Name:____________________________Student number:___________________

MID-TERM BIOL 300: March 2010

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

Some questions have a box for the final answer. Please put the final answer in this box, and show all work in the other space provided, including the back of the page if necessary.

By taking this test and putting your name above, you are declaring that your answers on this test are all your own work.

Make sure that your copy of the test includes 4 pages, including this one.

1. What statistical test would you use in each of the following situations? Give the full name of the test. (For this question, you do NOT need to do the tests.)

a. You measure the flight speeds of 25 swallows, and then want to use the data to test the null hypothesis that the mean speed of all swallows is 11 m/s. (*4 points*)

1a. one-sample t-test

b. You have two categorical variables, and you want to test the null hypothesis that the two variables are independent. Assume that you have quite a large sample size. (*4 points*)

1b. χ^2 contingency test (NOT goodness-of-fit)

c. You have two categorical variables, each with two categories, and you want to test the null hypothesis that the two variables are independent. Assume that your total sample size is only 15 individuals. (*4 points*)

1c. Fisher's exact test ("Fisher's test" is OK)

d. Hockey team A beats hockey team B by a score of 6 to 2. You want to test the null hypothesis that the two teams are in fact equal in ability and have equal chances of scoring each goal. (*4 points*)

1d. Binomial test

Name:______________________________Student number:____________________

2. In birds, beak length shows a strong relationship with characteristics of food. In a sample of 5 birds of species A, you measure the following beak lengths, in millimeters:

7 10 8 6 13

For fun, we calculated the sum of these beak lengths ((sum of Y_i) = 44 mm) and the sum of the squared beak lengths ((sum of (Y_i^2)) = 418 mm^2). Using this sample, calculate the following summary statistics (remember to include a few decimal places in your calculations, to avoid rounding error):

a. Median (*4 points*)

2a. 8

b. Mean (*4 points*)

2b. 8.8

c. Variance (*4 points*)

2c. 7.7

d. Standard deviation (*4 points*)

2d. 2.77 (2.8 is OK)

e. Standard error of the mean (estimated from the sample) (*4 points*)

$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}} = \frac{2.77}{\sqrt{5}} = 1.24$$

1e. 1.24 (1.3 is OK)

f. 95% confidence interval for the population mean (*12 points*)

$$df = 5 - 1 = 4$$
$$\bar{Y} \pm \left(SE_{\bar{Y}}\right)\left(t_{0.05(2),4}\right) = 8.8 \pm (1.24)(2.78) = 8.8 \pm 3.4$$
$$or$$
$$5.4 \le \mu \le 12.2$$

1f. 5.4 < population mean < 12.2 (or 8.8 plus or minus 3.4)

Name:______________________________Student number:____________________

3. You make blueberry pancakes one morning, and collect data on the number of blueberries in each pancake. You want to test the null hypothesis that blueberries were distributed among pancakes randomly and independently. You test this null hypothesis by comparing the observed distribution to a null distribution, which I provide below (note that I used the sample data to calculate the mean of this null distribution). For fun, I also calculated the square of the difference between observed and expected in each category.

Number of blueberries per pancake	Observed number of pancakes with given number of blueberries	Expected number of pancakes with given number of blueberries, according to null	(Observed – Expected)^2
1	5	8.7	13.69
2	17	10.6	40.96
3	15	8.7	39.69
4	4	5.3	1.69
5 or more	0	7.7	59.29

a. What is the name of the distribution I used to calculate the expected (null) distribution above? (*4 points*)

3a. Poisson

b. Do a hypothesis test to determine whether you can reject the null hypothesis. Give the P-value as precisely as possible from the statistical tables, and interpret your results in words. Do the blueberries appear to be clumped, randomly distributed, or dispersed? (*18 points*)

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} = \frac{13.69}{8.7} + \frac{40.96}{10.96} + \frac{39.69}{8.7} + \frac{1.69}{5.3} + \frac{59.29}{7.7} = 18.02$$

$$df = n - 1 - 1 = 5 - 1 - 1 = 3$$

$$\chi^2_{0.05,3} = 7.81$$

$$18.02 > 7.81$$

$$18.02 > \chi^2_{0.001,3} = 16.27$$

So we reject the null model that blueberries are distributed randomly. We can reject this with a P-value of less that 0.001.

By looking at how the observed distribution differs from the expected (the Poisson distribution), we can see that there are more pancakes in the middle of the distribution (with 2 or 3 blueberries), and fewer in the sides of the distribution (with 1 or 5), so the blueberries are dispersed (not clumped). The blueberry maker perhaps tried to put roughly an even number of blueberries in each pancake.

Additional note: No one else noticed it, but I actually made an error in using the Poisson to calculate a null distribution. I should have included a category of zero blueberries per pancake, as some of the Poisson distribution would fall into the zero category. Doing this correctly, we would expect 3.6 pancakes with zero blueberries, and 4.1 pancakes with five or more blueberries (the other categories remain the same as above). Chi-square then becomes 14.4, which allows us to reject the null at a P value between 0.005 and 0.001. (I marked myself down 5 points for making this mistake, but rewarded myself 2 points for discovering the error)

Name:_____________________________Student number:____________________

4. a. You plan to conduct a hypothesis test using a significance value of $\alpha = 0.05$. What is the probability of committing a Type I error? (*4 points*)

4a. 0.05

4. b. Still using a significance value of $\alpha = 0.05$, you plan to conduct hypothesis tests on five independent null hypotheses, using independent data sets every time. If all the null hypotheses are true, what is the probability that at least one will be rejected? (*10 points*)

$$\Pr[at\_least\_one\_rejected] = 1 - \Pr[all\_not\_rejected]$$
$$= 1 - (0.95)^5$$
$$= 1 - 0.774$$
$$= 0.226$$

4b. 0.226 (or 22.6%)

5. The mean height of the population of dandelions is 9.5 cm, with a variance of 4.0 cm^2. Assume that dandelion height is normally distributed. What percentage of dandelions would be cut by a lawn mower that cuts at a height of 7.0 cm? (*16 points*)

$$Z = \frac{Y - \mu}{\sigma}$$
$$\sigma = \sqrt{\sigma^2} = \sqrt{4.0} = 2.0$$
$$Z = \frac{Y - \mu}{\sigma} = \frac{7.0 - 9.5}{2.0} = \frac{-2.5}{2.0} = -1.25$$
$$\Pr[height > 7.0] = \Pr[Z > -1.25]$$
$$\Pr[Z > -1.25] = \Pr[Z < 1.25]$$
$$\Pr[Z < 1.25] = 1 - \Pr[Z > 1.25]$$
$$= 1 - 0.106$$
$$= 0.894$$
$$= 89.4\%$$

5. 89.4% (or 0.894)