

Name: Key

Student Number:

**Exam 1B**

**Introductory Statistics  
Resource Economics 211**

**Please read the following rules and sign below to indicate that you do understand our rules. Thanks!**

- ❖ Please have nothing but a pencil/pen and your calculator on your desk. All book bags, backpacks need to be stored at the front of the class.
- ❖ You cannot share calculators.
- ❖ You cannot talk with those around you. Raise your hand, if you have any questions or problems.
- ❖ No cell phones. If you are seen with a cell phone, we will take your exam.
- ❖ Please check your Exam; there should be **8 pages** and three parts:
  - A – information for us.
  - B – 15 MC questions - answer on the “bubble sheet.”
  - C – Problems, **questions 1-6**. You must **provide exact formulas and show all work. You must indicate clearly what you are calculating – the sample mean and population mean, for example, are different.**
- ❖ We will provide you with a formula sheet and a bubble sheet. **Only the bubble sheet answers count for the MC questions.**
- ❖ You will need a #2 pencil for the bubble sheet.
- ❖ Neatness counts! **If we cannot read your answers, then we cannot give you credit.**
- ❖ Please keep the parts stapled together! To be safe, please put your name on all pages!
- ❖ Relax, you'll think better. Stress is bad. Good Luck.

**Okay, Dan. I've read and agree to these rules.** \_\_\_\_\_

**Part A. No points here, just necessary information. Declare your weights.** You can revise these weights at Exam 2 and again at the Final. Answer the following questions on the ScanTron Bubble Answer sheet with a number 2 pencil.

1. What is the color of your exam?
  - a. Yellow
  - b. Blue
  
2. What weight do you wish to place on Exam 1?
  - a. 10%
  - b. 20%
  - c. 30%
  
3. What weight do you wish to place on Exam 2?
  - a. 10%
  - b. 20%
  - c. 30%
  
4. What weight do you wish to place on the Final Exam?
  - a. 20%
  - b. 30%
  - c. 40%
  
5. What weight do you wish to place on OWL quizzes?
  - a. 10%
  - b. 20%

Name: Key

Student Number:

**Part B: Multiple Choice – answer on the “bubble sheet.” Answer all questions, 2 points each.**

6. The median is sometimes preferred to the mean because it ...
- balances the data distribution.
  - is resistant.**
  - uses all data values.
  - is a measure of center.
  - can be used with qualitative data.
7. The data on “favorite athletic shoe” from the OWL survey should be graphically displayed using ...
- A grouped data table with single value classes.
  - A relative frequency bar chart.**
  - A frequency histogram.
  - A relative frequency histogram.
  - None of the above.

8. The stem and leaf diagram at the right shows the number of days to maturity for 40 short-term investments. How many investments have no more than 60 days to maturity?

- 12
- 13**
- 18
- 22
- 28

Stem-and-leaf of No. Days N = 40  
Leaf Unit = 1.0

```
3 3 689
4 4 7
12 5 01135567
(10) 6 0234456789
18 7 0001589
11 8 0135679
4 9 5899
```

9. A variable with an unknown distribution has a mean of 500 and a standard deviation of 30. At least 75% of the data will fall between what two values?

- 520 and 480
- 540 and 580
- 440 and 560**
- 460 and 540
- 447.5 and 553.5

10. Suppose we know that a variable is normally distributed with mean of 110 and a standard deviation of 24. Approximately what percent of the data fall between the two values 86 and 182?

- 16
- 34
- 68
- 80
- 84**

Name: Key

Student Number:

11. The portion of the universe that has been selected for analysis is called a...
- parameter.
  - statistic.
  - frame.
  - sample.
  - population proportion
12. "According to new research, just reported in the Journal of Hypertension, the number of American adults with high blood pressure has jumped 30 percent in the past decade, to 65 million ..."
- This statement is an example of:
- descriptive statistics
  - a sample statistic
  - a population parameter
  - inferential statistics
  - regression.
13. For a set of data, the number of observations that fall into one of the distinct classes for the distribution is called a
- frequency.
  - descriptive statistic.
  - relative frequency.
  - class boundary.
  - None of the Above
14. A homeowners association considering legal action against a homeowner conducted a survey of homeowners on two streets. Six of 8 homes on the same street as the problem house called for legal action; 50 % on the second street favored legal action. Which type of chart might be most effective for conveying the information about percentage of residents favoring legal action by street?
- histogram.
  - pie chart.
  - bar chart.
  - scatter diagram.
  - stem and leaf diagram.
15. A histogram is used to display which of the following characteristics for a quantitative variable?
- The spread of the data.
  - The center of the data.
  - The shape of the distribution.
  - All of the above.
  - None of the above.

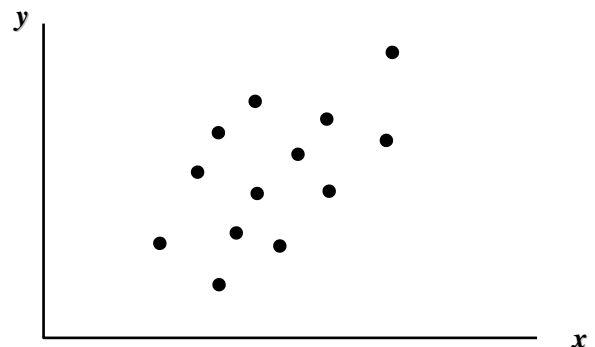
Name: Key

Student Number:

16. The interquartile range is sometimes preferred to the standard deviation as a measure of variation because it:
- is easier to compute.
  - utilizes all the data in its computation.
  - is the appropriate measure of variation to use with the sample mean.
  - is resistant.
  - It is never preferred to the standard deviation.
17. An “error,” in the context of linear regression, is the
- difference between two observed y-values.
  - vertical difference between the x-value and the y-value.
  - perpendicular distance from the regression line to an observed y-value.
  - vertical distance from the regression line to an observed y-value.
  - difference between the sample mean and the observed y-value.
18. According to the least squares criterion or principle, we minimize . . .
- the sum of the errors squared by choosing values for y and x.
  - the sum of the errors by choosing the population parameter values  $b_0$  and  $b_1$ .
  - the sum of the squared errors by choosing estimates of the population parameter values  $b_0$  and  $b_1$ .
  - the sum of the squared errors by choosing the population parameter values  $b_0$  and  $b_1$ .
  - the sum of the squared errors by choosing values for x and y.

19. Given our discussion of correlation coefficients, what value for r would you guess for the scatter diagram at the right?

- $r = -0.95$
- $r = -0.50$
- $r = 0.00$
- $r = 0.50$
- $r = 0.95$



20. Given the Minitab descriptive statistics for number of siblings shown below, which **one** of the following statements is *definitely* correct?

Variable	N	N*	Mean	Median	TrMean	StDev
Siblings	249	0	1.8153	2.0	1.7067	1.3068

Variable	SE Mean	Minimum	Maximum	Q1	Q3
Siblings	0.0828	0.00	8.00	1.00	2.00

- Fifty percent of the class has 2 or fewer siblings.
- Fifty percent of the class has more than 2 siblings.
- Twenty five percent of the class has between 1 and 2 siblings.
- At least twenty five percent of the class has 2 siblings.
- Twenty five percent of the class has more than 2 siblings.

Name: Key

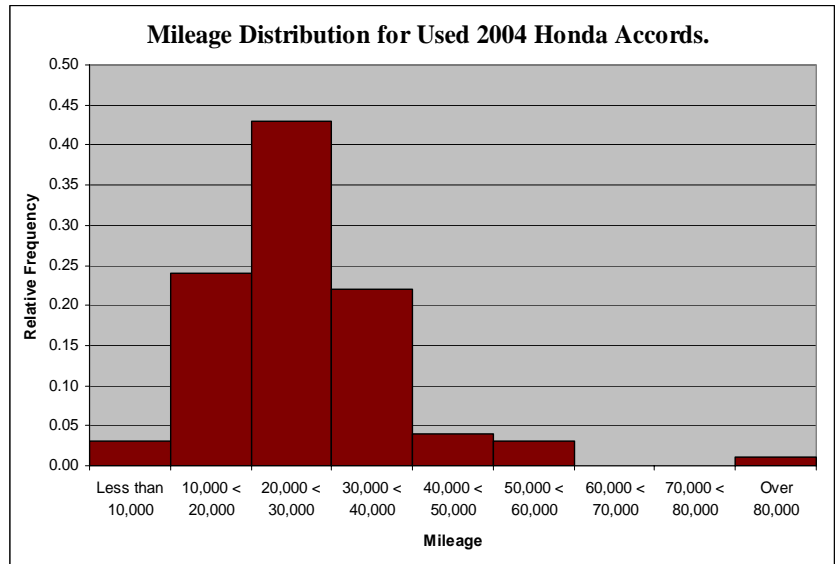
Student Number:

**Part C: Please answer all questions in this section.** Point totals are shown in the left margin.  
**Show all work: No formulas, no credit. No calculations, no credit.**

1. On January 21, 2005, there were 100 Used 2004 Honda Accords available through AutoTrader.com. Data for the number of miles on the cars (mileage) available are presented in the graph below. Answer the question below using the graph at the right.

- (3) a. How would you characterize the shape of the distribution of used 2004 Accord mileages?

**The shape is quite symmetric, with the exception of the one car with over 80,000 miles. That may be an outlier, so we would conclude the distribution is "symmetric," perhaps even "Bell-Shaped."**



- (3) b. About what percentage of Used 2004 Honda Accords had fewer than 30,000 miles on them?

**Add the heights of the bars for the classes: Less than 10,000; 10,000 < 20,000; and 20,000 < 30,000. The sum of the bar heights is about 0.70. So about 70 % had fewer than 30,000 miles on the odometer.**

- (3) 2. The variable  $x$  is known to have a population mean of **3,250**, but we have no information about the shape of the distribution. *At least what percentage of the data will fall within 3.17 standard deviations of the population mean? (Round your answer to 1 decimal place.)*

**Regardless of how the data are distributed, at least  $\left(1 - \frac{1}{k^2}\right) \times 100\%$  of the values will fall within  $k$  standard deviations of the mean.**

**For this problem:  $\left(1 - \frac{1}{k^2}\right) \times 100\% = \left(1 - \frac{1}{(3.17)^2}\right) \times 100\% = \left(1 - \frac{1}{10.0}\right) \times 100\% = 90\%$**

Name: Key

Student Number:

3. The data in the array at the right are a sample of summer earnings by ResEc 211 Students for 2007 in dollars. Use these data to complete the following questions.

- (12) a. **Create a grouped data table showing frequency and relative frequency distributions.** Your classes are to have an equal width of \$1,000. Provide “titles” in the first row for each of the three columns in the table. (You may not have to use all the rows provided in the table.)

Summer Earnings	Frequency	Relative Frequency
\$ 0 < \$1,000	6	0.176
\$1,000 < \$2,000	4	0.118
\$2,000 < \$3,000	8	0.235
\$3,000 < \$4,000	4	0.118
\$4,000 < \$5,000	5	0.147
\$5,000 < \$6,000	3	0.088
\$6,000 < \$7,000	1	0.029
\$7,000 < \$8,000	0	0.000
\$8,000 < \$9,000	1	0.029
\$9,000 < \$10,000	2	0.059
	34	1.000

SumSal
0
0
0
300
500
500
1000
1500
1500
1790
2000
2000
2000
2000
2000
2100
2200
2500
3000
3000
3000
3000
4000
4000
4000
4000
4000
4000
4200
5000
5000
5000
6000
8760
9000
9800

- (8) b. **Complete a 5-Number summary** for the sample of summer earnings. **Report 2 decimal places and provide all work. Identify the 5 numbers** and be sure your methods are clear.)

**The 5-Number Summary includes: the Minimum, Q1, Q2, Q3, and the Maximum.**

**Minimum = \$ 0**

$$Q_1 @ \frac{(34+1)}{4}; Q_1 @ \frac{35}{4} = 8.75; Q_1 = X_8 + 0.75(X_9 - X_8) = 1500 + 0.75(1500 - 1500) = \$1,500$$

$$Q_2 @ \frac{2(34+1)}{4}; Q_2 @ \frac{35}{2} = 17.5; Q_2 = X_{17} + 0.5(X_{18} - X_{17}) = 2200 + 0.5(2500 - 2200) = \$2,350$$

$$Q_3 @ \frac{3(34+1)}{4}; Q_3 @ \frac{105}{4} = 26.25; Q_3 = X_{26} + 0.25(X_{27} - X_{26}) = 4000 + 0.25(200) = \$4,050$$

**Maximum = \$ 9,800.**

- (6) c. Are there any potential outliers? Be sure to show how you made your determination.

**Lower Limit =  $Q_1 - 1.5 IQR = \$1,500 - 1.5(\$2,550) = -2,325$ ; because this is not a reasonable limit, we use a lower limit of 0.**

**Upper Limit =  $Q_3 + 1.5 IQR = \$4,050 + 1.5(\$2,550) = 7,875$ .**

**The data values in the sample, \$8,760, \$9,000 and \$9,800 would all be considered a potential outliers.**

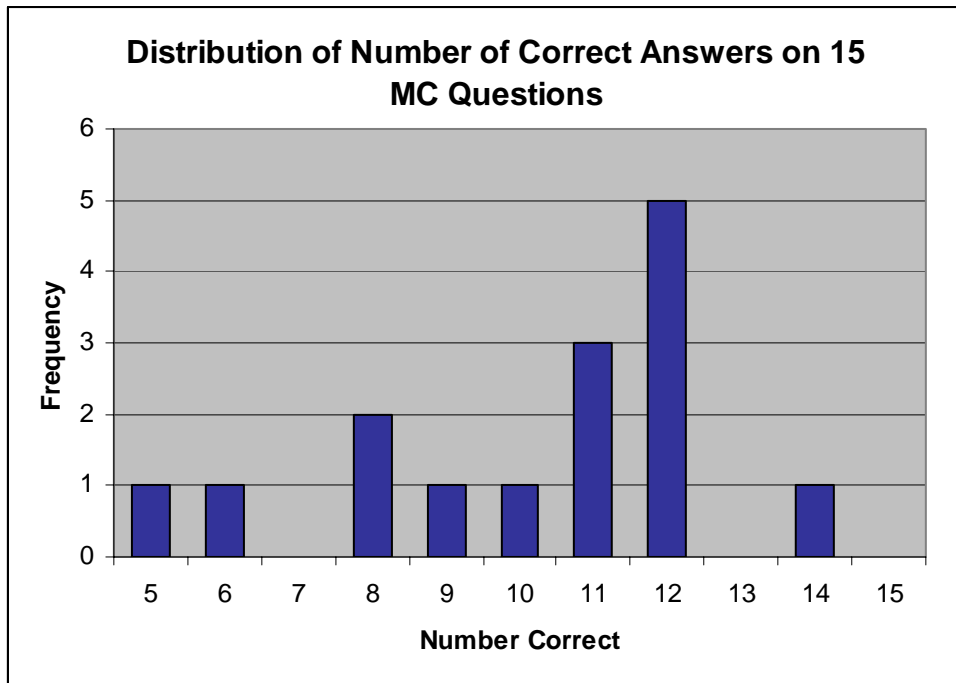
Name: Key

Student Number:

4. The variable “Ncorrect” is the number of multiple choice questions answered correctly on last year’s Exam 1. To anticipate the possible distribution for this year’s Exam 1, a sample of 15 observations was drawn from last year’s class.

Ncorrect
6
12
12
14
9
10
11
11
11
12
8
12
8
12
5

- (8) a. Use the template below to create a graphic representation of the distribution for the number of questions answered correctly (Ncorrect).



- (3) b. What kind of graph did you create? Why?

The graph should be a “bar graph.” Because these are discrete data, the bars should not touch. Of course, you could put together a proper dotplot or stem and leaf diagram.

- (3) c. What is the median for this sample of data? Be sure to report units.

$$Q_2 @ \frac{2(n+1)}{4}; Q_2 @ \frac{2(15+1)}{4}; Q_2 @ 8.$$

You must first sort the data. Then, we find the median at observation 8 in the ordered list.

$Q_2 = 11$  questions correct.

- (3) d. What is the mean for this **sample** of data? Use two decimal places and be sure to report units.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum x}{n} = \frac{153}{15} = 10.20 \text{ questions correct.}$$

Name: Key

Student Number:

5. Data were collected for a sample of 30 used 2004 Honda Accords. The relationship between the price of the used car and the mileage was estimated by OLS to be:

$$\hat{y} = 21640 - 0.079x.$$

- (2) a. Which variable, is the dependent variable,  $y$ , for this relationship? **Price is the dependent variable. The Price of a used car “depends” on the number of miles on the odometer.**
- (4) b. Dan is considering buying the used cars shown in the table below. Which appears to be the best deal? Complete the table and tell Dan **which car appears to be the best deal** and **why** it appears to be the best deal.

<i>Mileage</i>	<i>Price</i>	$\hat{y}$	$e$
24680	\$19,495	<b>\$19,690.28</b>	<b>-\$195</b>
20027	\$19,995	<b>\$20,057.87</b>	<b>-\$63</b>
44273	\$18,485	<b>\$18,142.43</b>	<b>\$343</b>

**Choose the car that has the actual price that is lower than the predicted price ( $\hat{y}$ ). That would also be the car with the “greatest negative error.” In this case, it is the first car.**

6. The grades (in percent correct) for all six students in a small seminar course are shown in the table below.

- (4) a. Calculate the class mean grade for the course (round to 2 decimals – report units).

$$\mu = \frac{\sum_{i=1}^N x_i}{N} = \frac{492.7}{6} = 82.12 \text{ percent}$$

$x$		
76.1	<b>-6.02</b>	<b>36.200</b>
92.5	<b>10.38</b>	<b>107.814</b>
97.7	<b>15.58</b>	<b>242.840</b>
71.5	<b>-10.62</b>	<b>112.714</b>
60.3	<b>-21.82</b>	<b>475.967</b>
94.6	<b>12.48</b>	<b>155.834</b>
<b>492.7</b>	<b>0.00</b>	<b>1131.368</b>

- (8) b. Determine the **standard deviation** for this data set **using the defining formula**. Complete the table provided to show the steps required in calculating the standard deviation. (Round your answer to 2 decimal places and include the units.)

$$\sigma_x = \sqrt{\frac{\sum (x - \mu_x)^2}{N}} = \sqrt{\frac{1131.368}{6}} = \sqrt{188.5614} = 13.73 \text{ percent}$$