1999-1 Session - Stat403 - Final Exam

This was the final exam for Stat-403 and was administered on 13 April 1999.

Please answer all questions in the examination booklet. Be concise and succinct - marks will be deducted for excessive wordiness. You are permitted to refer to your notes. Each question is worth equal marks.

---

1. **Estimating salmon catch**

   DFO needs to monitor the catch of sockeye salmon as the season progresses so that stocks are not over fished.

   The season in one statistical sub area in a year was a total of 2 days. On the first day, there were a total of 250 vessels participating, but on the next day only 125 were present.

   In this particular year, observers were randomly placed on selected vessels and at the end of each day the observers contacted DFO managers with a count of the number of sockeye caught on that day. Here are the summary statistics

date	Observers	Mean Catch	std dev of catch
29-Jul-98	15	322	226
30-Jul-98	15	205	135

   (a) What is the population of interest? What is the frame?

   Solution: *The population of interest are the boats participating in the fishery. The response variable is the amount of salmon that each boat catches. The frame is the list of boat licenses or boat ids that participated in the fishery from which the sample was selected. This list may not match the population, if, for example, the sampled boats were selected from those that left from a certain harbor.*

   (b) Estimate the total catch with its standard error.

   Solution:

date	Observers Catch	Mean of Catch	std Dev	Total Boats	Est Total Catch	se total catch
29-Jul-98	15	322	226	250	80,500	14,144
30-Jul-98	15	205	135	125	25,625	4,087
				Total	106,125	14,723

(c) The above survey had an equal allocation of effort among the two dates. How many observers would be allocated to each date under a proportional and optimal allocation?

Solution:

date	Current Allocation Observers	std dev of Catch	Total Boats	NsSh	Allocation of observers under Prop	under Optimal
29-Jul-98	15	226	250	56,500	20	23.1
30-Jul-98	15	135	125	16,875	10	6.9
	30		Total	73,375	30	30.0

(d) Explain why the optimal allocation is based on both the total number of survey units in each stratum and the within stratum variability?

Solution: *The optimal allocation depends upon the total number of survey units in the stratum because strata that have larger number of units presumably contribute more to the total - hence it makes sense to allocate more effort there.*

Strata with a large within stratum standard deviation also receive more effort because a large standard deviation implies that units within the stratum are not very consistent and hence you will need a larger sample size to determine the mean for that stratum to a certain precision than for a stratum with a smaller within stratum standard deviation.

(e) What are the primary reasons for stratification. Under what conditions will stratification be beneficial?

Solution: *The primary reasons are stratification are: (a) sometimes estimates are required for each stratum (b) precision can be increased for the same sample size by stratification (c) sometimes different strata require different survey methods (d) administrative convenience.*

Stratification will work well if the units within a stratum are homogeneous, i.e. have approximately the same value and if the mean strata values are quite different among themselves.

---

2. **Oil spill in the Burrard Inlet**

Did an oil spill on the south side of the Burrard Inlet near the Petro-Canada refinery affect the clam populations. The oil spill occurred in early January. In early February, samples were taken at several random points on the south shore from the soiled sites and samples were taken from several sites on the north shore with the following results:

```
South Shore (oiled) clams/sq m       10, 12, 14, 16, 18
North Shore (control) clams/sq m     22, 24, 26, 28, 30
```

Is this evidence of an environmental impact? Why or why not? How could the design be improved?

Solution:

The major problems in this design are the lack of before measurements, lack of control replicates, and lack of long-term monitoring.

The lack of before measurements implies that the difference between the two sites could have existed before the oil spill took place.

There are no replicates (the repeated values above are pseudo-replicates). Hence we don't know if the observed differences are within normal random variation.

The lack of long-term monitoring implies that we don't know if the observed difference is acute or chronic.

The design could be improved by trying to be before-measurements. [This could be problematical as the impact was no planned.] As well, more control sites should be used including some on the north shore to alleviate criticism that the observed difference is simply a north vs. south shore effect. The monitoring program should continue for several months to see if this a chronic or acute effect.

3. **Some questions on experimental design**

   (a) What considerations would you employ to determine if a factor is a fixed or random effect? Give examples of each.

   Solution: *A factor is treated as fixed if the levels used in the experiment are of primary interest and if the experiment were to be repeated, the same levels would be chosen. For example, temperature (5C or 10C) may be a fixed effect.*

   A factor is treated as random, if the levels used in the experiment are not of primary interest, were chosen at random from a larger set of levels, and if new levels would typically be used in a new experiment. For example, people are often random effects as the actual people used in an experiment are not likely of primary interest, they were chosen from a large set of subjects, and likely new people would be used in a new experiment.

   (b) What is the difference between a "block" and a "factor".

   Solution: *Blocks are typically not manipulated but rather are a collection of experimental units that are similar. You typically don't randomize the block assemblies, and you usually assume that blocking effects don't interact with Factor effects.*

A factor has levels that are manipulated and randomized over the experimental units.

(c) Hurlbert (1984) wrote an important paper on the dangers of pseudo-replication. Therefore:

- Give an example of pseudo-replication
- What are the two primary dangers introduced by pseudo-replication?

Solution: *An example of pseudo-replication is the fish within a tank illustration used many times in class. The treatment levels are applied to the tank, while the fish within a tank would be pseudo-replicates.*

The primary dangers introduced by pseudo-replication are that the experimenter will treat the pseudo-replicates as real, independent replicates when they clearly are not independent. This typically leads to underestimates of the true variations present in the experiment, underestimates of the true standard errors of the estimates, increased Type I error rates over the nominal alpha-value, and attempts to make inferences outside the true scope of the experiment, i.e. there may be no real replicates.

(d) Day and Quinn (1989) wrote a nice review paper on multiple-comparison techniques. Therefore:

- Why is a multiple-comparison technique needed after an ANOVA?
- What does a multiple-comparison technique control?
- Explain the logic behind the Bonferonni method for multiple comparisons.
- Explain how the following results could be obtained: "After the ANOVA showed that there was strong evidence of a difference in the treatment means, a multiple comparison technique showed that the mean of level a1 did not appear to be different than the mean of level a2; the mean of level a2 did not appear to be different than the mean of level a3; yet there was strong evidence the means of levels a1 and a3 differed."

Solution: *A multiple comparison procedure is needed after an ANOVA because the ANOVA test-statistic just indicates that there is evidence that some of the means differ from each other - it does not give any indication which mean may be different from what other mean.*

The multiple-comparison procedures control the experimentwise-error rate, i.e the probability that at least one Type I error will be made in all the comparisons made from the experiment.

The Bonferonni procedure uses that if you make k comparisons, each with a Type I error rate of α, then the approximate probability of at least one Type I error among all k comparisons is $k \times \alpha$. Hence, to control the overall Type I error rate to α, you need to make each comparison at the α/k level.

The last puzzle is caused by the fact that the means are compared only two at a time. Consider the following diagram of the three means with 95% confidence intervals also shown:

```
<----- a1 ----->
      <------a2  ----->
                  <----- a3 ------>
|-----|-----|-----|-----|-----|-----|
```

Because the c.i. for a1 and a2 overlap considerably, the two means could be the same. Similarly, the ci for a2 and a3 also overlap, so their respective true population means could be the same. However it is quite clear that the means for a1 and a3 are quite different.

4. **Experimental Design**

There are two primary costs in fish farming - labour to tend the fish and feed for the fish. Naturally, fish farmers with to maximize feeding efficiency, i.e. what fraction of the food fed to the fish is turned into sellable flesh.

Several factors control the growth of fish, many of which are uncontrollable by the fish farmer. However, two factors can be easily controlled - the stocking density and the type of food. The BC Association of Fish Farmers wishes to conduct an experiment to investigate these two factors.

There are three traditional stocking densities - for convenience call these high (H), medium (M), and low (L). There are two types of fish food - for convenience call these seal-based (s) and beef-based (b).

The association has two test sites off the west coast of Vancouver Island. At each site, there are 3 separate cages that have three internal partitions in each cage. The density and feed type can be manipulated separately for each partition of the cage.

Design the experiment as a CRD, RCB, split-plot with main plots in a CRD, and a split-plot with main plots as an RCB. Be sure to describe in sufficient detail how you randomized so that it is clear how each experiment was done. Note that for some designs, you many not be able to use all the partitions in all the cages. Use the following template when showing the experimental layouts.

```
                            Site 1                  Site 2
                Cage 1  Cage 2  Cage 3  Cage 1  Cage 2  Cage 3
Partition 1
Partition 2
Partition 3
```

Solution: CRD

There are 6 treatment combinations. Randomize these completely at random over the 18 cage-partitions among the two sites.

		Site 1			Site 2	
	Cage 1	Cage 2	Cage 3	Cage 1	Cage 2	Cage 3
Partition 1	Hb	Mb	Ms	Hs	Ms	Lb
Partition 2	Hs	Lb	Hb	Mb	Ls	Ls
Partition 3	Hs	Lb	Ls	Hb	Mb	Ms

RCB

A reasonable blocking factor would be site. In a traditional RCB, each treatment combination appears only one in each block - hence there will be three vacant cage-partitions vacant in each block. Each block is randomized separately.

		Site 1			Site 2	
	Cage 1	Cage 2	Cage 3	Cage 1	Cage 2	Cage 3
Partition 1	Hb	Mb	Ms	Hs	Ms	
Partition 2		Ls			Hb	Lb
Partition 3	Hs		Lb	Mb	Ls	

Split-plot design with main-plots as a CRD

This can be done in a number of ways. The main-plot experimental unit is typically the cage.

In this first solution, the stocking density is assigned complete at random to the six cages and the two food types randomized to the partitions within a cage. The main-plot factor is stocking density, the sub-plot factor is food-type. Because there are only two food types, one partition in each cage is left blank.

		Site 1			Site 2	
	Cage 1	Cage 2	Cage 3	Cage 1	Cage 2	Cage 3
Partition 1	Hb	Lb	Ls		Hs	Mb
Partition 2		Ls		Mb	Hb	Ms
Partition 3	Hs		Lb	Ms		

In this second solution, the food type is randomized to cage, and the stocking densities within each cage to the partitions. The main-plot factor is food type, the sub-plot factor is stocking density.

		Site 1			Site 2	
	Cage 1	Cage 2	Cage 3	Cage 1	Cage 2	Cage 3
Partition 1	Hb	Ms	Ls	Lb	Hb	Hs
Partition 2	Mb	Hs	Ms	Hb	Lb	Ls
Partition 3	Lb	Ls	Hs	Mb	Mb	Ms

Split-plot design with main-plots in an RCB

This can be done in a number of ways. The main-plot experimental unit is typically the cage. The blocks are the sites.

In this first solution, the stocking density is assigned at random to a cage within each site. The two food types are randomized to the partitions within a cage. Each site was all stocking densities.

	Site 1			Site 2		
	Cage 1	Cage 2	Cage 3	Cage 1	Cage 2	Cage 3
Partition 1	Hb	Lb		Ls	Hs	Mb
Partition 2		Ls	Mb		Hb	Ms
Partition 3	Hs		Ms	Lb		

In this second solution, the food types are randomized to cages within each site and the stocking densities randomized within each cage. Because there are only two levels of food type, one cage is left vacant in each site.

	Site 1			Site 2		
	Cage 1	Cage 2	Cage 3	Cage 1	Cage 2	Cage 3
Partition 1		Hb	Ms	Lb	Hs	
Partition 2		Mb	Hs	Hb	Ls	
Partition 3		Lb	Ls	Mb	Ms	

5. **Effect of gender and handedness on completing a maze**

   Cast your mind back to the tutorial where members of both genders completed mazes using both left and right hands.

   In this experiment, students worked through a set of pencil and paper mazes two times - once with their left hand and the other time with their right hand. The order of the mazes was randomized.

   The time to complete the maze was recorded in seconds.

   Here is some output from the experiment:

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob>F
Gender	275.009	275.009	1	39.2642	0.0002
subject[Gender]	56.0326	7.00407	8	0.5370	0.8012
hand	277.148	277.148	1	21.2497	0.0017
hand*Gender	31.6509	31.6509	1	2.4268	0.1579

Least Squares Means

Level	Least Sq Mean	Std Error	Mean
f	31.09047363	0.8369031852	31.0905
m	38.50679595	0.8369031852	38.5068

Warning: Std Err calculated with respect to Synthetic Denominator.

Tests wrt Random Effects

Source	SS	MS Num	DF Num	F Ratio	Prob>F
Gender	275.009	275.009	1	39.2642	0.0002
subject[Gender]	56.0326	7.00407	8	0.5370	0.8012
hand	277.148	277.148	1	21.2497	0.0017
hand*Gender	31.6509	31.6509	1	2.4268	0.1579

LSMeans Differences Tukey HSD

Alpha= 0.050 Q= 2.30593

LSMean[j]

LSMean[i]	Mean[i]-Mean[j] Std Err Dif Lower CL Dif Upper CL Dif	f	m
	f	0 0 0 0	-7.4163 1.18356 -10.146 -4.6871
	m	7.41632 1.18356 4.68712 10.1455	0 0 0 0

Level			Least Sq Mean
m	A		38.506796
f		B	31.090474

Levels not connected by same letter are significantly different

(a) Write the statistical model for this experiment.

Solution:

$$Y_{ijk} = \mu + g_i + h_k + (hg)_{ik} + id(g)_{ij} + epsilon_{ijk}$$

where μ is the overall mean
g_i is the effect of gender i - a treatment factor
h_k is the effect of handedness - a treatment factor
$(hg)_{ik}$ is the effect of the handedness-gender interaction - an interaction between treatment factors
$id(g)_{ij}$ is the effect of person nested within each gender - random effect - an experimental unit term
ϵ_{ijk} is the unexplained random variation.

This can also be written in the simplified syntax as:

$$Y = Gender \;\; Handedness \;\; Gender * Handedness \;\; id(Gender) - R$$

(b) Perform a test for the main effects of gender. Indicate the null and alternate hypothesis, the test-statistic, the p-value, and your conclusion.

Solution:
H: $\mu_f = \mu_m$
A: $\mu_f \neq \mu_m$
The test statistics is F=39.3, the p-value is .0002.

We conclude that there is very strong evidence that the mean response for both genders are not equal.

(c) Interpret the p-value from (b).

Solution: *The p-value measures the probability of observing this set of data if the null hypothesis is true, i.e. the probability if only .0002 of observing this large (or larger) of a difference in the sample mean time to complete the maze between the two genders if, in fact, the two genders has the same mean time to complete the maze.*

(d) Estimate the size of the gender effect and find an approximate 95% confidence interval for it. Interpret the confidence interval.

Solution: *The estimated difference in the mean time to completion is 7.4 seconds with a se of 1.18 minutes. An approximate 95% confidence interval is found as* $7.4 \pm 2(1.18)$ *seconds. We would be 95% confident that the true difference in the mean time to completion between the two genders.*

(e) The test for gender used subject(gender) as the denominator for the test-statistic, while the tests for handedness or for the interaction used MSE as the denominator of the test-statistics. Give an intuitive explanation for this.

Solution: *In general, every effect is tested using an F-test whose denominator corresponds to the experiment unit upon which the factor operates. In this case, the factor gender operates on the whole-plot unit of people.*

The reason for this is that the F-test measures the ratio of the variation among the experimental unit explainable by the factor to that *not explained. As all subjects did both mazes with both hands, the handedness is not an issues to see if differences exist between the two genders.*