Name ____Answer Key_____

When you have completed your test make sure you email your Excel file containing all your work—before you leave class—to bailey.norwood@okstate.edu, with the subject line "<your last name> Exam 2 work".

All the data for this exam are in one spreadsheet at
http://seeds.okstate.edu/SeedsPPP/fall2013/Exams/Exam2/Exam2Data.xlsx

Section 1

Download the data containing 3,000 survey responses detailing the value each respondent places on (a) eggs raised in a cage system and (b) eggs raised in a cage-free system. The sheet *Original Data, Eggs* contain these data. However, I want you to evaluate the values only for a subset of these individuals. The sheet *Filtered Data, Eggs* have the IDs for this subset, and only these individuals should be used in the data you analyze.

(1) **[VLOOKUP function from CN.2] (worth 2 points)** Using the VLOOKUP function, retrieve the values for both eggs for each ID number in the sheet *Filtered Data*.

(2) **[CN.2]** Calculate the average values for each egg type.

(2.a) **[CN.2](worth 1 point)** The average value for cage eggs is $ 1.56 per dozen.

(2.b) **[CN.2] (worth 1 point)** The average value for cage-free eggs is $ 1.61 per dozen.

(3) **[TAN.3, Section A](worth 1 point)** Using the information in the previous question, which egg type has the highest *sample* average?

a) Cage eggs
b) Cage-free eggs
c) Unknown. Must use statistical test to determine if they are statistically different.

(4) **[TAN.3, Section A, hypothesis test, Homework 7]** Use a statistical test (the T.TEST function) to determine whether the value for one egg type is greater than another.

(4.a) **(worth 3 points)** What is the p-value of the test? Tell me precisely, using four decimal places.

p-value is 0.245577

(4.b) **(worth 3 points)** This p-value tells us the probability we are wrong if we say...? If we say what? This answer tells us the precise definition and interpretation of a p-value.

If we say that the population averages for the value of cage and cage-free eggs are different (or if we say cage-free eggs have a higher population average).

(4.c) **(worth 1 point)** What does the p-value tell us about the relative value of the two egg types?

That there is not much evidence that people in the population really value cage-free eggs more than cage eggs. If we go out and claim the value of cage-free eggs is larger there is a 25% chance we will be wrong, and that is too high of a risk to take.

12

Section 2

Use the salary data in the *Salary Data* sheet to answer the following questions. However, before you use Pivot-Tables to compute average salaries, use DATA FILTER or DATA SORT to remove all individuals with less than 10 years of experience. Round to zero decimal places

(5) **[CN.2 Pivot-Tables, Homework 6] (worth 1 point)** What is the average salary for an unmarried male, living in a rural area, with an ag econ degree, ~~and no master's degree~~?

$49,375

(6) **[CN.2 Pivot-Tables, Homework 6] (worth 1 point)** What is the average salary for a married male, living in a rural area, with an ag econ degree, ~~and no master's degree~~?

$49,054

(7) **[CN.2 Pivot-Tables, Homework 6] (worth 1 point)** What is the average salary for a married male, not living in a city, with an ag econ degree, ~~and no master's degree~~?

$50,255

~~(8) **[CN.2 Pivot-Tables, Homework 6] (worth 1 point)** What is the average salary for a married male, not living in a city, with an ag econ degree, and a master's degree in ag econ?~~

(9) **[CN.2 Pivot-Tables, Homework 6] (worth 1 point)** What is the average salary for a married female, not living in a city, with an ag econ degree, ~~and a master's degree in ag econ~~?

$39,167

(10) **[TAN.3, Homework 7] (worth 1 point)** The averages in questions 5-9 are

__sample_____ averages, which we use to make inferences about the

____population_____ <should have inserted "averages"> of ag econ graduates.

18

Section 3

Suppose you collect the following auction bids in a second-price auction.

Bidder	Bid
Franklin Norwood	$200
Tajh Boyd	$158
Beth Murphy	$157
Diamond Dallas Page	$202
Jon Stewart	$203
Mike Gundy	$5
Marcus Aurelius	$4

(11) **[TAN.3, Section A, 2nd price auction](worth 1 point)** The winner of the auction is _____Jon Stewart_____

and that person pays a price of $__202_____.

(12) **[TAN.3, Section A, 2nd price auction](worth 1 point)** The reason researchers use a second-price auction is that it gives rational people the incentive to submit a bid equal to their

_true value for the good_____.

20

Section 4

In the sheet *WTP Humane Pork* are data I have collected on the value people place on gestation-crate free pork, above the value they place on regular pork. That is, the numbers tell us how much more they will pay in addition to the price of a regular pork chop to know the sows were not confined in gestation-crates. Use these data to answer the following questions.

Please recognize that entries with the value that says, "$ - ," are <u>not</u> missing values, but instances when the person places zero value on animal welfare, and will not pay a higher price for gestation-crate pork. These values of zero should be included in your analysis.

(13) **[TAN.3, Section D](worth 1 point)** Based on the sample data (no weighted averages), the average premium people in the sample

placed on gestation-free pork was $_____1.49_____ per pork chop.

Make a new variable for demographic groups, where Group 1 is low income females, Group 2 is low income males, Group 3 is high income females, and Group 4 is high income males. High income is defined as household income of $60,000 or more. In column D, give Group 1 a value of 1, Group 2 a value of 2, and so on. You may need to sort the data first to make this easier.

(14) **[TAN.3, Section D] (worth 1 point)** The average premium for people in Group 1 was (round to two decimal places throughout)

$_____0.95_____ per pork chop, and Group 1 is ___27_____% of the sample.

(15) **[TAN.3, Section D] (worth 1 point)** The average premium for people in Group 2 was

$_____1.81_____ per pork chop, and Group 2 is _____25_____% of the sample.

(16) **[TAN.3, Section D] (worth 1 point)** The average premium for people in Group 3 was

$_____0.86_____ per pork chop, and Group 3 is _____21_____% of the sample.

(17) **[TAN.3, Section D] (worth 1 point)** The average premium for people in Group 4 was

$_____2.19_____ per pork chop, and Group 4 is ___28_____% of the sample.

25

We use this sample of people to make inferences about the population of Americans. Suppose we know that the distribution of each group in the American population is given by the table below.

Group	Percent of Population
1	40%
2	30%
3	15%
4	15%

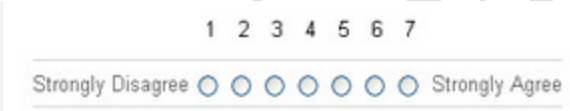
(18) **[TAN.3, Section D] (worth 3 points)** Using weighted averages to correct for the fact that our sample is not representative of the American public, we say that the average value for Americans as a whole is

$_____1.38_____ per pork chop.

Section 5

When I interview and survey people regarding their views on animal welfare issues, I suspect that those with the most empathy for animals are more likely to give me information, while those who think animals have no rights, and that farmers should be able to do whatever they want with animals, are reluctant to chat with me. Suppose that my suspicion is true, and suppose I am particularly interested in people's answer to the question…

To what extent do you agree or disagree with the following statement (1 = strongly disagree, 7 = strongly agree): ***Low meat prices are more important than the welfare of farm animals****.*



(19) **[TAN.3, Section C] (worth 2 points)** Because my sample is missing those who care least for animals, my sample average (average of the scale from 1-7 across all people) is probably

(*circle one*) SMALLER / LARGER than the population average. This means the sample

(if I do not use weighted averages) will imply that the average Americans care

(*circle one*) MORE / LESS about farm animals than they really do.

30

Section 6

(20) **[TAN.3, Section C] (worth 2 points)** What can I do to increase the likelihood that I will get survey responses from those who care least about animals? That is, if I'm missing the least empathetic Americans from, say, a telephone survey, what else can I do besides a telephone survey that will allow me to get their information? Tell me in an articulate paragraph.

One could begin offering to pay people money to take the survey. The people who care little about farm animals are choosing not to take the survey because there is nothing in it for them (those who care a lot have ample motivation, as they want to express their support for high animal welfare). By offering people, say, $30 to take the survey, then you are more likely to have a sample which includes the full diversity of attitudes towards farm animal welfare issue. Also, you could increase your sampling in places where meat-eaters who care little about farm animal welfare tend to go, like sampling from Walmart instead of Whole Foods, and sampling people at places like gun shows, farm shows, rodeos, and the like. Another alternative is to target websites for hunting and fishing with invitations to take an internet survey.

32

(21) **[TAN.3, Section C] (worth 2 points)** Alfred Kinsey's research was controversial for a number of reasons, one being that he was postulating things about the sexual behaviors of all Americans that just didn't seem accurate to many Americans. This was due to a flaw in his statistical methodology. What was this flaw and to what extent was it compensated for by a large sample size? Tell me in an articulate paragraph.

The flaw in Kinsey's research methodology was a biased sample. He could only study people who voluntarily agreed to share their sexual history, and it is likely that those who had the most unusual sexual predilections were more likely to be willing to share their sexual past, perhaps because they wanted to know how "normal" they are. (Remember from the video how all the interviewees kept asking, "Am I normal?") This means that Kinsey's sample was dominated by people with unusual sexual habits, and thus this sample was not representative of the American public. For instance, the sample probably displayed more homosexual tendencies than the American public, which is why the public felt Kinsey's work to be biased. Although a larger sample size does provide us with more information on the *sample* of people, it does not make the sample more representative.

34

(22) **[TAN.3, Section C] (worth 2 points)** You have 5 observations of Y for males whose values are: Y_{males} = 1, 2, 3, 4, 5; you also have 3 observations for females: $Y_{females}$ = 6, 7, 8. Below is an attempt to prove that the average $\left(\frac{\Sigma_{i=males,females} Y_i}{8}\right)$ can be written as the weighted average of the males and females, where the weight equals the percent of the sample comprised of each gender. But there is (are) a mistake(s) in the proof. Identify all mistakes by circling then, drawing an arrow to the circle, and indicating the correct numbers that should have been used.

$$\text{Average} = \left(\frac{\Sigma_{i=males,females} Y_i}{8}\right) = \left(\frac{\{1+2+3+4+5\}+\{6+7+8\}}{8}\right) =$$

$$\left(\frac{\{1+2+3+4+5\}}{8}\right) + \left(\frac{\{6+7+8\}}{8}\right) =$$

$$\left(\frac{3}{3}\right)\left(\frac{\{1+2+3+4+5\}}{8}\right) + \left(\frac{5}{5}\right)\left(\frac{\{6+7+8\}}{8}\right)$$

$$\left(\frac{5}{5}\right)\left(\frac{\{1+2+3+4+5\}}{8}\right) + \left(\frac{3}{3}\right)\left(\frac{\{6+7+8\}}{8}\right)$$

36

(23) **[TAN.3, Section C] (worth 2 points)** Suppose that it is 2016 and the presidential election is being held between Hillary Clinton and Ted Nugent. I conduct a survey where I sampled 2,000 Democrats and 1,000 Republicans, and I found Clinton would win with 57% of votes, and this 57% was calculated using weighted averages to make the sample more representative of Americans. **Fox News then accused me of over-sampling Democrats so that I could make Clinton look better. How could I respond and defend myself? What would you recommend?** Tell me in an articulate paragraph.

I would explain to them that, although the sample was not representative of the American public, as the ratio of Democrats to Republicans in the sample was higher than the ratio of the voting public, the use of a weighted average makes the non-representative sample act as if it *was* representative. The weighted average would make it such that the average vote for Democrats counted about the same as the average vote for Republicans (assuming there are about the same number of Democrats and Republicans), even if there were twice as many Democrats in the sample as Republicans. Thus, Fox News' critique is not valid, and really represents their lack of understanding about polling and/or their willingness to deceive the public.

38