KEY: MID-TERM BIOL 300: October 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.

1. (*10 points*)A recent issue of the Globe and Mail discussed a scientific study as follows:

> ' "Fatherhood, as Miller-McCune magazine reported earlier this year, alters a man's neurochemistry," Tom Jacobs writes for Miller-McCune.com, "increasing his ability to cope with stress and generally making him a better mate. Just-published research suggests the benefits of this transformation extend far beyond one's immediate family and remain robust as the years go by. … 'We find that middle-aged men who at some point in their lives become fathers are significantly more likely to have altruistically oriented social relationships and be involved in service organizations compared with men who never become fathers,'" a Penn State University research team writes in the Journal of Family Issues.'

Are the conclusions drawn by Tom Jacobs defendable? Why or why not?

No. This interpretation confuses causation and correlation. It is very possible that the men who are likely to later have more altruistic social relationships are also those who are more likely to have children.

2. (*7 points*) Define Type I error. If the significance level of a test is increased, will the Type I error increase, decrease, or stay the same? Explain.

Type I error is rejecting a true null hypothesis. The Type I error rate is equivalent to the significance level α, and so higher significance levels correspond to higher Type I error.

3. (*20 points*) A study of 515 patients recently diagnosed with extremely serious diseases (such as terminal cancer, MS, or a brain tumor). Each of the patients were married to a member of the opposite sex; 254 of the patients were female and 261 were male. The researchers followed the couples after the diagnosis, specifically to ask whether the couple divorced soon after the bad diagnosis. Of the couples where the male was ill, 7 resulted in divorce. When the female was the patient, 53 resulted in divorce. Test for a difference in the proportions of divorce after serious diagnosis as a function of whether the partner is male or female.

H_0: Divorces happen with equal probability when the woman is sick as when the man is sick.
H_A: The proportion of divorces differs depending on which sex is ill.

Here is the contingency table with observed and expected values.

	male patient		female patient		total
divorce	7	30.4	53	29.6	60
no divorce	254	230.6	201	224.4	455
total	261		254		515

The expected value for divorce/male patient was calculated from 60/515 (estimated probability of divorce) times the probability of male patient (261/515), multiplied times the 515 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.

$$\chi^2 = \frac{(7-30.4)^2}{30.4} + \frac{(53-29.6)^2}{29.6} + \frac{(254-230.6)^2}{230.6} + \frac{(201-224.4)^2}{224.4} = 41.3$$

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 $P<0.05$, and we can reject the null hypothesis. Female patients are more likely to be divorced by their spouses after a serious diagnosis than are male patients.

4. (*7 points*)You draw two cards without replacement from a deck of 52 cards. If the first card is not a spade, find the probability that the two cards drawn are both diamonds.

The probability that the first card is a diamond, given that it is not a spade, is 1/3. The probability that the second card is a diamond given that he first was a diamond is 12/51. So the probability that they are both diamonds is (1/3)(12/51) = 0.0784

5. (*7 points*) My son mentioned last night that none of the 21 kids in his 3rd grade class had had a birthday since school started 56 days previously. What is the probability that a class of this size drawn from a population where birthdays were equally likely on all days of the year would not yet have a birthday for any kid?

The probability that a given kid was *not* born during this time period is (365-56)/365 = 0.8466. The probability that all 21 have a birthday outside of this interval is $(0.8466)^{21} = 0.03$.

6. (*5 points each*)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 whether stickleback fish occur with equal probability through all areas of a pond.

χ^2 Goodness of fit test

H_0: Fish numbers per unit area follow a Poisson distribution.

b. Asking whether Douglas fir trees or Western hemlocks were more likely to be infested by pine beetles, assuming that the sample sizes in the study were very large.

χ^2 contingency analysis

H_0: Equal proportions of Douglas fir and Western hemlocks are infested by pine beetles.

c. Asking whether Douglas fir trees or Western hemlocks were more likely to be infested by pine beetles, where the number of fir in the sample expected to get infested was only 2.3.

Fisher's exact test

H_0: Equal proportions of Douglas fir and Western hemlocks are infested by pine beetles.

d. Asking whether the number of smokers in a list of cities is proportional to the population size of those cities.

χ^2 Goodness-of-fit test

H_0: The number of smokers in a city is proportional to the number of people in that city.

e. Testing whether patients change in body mass during a hospital stay.

One-sample *t*-test

H_0: Mean change in patient body mass during hospital stay is zero.

7. (*24 points*) Weather Canada records temperature on at least an hourly basis for Vancouver. The temperature at noon on Oct 28 for the last several years is given below in Celsius:

a. (*5 points*) What is the best estimate of the mean recent temperature in Vancouver at noon on this date? Give a 95% confidence interval for this estimate.

$$\bar{Y} = 12.82$$

$$\bar{Y} \pm SE_{\bar{Y}} \quad t_{\alpha(2),df}$$

$$df = 9 - 1 = 8; s = 2.25$$

$$12.82 \pm \frac{2.25}{\sqrt{9}}(2.31)$$

$$11.09 < \mu < 14.55$$

b. (*2 points*) What is the median relative temperature from this sample?

Putting the numbers in order:

10.0, 10.5, 10.5, 11.3, 14.1, 14.1, 14.2, 14.5, 16.2

The middle number is the median: 14.1.

c. (*5 points*) What is the variance of temperature? Give a 95% confidence interval for this estimate.

The variance is $s^2 = (2.25)^2 = 5.057$.

$$\frac{df\ s^2}{\chi^2_{\frac{\alpha}{2},df}} \le \sigma^2 \le \frac{df\ s^2}{\chi^2_{1-\frac{\alpha}{2},df}}$$

$$\frac{8(5.057)}{\chi^2_{0.025,8}} \le \sigma^2 \le \frac{8(5.057)}{\chi^2_{0.975,8}}$$

$$\frac{8(5.057)}{17.53} \le \sigma^2 \le \frac{8(5.057)}{2.18}$$

$$2.31 \le \sigma^2 \le 18.56$$

d. (*4 points*) What is the standard deviation of this temperature? Give a 95% confidence interval for this estimate.

$s = 2.25$

$$\sqrt{2.31} \le \sigma \le \sqrt{18.56}$$

$$1.52 \le \sigma \le 4.31$$

e. (*2 points*) What is the coefficient of variation of this temperature?

$$CV = 100\% \frac{s}{\overline{Y}} = 100\% \frac{2.25}{12.82} = 17.5\%$$

f. (*6 points*) In some parts of the previous calculations you may have used a "standard error." Describe what this phrase means.

The standard error is the standard deviation of the distribution of sample estimates. It is an indicator of how reliable the estimate is.