MID-TERM BIOL 300: February 2005

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 6 pages, including this one.

	Points
Q1	35
Q2	15
Q3	15
Q4	10
Q5	10
Q6	15 -OPTIONAL

1. In the northern hemisphere, dolphins swim predominantly in a counter-clockwise direction while sleeping. A group of researchers with nothing better to do wanted to know whether the same was true for dolphins in the Southern hemisphere. They watched eight sleeping dolphins, and the variable that they recorded was the percentage of time that the dolphins swam clockwise. Assume that this is a random sample, and that this variable has a normal distribution. These data are here:

77.7 84.8 79.4 84.0 99.6 93.6 89.4 97.2

a. (*5 points*)What is the mean percentage of clockwise swimming for Southern hemisphere dolphins?

b. (*5 points*) What is the 95% confidence interval for the mean time going clockwise in the Southern hemisphere dolphins?

c. (*5 points*) What is the 99.8% confidence interval for the mean time going clockwise in the Southern hemisphere dolphins?

d. (*5 points*) What is the best estimate of the standard deviation percent clockwise swimming?

e. (*5 points*) What is the median of the percentage clockwise swimming?

f. (*10 points*) Test the hypothesis that Southern hemisphere dolphins spend half of their time while asleep swimming clockwise.

Answers for 1a-e:

a. 88.2	b. $81.4 < \mu < 94.95$	c. $74.5 < \mu < 101.9$
d. 8.09	e. 87.1	

b.

$$\bar{x} \pm t_{\alpha(2),\nu} \frac{s}{\sqrt{n}} = 88.2 \pm t_{0.05(2),7} \frac{8.09}{\sqrt{8}}$$

$$df = 8 - 1 = 7$$

$$t_{0.05(2),7} = 2.36$$

$$\bar{x} \pm t_{\alpha(2),\nu} \frac{s}{\sqrt{n}} = 88.2 \pm 2.36 \frac{8.09}{\sqrt{8}} = 88.2 \pm 6.75$$

c. Same as above, except now $\alpha = 0.002$:

$$t_{0.002(2),7} = 4.79$$

$$\bar{x} \pm t_{\alpha(2),\nu} \frac{s}{\sqrt{n}} = 88.2 \pm 4.79 \frac{8.09}{\sqrt{8}} = 88.2 \pm 13.7$$

e. The middle two numbers of the ordered data are 84.8 and 89.4. Averaging these gives 87.1.

f. H_0: The mean proportion of the time that Southern hemisphere dolphins spend swimming clockwise is $\mu = 0.5$.

H_A: The mean proportion of the time that Southern hemisphere dolphins spend swimming clockwise is not 0.5.

$$t = \frac{\overline{X} - \mu}{s/\sqrt{n}} = \frac{88.2 - 50}{8.09/\sqrt{8}} = 13.35$$

The critical value is $t_{0.05(2),7} = 2.36$. $t < t_{0.05(2),7}$, *so we can reject the null hypothesis*

Southern hemisphere dolphins swim clockwise significantly more than half of the time.

By the way, these are real data.

2. A common perception is that people may be able to delay their deaths from chronic illness until after a special event, like Christmas. Out of 12028 deaths from cancer in either the week before or after Christmas, 6052 happened in the week before.

a. (*5 points*) What is the best estimate of the proportion of deaths out of this time interval that occurred in the week before Christmas?

b. (*5 points*) What is the 95% confidence interval for this estimate?

c. (*5 points*) Us this confidence interval to ask "Are these data consistent with a true value of 50%?"

Answers for 2a-c:

a. 0.503	b. 0.494 < p < 0.512	c. yes

a. 6052/12028=0.503

b. $p' = \frac{X+2}{n+4} = \frac{6054}{12032} = 0.503$,

$$\left(p' - Z\sqrt{\frac{p'(1-p')}{n+4}}\right) \leq p \leq \left(p' + Z\sqrt{\frac{p'(1-p')}{n+4}}\right) =$$

$$\left(0.503 - 1.96\sqrt{\frac{0.503(1-0.503)}{12032}}\right) \leq p \leq \left(0.503 + 1.96\sqrt{\frac{0.503(1-0.503)}{12032}}\right)$$

$$0.494 < p < 0.512$$

c. 0.5 is inside the confidence interval give in b., so we can say that the data are consistent with a true value of 0.5.

3. (*15 points*) A study of 3000 people compared the rates of liver cancer between 2000 people who drank one or more cup of coffee per day and 1000 people who drank less coffee than that. Of the coffee drinkers, 21 got liver cancer, while of the people who didn't drink coffee, 20 got liver cancer. Do a hypothesis test for the question: "Is there an association between coffee drinking and liver cancer? " Interpret your result. Does coffee drinking increase, decrease, or not affect the probability of getting liver cancer?

We want to test for a relationship between two categorical variables, so χ^2 contingency analysis is possibly appropriate.

H_0: Coffee drinking and liver cancer are independent.

H_A: Coffee drinking and liver cancer are not independent.

Our observed:

Observed	coffee drinker	non-coffee drinker	Total
Liver cancer	21	20	41
No liver cancer	1979	980	2959
Total	2000	1000	3000

We can estimate the rate of coffee drinkers as 2000/3000 = 2/3, and the rate of liver cancer as 41/3000 = 0.0137.

From these generate the expected values:

Expected	coffee drinker	non-coffee drinker	Total
Liver cancer	27.33	13.67	41
No liver cancer	1967.33	986.67	2959
Total	2000	1000	3000

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = \frac{(21-27.33)^2}{27.33} + \frac{(20-13.67)^2}{13.67} + \frac{(1979-1976.67)^2}{1976.67} + \frac{(980-986.33)^2}{986.33}$$
$$= 4.41$$

There are (r-1)(c-1) = (2-1)(2-1)= 1 degrees f freedom, so the critical value of χ^2 = 3.84 for α=0.05. χ^2 =4.41 is greater than the critical value, so we can reject the null hypothesis. Coffee drinking and liver cancer are not independent.

Coffee drinkers develop liver cancer at frequency 21/2000 = 0.0105, while non-coffee drinkers developed cancer at frequency 0.02. Coffee drinking reduced the rate of liver cancer.

4. (*10 points*) A bag contains 7 coins. 4 of these coins are normal fair coins, with one side heads and one side tails. 2 of the coins are trick coins with heads on both sides, and the last one coin has tails on both sides. Each coin is equally likely to show either side when flipped. A coin is drawn at random, and it is flipped. What is the probability that is comes up heads?

Answer 4:

4/7 = 0.57

There's more than one way to answer this question. One way is by saying there are three possible conditions, which correspond to which coin is drawn from the bag. We can then average the probability of heads across thee three conditions:

$$\begin{aligned}\Pr[Heads] = {} & \Pr[Heads \mid 2\ \ headed\ \ coin]\Pr[2\ \ headed\ \ coin] + \\ & \Pr[Heads \mid 1\ \ headed\ \ coin]\Pr[1\ \ headed\ \ coin] + \\ & \Pr[Heads \mid 0\ \ headed\ \ coin]\Pr[0\ \ headed\ \ coin] \\ = {} & (1)(2/7) + (1/2)(4/7) + (0)(1/7) \\ = {} & 4/7\end{aligned}$$

Alternatively, it may be simpler to think that each of the 14 sides of the coins in the bag has an equal chance of being the answer. In the bag 8 sides are heads, and there are 14 sides in total. So the probability of heads is 8/14 = 4/7.s

5. (*10 points*) At the end of a hypothesis test, we have concluded that $P < \alpha$. What do we conclude? Why?

We reject the null hypothesis. P is the probability of getting the data that we got, or something more extreme, assuming that the null hypothesis were true. If N_e small, then there is a low probability of getting the data assuming the null hypothesis, and therefore we can reject it. α is the rather arbitrary point set to decide when the match between null hypothesis an data is too low. If P is less than α, then we reject H_0.

6. Three dice have six sides, numbered 1 to 6, and the probability of rolling any integer between 1 and 6 is 1/6 for each die.

a. (*5 points*) If all three dice are rolled together, what is the probability of rolling exactly two 3's and one die that is not a 3?

This can be answered from the binomial distribution. There are three trials (each die) and for each trial the probability of a three is 1/6:

$$\Pr[two\ \ 3's] = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(1-\frac{1}{6}\right) = 0.069.$$

b. (*5 points*) If all three dice are rolled together, what is the probability of rolling exactly two 3's and one 4?

We can solve this based on the answer to the last question. The probability that two of the dice show 3's is 0.069. This tells us the probability of exactly two 3's , so we know that the last die is not a 3. The probability that the last die is a 4, given that it is not a three, is 1/5. Therefore the probability of two 3's and one 4 is 0.069 (1/5) = 0.0139.

c. (*5 points*) If all three dice are rolled together, what is the probability that none of the three show a number greater than or equal to 3?

The probability that a single die rolls a number greater than or equal to 3 is 2/3. The probability that it does not is 1-2/3 = 1/3. The probability that three independent dice do not is therefore $(1/3)^3 = 0.037$.