SAMPLE SIZE DETERMINATION FOR AGRONOMIC EXPERIMENTS

Medhat M. Elsahookie
Prof.
Univ. of Baghdad
Coll. of Agric. Eng. Sci.

A. A. Dawood
Researcher
General Board of Agric. Res.
Ministry of Agric., Iraq

ABSTRACT

Determining sample size to estimate the mean value of a target (infinite) population is of prime importance to have correct results and conclusions. In general, sample sizes differ as the studied objects differ. Another factor affect sample size is number of replicates or groups. Small objects, such as seeds, grains, and bacterial cell sizes require in general larger sample sizes as compared to large objects, such as human patients, large animals, and large trees. In this short article, a mixture of 12 crosses, maize kernels (k) were mixed to have 12,000 kernels. Sample sizes of 50,100, 150, 200, 250, and 300k were taken to find out which sample size fits the mean of target population (12,000 k). Sample size of at least 100 k and up gave the best good fit as compared by Chi-square, *P*=0.10 and 0.05 at one of degree of freedom. This result was on a heterogenous population of different maize crosses kernel sizes. On the other hand, a population of kernels of a maize cross (32x19), as homogenous population was studied by taking sample sizes of 25, 50, 100, 200, 300, 500, and 1000 k. Depending on Chi-square test, sample sizes of 50 and 100 k gave good estimate to the mean value of kernel weight. A previous research by Elsahookie et al (6) studied the best combination of sample sizes and number of replicates to estimate mean value of plant seed yield of maize and sunflower. They found that the best recommended combination was using sample size of 10 plants with 4 replicates. Hertzog (10) reported that sample size should be over 40, while Julious (12) recommended sample size of 12 per group. More studies on other objects of sample sizes are really required.

***Key words: kernel weight, plant seed yield, sunflower, maize.**

تحديد حجم العينة لتجارب المحاصيل

مدحت الساهوكي
أستاذ
كلية الزراعة/ جامعة بغداد

عبد الباسط عبد الرزاق داود
باحث
دائرة البحوث الزراعية/ وزارة الزراعة

المستخلص

إن حجم عينة الدراسة، أمر هام جدا في الأبحاث العلمية كافة، وذلك لإيجاد معدل الصفة المدروسة، للمجتمع الهدف (اللانهاية) وذلك للحصول على نتائج موثوقة واستنتاجات علمية. بشكل عام، يختلف حجم العينة باختلاف الأشياء المدروسة وعدد المكررات. إن الاجسام الصغيرة مثل البذور والحبوب والخلايا البكتيرية تحتاج بطبيعتها الى حجم عينة أكبر مقارنة بالأجسام الكبيرة مثل الانسان والحيوانات الكبيرة والأشجار. تم في هذا البحث دراسة احجام مختلفة من حبوب الذرة الصفراء المخلوطة من اثني عشر هجينا، وذلك بخلط ألف حبة من كل هجين ليكون المجتمع الخليط 12 ألف حبة. اخُذت عينات 50 و 100 و 150 و 200 و 250 و 300 حبة واستخرجت أوزانها لمقارنتها مع معدل عينة المجتمع الخليط. اظهر البحث ان اخذ 100 حبة في الأقل ضرورية لتقدير معدل القيمة المدروسة في المجتمع، وذلك باعتماد اختبار مربع كاي لدرجة حرية واحدة ومستوى احتمال 0.05. ثم اجراء اختبار آخر مع اوزان حبوب هجين واحد (32x19) باعتبار بذوره متماثلة بعكس المجتمع الأول الخليط. اخُذت عينات حبوب 25 و 50 و 100 و 200 و 300 و 500 و 1000 حبة للمقارنة. بعد مقارنة المتوسط باعتماد مربع كاي ومع مستوى احتمال 0.05 ظهر ان عينة 50 حبة فأكثر هي الأفضل لتمثيل متوسط المجتمع. أوضحت دراسات سابقة من قبل الساهوكي وآخرون (4) ان أفضل توليفة بين عدد المكررات واحجام العينات لحاصل نباتات الذرة الصفراء وزهرة الشمس هي 10 نباتات من كل مكرر وبعدد 4 مكررات في الأقل. فيما ذكر Hertzog (10) ان حجم عينة اعلى من 40 هو الأفضل، بينما ذكر Julious (12) ان حجم عينة 12 يمكن اعتماده في البحث بشرط وجود مجاميع او مكررات. ان هذا الموضوع يحتاج دراسات موسعة حول دراسة مختلف الاحياء بسبب أهمية الموضوع ولوجود تغايرات كثيرة مقارنة مع بعضها البعض.

***كلمات مفتاحية: ذرة صفراء، زهرة الشمس، حاصل النبات، وزن الحبة.**

Received:11/1/2024, Accepted:22/3/2024

INTRODUCTION

Sample size determination is one of the most important procedures in experimention. Sample size differs according to design used and type of material studied (1). Studying seed size or weight is different than studying human patients or animals (2). Using Chi-square test does not require large sample size, while using replicated factorial experiments need larger sample size (4). However, homogenous populations such as industrial objects need smaller sample sizes (8, 9), while studying heterogenous populations need large sample size because of high heterogencity in the population. Singh and Masuku (18), Jenkins (11) and Rayan (17) showed in their article two tables to choose sample size: one of them suggests sample size of population is 222, if the population 500, while second table shows that sample size is 212 for a population of 450. Several researchers such as Rutterford et al (16), Adcock (3), Guadagnoli and Velicer (7), Rahman (14) and Wisz et al (20) used lengthy equations to determine the right sample size in the study. Meanuhile, Dell et al (5) used logarethmic equation to determine sample size.
On the other hand, Ranatunga et al (15) used R^2 and the structural equation model to find out the minimum sample size required in the study. Another researcher (19) compared two equations; Dahlberg's equation and method of moment's estimator, and reported that a sample size of 20 or less is not adequate, but recommended sample size of 25 to 30 when the trial is replicated. This recommended result is with coincidence with that of Hertzog (10) who reported that sample size of 10 to 40 participants are usually used in references. Julious (12) stated that sample size of 12 per group rule of thumb for a pilot study counting on mean, variance and regularity considerations. Finally, Krejcie and Morgan (13) recommended using the formula;

Sample size: $\frac{\chi^2 NP\,(1-P)}{d^2(N-1)+\chi^2 P(1-P)}$ **when**

χ^2 = value of Chi-square 3.84 for 1 d.f
N = population size
P= population proportion (assumed to be 0.50)
d = degree of accuracy (*P*= 0.05).

They showed a table with population size up to 1,000,000, and that the sample size of a population of 500 is 217. This number is still high for many experimental uses. The objectives of this experiment was to shed some light on sample sizes used by researchers and to show new results on sample size used on maize (*Zea mays* L.) kernels.

MATERIALS AND METHODS

Dry kernels of 12 crosses of maize (*Zea mays* L) were taken and 1000 kernels of each were counted and weighed. Then all kernels of the crosses were mixed together to make a heterogenous population. Means of single kernels were calculated (Table 1). The overall mean of all kernels was 212.5 mg. Samples of 50, 100, 150,200,250 and 300 kernels were taken from the main mixed population (12000 kernels). Data in Table 2 showed the sample size weights and single kernel weights. On the other hand, sample sizes of 25, 50, 100, 200, 300, 400, 500, and 1000 kernels were counted from maize cross (32×19). Table 3 shows sample sizes, sample weights, and single kernel weights. The Chi-square was used to compare means of data to the population mean (1000 k).

RESULTS AND DISCUSSION

Data on maize kernels weights of the 12 crosses (1000k each) are shown in Table 1. Variation in kernel weights of maize crosses was ranging from 140.2 mg/k to 264.2 mg. These kernels were mixed together to have a heterogenous population of kernel weights. Samples of 50, 100, 150, 200, 250, and 300k were taken from mixed kernels of Table 1 and shown in Table 2.

Table 1. Weights of 1000 kernels of twelve crosses of maize and their single kernel weight

Cross	1000k (g)	single k (mg)
1- 32×49	240.22	240.2
2-32×21	264.24	264.2
3-4×51	217.40	217.4
4- 60×49	220.02	220.2
5- 49×17	140.20	140.2
6- 17×4	195.38	195.4
7- 32×60	208.10	208.1
8- 32×51	243.14	243.1
9- 60×51	203.60	203.6
10- 4×21	216.72	216.7
11- 61×49	187.34	187.3
12- 32×19	213.15	213.2

Overall average 212.5

Here we intended to check which sample size will represent the original population shown in Table 1 counting on Chi-square value for p = 0.10 and .05 of one degree of freedom. The results show that sample size of 50 k did not gave a good estimate. However, samples of 100 to 300k all gave a good fit to the value of single kernel weight shown in Table 1(213.2 mg/k). This is because of different position of kernels on maize cob. So, we can recommend using at least 100 k to estimate kernel weight of maize with or without replication. Julious (12) reported using sample size of 12 per group as a rule of thumb. However, small objects under study are of prime importance in sampling technique. Suppose we work on human patients, they are similar in many things, same thing will be correct on animals or trees, that is to say that variation in human patients responding to a drug could be very similar. Otherwise, we can not recommend a drug to treat a disease. Chi- square could be used to compare a group of patients to another, one control, and the other taking the drug. Meanwhile, we can notice the recommended result previously reported by Krejcie and Morgan (13) which they gave a table prepared by using a formula that gave high sample sizes of large population which we can not handle in experimentation.

Table 2. Sample sizes of maize kernels taken randomly after mixing all the 12 crosses kernels and their corresponding weights (mg).

Sample size (k)	Sample wt. (g)	Single k (wt. mg)
50	10.56	209.2
100	21.33	213.3
150	32.02	213.5
200	42.44	212.8
250	52.98	211.9
300	63.62	212.1

χ^2 at p=0.10 and 1 d f = 2.70 and p=0.05 = 3.84

The second test in this article was on kernel weight estimation of a maize cross (32× 19) which we could consider as homogenous population in term of kernel size (Table 3). Sample sizes of 25, 50, 100 200, 300, 400, 500, and 1000 k were taken with their weight and single kernel weights. Chi-square test of probability 0.10 and 0.05 of one degree of freedom was used for comparison. Sample size of 25k should be avoided for it did not give a good fit, while sample size of 50 k and up showed best fit to estimate the 1000 k population (213.2 mg/k) as the infinite or target population. In general, sample sizes of 100 up to 500 k all gave almost the same estimate of kernel weight of this maize cross.

Table 3. Sample size, total weight (g) and single kernel weight (mg) of the maize cross 32×19

Sample size (k)	Sample (wt.g)	Kernel (wt.mg)
25	5.40	216.2
50	10.74	214.8
100	21.31	213.1
200	42.6	213.0
300	63.93	213.1
400	85.31	213.3
500	106.50	213.0
1000	213.15	213.2

χ^2 for ⋉ 0.10 and 1 d.f = 2.71, ⋉ 0.05 =3.84

Hertzog (10) used samples of 10 to 40 per group and evaluated them, and concluded that sample size of 40 participants was not enough adequate for high confidence, then sample sizes of over 40 are recommended in that study of participants. On the other hand, sample sizes of 10-20 are very common in agronomic studies. Elsahookie et al (6) estimated plant seed yield of maize and sunflower (*Helianthus annuus* L.). They reported using replicates 3 to 6 with sample sizes of 5, 10, and 25 plants at each replicate. That was to find out the best combination of replicate number with a specific sample size. They found in maize plant grain yield that using 10 plants at each of 4 replicates gave the best grain yield estimate of target population of 630 plants population as compared by LSD, 0.05 probability level. Similar result was obtained on sunflower with a little difference between the two years of study. We can conclude and recommend using at least 100 kernels or seeds to estimate their weight of a cultivar, and using 10 plants with at least 4 replicates to estimate sunflower and /or maize plant seed yield. Studies on similar crops can use this sample size, but studies on human patients, large animals and large trees should consider other sample sizes.

REFERENCES

1.Abd, Z. F., M. M. Sultan, and S. S. Abd, 2020. Impact of herd size on the productive efficiency of sheep breeding projects at the kokjali region in Nineveh governorate for the production season 2018. The Iraqi Journal of Agricultural Science, 51(6):1613-1622. https://doi.org/10.36103/ijas.v51i6.1188

2.Abdullah, S. A. and A. S. Khalaf, 2023. Effect of seed size and seeding rates on yield attributes of triticale var. Admiral. Iraqi Journal of Agricultural Sciences, 54(3):820-836. https://doi.org/10.36103/ijas.v54i3.1766

3.Adcock, C. J., 1997. Sample size determination: a review. Journal of the Royal Statistical Society: Series D (The Statistician), 46(2): 261-283.

4.Al-Bahadely, F. H. N. and O. K. Al-Ukeili, 2018. Economies of potato production (Baghdad Province as a Case Study). The Iraqi Journal of Agricultural Science, 49(4):551-559. https://doi.org/10.36103/ijas.v49i4.62

5. Dell, R. B., S. Holleran, and R., Ramakrishnan. 2002. Sample size determination. ILAR Journal, 43(4):207-213.

6.Elsahookie, M. M., A. Shehab, and A. Mahmoud. 2001. Optimum sample size and number of replications for row crops plant yield evaluation. The Iraqi J. Agric. Sci. 32(6): 95-100.

7.Guadagnoli, E., and W. F. Velicer. 1988. Relation of sample size to the stability of component patterns. Psychological Bulletin, 103(2):265.

8.He, Y., Chen, J., Shen, J., P. Fournier-Viger, and J. Z.,Huang, 2024. Density estimation-based method to determine sample size for random sample partition of big data. Frontiers of Computer Science, 18(5), p.185322.

9.Hernandez, P. A., C. H., Graham, L. L. Master, and D. L. Albert, 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29(5):773-785.

10. Hertzog, M. A., 2008. Considerations in determining sample size for pilot studies. Research in Nursing & Health, 31(2):180-191.

11. Jenkins, D. G. and P. F Quintana-Ascencio,. 2020. A solution to minimum sample size for regressions. Plos. one, 15(2), p.e0229345.

12. Julious, S. A., 2005. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry, 4(4):287-291.

13.Krejcie, R. V. and D. W. Morgan, 1970. Determining sample size for research activities. Educational and Psychological Measurement, 30(3):607-610.

14..Rahman, M. M., 2023. Sample size determination for survey research and non-probability sampling techniques: A review and set of recommendations. Journal of Entrepreneurship, Business and Economics, 11(1): 42-62.

15.Ranatunga, R., H. M. S. Priyanath, and R. G. N., Megama,. 2020. Methods and rules-of-thumb in the determination of minimum sample size when applying structural equation modelling: A review. J.of Social .Sci Res, 15(2):102-109. https://doi.org/10.24297/jssr.v15i.8670

16.Rutterford, C., A. Copas. and S., Eldridge, 2015. Methods for sample size determination in cluster randomized trials. International Journal of Epidemiology, 44(3):1051-1067.

17.Ryan, T. P. 2013. Sample size determination and power. John Wiley and Sons. Wily Series in Probability and Statistics.

18.Singh, A. S. and M. B., Masuku, 2014. Sampling techniques & determination of sample size in applied statistics research: An

overview. International Journal of Economics, Commerce and Management, 2(11):1-22.
19.Springate, S. D., 2012. The effect of sample size and bias on the reliability of estimates of error: a comparative study of Dahlberg's formula. The European Journal of Orthodontics, 34(2):158-163.
20.Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. and NCEAS Predicting Species Distributions Working Group, 2008. Effects of sample size on the performance of species distribution models. Diversity and distributions, 14(5):763-773.