

**DRAFT**  
**Prairie Falcon Fecundity Legacy Data Analysis, Initial Power and  
Sample Size Estimates**

**Gretchen LeBuhn and Edward Connor**

**2007**

**Submitted to:**

**San Francisco Bay Area Network Inventory and Monitoring Program  
Fort Cronkhite, Bldg 1063  
Sausalito, CA 94965**

**under Cooperative Agreement H8C07030001 with the California Cooperative  
Ecosystem Studies Unit under Task Agreement J8C07060015**

## **Summary of recommendations**

We recommend continuing to monitor all 30 territories in an effort to detect trends toward a decline in the fecundity of Prairie Falcons. At this level of sampling, a 15% difference between core and non-core annual decline in fecundity over 10 years ( a cumulative decline of 80% can be detected with no greater risk of failing to detect such a decline ( $\beta = 30\%$ , power = 70%) than that of claiming a decline is occurring when there actually is no decline ( $\alpha = 30\%$ ). Because the power to detect a change in fecundity is low, we suggest that management action be taken when a decline is combined with either multiple years where there is complete reproductive failure and the failure to establish nests.

## **Sampling Design**

The sampling design to estimate trends in Prairie Falcon fecundity will involve sampling territories found in core and non-core habitat and monitoring fecundity for raptor pairs on those territories annually. This type of sampling design is a repeated measures analysis of variance. It combines both a within-subjects design (also called repeated-measures or split-plot designs) since each territory is considered a statistical subject and the temporal trends are estimated by comparing each territory to itself over time - that is treatment effects are estimated within subjects (Winer et al. 1991) and a between subjects design where the values for one set of sites (in this case, core) are compared to another set of sites (in this case, non-core).

A within-subjects design requires that each subject be monitored continuously. Any subject not monitored in a sampling period will be lost from the data set when the data are analyzed for trend, unless the analyst is willing to interpolate the missing data. If few sample periods are missed in the aggregate data then such interpolation should provide reasonable means to retain all subjects for the data analysis. Furthermore, the sensitivity of the conclusions to the interpolated values can be assessed by analyzing the data with high or low estimates to see if the conclusions are robust (unaffected by a range of reasonable interpolated values).

A between-subjects design does not have the same requirement for continuous monitoring.

## **Analysis of Legacy data**

### **Actual Fledglings per Territory**

We analyzed the legacy data from 1989-2006. We assumed that actual fledglings seen was the best measure of fecundity so, we based our first analysis on those data. We chose to start with the data from 1989 because this was the first year that had fairly complete core and non-core samples. For any territory where no nest was detected or there was no sampling data, we made the assumption that fecundity on that territory was zero. We also chose to assign the fecundity of a pair that used multiple territories to a

single territory rather than dividing it between two territories. There were so few examples of this that we don't expect it to change the results in any way.

There are two different ways to test the within-subjects effects of a repeated measures design: a multivariate approach and a univariate approach. In the multivariate model, the multiple estimates of fecundity on a territory over several years are viewed as *different response variables* and one assumes the data are multivariate normal. Under these assumptions, one analyzes the data using a multivariate analysis of variance (MANOVA) and the test statistic is a multivariate statistic. The power of this test can also be assessed directly. Under the univariate model, each estimate of fecundity on a territory is viewed as different values of a single response variable and the data are assumed to be univariate normal. Under these assumptions, one analyzes the data using a univariate analysis of variance and the test statistic is a univariate *F*-statistic. However, in the univariate analysis both hypothesis tests and power depend on an additional assumption that the *var/cov* matrix is spherical. This assumption amounts to conjecturing that the variances and covariances are equal among years. Hypothesis tests and estimates of power can be adjusted for the degree of departure from meeting the assumption of sphericity using the recommended Greenhouse-Geisser epsilon (O'Brien and Kaiser 1985, Winer et al. 1991, Muller and Barton 1989, 1991, Muller et al. 1992). Unadjusted tests tend to make too many Type I errors, and to over estimate power. Both models are appropriate and usually yield similar results, if the univariate tests are adjusted using the Greenhouse-Geisser epsilon.

We tested for sphericity using Mauchly's test which is printed below. Since the Prairie Falcon data does not meet this assumption, we report the statistics from the multivariate analysis.

Sphericity Tests				
Variables	DF	Mauchly's Criterion	Chi-Square	Pr > ChiSq
Transformed Variates	119	5.9689E-9	431.96706	<.0001
Orthogonal Components	119	0.0001001	210.0769	<.0001

**Within-subjects effects:**

The first multivariate test of the within subjects effects tests the main effect: Does year influence fecundity? The null hypothesis is that fecundity does not change across years. To evaluate these data, the most commonly used multivariate test statistic is the Wilks' Lambda.

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no year Effect H = Type III SSCP Matrix for year E = Error SSCP Matrix					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.18464255	3.12	17	12	0.0254

Since the F statistic is rather large, we can reject the null hypothesis that there is no variation among years in fecundity. There were significant differences between years.

The next analysis looks at within subjects by between subjects interaction effect: Does the influence of habitat use depend on the year ?

```

The GLM Procedure
Repeated Measures Analysis of Variance

MANOVA Test Criteria and Exact F Statistics
for the Hypothesis of no year*Core Effect
H = Type III SSCP Matrix for year*Core
E = Error SSCP Matrix

S=1    M=7.5    N=5

Statistic              Value    F Value    Num DF    Den DF    Pr > F
Wilks' Lambda          0.31407695    1.54        17        12    0.2254

```

Looking at the Wilks' Lambda and its associated  $F$ -value (0.314) and p-value (0.2254), suggests that differences between core and non-core areas do not depend on the year sampled.

### Between-subjects effects:

To examine the between subjects effects, there is only one approach.

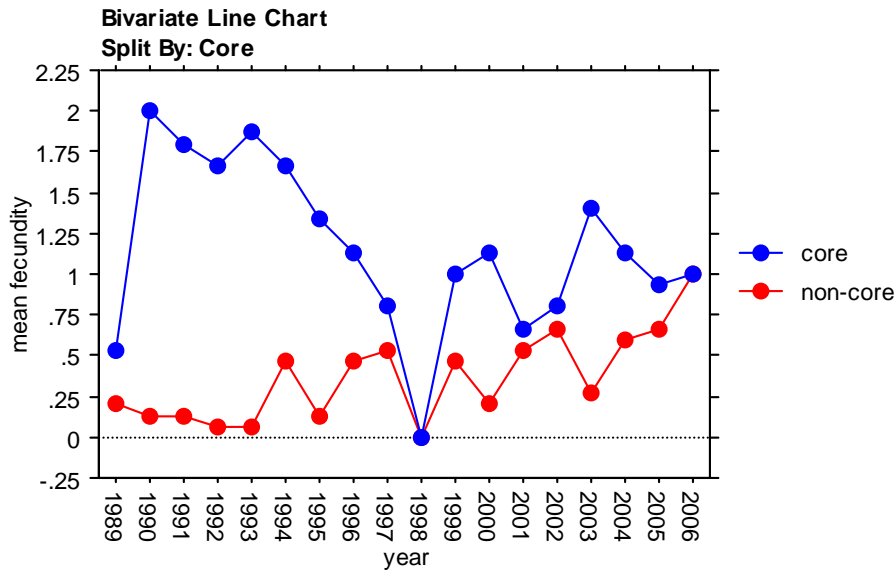
```

The GLM Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source    DF    Type III SS    Mean Square    F Value    Pr > F
Core      1     84.8074074    84.8074074     9.65     0.0043
Error    28    246.0740741    8.7883598

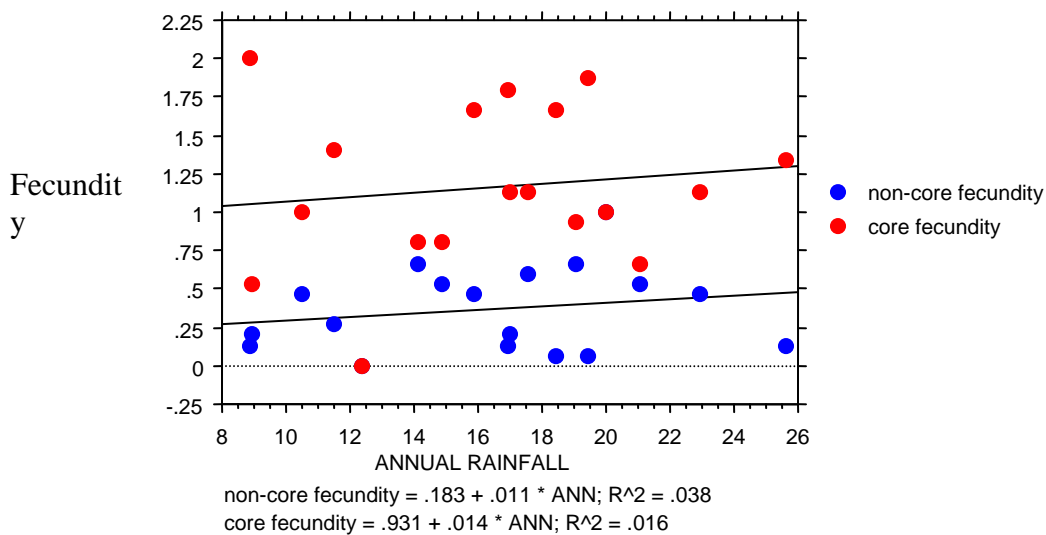
```

The large F-value (9.65) and associated p-value (0.0043), there is a significant difference in fecundity between core and non-core areas. The core areas have higher fecundity.



A graph of the difference between the core and non-core areas by years shows that the core areas have higher fecundity. It would be useful to know something about 1990-1995 since during these years the fecundity of the core areas is so much higher than the non-core. The fecundity in 1989 makes us think that the current levels of fecundity are closer to normal and that the early 1990's data were unusual for the core area.

We explored whether there was a strong relationship between fecundity and annual rainfall, especially El Nino events and found no significant relationship.



To get a better idea of the influence of climbing on prairie falcon fecundity, it might be worth trying to look at how many of the core territories are in known climbing routes.

We were surprised at the lower fecundity of the non-core territories. One possibility is that there is a difference in predation between core and non-core territories. Great horned owls, golden eagles and red tail hawks may be able to access more territories in the non-core areas if there is a difference in the size of cliff and aspect in core and non-core areas. The East Bay Regional Parks have been finding tremendous great horned owl predation on their Prairie Falcons (Doug Bell, pers. comm.) Since ground squirrels are important prey for prairie falcons, it would be interesting to know whether there was ground squirrel control from 94-98 in local areas that might have influenced fecundity? Since golden eagles also consume many ground squirrels, we would expect a similar decline in their fecundity.

**Eggs, Hatchlings and Potential Fledglings per Territory**

To see whether the patterns detected in actual fledglings were reflected in the other measures of fecundity, we did the same analysis looking for the effects of core, year and interactions on the number of eggs, the number of hatchlings and the possible fledglings. The effects were very similar to the actual fledgling results so, we suggest that the actual fledgling data be used as the measure of fecundity. Once again, there was a consistent effect of core vs. non core on each of the measures of reproduction. Core habitat had higher fecundity. There was also an effect of year on Hatchlings. The number of Hatchlings varied among years.

**Effect of Year**

**Number of eggs:**

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.25237527	1.81	18	11	0.1582

**Number of Hatchlings:**

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.14514738	3.10	19	10	0.0354

**Number of Potential Fledglings:**

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.26751204	1.44	19	10	0.2815

**Effect of Interaction between Year and Core**

**Number of eggs:**

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.44561272	0.76	18	11	0.7075

**Number of Hatchlings:**

Statistic	Value	F Value	Num DF	Den DF	Pr > F
-----------	-------	---------	--------	--------	--------

Wilks' Lambda	0.36708204	0.91	19	10	0.5914
---------------	------------	------	----	----	--------

**Number of potential Fledglings:**

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.43853151	0.67	19	10	0.7799

**Effect of Core**

**Number of Eggs**

Repeated Measures Analysis of Variance  
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Core	1	12.90666667	12.90666667	5.54	<b>0.0259</b>
Error		2865.28666667	2.33166667		

**Number of Hatchlings:**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Core	1	14.10666667	14.10666667	6.89	<b>0.0139</b>
Error	28	57.28666667	2.04595238		

**Number of Potential Fledglings**

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Core	1	14.72666667	14.72666667	6.62	<b>0.0157</b>
Error	28	62.31333333	2.22547619		

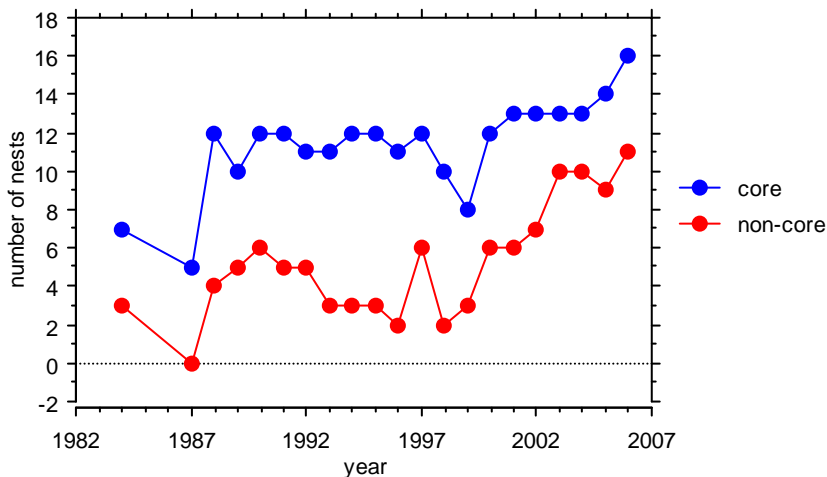
## Number of Nests in core and non-core areas

We analyzed the legacy data from 1984 to 2006 for the number of nests in core versus non-core data. We assumed that sampling effort in each area had been equal during each year. To be conservative, we did a non-parametric test, the Wilcoxon signed rank test which can be used to compare two related samples or repeated measurements on a single sample. The analysis was done in Statview.

This analysis asks whether there is a significant difference in the number of nests per year between core and non-core areas. There was a significant difference in the number of nests between core and non-core areas. The core areas had significantly more nests. It is interesting that nesting in the core and non-core areas cycle together much more than fecundity. This does suggest that the variation in fecundity and nesting is not due to individual pairs switching between core and non-core habitats.

### Wilcoxon Signed Rank Test for non core nests, core nests

# 0 Differences	0
# Ties	6
Z-Value	-3.920
P-Value	<.0001
Tied Z-Value	-3.933
Tied P-Value	<.0001



We also examined per nest fecundity (fecundity/established nest) and found it to be highly variable. There was a significant difference between core and non-core per nest fecundity. The core areas had significantly higher fecundity.

**Wilcoxon Signed Rank Test for core, non-core**

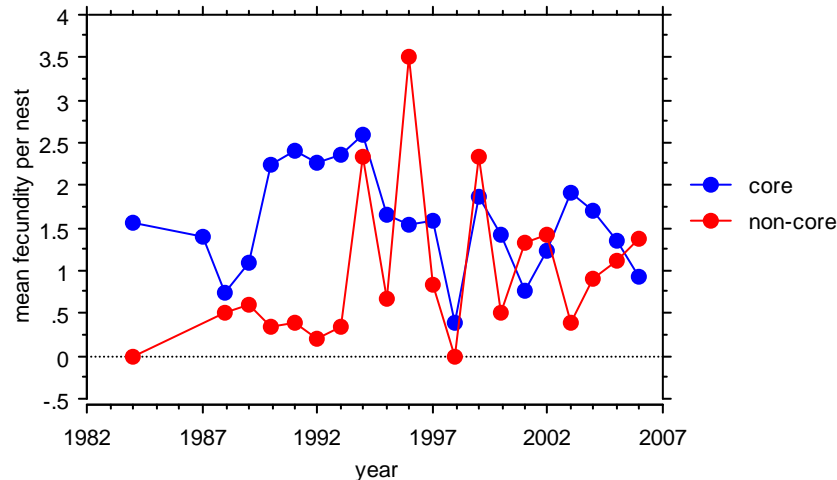
# 0 Differences	0
# Ties	1
Z-Value	-2.427
P-Value	.0152
Tied Z-Value	-2.427
Tied P-Value	.0152

One case was omitted due to missing values.

**Wilcoxon Rank Info for core, non-core**

	Count	Sum Ranks	Mean Rank
# Ranks < 0	5	40.000	8.000
# Ranks > 0	15	170.000	11.333

One case was omitted due to missing values.



---

**Power and Sample Size Analysis**

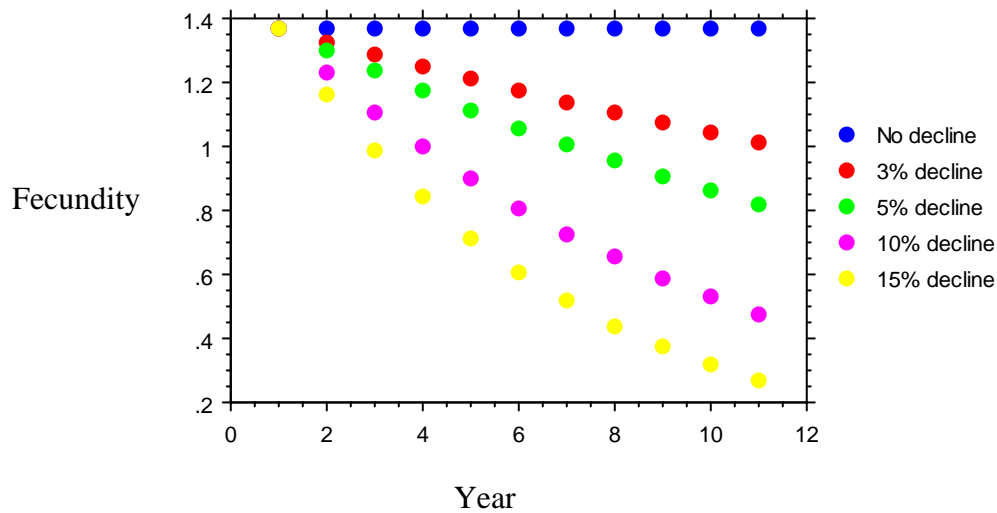
*Background and Justification* - An examination of the sample size needed to detect a specific magnitude of decline in fecundity over a specified time frame can be accomplished using the legacy data on Prairie falcon fecundity available as of 2006. This legacy data involves fecundity estimates for owls occupying 30 distinct territories during the period 1989-2006. We did not use the first few years of sampling for this analysis because we were not certain sampling effort had been equal across core and non-core areas. Fledglings were produced for these territories from as few as 0 or as many as 16 years during that time frame. Using the complete data set estimates of the annual mean fecundity per territory, the variances in fecundity among territories with each year, and the covariances in fecundities between years can be estimated.

Power in the within-subjects design depends on the four common parameters,  $n$  – sample size,  $\alpha$  – Type I error rate,  $\sigma$  – the variability among subjects (sites), and  $\delta$  – the trend magnitude one desires to detect. More powerful designs involve using higher sample

sizes, lower variances, or larger Type I error rates and trend magnitudes. However, in the within-subjects design, power also depends on the covariances (related to correlations) within – subjects, but between years, and in the heterogeneity of the variances and covariances among years (Winer et al. 1991, Muller and Barton 1989, 1991, Muller et al. 1992). Lower, less heterogeneous variances and higher, less heterogeneous covariances leads to higher power. Not only does power depend on the variances and covariances (*var/cov* matrix) and their heterogeneities, so does the *F*- statistic used in testing the null hypothesis of no difference in fecundity between years (Winer et al. 1991). Since the data collected to date does not meet the assumption of sphericity, we suggest all analyses are done using the multivariate tests.

*Determining Detectable Trend Magnitude* - Using the legacy data, we estimated the annual mean fecundity per territory to be 1.369 fledglings. We recommend using this long-term mean as a baseline against which to conjecture declines of specific magnitudes over specific time frames. For example, one might conjecture that Prairie Falcon fecundity will decline 3%, 5%, 10%, or 15% per year over a 5, 7 or 10 year time frame. Such declines are essentially the detectable trend magnitude for which we calculate power or for specified power we calculate sample size. A 3% annual decline in fecundity over 7 years amounts to a cumulative reduction in fecundity of approximately 17% or to 1.14 fledglings, a 5% annual decline results in a cumulative reduction in fecundity of approximately 26% or to 1.01 fledglings, a 10% annual decline results in a cumulative reduction in fecundity of approximately 47% or to 0.727 fledglings, and a 15% annual decline results in a cumulative reduction in fecundity of 62% or to 0.516 fledglings (Figure 1). We use a ten year trend magnitudes in our preliminary analyses.

Figure 1. Potential detectable trend magnitudes over 10 years.



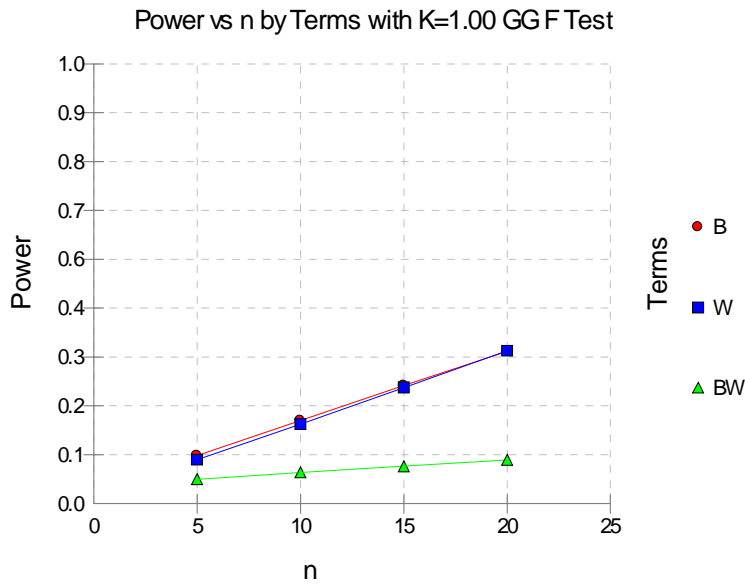
*Estimating Variances and Covariances* – We estimated the variances and covariances for a 21-year period using the legacy data on the Prairie Falcon. Using the statistical package SAS, we calculated the variances with years and the correlations between years. From the between year correlations ( $cor(i,j)$ ) we calculated the covariances ( $cov(i,j)$ ):

$$cov(i, j) = cor(i, j) \times \sqrt{var(i) \times var(j)}$$

The resulting  $var/cov$  matrix is included as an appendix with the within-year variances on the diagonal, and the covariances as the off-diagonal elements.

*Power and Sample Size Calculation* – We estimated sample size and power using the software package *PASS* (Power and Sample Size) (Hintze 2001) with the trend magnitudes depicted in Figure 1 and the estimated sample  $var/cov$  matrix of Appendix 1. We started by setting alpha to 0.05. We ran this analysis for detecting a 3%, 5%, 10% and 15% decline in one population and a stable (0% decline) alternative population. As you can see from figure 1 which evaluates our ability to detect the case when one population has remained stable and the other one population decline to 80% of its original size, the power to detect difference between core and non-core areas (B) and through time (w) or an interaction is less than 0.25 with 15 territories in core and 15 territories in non-core areas.

Since you are limited by the actual number of territories that are in the core habitat, increasing the number of territories monitored is not possible. If we set  $\alpha$  and  $\beta = 0.2$ , our power to detect differences between core and non-core areas (B) and differences across years (W) becomes more acceptable but still low. With the current sampling of 15 core and 15 non-core territories, you have a about a 25% chance of detecting a decline of 15% per year over 10 years.



If we set  $\alpha$  and  $\beta = 0.3$ , our power to detect differences between core and non-core areas (B) and differences across years (W) becomes more acceptable. With the current sampling of 15 core and 15 non-core territories, you have a 70% chance of detecting a decline of 15% per year over 10 years with no greater risk of failing to detect such a decline than claiming a decline is occurring when there actually is no decline. Traditionally statistical hypothesis tests are performed with  $\alpha$  kept at a minimum level. However, given that the cost of failing to detect a decline in prairie falcon fecundity and responding with appropriate conservation measures is greater than the cost of responding to a false detection of a decline (more expensive to recover a small population than to prevent its decline), we recommend setting  $\alpha$  to be a larger value. Our recommendation is to set  $\alpha$  and  $\beta$  to be approximately equal, but one could argue for  $\alpha$  to be even larger (Elzinga *et al.* 2001). Therefore, we recommend continuing monitoring 30 territories.

We still think the power of this monitoring is low. If you examine the actual fecundity data between 1990 and 1998 in the core area, you can see there was a massive decline in fecundity and this makes the data variable. However, the fecundity in those early years was much higher than it has been before and after that period. If a manager was solely tracking fecundity data, this type of decline would undoubtedly trigger management action. However, it is clear from the subsequent years that the population continued and the fecundity returned to a more normal level without intervention. Since fecundity is likely to be one of the most sensitive indicators of threat, we suggest that fecundity data be used for monitoring but be evaluated in a broader context. For example if there was reproductive failure for two years, or if in a single year there was reproductive failure and no nest establishment or a significant decline in the number of nests, intervention would be appropriate.

## Resources

Analysis of within-subjects design: Chapter 4 of Winer *et al.* 1991

## Literature Cited

- Elzinga, C. L., D. W. Salzer, J. W. Willoughby, and J. P. Gibbs. 2001. *Monitoring Plant and Animal Populations*. Blackwell Science. Malden, MA.
- Hintze, J. 2001. *NCSS and PASS*. Number Cruncher Statistical Systems. Kaysville, Utah
- Muller, K.E. and C.N. Barton. 1989. Approximate power for repeated-measures ANOVA lacking sphericity. *Journal of the American Statistical Association* 84: 549-555.
- Muller, K.E. and C.N. Barton. 1991. Correction to "Approximate power for repeated-measures ANOVA lacking sphericity." *Journal of the American Statistical Association* 86: 255-256.
- Muller, K.E., L.M. LaVange, S. Landesman, and C.T. Ramey. 1992. Power calculations for general linear multivariate models including repeated measures applications. *Journal of the American Statistical Association* 87: 1209-1226.
- O'Brien, R.G. and M.K. Kaiser. 1985. MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin* 97: 316-333.
- Urquhart, N.S. and T.M. Kincaid. 1999. Designs for detecting trends from repeated surveys of ecological resources. *Journal of Agricultural, Biological, and Environmental Statistics* 4: 404-414.
- Winer, B.J., D.R. Brown, and K.M. Michels. 1991. *Statistical Principles in Experimental Design*. McGraw-Hill, Inc.: New York.

**Open Questions:**

- 1. Is the increase in the number of nests detected in the non-core area due to more sampling?**
- 2. Is there something that happened in 1990-1996 that might explain the high fecundity in core areas (e.g. no more ground squirrel control or a decrease in predation due to the absence of owls...)?**
- 3. Is there any ground squirrel data that could be related to fecundity?**