March 12, 2014
Quiz 5

Your name: ____

Quiz rules:

(a) This quiz is closed book, but you are allowed a two-sided sheet of paper of notes and a calculator.

(b) Each question is worth 6 points.

(c) A normal table is provided on the last page

(d) You have 50 minutes to complete this quiz.

(e) If you fail to show your work and/or explain how you arrived at your answer then no points will be awarded.

On Quiz 2 we heard about a study to determine the effect of a new dog food on the growth of puppies over the first year of life.

Recall: Researchers decide to run a randomized study. The researchers recruit 36 dogs – twelve large (Bernese Mountain Dogs), twelve medium (English Bulldog) and twelve small (West Highland Terrier). Half the dogs were randomly assigned to treatment, the other half to control.

		Baseline (total lbs)	Final (total lbs)
Treatment Group	• 8 Bernese • 8 Bulldogs • 2 Westies	332	1318
Control Group	• 4 Bernese • 4 Bulldogs • 10 Westies	220	812

1. On Quiz 2, we discussed how there was a noticeable imbalance between the types of dogs in the treatment vs. control. Looking at this data, would you reject the assertion that the dogs were randomly assigned to treatment and control? Provide a statistical test justifying your answer.

If the dogs were randomly assigned then the categorical variable "treatment type" would be independent of the categorical variable "breed." If they were assigned nonrandomly to treatment or control then we would be using information about the breeds to help assign to treatment type. Thus this is a Chi-squared test of independence. You could also call this a 3x2 contingency table.

H0: The categorical variables treatment type and breed are independent
HA: There is some dependency.

Under the null we would expect to see 6 of each type of dog in the treatment and control groups.

$$X^2 = \frac{(8-6)^2}{6} + \frac{(8-6)^2}{6} + \frac{(2-6)^2}{6} + \frac{(4-6)^2}{6} + \frac{(4-6)^2}{6} + \frac{(10-6)^2}{6} = \frac{46}{6} = 8$$

With (3-1)(2-1) = 2 degrees of freedom. The p-value is between 1% and 5%, so we would reject that hypothesis that the dogs were chosen randomly.

2. Continuing with the dog food example: A statistician suggests they run a new study using a matched pair design. In a matched pair design, each dog is paired with another dog such that they are of the same breed and had an identical birth weight. The observational unit becomes the pair, not an

individual dog. The pairs are then tested using a one-sample test of the null that the average difference is zero. The data for 6 of the 18 pairs are reported below.

Pair	Breed	Control (lbs increase)	Treated (lbs increase)	Difference
1	Bernese	68	73	5
2	Bernese	69	70	1
3	Bulldog	42	47	5
4	Bulldog	44	45	1
5	Westie	21	22	1
6	Westie	17	22	5
SD		20.3	20.2	2
AVE		43.5	46.5	3

(a) What would the SE be if you compared treatment vs control using a two-sample z-test? What about if you used a matched pairs z-test?

$$SE(two\ sample) = \sqrt{\frac{20.3^2}{6} + \frac{20.2^2}{6}} \cong 11.7$$

$$SE(matched\ pairs) = \sqrt{\frac{2^2}{6}} \cong 0.816$$

(b) Why are the two SEs you calculated in (a) so different?

The variation from dog to dog is quite large (likely because the breed types are so different in size). This makes the two-sample SD quite large because it is calculated across the dogs. But, within a matched pair, there is not much variation between the two dogs. Thus across pairs there is not as much variation as we were seeing across dogs.

3. Now let's analyze the data from Question 2 using a matched pairs test. Although you calculated the SE assuming a z-test in the previous question, you should really use a t-test here because the sample size is small. Be sure to state the null and alternative hypotheses.
(If you weren't able to calculate the SD of the differences in Question 2, please assume SD=7.3.)

Because the sample is small, we need to use a t-test and thus we need$SD+= \sqrt{\frac{n}{n-1}} * SD = \sqrt{\frac{6}{5}} * 2 \cong 2.2$

So the t-stat is $\frac{3}{2.2/\sqrt{6}} \cong 3.34$, which follows a t distribution with 5 degrees of freedom. Doing a one-sided test with the alterative that the difference is positive (because we believe the treatment will increase weight) we get a p-value of 1%. Therefore, we believe that the dog food does increase growth in puppies during the first year.

If you use SD=7.3, then you'll get a t-stat of 0.92 and the p-value is between 10% and 25%.

4. A parking lot has 200 cars. All of the cars have the same kind of car alarm. You may assume that each car runs a hypothesis test (independently of all the other cars) before deciding whether to sound its alarm.

   (a) Unfortunately, the car alarms go off quite frequently. The neighbors are honked off and request that a local ordinance be created modifying the threshold at which the alarms go off. Assuming this is a no-crime area, which one of the following is the **most** appropriate statistical term to indicate what the ordinance is seeking to modify: (i) observed significance level, (ii) the P-value, or (iii) the alpha level? Explain your selection in a few sentences.

The alpha level, (iii), is the most appropriate choice. To reduce the number of car alarms falsely sounding, the ordinance should control the frequency of type I errors. Since alpha level is the probability of a type one error, modifying alpha would change the expected number of type I errors.

   (b) Explain the statistical reason why the car owners may not like this new ordinance. Full points will be awarded to only those solutions correctly using hypothesis testing terminology.

Typically, when type I error is increased, type II error is decreased. The type II error in this setting corresponds to a car alarm not going off when there is crime (which is a false negative). Car owners might not like the new ordinance because their car alarms will go off less frequently when there is crime.

5. An instructor would like to know if grades on an exam follow the normal curve. In a class with 241 students, he finds that 157 students scored within 1 SD of the average and 230 students scored within 2 SDs of the average. Propose a hypothesis test and carry it out. Can you conclude that the data do not follow the normal curve?

If the students' grades follow a normal curve, we would expect that 68% of students fall within 1 SD of the mean, 27% fall between 1 and 2 SD's from the mean, and 5% fall more than 2 SD's from the mean

So a chi-squared test statistic for goodness of fit to a normal distribution is given by

Distance from Mean	**0 – 1 SD's**	**1-2 SD's**	**>2 SD's**
observed	157.0	230-157 = 73.0	241-230 = 11.0
expected	.68*241 = 163.9	.27*241 = 65.1	12.1

$$X^2 = \frac{(157 - 163.9)^2}{163.9} + \frac{(73 - 65.1)^2}{65.1} + \frac{(11 - 12.1)^2}{12.1} \cong 1.35$$

This follows a chi-squared distribution with 3-1 = 2 degrees of freedom. The p-value is between 50% and 70%, so we cannot reject the null hypothesis that the students' grades are normally distributed.

6. Tom and Joe are twins. Recently Tom has been living in Australia and Joe here in the USA. Given their divergent lifestyles, they're curious if they still are so "identical." To assess this they decide to measure their weights. Naturally, they have to use different scales. Both scales are known to have measurement error that is well approximated by the Gauss model. The SDs of the chance errors are known to be 1lb for Tom's scale and 2lbs for Joe's scale.

   Each twin does 36 measurements on his own scale. Tom's measurements have an average of 143lbs and an observed standard deviation of 2lbs. Joe's measurements have an average of 142lbs and an observed standard deviation of 2lbs.

   (a) State the null and alternative hypotheses for this test. Run the test at the 0.05 level and state your conclusion.

H0: Joe and Tom have the same weights (and thus have the same average measurements)
HA: they are not equal

This is a two-sided test because HA is "not equal" rather than directional. We have two-samples (i.e., Tom's and Joe's measurements). The scales have Gaussian measurement error with known SD so a z test statistic is the best test. (A t-test, using the observed standard deviations, would be less correct.) Thus we will run a two-sample z-test:

$$z = \frac{(143 - 142) - 0}{\sqrt{\frac{1}{36} + \frac{2^2}{36}}} = 2.7$$

The p-value for the two-sided test is 0.7%, so we reject the null hypothesis that they have the same weight. This may give evidence that Tom and Joe are not truly identical.

   (b) The test for the difference between the averages was highly significant. State in plain English what this P-value means in the context of this problem. How important is the observed difference?

Technically: The p-value gives chance that the observed difference in Tom and Joe's is as large as one pound under the assumption that they actually weight the same amount.

Informally: The p-value tells us that, yes, there is probably a real difference in their weights (i.e., it's not just a fluke due to the scales having measurement error). Maybe you could say: "Using these scales, we can tell them apart."

Although the observed difference in weight is quite statistically significant, it is only one pound. Realistically, most people would not say this difference means Tom and Joe are not identical. This is a statistically significant result that is not that important or meaningful in this context.