Midterm GUIDE for Biology 300: October 2011
Instructor: Greg Bole

The mean of the midterm was 41.36 (68.9% or B-) with a variance of 92.46 and a standard deviation of 9.62. I'm guessing it was a pretty bimodal distribution but I haven't graphed it out yet.

Please note that this Answer Guide is posted to give you a better idea what kind of answers we were looking for. Other examples are possible, but for full marks the answer would have to contain all of the same major concepts, calculations and statements.

If you still have questions about your exam, please get in touch with me to address these questions. Come to my office hours or let me know what days and times you would be available to meet. You are required to go through the entire guide carefully before you ask about answers. Each exam will be considered on their own based on the answer guide and marking scheme.

If you still feel that you were incorrectly marked, you can resubmit your entire exam with a separate written note explaining specifically which answers you believe deserve more marks and why. Then the entire exam will be remarked, the new mark will be final no matter if it goes up or down.

Please note these rules do not apply to addition errors of marks, those can be submitted to Greg before or after lecture for correction.

For all statistical tests, make sure that you clearly state your hypotheses. Unless otherwise stated, assume 〈 = 0.05. Show your work. Be as precise as possible about P-values.

Mark allocation:

Question	Marks possible	Your mark
1.	8	
2.	6	
3.	16	
4.	10	
5.	10	
6.	10	
Total	**60**	

1. (8 marks total) In Victoria, British Columbia, the probability of rain during a winter day is 0.12, for a spring day is 0.07, for a summer day is 0.05, and for a fall day is 0.12. Each of these seasons lasts one quarter of the year.

a) What is the probability of rain on a randomly chosen day in Victoria? (4 marks)

Pr[rain on random day in Victoria] = (0.25 × 0.12) + (0.25 × 0.07) + (0.25 × 0.05) + (0.25 × 0.12) = 0.09

b) If you were told that on a particular day it was raining in Victoria, what would be the probability that this day would be a fall day? (4 marks)

Pr[fall | raining] = Pr[raining | fall] × Pr[fall] / Pr[raining] = 0.12 × 0.25 / 0.09 = 0.333

(The correct application of Bayes' Theorem is required for full marks)

2. (6 marks total)
a) What effect does increasing the sample size have on the probability of committing a Type I error? Explain. (3 marks)

No effect. No matter what the sample size is, the alpha determines the probability of committing a Type I error, and that has remained unchanged.

b) What effect does reducing the value of the significance level from 0.05 to 0.01 have on the probability of committing a Type II error? Explain. (3 marks)

Decreasing the significance level (〈) of a test will **increase** the probability of committing a Type II error.

Reducing alpha makes the null hypothesis more difficult to reject when true, but it also makes the null hypothesis more difficult to reject when false.

3. (4 marks each, 16 total) For each of the following scenarios **identify the best statistical test to use** and **state the null hypothesis**. (Please note, do not give the answer to the specific question, but simply state the best test to use and the null hypothesis for the scenario.)

a. Asking if the number of students in this Biology 300 class that are left-handed, is the same as you would expect based on the reported number of left handed students at UBC.

Binomial test
H_0: The proportion of left-handed students in Bio 300 is equal to the proportion at UBC.

(Please note that a χ^2 goodness-of-fit test fit to a binomial distribution would be easier to calculate by hand, but it is not the BEST STATISTICAL TEST do use, since it is an approximation)

b. Asking if the height of all Canadian female UBC professors is consistent with a Canadian female population mean height of 161 cm.

One-sample t-test
H_0: The mean height of Canadian female UBC professors is 161 cm.

c. Asking whether there is a relationship between which of two drugs a patient is given and whether they survive for a month after treatment, in a small sample where the expected number of patients who die on one of the drugs is 2.3.

Fisher's exact test (the expected values of one of the cells are too low to use the χ^2 test.)
H_0: Survival and drug treatment are independent.

d. Asking whether the WebCT Vista system crashes each month have been equally likely to have happened over each month of the last three years.

χ^2 goodness-of-fit test
H_0: The number of system crashes per month follows a Poisson distribution.

3. (10 marks total) Researchers in the Okanagan, BC asked a few hundred people about their daily wine intake. Of the 22 people from that group who had the highest wine intake, nine of them reported seeing Ogopogo, a rumoured "lake monster" in Okanagan Lake. Of the 22 people with the lowest wine intake, three people reported seeing Ogopogo.

a. Do an appropriate hypothesis test to investigate a possible link between wine intake and seeing Ogopogo. (8 marks)

H_0: Wine intake and reporting to see Ogopogo are independent.
H_A: Wine intake and reporting to see Ogopogo are not independent.

Here is the contingency table with observed and expected values.

	Reported sighting	No sighting	
Highest wine	9 / 6	13 / 16	22
Lowest wine	3 / 6	19 / 16	22
	12	32	

The expected value for Wine/Seeing Ogopogo was calculated from 12/44 (estimated probability of seeing Ogopogo) times the probability of drinking wine (22/44), multiplied times the 44 total observations. The others can be calculated by subtraction from the row and column totals.
We can calculate a χ^2 contingency test, because all of the cells have *expected* values of 5 or more.

$$\chi^2 = (9-6)^2/6 + (3-6)^2/6 + (13-16)^2/16 + (19-16)^2/16 = 4.125$$

We have (r-1)(c-1) = 1 degrees of freedom,
so we compare this to the critical value of $\chi^2_1 = 3.84$.
Therefore $P < 0.05$ (1) and we can reject the null hypothesis.
Heavy wine drinkers are more likely to report seeing Ogopogo than light wine drinkers.

b. Are the results from part (a) sufficient evidence to support the statement that wine drinking sometimes causes people to see Ogopogo? Why, or why not? (2 marks)

No, because the data merely show an association. It is possible that seeing Ogopogo causes people to drink wine more often, or that a third factor may cause both.

4. (10 marks) The following description of an experiment comes directly from a paper by Mantonakis *et al*. (2009):
In many competitions, the competitors are judged by subjective criteria one after the other. (Think for example of an audition or musicians in a talent contest.) Does the order in which the choices appear affect the decisions of the judges? An experiment to look for these order effects gave volunteers 4 glasses of wine, and the volunteers were asked to say which of the four was the best wine. In actual fact, all 4 glasses were poured from the same bottle.

This experiment was done on 33 volunteers, of whom 15 preferred the first glass, 5 preferred the second glass, 2 preferred the third glass, and the other 11 preferred the last glass.

Is there evidence from these data that the order in which these were presented had an effect on the preference of the volunteers?

H_0: There is no preference for any of the glasses OR The probability of a wine being chosen is the same for every glass OR p=0.25 for each glass

H_A: There is a preference for one of the glasses OR The probability of a wine being chosen is not the same for every glass OR p does not equal 0.25 for each glass

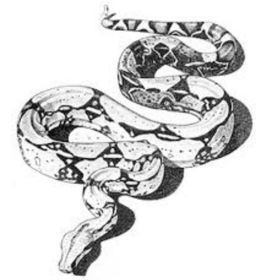
Glass	Observed	Expected	$\frac{(obs-exp)^2}{exp}$
1	15	8.25	5.522727272
2	5	8.25	1.28030303
3	2	8.25	4.73484848
4	11	8.25	0.9166666
Total	33	33	33

(1)

$\chi^2 = (15-8.25)^2/8.25 + (5-8.25)^2/8.25 + (2-8.25)^2/8.25 + (11-8.25)^2/8.25 = 12.45$

We have 4 - 1 = 3 degrees of freedom,
so we compare this to the critical value of $\chi_3 = 7.81$
12.426 > 7.81 therefore $P < 0.05$ and we can reject the null hypothesis
The order presented did have an effect on the preference of these volunteers.

5. (10 marks) A random sample of 200 Boa constrictors (a kind of snake) have a length of 1.5 meters. Boa lengths are normally distributed, and 95% of the Boas in this sample are between 1.0 and 2.0 meters. Using the information from this sample, what is the 95% confidence interval for the mean length of Boa constrictors?



Because Boa length is normally distributed, 95% of values should fall between the mean minus 1.96 standard deviations and the mean plus 1.96 standard deviations.

Since the mean is 1.5 and the upper range is 2.0, the 0.5 should be equal to 1.96 standard deviations.

So the standard deviation is approximately 0.5/1.96 = 0.255 in this sample.

Therefore the standard error is approximately $SE_Y = 0.255$ / sq.rt.200 = 0.018

Thus the confidence interval for the mean is approximately
$Y \pm SE_Y\, t_{0.05(2),199} = 1.5 \pm (0.018)(1.97) = 1.5 \pm 0.036$