WG-CEMP-90/9

# OPTIMIZATION OF SURVEY SAMPLING DESIGN IN THE DETECTION OF INTERANNUAL VARIABILITY AND PREY SIZE SELECTIVITY IN THE DIET OF PENGUINS

E. Marschoff and B. Gonzalez*

Abstract

Based on the preliminary results of a survey of prey size in Adélie penguins, sampling designs intended for the detection of interannual variations in mean prey size and penguin size-related selectivity were analyzed in order to find the optimal allocation of observations at each level in nested ANOVA designs. The optimal numbers of penguins and krill to be used in the detection of interannual variability in the parameters concerned are recommended and a required sampling protocol is suggested.

A table of minimal numbers of penguins to be sampled as a function of number of sampling dates and magnitude of the variance component is given, under the assumptions of balanced design and the variance component of "penguin-by-date" interaction equal to the variance component of "penguin-within-dates" obtained from a previous study in Bahía Esperanza.

Restrictions have been fixed in accordance with the recommendations of the fourth meeting of WG-CEMP: minimum interannual variation for mean prey size is fixed at 10%; the test power is taken as 0.9 and Type I errors fixed at 0.1.

Cost values of manpower required are based on Argentinian logistics.

Résumé

Les modèles d'échantillonnage destinés à la détection des variations interannuelles de la taille moyenne des proies et de la sélectivité en fonction de la taille des manchots sont analysés sur la base des résultats préliminaires de la taille des proies chez les manchots Adélie, pour une allocation optimale des observations à chaque niveau des modèles ANOVA à emboîtement. Le nombre optimal de manchots et d'individus de krill à utiliser pour la détection de la variabilité interannuelle est recommandé, et un protocole d'échantillonnage suggéré.

Un tableau du nombre minimal de manchots à échantillonner en fonction du nombre de dates d'échantillonnage et de l'ampleur de la variance est donné, en présumant : un modèle équilibré et l'interaction d'une variance des "manchots-par-dates" égale à la variance des "manchots-individuels-sur-un-même-jour" obtenue à Bahia Esperanza, lors d'une étude antérieure.

* Instituto Antártico Argentino, Dirección Nacional del Antártico, Cerrito 1248, 1010 Buenos Aires, Argentina

Des restrictions ont été établies conformément aux recommandations de la quatrième réunion du WG-CEMP : la variation interannuelle minimale de taille moyenne des proies est fixée à 10%; la puissance du test est fixée à 0.9 et les erreurs de type I à 0.1.

Les coûts de la main-d'œuvre sont fonction des normes argentines.

## Резюме

На основании предварительных результатов исследования размерного состава видов, потребляемых пингвином Адели, схемы сбора проб в целях выявления межгодовых изменений среднего размера потребляемых видов и размерной селективности пингвинов анализируются с точки зрения оптимальной интенсивности наблюдений на каждом уровне гнездовых схем. Предлагаются оптимальные размеры проб криля и выборок пингвинов для выявления межгодовой изменчивости рассматриваемых параметров и схема взятия проб.

Приведена таблица, в которой указывается зависимость минимального размера выборки пингвинов и количества дней сбора проб, а также величина коэффициента изменчивости, исходя из того, что схема сбалансирована и изменчивость связи "количество пингвинов в выборке - количество дней сбора проб" равна изменчивости элемента "количество пингвинов, обследованное в течение определенного периода", выявленной в ходе предыдущего исследования в заливе Эсперанза.

Ограничения были установлены в соответствии с указаниями Четвертого совещания Рабочей группы по СЕМР: минимальное межгодовое изменение среднего размера потребляемых видов было установлено на уровне 10%; чувствительность принята за 0,9 и уровень погрешностей типа I - 0,1.

Стоимость необходимой рабочей силы основана на данных по аргентинскому материально-техническому обеспечению.

## Resumen

Basándose en los resultados de estudios preliminares del tamaño de las especies-presa en la dieta de los pingüinos adelia, para encontrar la asignación óptima de las observaciones para cada nivel de los diseños anidados ANOVA, se han analizado los diseños de muestreo dirigidos a la detección de las variaciones interanuales entre el tamaño medio de la presa y la selectividad correspondiente de acuerdo al tamaño del pingüino. Se recomiendan las cantidades óptimas de pingüinos y de krill que han de emplearse para la detección de la variabilidad interanual en los parámetros relacionadso y se sugiere un protocolo de muestreo.

Se presenta una tabla detallando la cantidad mínima de pingüinos de los cuales se deben tomar muestras en función del número de fechas de muestreo y de la magnitud del componente de varianza, suponiendo que: el diseño está equilibrado y el factor de varianza de la interacción del "pingüino a la fecha" es igual al factor varianza del "pingüino entre fechas" obtenido de un previo estudio realizado en la bahía Esperanza.

Las limitaciones han sido fijadas de acuerdo a las recomendaciones de la cuarta reunión del WG-CEMP: la variación mínima interanual para el tamaño medio de las especies-presa se fija en un 10%; la prueba de potencia se considera en 0.9 y los errores Tipo I se fijan en 0.1.

El costo del personal requerido se basan en la logística argentina.

---

## 1. INTRODUCTION

In a previous paper (Marschoff and Gonzalez, 1989) a preliminary analysis of stomach contents of Adélie penguins from Bahía Esperanza was made using a nested ANOVA design with data grouped by dates, "penguins-within-dates" and "krill-within-penguins". These data do not allow detection of the variability due to penguin selectivity because the design used was not intended to calculate the "penguin-by-dates" interaction. Taking account of the 1989 deliberations of the Working Group for the CCAMLR Ecosystem Monitoring Program (WG-CEMP), the present paper develops a sampling protocol for the optimal allocation of observations at each sampling level.

Noting that penguin selectivity not only might affect monitoring results but also has a biological interest of its own, a suitable sampling design for its detection is also discussed.

## 2. DETECTION OF INTERANNUAL VARIATIONS IN THE MEAN SIZE OF KRILL EATEN BY PENGUINS

At the 1989 Meeting of WG-CEMP, the desirable minimum size of changes to be detected was set at 10% of the parameter values. A value of 0.9 was suggested for the test power (probability of obtaining a significant result).

In the nested ANOVA design, discussed in Marschoff and Gonzalez (1989), comparison between years results naturally in a new level added at the top. Exact **F** tests are obtained dividing the mean square for "years" group by the mean square for "dates-within-years" group.

Given the costs of sampling at each level, it is possible to define a sampling protocol achieving the desired test power while minimizing cost. We will present the calculations made for the completely balanced case, along the lines discussed in Bliss (1967). The resulting design should be adjusted for different cost structures and ancillary objectives of the field work. In the present calculations manpower in Antarctica was estimated as four times as expensive as on the mainland.

In our design (4-level nested Model II ANOVA), the expected mean square for "years" group is:

$$E(MSY) = \sigma^2_{KCP} + n_K \cdot \sigma^2_{PCD} + n_p \cdot n_K \cdot \sigma^2_{DCY} + n_D \cdot n_P \cdot n_K \cdot \sigma^2_Y \quad (1)$$

where $\sigma^2_{KCP}$ is the variance component of the "krill-within-penguins" group,

$n_K$ is the number of krill measured in each penguin sampled,
$\sigma^2_{PCD}$ is the variance component of "penguins-within-dates" group,

$n_P$ is the number of penguins sampled each day,
$\sigma^2_{DCY}$ is the variance component of "dates-in-years" group,

$n_D$ is the number of sampling days, and
$\sigma^2_Y$ the variance component due to differences between years.

The expected values for the variance components and their 80% confidence intervals have been extracted from Table 3 of Marschoff and Gonzalez (1989):

| Variance Component | Expected Value | 80% Confidence Lower | Interval Limits Upper |
|---|---|---|---|
| $\sigma^2_{KCP}$ | 44.1661 | 42.54 | 45.93 |
| $\sigma^2_{PCD}$ | 10.2986 | 7.33 | 14.98 |
| $\sigma^2_{DCY}$ | 32.6738 | 12.91 | 51.62 |

The cost associated with each determination of the annual mean results from the sum of costs at the different levels. We included the additional cost for the first unit of sub-samples (first sample from a penguin during the day or first sample of a "krill-within-a-penguin" group) at the higher sampling level.

The yearly cost (C) of estimating the mean size of krill eaten by penguins at a given site is:

$$C = n_D . C_D + n_D . n_P . C_P + n_D . n_K . C_K$$

where $C_D$ is the cost associated with each sampling day; not related to the number of penguins sampled. It is dependent on the location of colonies, other sampling programs to be carried out on the same colonies, etc. For Esperanza Base it was estimated on the basis of two people working for two and a half hours;

$C_P$ is the cost of catching, sampling of a stomach content and labelling a penguin. Depending on methods, experience, etc, for Esperanza Base it was estimated based on two people working for 20 minutes; and the laboratory time was one person working for one hour to sort and clean the well-preserved fraction of stomach content; and

$C_K$ is the cost of measuring and processing the measurements of one krill specimen. Actually it takes us about four to five minutes, but by means of automating the recording procedure reduces this to two and a half minutes.

C should also include costs which are not directly dependent on the design, e.g. transportation, food, etc.

The optimal numbers of penguins and krill (those yielding maximum information at a fixed cost) were calculated from:

$$n_P^* = \sqrt{\frac{C_D \cdot \sigma^2_{PCD}}{C_P \cdot \sigma^2_{DCY}}} = 1.89$$

$$n_K^* = \sqrt{\frac{C_P \cdot \sigma^2_{KCP}}{C_K \cdot \sigma^2_{PCD}}} = 20.71$$

These optimal numbers are sensitive to errors in a ratio between variance components. For this reason, 80% confidence limits were used in order to achieve the desired power of the test. In our case, the upper limit for optimal numbers has been calculated using the following expressions, where sub-indices **u** and $l$ indicate the upper and lower limits of confidence intervals respectively:

$$n_P^* \leq \sqrt{\frac{C_D \,.\, \sigma^2_{PCD,u}}{C_P \cdot \sigma^2_{DCY,l}}} = 2.63 \leq 3$$

$$n_K^* \leq \sqrt{\frac{C_P \,.\, \sigma^2_{KCP,u}}{C_K \cdot \sigma^2_{PCD,l}}} = 25.03 \leq 26$$

A 10% change in the mean of two years results in the following variance component due to interannual differences:

$$\sigma^2_Y = 2(0.05\,\hat{\mu})^2$$

The parametric mean length ($\mu$) of "krill-within-penguins" group, was estimated for Bahía Esperanza as 38.41 mm. This means that the value of $\sigma^2$ which we wish to detect will be 7.350.

From (1) the $\sigma^2_Y$ can be expressed as a function of $n_D$ and the expected mean square for "dates" group E(MSD):

$$E(MSD) = \sigma^2_{KCP} + n_K \cdot \sigma^2_{PCD} + n_P \cdot n_K \cdot \sigma^2_{DCY}$$

Using the formulae for confidence intervals of variance components in an iterative procedure (Bliss, 1967, p. 259), the minimal value of $\mathbf{n_D}$ yielding a 90% confidence interval for $\sigma_Y^2$ resulted in n=16. This value defines the size of sub-samples at the year level assuring a test power of 0.9 for the detection of a 10% variation in the mean size of krill eaten by penguins. Finally, the suggested sampling protocol is:

> Sample during the crèche period on 16 selected days at random. On each day, three stomach contents of randomly selected penguins should be obtained including at least 26 well-preserved krill specimens the measurements of which are to be recorded.

The value obtained for $\mathbf{n_K}$ (number of krill in a stomach) seems to be relatively low. As these data might be used for other purposes, we recommend that either all krill present be recorded or that the sample be kept for further use. This model should be revised in view of the results of the penguin selectivity analysis.

## 3. TESTING PENGUIN SELECTIVITY

If extant, size selectivity of krill by penguins will result in a significant random factor. When isolated in an appropriate Model II design, this factor will yield a variance component for penguins significantly different from zero ($\sigma_P^2$).

A suitable design for such analysis is the two way Model II ANOVA where dates and penguins are crossed; the mean squares of factors should be tested against the mean squares of the "penguin-by-dates" interaction. The expressions for the expected mean squares are:

$$E(MSD) = \sigma_K^2 + n_K \cdot \sigma_{PXD}^2 + n_P \cdot n_K \cdot \sigma_D^2$$

$$E(MSP) = \sigma_K^2 + n_K \cdot \sigma_{PXD}^2 + n_D \cdot n_K \cdot \sigma_P^2$$

$$E(MSPxD) = \sigma_K^2 + n_K \cdot \sigma_{PXD}^2$$

$$E(MSE) = \sigma_K^2$$

We estimated that the variance component due to "krill-within-penguins" interaction as $\sigma_K^2$=44.1661 (Marschoff and Gonzalez, 1989) and the upper limit for the "penguins-by-dates" variance component as:

$$\sigma_{PXD}^2 \leq \sigma_{PCD}^2$$

We assumed that $\sigma_{PXD}^2 \leq \sigma_{PCD}^2$ in the estimation of sample sizes would result in a test power at least equal to that required. The sample sizes obtained should therefore be considered as upper limits for the optimal values. Table 1 was constructed for different values of the penguin variance component based on the abovementioned assumption.

It is to be noted that this design requires all penguins involved to be sampled on the same dates. As this is not possible because of the differences in penguin return dates, the actual design will be incomplete. This results in a test power loss which cannot be estimated without the actual values (return date is in itself a random variable) and using more complicated designs. Therefore, $\mathbf{n_P}$ and $\mathbf{n_D}$ values should be considered only as a guide for field work.

The cross-classification Model II ANOVA might also be used for the comparison between fresh and digested fractions of stomach contents. In this case we have an extra component of variance coming from variation in meal times. This component will be a part of the interaction and cannot be assessed without direct information on its magnitude. Thence, at least a preliminary set of field data is needed to proceed with this line of investigation.

## REFERENCES

BLISS, C.I. 1967. *Statistics in Biology. Vol I.* McGraw Hill Book Company.

MARSCHOFF, E. and B.N. GONZALEZ. 1989. The use of penguin stomach contents for the simultaneous study of prey and predator parameters. In: *Selected Scientific Papers, 1989 (SC-CAMLR-SSP/6).* Hobart, Australia: CCAMLR. pp. 367-376.

Table 1: Numbers of penguins to be sampled as a function of number of sampling dates and variance component due to differences between penguins. Test power fixed at 0.9.

| Variance Component | Number of Sampling Dates | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 10.298 | 13 | 6 | 4 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 7.3750 | 19 | 8 | 5 | 4 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3.2674 | 60 | 22 | 13 | 8 | 6 | 5 | 4 | 4 | 3 | 3 | 3 | 3 | 2 | 2 | 2 |
| 1.0298 | 434 | 150 | 78 | 48 | 34 | 25 | 19 | 16 | 13 | 11 | 10 | 9 | 8 | 7 | 6 |

Liste des tableaux



Список таблиц



Lista de las tablas