

Name: KEY

TA's name: (Alathea, Becca, Kate, Kim, Kira)

Student number:

MID-TERM BIOL 300: October 2013

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

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 7 pages, including this one.

	Points
Q1	8
Q2	9
Q3	20
Q4	16
Q5	25
Q6	18
Q7	4
	100

1. (8 points) a. Imagine that cases of tuberculosis occur completely independently of each other and with equal probability in any month, in a city like Vancouver. If we know that there are an average of 1.3 cases per month, what is the probability that a given month will have 3 cases of tuberculosis in this city?

By the Poisson distribution ,

$$P[3] = \frac{\lambda^3 e^{-\lambda}}{3!} = \frac{(1.3)^3 e^{-1.3}}{3!} = 0.09979$$

b. Tuberculosis cases actually do not occur independently of each other, but instead when there is one case it is more likely for there to be more because of contagion. Does this increase or decrease the variance among months in the number of cases?

Increase.

2. (9 points) John has calculated a sample mean based on 15 individuals. Mary has collected an independent random sample from the same population, and has calculated a sample mean on the same variable from 150 individuals.

a. Which person is most likely to have a smaller standard error?

Mary, because she has a larger sample size.

b. Which person is most likely to have a sample mean closest to the true mean?

Mary, because she has a smaller standard error.

c. Assume that they both test a null hypothesis about the mean of this population using $\alpha = 0.05$. The null hypothesis is in reality true. Which person is more likely to reject the null hypothesis?

They are equally likely to reject a true null hypothesis, because they used the same α value.

3. (20 points: a-f 2 point each, g and h 4 points each) As in other vertebrates, individual zebrafish differ from one another along the shy-bold behavioral spectrum. In addition to other differences, bolder individuals tend to be more aggressive, whereas shy individuals tend to be less aggressive. Norton et al. (2011) measured several behaviors associated with this syndrome in zebrafish that had the *spiegeldanio* (*spd*) mutant at the *Fgfr1a* gene (reduced fibroblast growth factor receptor 1a). The data below are measurements of the amount of time, in seconds, that individual zebrafish with this mutation spent in aggressive activity over 5 minutes when presented with a mirror image.

96 97 100 127 128 156 162 170 190 195

(For each of the following, remember to give units if appropriate.)

a. What is the mean of this sample?

142.1

b. What is the median of the measurements?

$(128+156)/2 = 142$

c. What is the variance of this sample?

1424.3

d. What is the standard deviation of this sample?

37.7

e. What is the coefficient of variation for this sample?

0.266

f. What is the standard error of the mean for this sample?

11.9

g. What is the 95% confidence interval for the mean for this sample? (Be as accurate as possible.)

$115.1 < \mu < 169.1$

h. Calculate a 95% confidence interval for the variance.

$df = 9$

$673.9 < \sigma^2 < 4747.0$

4. (16 points) Aging workers of the Neotropical termite, *Neocapritermes taracua*, develop blue, crystal-containing glands (“backpacks”) on their backs. When they fight intruding termites and are hampered, these “blue” termites explode, the glands erupting a sticky liquid (Šobotník et al. 2012). The data below are from an experiment that measured the toxicity of the blue substance. A single drop of the liquid extracted from blue termites was placed on individuals of a second termite species, *Labiotermes labralis*, and the number that were immobilized (dead or paralyzed) within 60 minutes was recorded. The frequency of this outcome was compared with a control treatment in which liquid from glands of “white” termites lacking the blue crystals was dropped instead. Is the blue liquid toxic compared to liquid from white termites?

	Number of <i>L. labralis</i> individuals	
	Unharmmed	Immobilized
Blue workers	3	37
White workers	31	9

H₀: The proportion of immobilized individuals is the same between liquid from blue and white workers.

H_A: The proportion of immobilized individuals is not the same.

Expected values:

	Unharmmed	Immobilized	Total
Blue workers	17	23	40
White workers	17	23	40
Total	34	46	80

$\chi^2 = 40.10$, $df = 1$. Critical value $\chi^2_{0.05(1)} = 3.84$. Since $40.10 > 3.84$, $P < 0.05$. Reject H₀. More precisely, because χ^2 is greater than 10.83, we know that $P < 0.001$. The contents of the blue backpacks are clearly toxic.

5. (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 the number of patients arriving at an emergency room is equal for each of the days of the week.

H_0 : The proportion of patients per day is $1/7$ for all days.

χ^2 goodness of fit test

b. Asking whether the mean amount of mercury contamination in seafood is greater than a standard set by the food protection agency.

H_0 : The mean mercury contamination is equal to the standard.

one-sample t -test

c. Asking whether first years or fourth year students have the same probability of getting a passing grade when given a BIOL 121 pretest.

H_0 : First and fourth year students have the same probability of passing the pretest.

χ^2 contingency analysis

d. Asking whether the proportion of first year students who pass a BIOL 121 pre-test is greater than 70%.

H_0 : Proportion that pass equals 0.70.

Binomial test

e. Asking whether the number of baby seahorses per father seahorse fits a Poisson distribution.

H_0 : Number of babies per father has a Poisson distribution.

χ^2 goodness of fit test

6. (a-c 4 points each, d 6 points) A roll of a fair six-sided die gives the numbers *one*, *two*, *three*, *four*, *five*, or *six* each with probability $1/6$. Each of the following questions is about a fair die like this.

a. What is the probability of rolling 4 *sixes* in row, out of a total of 4 rolls?

$$(1/6)^4 = 0.0007716$$

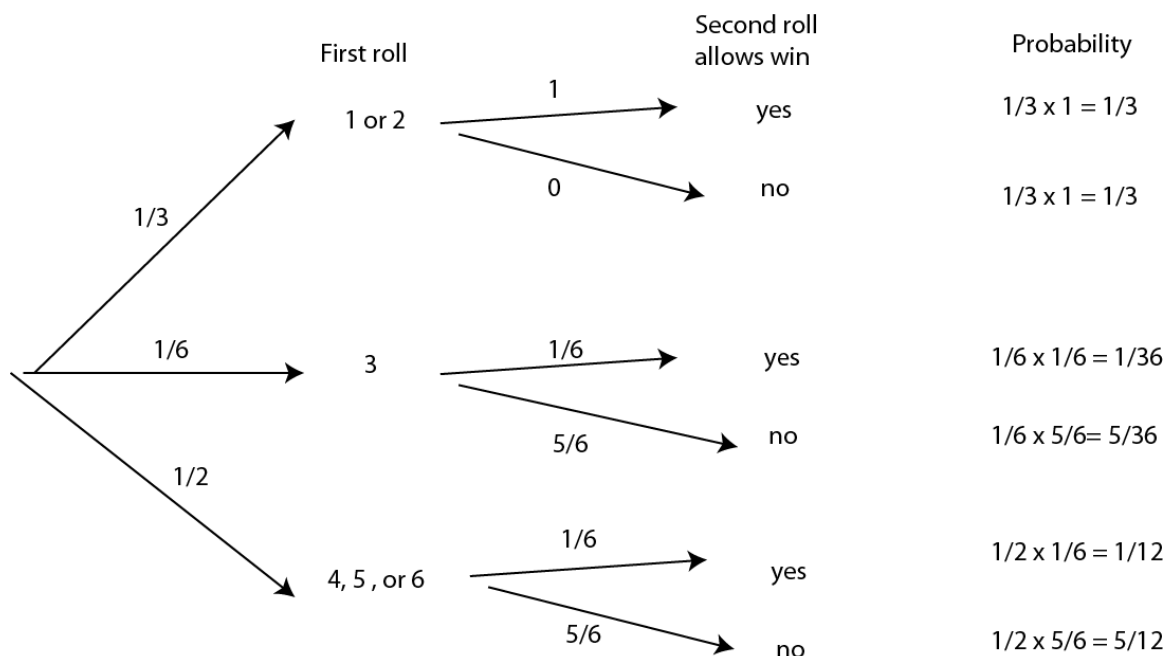
b. I roll a six-sided die, and I tell you (honestly) that it is not a *four* or a *six*. What is the probability that I rolled a *three*?

$$\Pr[\textit{three} \mid \textit{not four or six}] = 1/4$$

c. What is the probability that exactly 4 out of 20 rolls of this die are *threes*?

$$\Pr[4] = \binom{20}{4} (1/6)^4 (5/6)^{20-4} = 0.202204$$

d. I roll two dice, one after the other. If the first die is a *one* or *two*, then I will give you a prize (and ignore the second die). If the first die is greater than *two*, then if the second die matches the first I will give you a prize anyway. If after all this happens, and I gave you a prize, what was the probability that I rolled a three on the first die?



$\Pr[\textit{three on first roll} \mid \textit{prize}] =$

$$\frac{\Pr[\textit{prize} \mid \textit{three on first roll}] \Pr[\textit{three on first roll}]}{\Pr[\textit{prize}]} = \frac{(1/6)(1/6)}{1/3 + 1/36 + 1/12} = 1/16 = 0.0625$$

7. (4 points) Your colleague is conducting a study which compares a sample mean (of 47) based on a sample size of 22 to a hypothesized value for that mean (32). She is conducting a one-sample t -test, and has (correctly) calculated the value of her test statistic to be $t = -3.4$. As accurately as possible using the statistical tables, what is the P -value of her test? (She wants a two-tailed test.)

$P < 0.01$, based on the t value being further from 0 than the critical value of the t distribution for $t_{21,(0.01(2))} = 2.83$ and $t_{21,(0.001(2))} = 3.82$. Because the observed value of t exceeded the value for $\alpha = 0.01$ but not for $\alpha = 0.001$, so we know that P is less than 0.01, but not less than 0.001.

The numbers in this question turn out to be impossible, however. Because the observed mean is greater than the hypothesized mean, the t value would have been greater than 0. Pointing this out is worth full marks.