2016-1 Session - Stat-403/ 650/ 890 (BPK)
Final Exam

Answer all questions in the exam booklets. All parts have equal worth. Be succinct - if you are writing a thesis, you are doing far too much!

How does height in a tree influence the success of a nest?

Birds build nests at different heights based on availability of nesting platforms, nesting material etc. Does the height at which a nest is build, influence the probability that no eggs will hatch or chick fledge, due to predation. The theory is that that the higher the nest is above the ground, the less likely a predator (such as a rat) will be able to find the nest and eat the eggs.

Selected output is presented at the end of the exam.

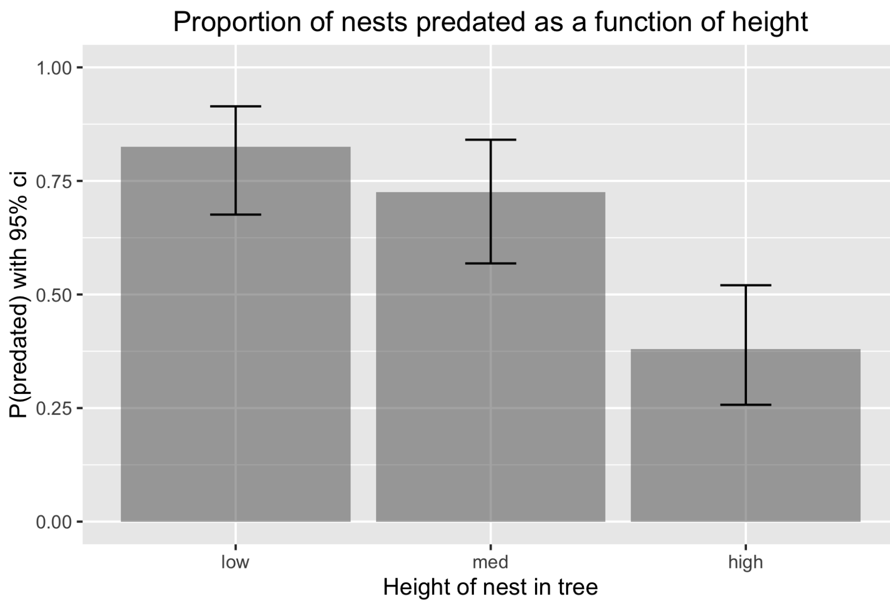
1. What is the experimental unit? What is the observational unit? Was there pseudo-replication and how was it accounted for?

   Solution: The experimental unit is the nest; the observational unit is the nest. There is NO pseudo-replication.

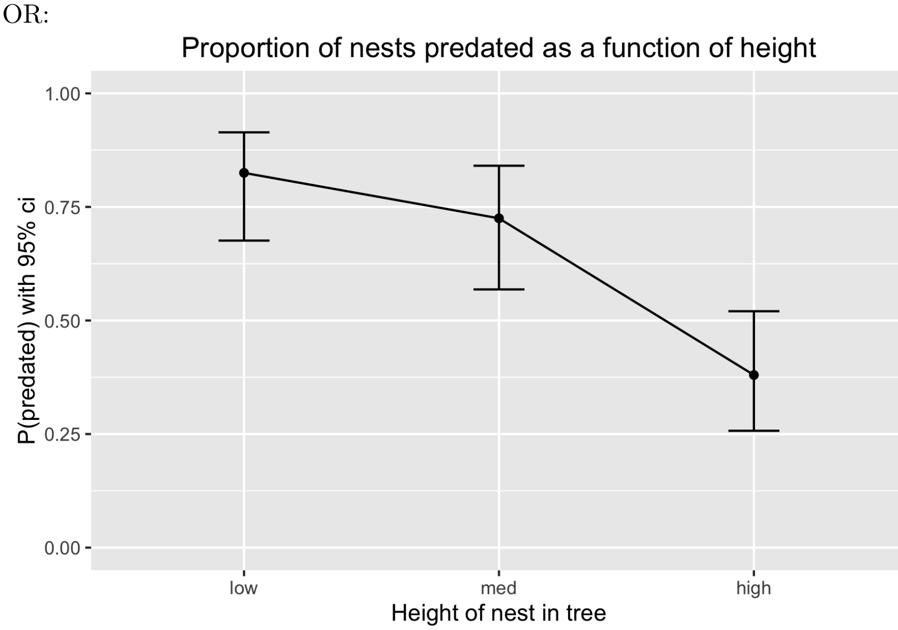
   If you wrote that the observations unit was the chick and concluded that pseudo-replication occurred that would be ok, BUT ONLY IF you then mentioned that analysis for the first 3 questions was at the nest levels, i.e. the values from the chicks were rolled up into a yes/no at the nest level.

2. Draw a suitable graph to display the results. Don't forget a measure of uncertainty.

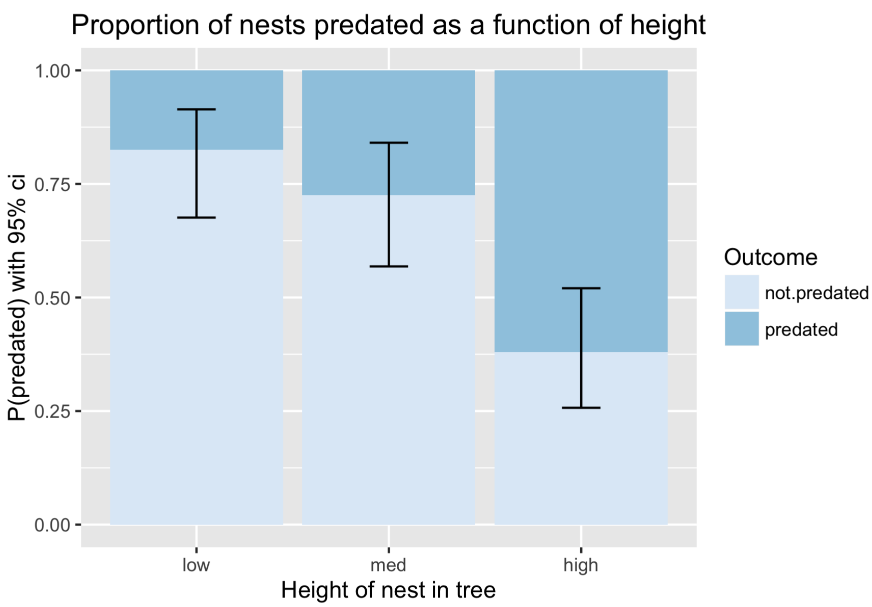
   Solution: You could draw a bar chart or a segmented bar chart. Don't forget to add the confidence interval bars.


Proportion of nests predated as a function of height
1.00
0.75
0.50
0.25
0.00
P(predated) with 95% ci
low
med
high
Height of nest in tree

OR:


Proportion of nests predated as a function of height
1.00
0.75
0.50
0.25
0.00
P(predated) with 95% ci
low
med
high
Height of nest in tree

OR:



Many students drew a mosaic chart but did not add confidence intervals to the bars.

3. What do you conclude based on your plot and the formal test in the output?

   Solution: There is evidence that the PROPORTION of nests that are predated varies among the different heights ($p < .0001$). Furthermore, there is no evidence that the proportion of nests that are predated is different at the two lowest levels of height, but there is evidence that the proportion predated is less at the highest elevation.

   Some students wrote the *mean proportion* varied among heights. You don't have to say "mean" and it makes no sense in this context.

   Some students didn't understand the *R* format for scientific notation and thought that a value of *3.2e-8* on the *R* output was $3.2 \times e^{-8}$ rather than 3.2×10^{-8} or .000000032.

4. The experimenter also measured the fate of the individual eggs within the nest. For example, not all eggs are necessarily predated in a nest. The experimenter repeated the analysis using the individual eggs as shown on the output. The reviewer wrote back "See Hurlbert (1984)". Why? Be specific. [Hint: what are the experimental and observations units and why?]

   Solution: The experimental unit is the *nest*, but the observations unit is the individual *egg*. There are multiple readings on each nest. This is either a case of simple pseudoreplication (but fixable) or sacrificial pseudoreplication (you threw away the information about which eggs is in which nest).

5. Explain how you would analyze the data at the egg level. Be specific on any data summarization that needs to be done. A small example with a few nests may be helpful. [It is not necessary to do the hand computations for the analysis, but give the appropriate analysis method and model, if needed.]

   Solution: Because of the pseudo-replication, you need to convert the egg measurements to a single number per nest. This is usually done by averaging, if the response variable is continuous. In the case when the response variable is categorical (i.e. predated/ not predated), the "averaging" is equivalent to finding a sample proportion for each nest.

   For example, here are some sample computations:

Height	Nest	Eggs	Predated	prop
High	1	4	2	0.50
	2	3	1	0.33
	3	2	0	0.00
...				
Medium	51	3	1	0.33
	52	4	2	0.50
...				

   Then the sample proportions (the last column in the above table) are now used in a single factor ANOVA using the model

$$prop \sim Height$$

   The followup analysis will be the standard Tukey multiple-comparison procedure.

   You also convert the last column to a logit (you would need to use the empirical logit transformation to avoid taking the logarithm of 0/1 or 1/0. Notice that this is NOT a logistic ANOVA, it is a regular ANOVA on the logits. A logistic ANOVA would be used if you actually modelled the individual eggs data (it would be a mixed effects logistic ANOVA – way beyond the scope of this course).

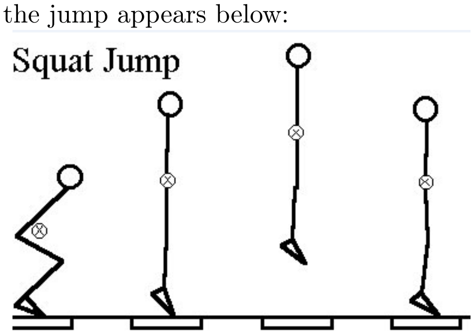
   You CANNOT use a simple chi-square test as shown on the output on the raw egg data because of the pseudo-replication. Nor can you use the simple logistic ANOVA for the same reason.

Does warming up before exercise improve performance?

In many sports, success is often highly related to muscular strength and power output. Participants who repeatedly perform at maximum power results in improved performance. For example, in basketball, if you can repeatedly jump higher than other participants, then you will do better!

What is the impact of different warm-up strategies on maximal strength and power? Do the different warm-strategies perform similarly for males and females?

The impact of three different warm up strategies (*none*, *static stretching*, or *dynamic stretching*) in the two sexes was assess using vertical jump heights. So for a particular trial, the participant does a warm-up strategy, and then does a vertical squat jump (similar to what you would do in basketball). A schematic of the jump appears below:



The height of jump is measured by looking at the difference in height of a tag on the person (indicated by an X in the above schematic) between the maximum during the jump and the height when the person is standing.

The actual experiment was done as a split-plot design.

6. Illustrate how this experiment could be run as a CRD, RCB, or split-plot design. How many participants will be needed to get 60 measurements under each design?

   Solution: In a CRD, each person would perform the jump after one type of stretching. So you would need 60 participants, 10 in each combination of sex and warm-up strategy.

   In an RCB, you could block by day of the week (e.g. Monday, Tuesday...) on the premise that perhaps people have more "energy" on Mondays after resting on the weekend, etc. Now on each day, you would test 6 people (both sexes x 3 strategies). You would stil need 60 people. You cannot block by subject directly because people cannot be both genders.

   In split-plot design, you would have test each subject using all three strategies. You need 10 males and 10 females, each tested three times. You cannot do the split-plot in the other direction (e.g. warm-up strategy at the main-plot level) because people cannot be both genders.

Some students said they would block by subject and have each subject do every warm-up method. While this is a form of blocking, it would NOT be an appropriate RCB design for this question because you have "lost" the gender factor. If you now classify each subject by gender, you are back to the split-plot design.

Some students said they would block by "sex". *Sex* is a factor; it cannot serve as a block.

7. Write the statistical model for the CRD, RCB, and split-plot designs.

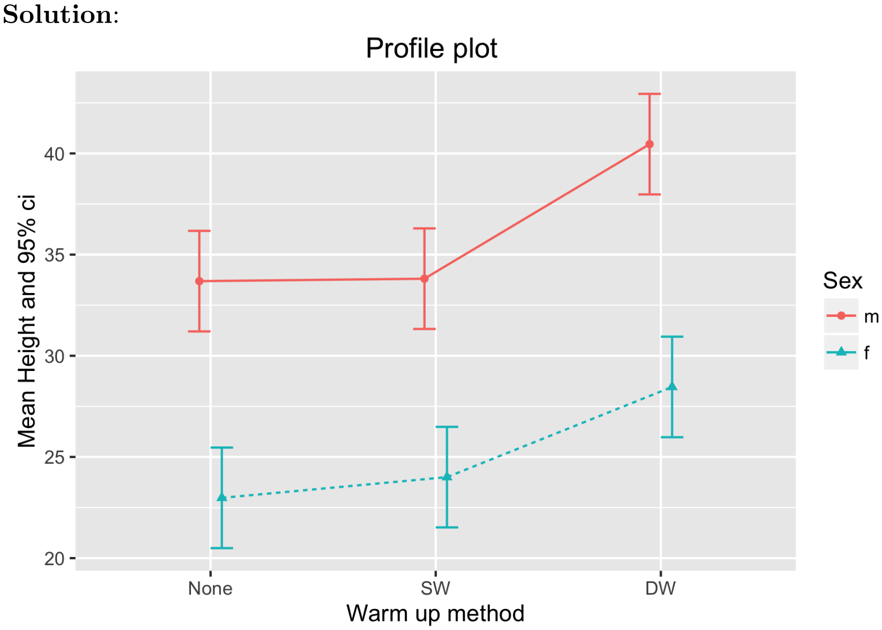
   Solution:
   CRD: $Height \sim Gender + Method + Gender : Method$
   RCB: $Height \sim Gender + Method + Gender : Method + Day(R)$
   SPD: $Height \sim Gender + Method + Gender : Method + Subject(R)$

   Declaring the *Day* as a random effect in the RCB model is optional if the blocks are complete.

   Subject must be declared as a random effect in the split-plot model because it is serves both a block for the methods (within a person) and as the experimental unit for *Gender*.

8. Draw a profile plot. Don't forget measures of uncertainty on the plot. What does the profile plot seem to indicate?

   Solution:



   There is clear gender effect; the effect of warmup method is less clear; the lines are nearly parallel so there does not appear to be any evidence of interaction.

There is evidence of a gender main effect (males tend to jump higher (on average) than females for all warm-up methods in a consistent amount).

There is evidence of a main effect of method with mean jumping heights after a dynamic warmup consistently higher than the mean heights under the other methods.

Some students drew the profile plot without confidence intervals; this makes it hard to interpret because you cannot tell without the confidence intervals if the apparent main effect of gender is "real" or not.

9. Show how the main effect of sex is estimated. Show how the main effect of *none* vs. *static warm-up* is estimated. It is not necessary to do the actual computations if you clearly show how they should be computed.

   Solution:
   We need the effect of sex averaged over all methods:

   $$((33.69 - 22.98) + (33.81 - 24.00) + (40.46 - 28.46))/3 = 10.84$$

```
 Sex   lsmean        SE df lower.CL upper.CL .group
 f   25.14752 0.9573865 18 23.13613 27.15892  1
 m   35.98597 0.9573865 18 33.97457 37.99736   2

Results are averaged over the levels of: Method
Confidence level used: 0.95
significance level used: alpha = 0.05
```

```
 contrast estimate       SE df t.ratio p.value
 m - f    10.83844 1.353949 18   8.005  <.0001

Results are averaged over the levels of: Method
```

   We need the effect of *none* vs. *static warm-up* averaged over both sexes

   $$((33.81 - 33.69) + (24.00 - 22.97))/2 = .57$$

```
 Method   lsmean        SE    df lower.CL upper.CL .group
 None   28.33397 0.8696068 40.46 26.57705 30.09089  1
 SW     28.90662 0.8696068 40.46 27.14970 30.66354  1
 DW     34.45966 0.8696068 40.46 32.70274 36.21658   2

Results are averaged over the levels of: Sex
Confidence level used: 0.95
P value adjustment: tukey method for comparing a family of 3 estimates
significance level used: alpha = 0.05
```

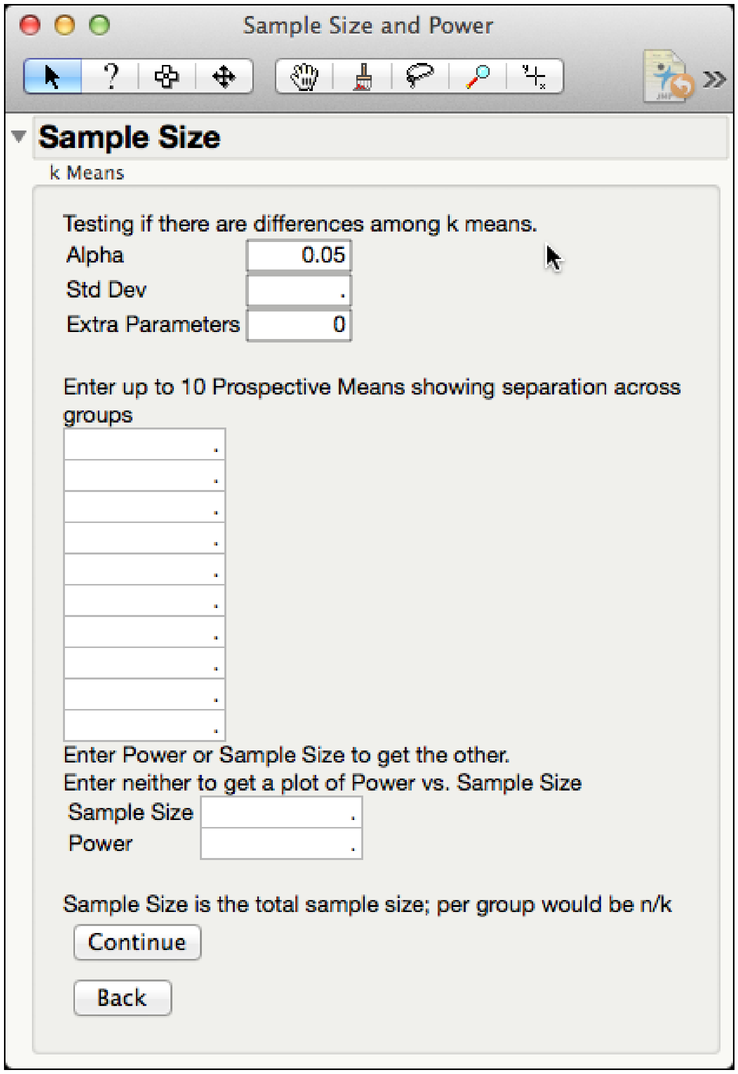
```
 contrast    estimate        SE df t.ratio p.value
 None - SW -0.5726501 0.9453913 36  -0.606  0.8179
 None - DW -6.1256910 0.9453913 36  -6.480  <.0001
 SW - DW   -5.5530410 0.9453913 36  -5.874  <.0001

Results are averaged over the levels of: Sex
P value adjustment: tukey method for comparing a family of 3 estimates
```

Some students said that you need to average all of the readings for *males* and all of the readings for *females* and then take the difference in the means. This will NOT work if the design is unbalanced. You need to say that you first find the means of each combination of *sex* and *method*; then take the means of the means for each *sex*; and lastly take the difference of the means of means.

10. Suppose that when subjects do a trial with a warm-up method, they were measured for five jumps. How would any analysis have to be modified and why?

    Solution: Pseudo-replication has again raised its ugly head. You would take the average (or maximum) of the heights over the five jumps as the response measurement.

11. A biologically important difference is 5 cm in the mean jump height among warm-up methods. Complete the power/sample size box or complete the appropriate *R* code below:



OR

```
group.means <- c( .... complete here ....)
sd <- .....
power <- power.anova.test(groups=length(group.means),
          within.var=sd**2,
          between.var=var(group.means),
          power=......,
          sig.level=......)

group.means <- c(10,15,12.5)
sd <- mean(report$sd)
```

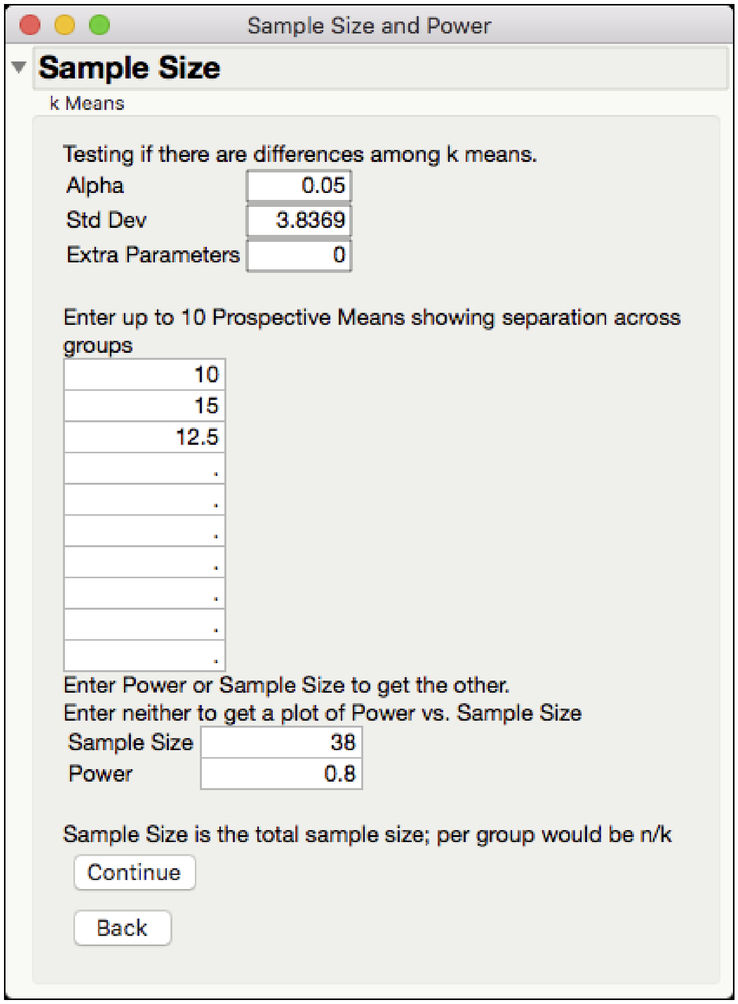
```
power <- power.anova.test(groups=length(group.means),
                          within.var=sd**2,
                          between.var=var(group.means),
                          power=0.8,
                          sig.level=.05)
power
```

which gives:

```
     Balanced one-way analysis of variance power calculation

         groups = 3
              n = 12.40268
    between.var = 6.25
     within.var = 14.72209
      sig.level = 0.05
          power = 0.8

NOTE: n is number in each group
```

Or a total of 13 subjects.

In *JMP*, we get:



Or, again 13 subjects (38/3) in total.

Some students gave three means that differed by 10? Some students gave more than 3 means – but there are only three warm-up methods.

12. A researcher noted that jumping height is related to the height of the subject.

    They fit the model

    $$JumpHeight \sim Method + Subject(R) + SubjectHeight$$

    and found strong evidence of a relationship between the jump height and the subject height.

    They the fit the model

    $$JumpHeight \sim Sex{+}Method{+}Sex : Method{+}Subject(R){+}SubjectHeight$$

where *SubjectHeight* is the height of the subject in cm. Rather surprisingly, once the covariate was added, the effects of both *sex* and *subject height* "disappeared". Why?

Solution This usually indicates something related to colinearity. While colinearity is not well defined when one variable is categorical (*sex*) and one variable is continuous (*subject height*), the same principles apply. There is a strong relationship between sex and height, i.e. females tend to be smaller than males. So when you add both variables to the model, each variable masks the effect of the other variable (remember everything is marginal).

13. Under what conditions is it better to run this as a split-plot design rather than a CRD and vice-versa.

    Solution: There are several reasons:

    - **Large subject-to-subject variation**. If there a wide range in jumping ability among subjects of the same sex (due to general athletic ability, weight, etc), then it is GOOD to block by subject as much as possible when comparing methods, and so a split-plot design is preferred. This way each person serves as his/her own control.

      If the among-subject variation is small, then there is little advantage to blocking by subject (but it seldom hurts you).
    - **Logistical concerns** A CRD might be easier to run to logistical reasons. For example, subjects may find it inconvenient to come in three times to the testing center. Or, if it is difficult for subjects to do all three methods, then a CRD would be preferred.
    - **Carry-over effects**. Subjects may "learn" to jump better over time. Or if the three methods occur too close together, the impact of a method may carry-over to the next trial. In this case, a CRD would be preferred.
    - **Interested in sex effect**. A split-plot design will result in better precision for the within-subject factor (*method*) rather than the among subject factor (*sex*). If you are really interested in the *sex* effect, then a CRD would results in equal precision for both factors.

    Some student said that a CRD would be more "representative" because it has a larger number of subjects compared to the split-plot design. Remember the RRRs – randomization is what controls representation; a larger sample size does NOT make a sample more representative.

14. Write a brief MATERIALS AND METHODS and RESULTS section using ALL of the analyses.

    Solution: *Methods* A split-plot design was used to investigate the impact of gender and warm-up method on the height achieved in squat jumps.

Twenty subjects (10 males; 10 females) were assigned three warm-up methods (none, static, and dynamic) in random order. In each trial, the subject warmed up, and then did a squat-jump. The height of the jump was measured. A split-plot ANOVA will be performed with sex as the main-plot factor and method as the split-plot factor. Subjects are both the experimental unit for the gender factor and the block for the method factor. A Tukey multiple-comparison procedure will be used to investigate differences in the means among factor levels.

Results All 10 subject completed the three methods of warm-up. Table 1 summarizes the results – standard deviations are comparable across all treatments. No outliers were detected and there were no missing values.

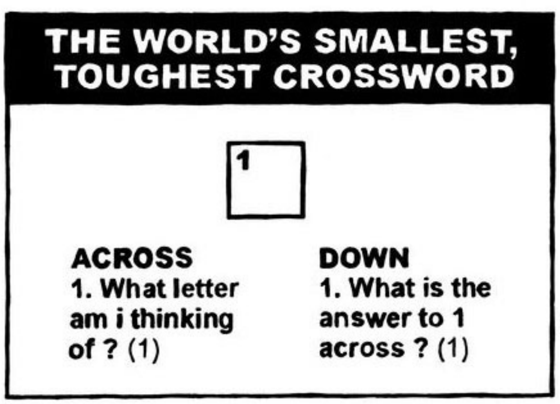
There was no evidence of an interaction between the gender and method of warm-up on the mean jump height ($p = 0.52$).

Males tended to jump an average of 11 (SE 1.4; $p < .0001$) cm higher than females when averaged over all three warm-up methods.

There was evidence that the mean jump height varied among warm-up methods ($p < .0001$). The Tukey comparison found no evidence of a difference in mean height between *no warm-up* and *static warm-up* ($p =$ .82; diff in means is 0.57 (SE .94) cm). There was evidence that the *dynamic warm-up* increases mean jump height compared to both of the other methods with an increase of 6.12 (SE .94) cm (*DW* vs. *none*), and an increase of 5.55 (SE 94) cm (*DW* vs. *SW*).

Some students didn't understand the *R* format for scientific notation and thought that a value of *3.2e-8* on the *R* output was $3.2 \times e^{-8}$ rather than 3.2×10^{-8} or .000000032.

15. The last question of the term.



Solution: The solution is *U*.

This page intentionally left blank

Nest Success Output at NEST level

CAUTION: *R* has a funny way to write small *p*-values. A value of *3.2e-8* is actually 3.2×10^{-8} or .000000032.

```
  Height Nests predated lower.conf.limit upper.conf.limit
1    low    40       33        0.6759031        0.9142123
2    med    40       29        0.5684091        0.8406989
3   high    50       19        0.2571948        0.5203646
```

```
        Pearson's Chi-squared test

data:  success[, c("success", "predated")]
X-squared = 21.293, df = 2, p-value = 2.378e-05
```

Tukey multiple comparison.
CAUTION: *R* uses LCL and UCL to refer to the lower and upper confidence limits respectively.

```
 Height  prob         SE df asymp.LCL asymp.UCL .group
 high   0.380 0.06864401 NA 0.2571948 0.5203646  A
 med    0.725 0.07060011 NA 0.5684091 0.8406989   B
 low    0.825 0.06007807 NA 0.6759031 0.9142123   B

Confidence level used: 0.95
Intervals are back-transformed from the logit scale
P value adjustment: tukey method for comparing a family of 3 estimates
Tests are performed on the log scale
significance level used: alpha = 0.05
```

Nest Success Output at EGG level

```
  Height Nests Eggs Eggs.predated eggs.lower.conf.limit eggs.upper.conf.limit
1    low    40   59            25             0.3049451             0.5520333
2    med    40   62            34             0.4240858             0.6669306
3   high    50   67            40             0.4762354             0.7070740
```

```
	Pearson's Chi-squared test

data:  success2[, c("Eggs.success", "Eggs.predated")]
X-squared = 3.9552, df = 2, p-value = 0.1384
```

Tukey multiple comparison.

```
 Height      prob         SE df asymp.LCL asymp.UCL .group
 low    0.4237288 0.06433264 NA 0.3049451 0.5520333  A
 med    0.5483871 0.06320202 NA 0.4240858 0.6669306  A
 high   0.5970149 0.05992383 NA 0.4762354 0.7070740  A

Confidence level used: 0.95
Intervals are back-transformed from the logit scale
P value adjustment: tukey method for comparing a family of 3 estimates
Tests are performed on the log scale
significance level used: alpha = 0.05
```

Jumping Output

```
  Sex Method n.values     mean       sd
1   m   None       10 33.68862 3.682664
2   m     SW       10 33.81018 4.083577
3   m     DW       10 40.45911 3.263790
4   f   None       10 22.97931 4.918817
5   f     SW       10 24.00305 4.106099
6   f     DW       10 28.46020 2.966675
```

```
Analysis of Variance Table of type III  with  Satterthwaite
approximation for degrees of freedom
           Sum Sq Mean Sq NumDF DenDF F.value    Pr(>F)
Sex        572.73  572.73     1    18  64.081 2.429e-07 ***
Method     457.92  228.96     2    36  25.618 1.205e-07 ***
Sex:Method  12.13    6.07     2    36   0.679    0.5136
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
 Sex   lsmean        SE df lower.CL upper.CL .group
 f   25.14752 0.9573865 18 23.13613 27.15892  1
 m   35.98597 0.9573865 18 33.97457 37.99736   2

Results are averaged over the levels of: Method
Confidence level used: 0.95
significance level used: alpha = 0.05
```

```
 contrast estimate       SE df t.ratio p.value
 m - f    10.83844 1.353949 18   8.005  <.0001

Results are averaged over the levels of: Method
```

Jumping Output - continued

```
 Method   lsmean        SE    df lower.CL upper.CL .group
 None   28.33397 0.8696068 40.46 26.57705 30.09089  1
 SW     28.90662 0.8696068 40.46 27.14970 30.66354  1
 DW     34.45966 0.8696068 40.46 32.70274 36.21658   2

Results are averaged over the levels of: Sex
Confidence level used: 0.95
P value adjustment: tukey method for comparing a family of 3 estimates
significance level used: alpha = 0.05
```

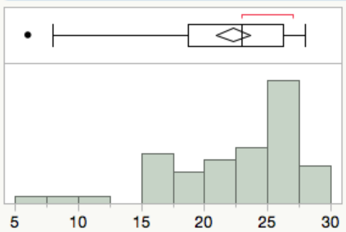
```
 contrast    estimate        SE df t.ratio p.value
 None - SW -0.5726501 0.9453913 36  -0.606  0.8179
 None - DW -6.1256910 0.9453913 36  -6.480  <.0001
 SW - DW   -5.5530410 0.9453913 36  -5.874  <.0001

Results are averaged over the levels of: Sex
P value adjustment: tukey method for comparing a family of 3 estimates
```

Tukey multiple comparison:

```
 Sex Method   lsmean      SE    df lower.CL upper.CL
 m   None   33.68862 1.22981 40.46 31.20396 36.17328
 f   None   22.97931 1.22981 40.46 20.49466 25.46397
 m   SW     33.81018 1.22981 40.46 31.32552 36.29484
 f   SW     24.00305 1.22981 40.46 21.51840 26.48771
 m   DW     40.45911 1.22981 40.46 37.97445 42.94377
 f   DW     28.46020 1.22981 40.46 25.97555 30.94486

Confidence level used: 0.95
```

Statistics about student performance

Final Exam - 403/ 650/ 890 - 2016 Spring - Out of 28



Quantiles

100.0%	maximum	28
99.5%		28
97.5%		28
90.0%		28
75.0%	quartile	26.25
50.0%	median	23
25.0%	quartile	18.75
10.0%		16
2.5%		6.95
0.5%		6
0.0%	minimum	6

Summary Statistics

Mean	22.275862
Std Dev	5.1291841
Std Err Mean	0.6734949
Upper 95% Mean	23.624512
Lower 95% Mean	20.927212
N	58

Multivariate

Correlations

	T01	T02	Final
T01	1.0000	0.7374	0.7797
T02	0.7374	1.0000	0.7944
Final	0.7797	0.7944	1.0000

There are 3 missing values. The correlations are estimated by REML method.

Scatterplot Matrix