

Using PROC SURVEYMEANS To Analyze A Creel Survey

Rebecca G. Frederick & E. Barry Moser

Department of Experimental Statistics

Louisiana State University AgCenter, Baton Rouge, LA

George Melancon

Rockefeller Wildlife Refuge, Grand Chenier, LA

ABSTRACT

Rockefeller Wildlife Refuge in Grand Chenier, LA, conducted a 3-year Creel Survey that had been designed by the Department of Experimental Statistics. The Creel Survey began in May 1997 and continued on until March of 2000. A multi-stage cluster-sampling scheme was used to survey recreational users of the Refuge. The design was modified mid-way through the 3-year survey resulting in 2 sampling periods: Period 1 is May, 1997- February, 1998 and Period 2 is March, 1998 – February, 2000. The survey was stratified by year, season, and day type, and sampling occurred at 1 of 4 locations with unequal probabilities. Information such as party-size, distance-traveled, hours fished, and numbers in harvest was recorded for each party. Interest is in estimating usage of the refuge by season and year. Additional information on the users and the harvest are also desired. Individual measurements were combined and then expanded to yield day-level estimates. Sampling weights were added to this primary sampling unit data and analyzed using PROC SURVEYMEANS.

INTRODUCTION

The Creel Survey was given to a total of 4849 parties of people visiting the Rockefeller Wildlife Refuge located in Cameron Parish. The elementary unit consisted of a car/boat of people. Important to this survey was the estimation of average party size, total visits and people per day, the catch during the day, and the number of hours fished. The variables of interest are:

Variables:

TypeOfDay:	W=Weekend Day, Saturday, Sunday and Holidays D=Day, Monday - Friday
ID:	Identification Number
Location:	PL= Price Lake Road, RO=Rollover Bayou EEL=East End Locks , EEB=East End Boat Launch

PL= Price Lake Road



RO=Rollover Bayou



EEL=East End Locks:



Finished, is a variable that describes whether the party was at the end of the trip or not.

VARIABLES: (continued)

Origin: City and State

Miles: Mileage of travel to Rockefeller Refuge

Hrs. Fished: Number of hours-fished

Man Hrs.: Number of man-hours in
Rockefeller Refuge

Finished: Y=yes, N=no

Species: Targeted species number

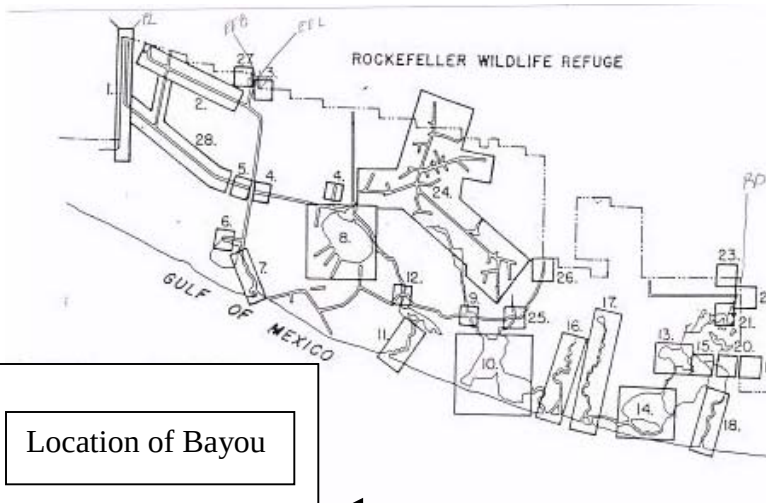
0. Non-Fishing activity (bird-watching)

1. Any Species 5. Trout

2. Crabs 6. Gar

3. Shrimp 7. Catfish

4. Redfish 8. Bass



- | | |
|--------------------------------|-------------------------------------|
| 1. Price Lake Road | 15. Flat Lake Canal |
| 2. Boat House Canal | 16. Pigeon Bayou Weir |
| 3. East End Locks | 17. Cop Cop Bayou |
| 4. Lake 4 | 18. Rollover Bayou |
| 5. Lake 5 | 19. Rollover Weir |
| 6. Miller Lake Weirs | 20. Flat Lake Weir |
| 7. Joseph Harbor Canal | 22. Pecan Island District |
| 9. Big Constance Structure | 23. Lake 14/Property Line Structure |
| 10. Big Constance Lake | 24. Superior Canal System |
| 11. Little Constance Canal | 25. Dyson Flapgate Structure |
| 12. Little Constance Structure | 26. Parish Line Flapgate Structure |
| 13. Flat Lake Structure | 27. East End Boat Launch |
| 14. Flat Lake | 28. Bertrand Canal Complex |

SAMPLE DESIGN

Rockefeller Refuge conducted a very involved sample survey. Interest is in estimating usage of the refuge by season and year. The sample design differed in two-periods:

Period 1 – Spring, 1997 – Winter, 1997

- **Four (4) Sampling sites are defined on the refuge, { East End Bayou(EEB), East End Locks(EEL), Price Lake Road(PL), Rollover(RO)}.**
- **Sampling days are stratified by**
 - Year,
 - Season, and
 - Type of Day.
- On the selected sampling day, each of the 4 sampling sites is visited for $\frac{1}{4}$ of the day in a randomly selected order. This is viewed as stratification on time-of-day, but where the strata (time-of-day) are dependent, since a sampling site selected for a specific time-of-day cannot be selected again on that same day.
- Only parties exiting the sampling site during the visited time are surveyed.
- This is a type of two-stage cluster sampling with stratification of Primary sampling units. Clusters are the orderings of visits to sites.

Sampling assumption issues:

- Time of entry and departure for parties at a site may depend upon the weather, time of day, type of day, and season. (for example, rainy weather or hurricane approaching)
- It is possible that most of the exits occur at a particular time of day, so that the proportion of usage recorded for each sampling site depends upon the order in which the sites are visited.
- The weather should reasonably affect all 4 sampling sites in a similar way on a given sampling day. (For example, flooding on Price Lake can cause the road to be closed for the day.)
- The proportion of users of the Rockefeller Refuge sampled at a sampling site on the sampled day is a random variable.

Period 2 – Spring, 1998 – Spring, 2000

- **Four (4) Sampling sites are defined on the refuge, {East End Bayou(EEB), East End Locks(EEL), Price Lake Road(PL), Rollover(RO)}.**
- **Sampling days are stratified by**
 - Year,
 - Season, and
 - Type of Day.
- On the selected sampling day only 1 of the 4 sampling sites is visited for the entire day. The site selected for sampling is randomly selected from a known distribution that varies depending upon the season, (spring, summer, fall or winter). This is two-stage cluster sampling but where the cluster sample size is 1.
- Only parties exiting the sampling site during the visited time are surveyed, but it is assumed that all parties at the sampled site exit during the survey period of the day.
- This is two-stage cluster sampling but where the cluster sample size is 1, again with stratification of the primary sampling units. Clusters are the sites.

Sampling assumption issues:

- Since a sampling site is visited for an entire day, time of entry/departure become less important than in period 1. There maybe exits prior to or after the sampling period for a day, and these are ignored in this analysis (a type of non-response bias).
- Since only 1 of the 4 sites is visited on a given sampling day, weather and/or other features that cause day-to-day variation in Rockefeller Refuge usage become very important.
- The proportion of users of the Rockefeller Refuge sampled at a sampling site on the sampled day is a random variable.

DESIGN-BASED ESTIMATION

In the design-based approach, the idea is to use the randomization process of the sampling design to provide a design-unbiased estimate of the quantities of interest, and to supply estimates of the sampling variance. Design-unbiased implies that if we were to have conducted all possible outcomes for the design, then the average over all of these outcomes would be exactly the population quantities of interest. Although unbiased, this approach sometimes leads to a larger variance, than what might be obtained from model-based approaches.

ALL PERIODS Spring 1997 – Spring, 2000

- Once the data for each period have been inflated over the secondary sampling units (adjusted to totals per day), the designs at the primary stage are identical. Thus, perform the inflations of the secondary units to the primary unit level, apply sampling weight, and then compute the overall analysis combining all data together.
- Only the variation among primary sampling units will be used in measures of precision. This is a fairly standard approach in cluster sampling since the largest variance component tends to be that of the primary stage.

USING SAS TO IMPLEMENT THE SURVEY

First, a calendar was constructed using Proc Calendar to indicate randomly selected days and sites. In the first period, the order of sampling the sites was randomized. All of these observations will be referred to as Creel_Survey_Before. Creel_Survey_Before included all observations that were made from March 1, 1997 – February 28, 1998. Rockefeller Refuge randomly needed 115 days per year to operate the survey; 30 days in the Spring, 40 days in the Summer, 30 days in the Fall and 15 days in the Winter. Nlaunch=4 is assigning the 4 locations to the calendar. The SAS[®] program was performed and the significant part of the program is included:

```
Data Date_Before; seed=0; nlaunch=4;
  do i = 1 to 115;
    y=round(nlaunch*ranuni(seed) + 0.5)
  output;
  end;
Run;
```

In the second sampling period, Rockefeller Refuge decided to use unequal probability sampling to better target seasonally used locations. By seasons, the following probabilities were used in selecting locations to visit.

- Spring
Price Lake Road = 3/8
East End Bayou = 2/8
Rollover Bayou = 2/8
East End Lake = 1/8

- Summer
Price Lake Road = 3/8
East End Bayou = 2/8
Rollover Bayou = 2/8
East End Lake = 1/8

- Fall
Price Lake Road = 1/9
East End Bayou = 3/9
Rollover Bayou = 3/9
East End Lake = 2/9

- Winter
Price Lake Road = 0/7
East End Bayou = 3/7
Rollover Bayou = 3/7
East End Lake = 1/7

Creel_Survey_After included all observations that were made from March 1, 1998 – March 31, 2000. The SAS[®] program below, was used to select the locations with unequal probabilities:

<pre> Data Spring; seed=0; nlaunch=8; do i = 1 to 30; y=round(nlaunch*ranuni(seed) + 0.5); If y>=1 and y<=3 then activity = "PL"; If y>=4 and y<=5 then activity = "EEB"; If Y>=6 and y<=7 then activity = "RO"; If Y=8 then activity = "EEL"; output; end; run; </pre>	<pre> Data Summer; seed=0; nlaunch=8; do i = 1 to 40; y=round(nlaunch*ranuni(seed) + 0.5); If y>=1 and y<=3 then activity = "PL"; If y>=4 and y<=5 then activity = "EEB"; If Y>=6 and y<=7 then activity = "RO"; If Y=8 then activity = "EEL"; output; end; run; </pre>
<pre> Data Fall; seed=0; nlaunch=9; do i = 1 to 30; y=round(nlaunch*ranuni(seed) + 0.5); If Y>=1 and y<=3 then activity = "RO"; If Y>=4 and y<=6 then activity = "EEB"; If Y=7 then activity = "PL"; If Y>=8 and Y<=9 then activity = "EEL"; output; end; run; </pre>	<pre> Data Winter; seed=0; nlaunch=7; do i = 1 to 15; y=round(nlaunch*ranuni(seed) + 0.5); If y>=1 and y<=3 then activity = "EEB"; If y>=4 and y<=6 then activity = "RO"; If Y=7 then activity = "EEL"; output; end; run; </pre>

Randomly, the SAS[®] program determines using a Proc Calendar, 30 days in the Spring, 40 days in the Summer, 30 days in the Fall and 15 days in the Winter. Occasionally, a selected date could not be sampled so the throw of a die was used to select an alternate day. For example, during the month of September 1998, the calendar depicted 9 days where TypeOfDay was Weekend Day or 8 Saturday & Sunday plus Labor Day Holiday. The rest of the month was TypeOfDay = Day. So, the September month displays each of the random locations according to their probabilities shown for the fall,(PL=1/9, RO=1/3, EEL=2/9, & EEB=1/3).

Proc SurveyMeans

The SurveyMeans procedure can produce estimates of survey population totals and variances from sample designs that have stratification, clustering, and unequal weighting. This design requires stratification and unequal weighting to accommodate the unequal probability sampling. It would be of interest to know, in general, how many people spent their day at Rockefeller or the total man-hours in fishing for a day at Rockefeller. Questions pertaining to the amount of time spent at Rockefeller **per day** require variable such as:

A sum of Hours Fished, A sum of Total Miles Driven, Number of Males and
Number of Females, Sum of Party Numbers, Sum of Person Hours Fished,
Average and Sum of Party Size of People.

These tables can be estimated using Proc SurveyMeans. Also, using the Ratio statement, the catch per unit of effort can be estimated. In fact, ratios of totals are the most important outcomes that we computed, according to Rockefeller. The Label statement describes the ratios, in detail:

```

Label  Hrs__Fished_Sum = "Total Party Hrs @ Rock/Day"
        Miles_Sum = "Party Miles Driven/Day"
        No__Male_Sum ="No. of Males/Day"
        Party_No_Sum ="No. of People/Day"
        No__Female_Sum="No. of Females/Day"
        Person_Hrs_Fished_Sum="Person Hrs @ Rock/Day"
        Party_Size_Sum= "No. Parties/Day"
        Person_Miles_Driven_Sum= "Person miles Driven/Day";

```

A Proc SurveyMeans requires a Total option, which lists the stratum sizes or numbers of days within each Year, Season, and Type Of Day. For example, there were 27 days including the holiday, July 3rd, during the weekend days within the Summer 1998. The complete Proc SurveyMeans code is:

```
Proc SurveyMeans Data=George5.Creel_Survey_Final
                  Total=George5.Stratumsizes_20
                  Mean CLM Sum CLSum CV;
Stratum Year2 Season1 TypeOfDay /list ;
By Year2 Season1 TypeOfDay;
Var Hrs__Fished_Sum Miles_Sum Party_No_Sum No__Male_Sum No__Female_Sum
    Person_Hrs_Fished_Sum Party_Size_Sum Person_Miles_Driven_Sum;
Ratio 'Avg. Person Fished HRS ' Hrs__Fished_Sum Miles_Sum / Party_No_Sum;
Ratio 'Avg. Miles per Party ' Miles_Sum /Party_Size_Sum;
Ratio 'Avg. Miles per Person ' Person_Miles_Driven_Sum / Party_No_Sum;
Ratio 'No. of People /Party ' Party_No_Sum /Party_Size_Sum;
Ratio 'Avg. prop. Males ' No__Male_Sum / Party_No_Sum;
Ratio 'Avg. prop. Females ' No__Female_Sum / Party_No_Sum;
Ratio 'Avg. Hrs @ Rock/Person' Person_Hrs_Fished_Sum /Party_No_Sum;
Ratio 'Avg. Hrs @ Rock/Party ' Hrs__Fished_Sum /Party_Size_Sum ;
Weight SamplingWeight;
ODS Output StrataInfo=George5.MyStrata_Creel_Survey_Final_4
            Statistics=George5.MyStat_Creel_Survey_Final_4
            Summary=George5.MySummary_Creel_Survey_Final_4;

Run;
```

The Sampling Weight in the survey is nothing more than the total stratum size divided by the sample size for that Year, Season and Type of Day. The weight for this survey is SamplingWeight. A sampling weight is added to each observation in the data set.

The first ratio gives a 95% confidence interval on the average hours spent fishing, which was between 1.44 hours and 1.59 hours, overall.

Ratio Analysis: Avg. HRS Person Fished					
Numerator	Denominator	Ratio	Std Err	95% Confidence Interval	
Hrs__Fished_Sum	Party_No_Sum	1.514721	0.038192	1.438716	1.590726

The second ratio gives a 95% confidence interval on the average miles traveled per person, which was between 33.66 miles and 39.07 miles, overall.

Ratio Analysis: Avg. Miles per Person					
Numerator	Denominator	Ratio	Std Err	95% Confidence Interval	
Miles_Sum	Party_No_Sum	36.361630	1.358691	33.657749	39.065512

The third ratio gives a 95% confidence interval on the average miles spent per Party, which was between 79.30 hours and 91.11 hours, overall.

Ratio Analysis: Avg. Miles per Party					
Numerator	Denominator	Ratio	Std Err	95% Confidence Interval	
Miles_Sum	Party_Size_Sum	85.203640	2.965576	79.301955	91.105325

The fourth ratio gives a 95% confidence interval on the average miles traveled per person, which was between 85.13 miles and 100.43 miles, overall.

Ratio Analysis: Avg. Miles per Person					
Numerator	Denominator	Ratio	Std Err	95% Confidence Interval	
Person_Miles_Driven_Sum	Party_No_Sum	92.779476	3.843096	85.131471	100.427480

The fifth ratio gives a 95% confidence interval on the No. of People per Party, which was between 2.26 people and 2.42 people, overall.

Ratio Analysis: No. of People /Party					
Numerator	Denominator	Ratio	Std Err	95% Confidence Interval	
Party_No_Sum	Party_Size_Sum	2.343229	0.040030	2.263567	2.422892

The eighth ratio gives a 95% confidence interval on the average hours a person spent at Rockefeller, which was between 3.39 hours and 3.67 hours, overall.

Ratio Analysis: Avg. Hrs @ Rock/Person					
Numerator	Denominator	Ratio	Std Err	95% Confidence Interval	
Person_Hrs_Fished_Sum	Party_No_Sum	3.528263	0.070281	3.388399	3.668127

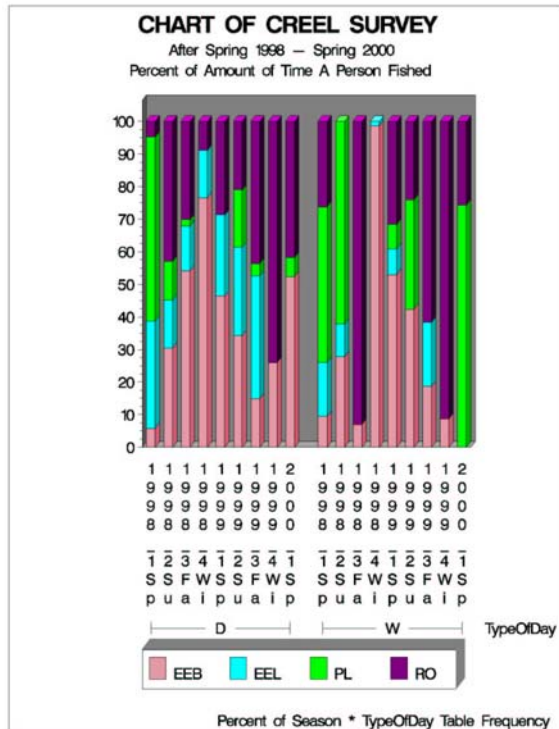
Lastly, the ninth ratio gives a 95% confidence interval on the average hours a party spent at Rockefeller, which was between 3.42 hours and 3.68 hours, overall.

Ratio Analysis: Avg. Hrs @ Rock/Party					
Numerator	Denominator	Ratio	Std Err	95% Confidence Interval	
Hrs_Fished_Sum	Party_Size_Sum	3.549339	0.066818	3.416367	3.682311

GRAPHICS

The Percent of Amount of Time a Person Fished, Percent of Amount of hours Fished per Person and Percent of Number of People in a Trip were displayed as bar charts. Basically, all the graphs, both Period 1 (Before) and Period 2 (After), were done similarly as the SAS® code given below for

the “Percent of Amount of Time A Person Fished”. The graphics include Title, Footnote, Axis, Legend and the Pattern statements. The Gchart vbar3D statement uses variables, Year and Season. The Group and Subgroup options are Type of Day, (W, D) and the Location, (EEB, EEL, PL, RO).



```

/***** ODS Print to my file in PDF format *****/
ods PDF File=
"C:\Sas_Seminar\SCSUG2004\Chart_Creel_Survey_After.pdf";
Title " CHART OF CREEL SURVEY";
Title2 "After Spring 1998 - Spring 2000";
Title3 "Percent of Amount of Time A Person Fished";
footnote1 h=2 j=r "Percent of Season * TypeOfDay Table Frequency";

axis1 label=none origin=(24,);
axis2 label=none order=(0 to 100 by 10);

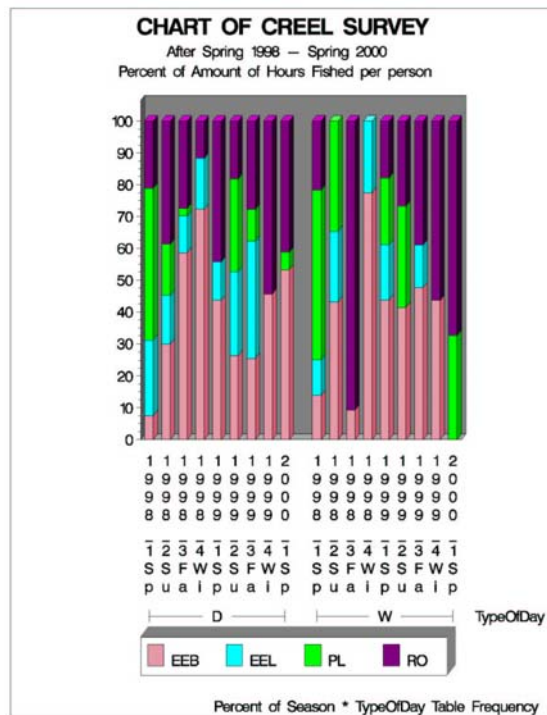
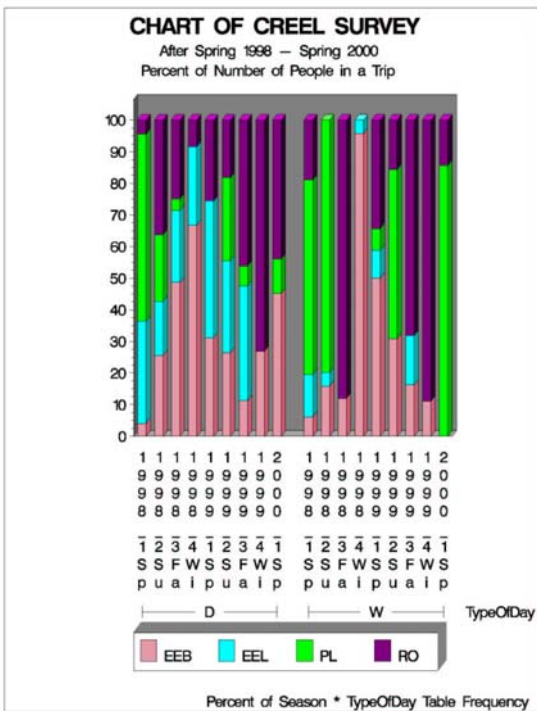
legend1 label=none shape=bar(3,3)
cborder=black cblock=gray origin=(24,);

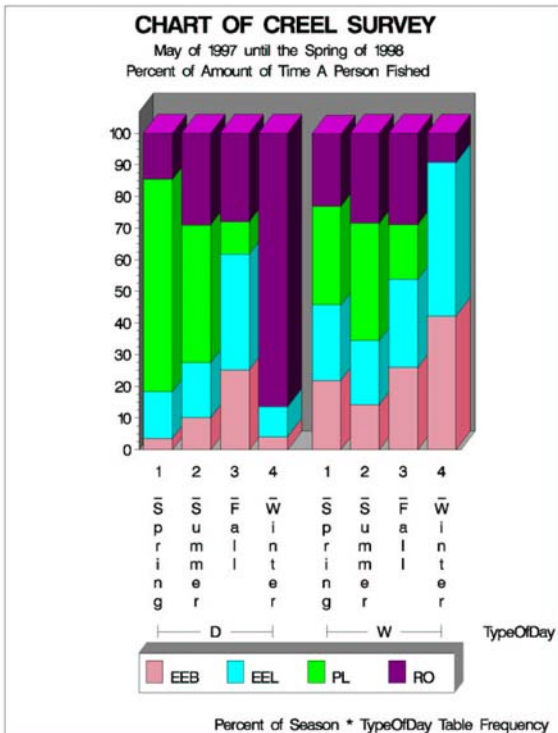
pattern1 color=lipk; pattern2 color=cyan;
pattern3 color=lime; pattern4 color=Purple;

Proc Gchart Data=George5.Creel_Survey_After_26;
vbar3d Year_Season /
group=TypeOfDay sumvar=Percent_Location_Fish
subgroup=Location inside=subject
outside=sum width=5
space=2 maxis=axis1
raxis=axis2 cframe=gray
coutline=black legend=legend1;

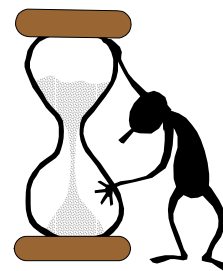
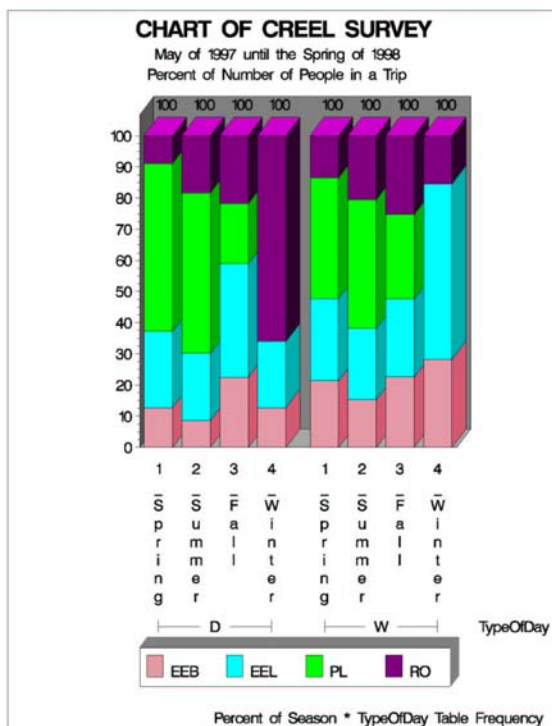
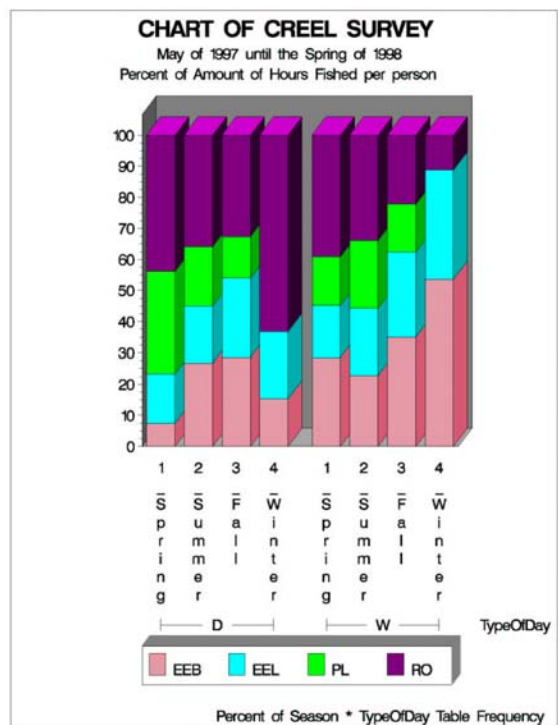
run;
quit;

```





1st Period of Creel Survey.



CONCLUSION

Producing results according to estimation of usage of the refuge by year, season, type of day, and location was accomplished through SAS. Usage of Proc SurveyMeans allowed us to apply sampling weights to the primary sampling unit data. Also, the Proc Gchart displayed how the proportion did not change from year to year but the seasons, type of day and locations were significant.

References

- Groves, Robert M., (1989), "Survey Errors and Survey Costs", New York: John Wiley & Sons, Inc.
- Lohr, Sharon, (1999), "Sampling: Design and Analysis", Brooks/Cole Publishing Company.
- SAS[®] Institute, Inc. SAS/GRAPH[®] Software: Usage, Version 9 Edition, First Edition. Cary, NC: SAS[®] Institute, Inc., 1999.
- SAS[®] Institute, Inc. SAS/STAT[™] Guide for Personal Computers, Version 8 Edition. Cary, NC: SAS[®] Institute, Inc., 1999.
- SAS[®] Institute, Inc. SAS[®] Procedures Guide for Personal Computers, Version 8 Edition. Cary, NC: SAS[®] Institute, Inc., 1999.
- SAS[®] Institute, Inc. SAS[®] Online Doc, Version 9 Edition. Cary, NC: SAS[®] Institute, Inc., 2002.
- Scheaffer, Richard L., Mendenhall, William and Ott, Lyman, (1996), "Elementary Survey Sampling", Fifth Edition, Duxbury Press.

ACKNOWLEDGMENTS

Special thanks to the Rockefeller Wildlife Refuge, especially George Melancon, Thomas J. Hess and the Survey Crew for their most important time the Biologist spent collecting data for the Creel Survey and making the Excel Sheets with all the raw data.

SAS[®] and all other SAS[®] Institute Inc. product or service names are registered trademarks or trademarks of SAS[®] Institute Inc. in the USA and other countries. ® indicates USA registration.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Rebecca Frederick or E. Barry Moser

Louisiana State University
Dept. of Experimental Statistics
161 Agric. Admin. Bldg
Baton Rouge, LA 70803
Work Phone: 225-578-8303
Fax Phone: 225-578-8344
Email: rfrederi@lsu.edu
bmoser@lsu.edu



George Melancon

Biologist 3
Rockefeller Wildlife Refuge
5476 Grand Chenier Hwy.
Grand Chenier, LA 70643
Email: melancon_ge@wlf.louisiana