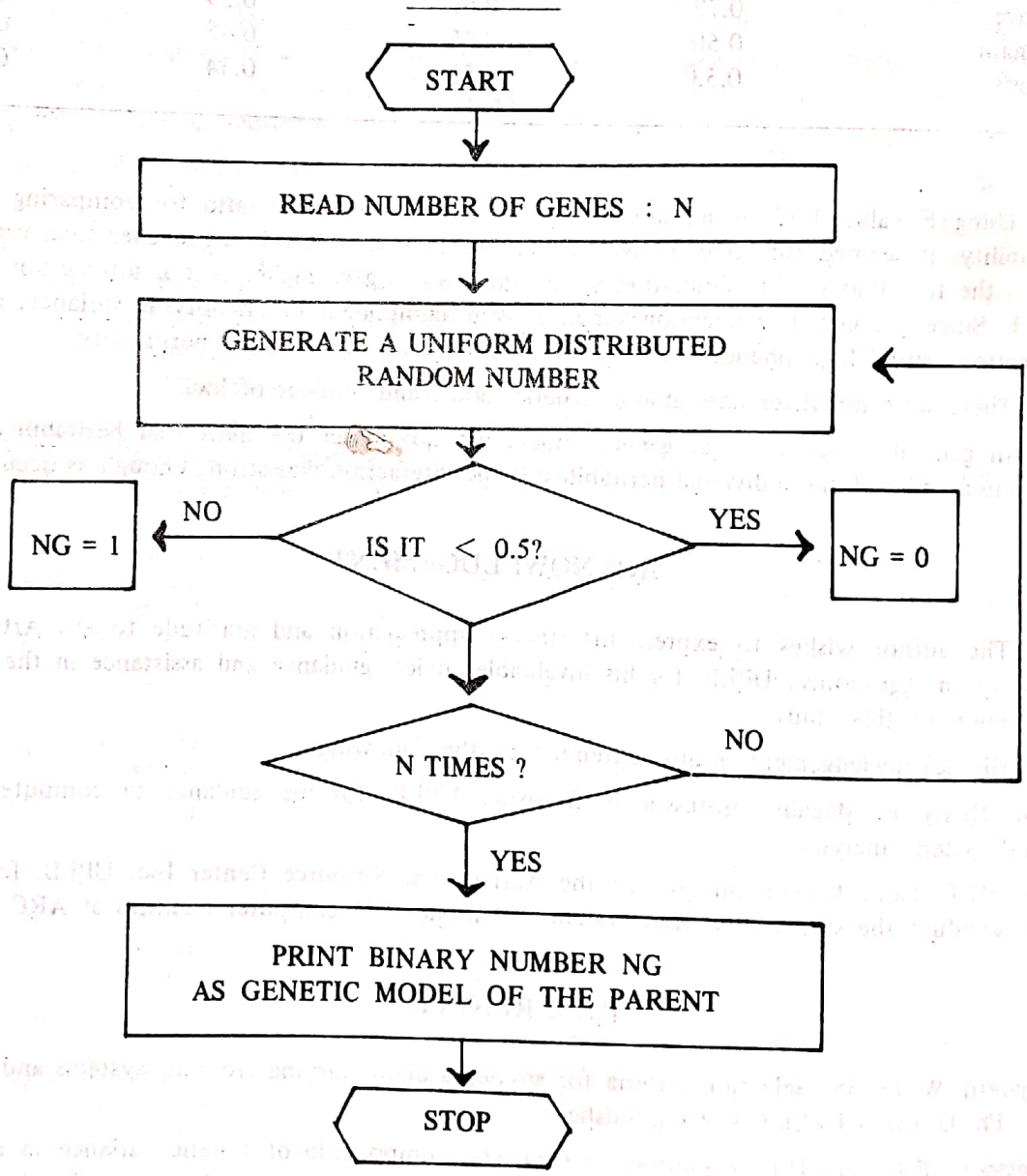


Figure 1. Flow chart of a subroutine for generating a random parent.

Scanned with CamScanner

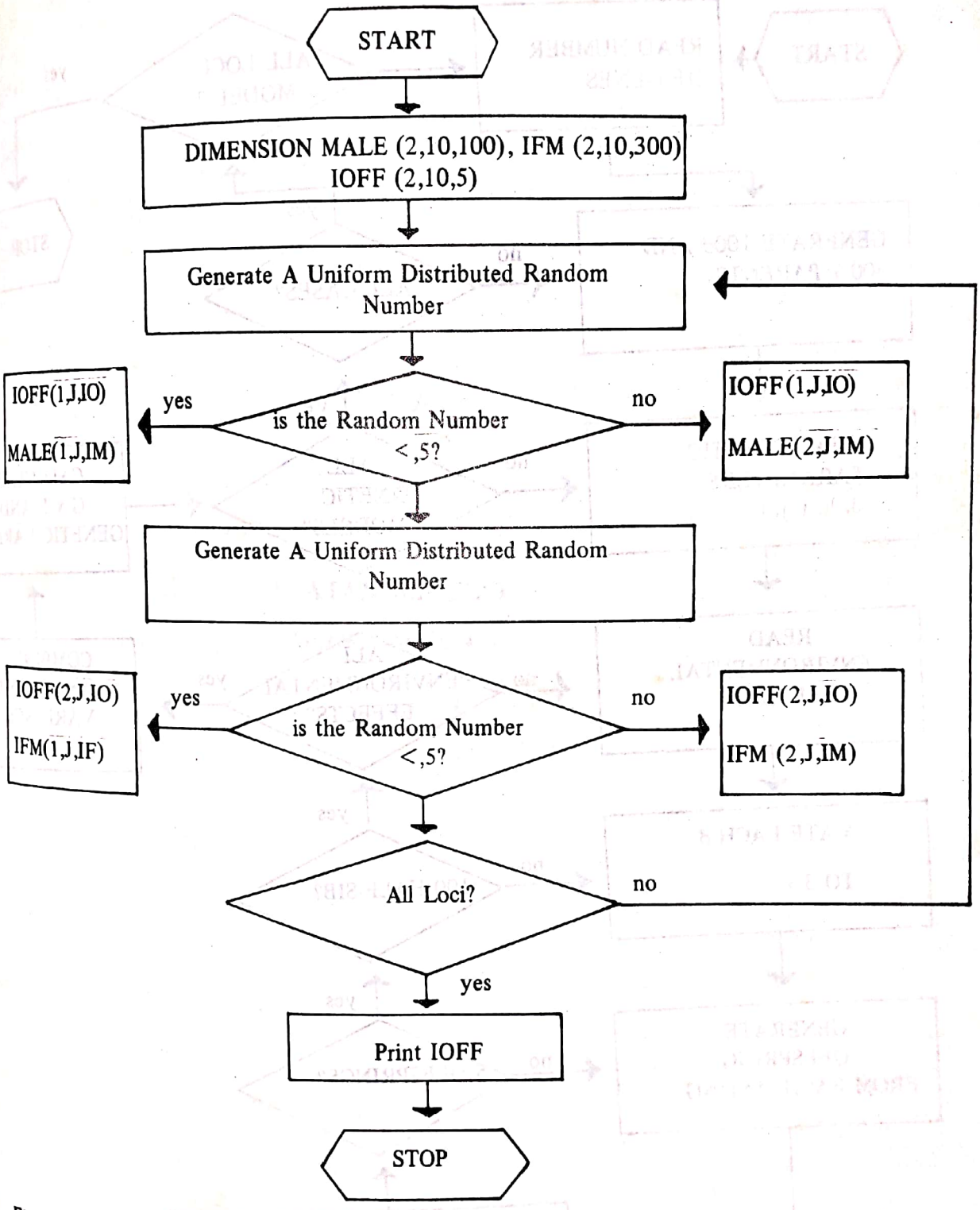


Figure 2. Flow chart of a subroutine for generating an offspring from mating between two parents.

Scanned with CamScanner

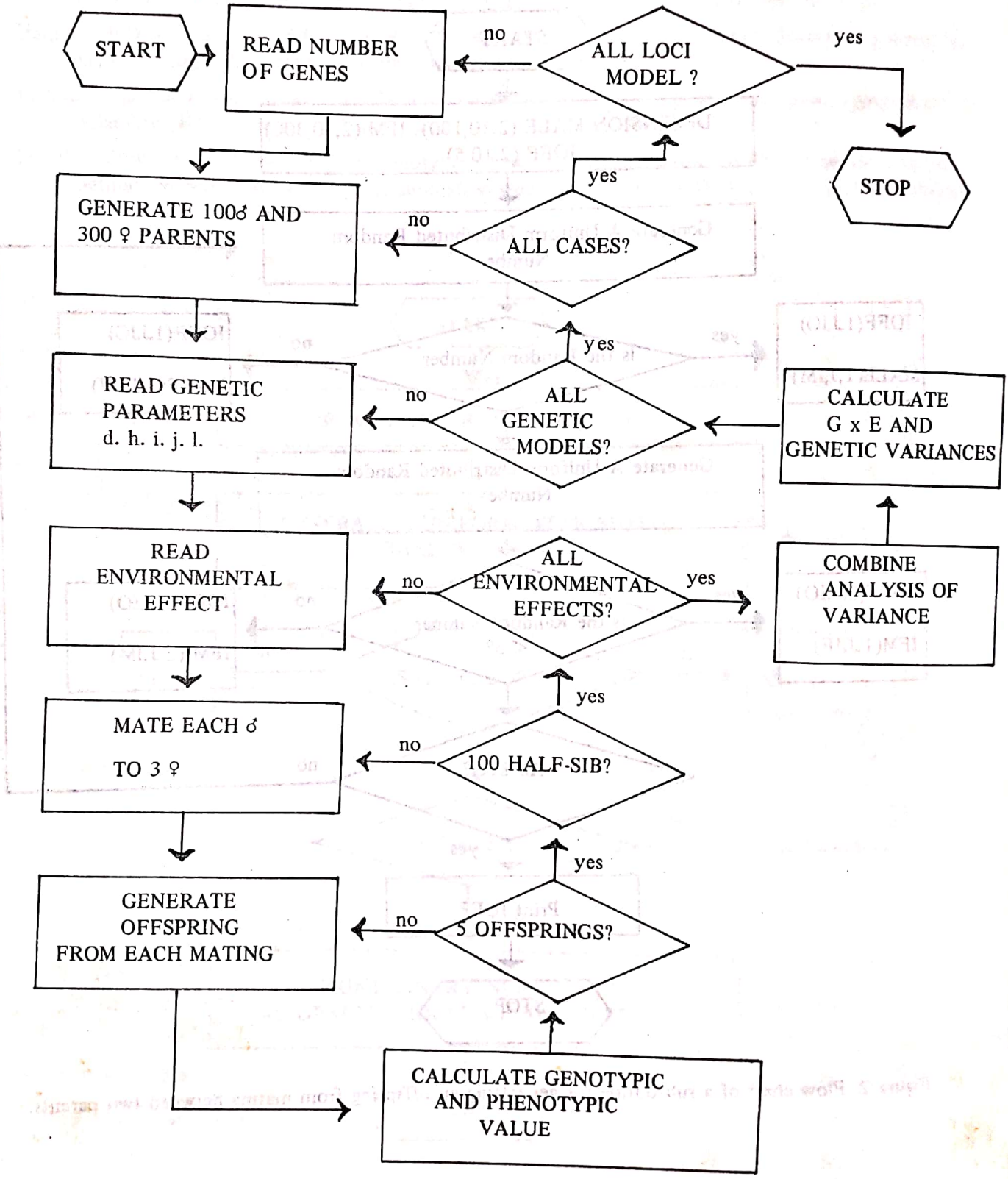


Figure 3. Flow chart of Design I simulation experiment.

Scanned with CamScanner