Exam 1
Statistics for Business (22S:008, Bognar) **September 21, 2006**

Name: Key.

TA: ____________________

Section: ____________________

- **Instructions:**
  - Use only a number 2 pencil.
  - Fill in your name, TA, and discussion section on this test form.
  - Fill in your name, student ID number, discussion code (see below), and test form (see below) on the answer sheet.
  - Read *all* of the answers, choose the *best* answer, grid your choice on the answer sheet, and circle your answer on this test form.
  - Turn in both the answer sheet and test form when you are finished.
  - Keep your crib sheet; you will be able to use it on the final exam.
- **Discussion Code:** Fill in your discussion code in columns O and P on the answer sheet. Example: Students in Joe's 8:30–9:20 section ($A03$) should fill in 0 in column O and 3 in column P. Students in Sorah's 3:30–4:20 section (SCA) should fill in 2 in column O and 1 in column P.
  - $A03$ = Joe (8:30–9:20) .... $O = 0, P = 3$
  - $A04$ = Greg (8:30–9:20) .... $O = 0, P = 4$
  - $A06$ = Zhanchong (8:30–9:20) .... $O = 0, P = 6$
  - $A07$ = Katherine (9:30–10:20) .... $O = 0, P = 7$
  - $A08$ = Greg (9:30–10:20) .... $O = 0, P = 8$
  - $A09$ = Joe (10:30–11:20) .... $O = 0, P = 9$
  - $A10$ = Wade (10:30–11:20) .... $O = 1, P = 0$
  - $A11$ = Barry (11:30–12:20) .... $O = 1, P = 1$
  - $A12$ = Wade (11:30–12:20) .... $O = 1, P = 2$
  - $A13$ = Katherine (12:30–1:20) .... $O = 1, P = 3$
  - $A14$ = Dan (12:30–1:20) .... $O = 1, P = 4$
  - $A15$ = Yiwen (1:30–2:20) .... $O = 1, P = 5$
  - $A16$ = Dan (1:30–2:20) .... $O = 1, P = 6$
  - $A17$ = Yiwen (2:30–3:20) .... $O = 1, P = 7$
  - $A18$ = Jie (2:30–3:20) .... $O = 1, P = 8$
  - SCA = Sorah (3:30–4:20) .... $O = 2, P = 1$
  - SCB = Jie (3:30–4:20) .... $O = 2, P = 2$
  - SCC = Jay (4:30–5:20) .... $O = 2, P = 3$
- **Test Form:** An test form is printed at the bottom of this page. Fill in the type of form (i.e. A, B, C, or D) on your answer sheet.
- **Scoring:** Each of the 25 questions on the exam is worth 4 points (100 points total). Incorrectly following these instructions will result in a 3 point deduction.

Test Form: A

1. The following data describe the daily gross income (in thousands of dollars) of a small restaurant over a three day period: 3.3, 2.8, 2.9. Find the standard deviation s of this dataset.

   (a) 0
   (b) 0.070
   (c) 0.047
   (d) 0.265
   (e) 0.216

2. Which of the following is/are true?

   (a) In general, surveys that use voluntary response samples yield results that accurately reflect the views of the population
   (b) A study is "skewed" if it favors a certain result
   (c) Suppose we wish to obtain a random sample from a population. If each individual in the population has the same chance of being chosen, then we must be performing simple random sampling (SRS)
   (d) Both (b) and (c)
   (e) None of the above

3. Consider the following stem and leaf plot. Which of the following is/are true? Note that the smallest number in the dataset equals 76.

Stem	Leaves
7	6
8	
9	
10	58
11	335

76 105 108 113 113 115

   (a) The distribution is skewed to the right
   (b) The median Q_2 is equal to 105.8
   (c) The interquartile range (IQR) is equal to 8.0
   (d) Both (a) and (c)
   (e) Both (a) and (b)

4. Which of the following is/are true?

   (a) Ones area code (e.g. 319) constitutes a quantitative variable
   (b) A histogram is commonly used to describe categorical variables
   (c) If the variance s^2 of a dataset is equal to 2, then the standard deviation is equal to 4
   (d) The standard deviation s describes a "typical" deviation from the median Q_2
   (e) None of the above

5. What is the best measure of variability for a large dataset that is strongly skewed to the right? Why?

   (a) Standard deviation, since the standard deviation is affected by outliers
   (b) Standard deviation, since the standard deviation is not affected by outliers
   (c) Interquartile range, since the IQR is affected by outliers
   (d) Interquartile range, since the IQR is not affected by outliers
   (e) Range, since the range is not affected by outliers

6. Suppose a manufacturer has a large warehouse containing 100,000 bolts. They want to estimate the average weight of the bolts so they are better able to estimate shipping costs. To this end, 100 bolts were randomly selected and weighed. The average weight of the selected bolts was 0.062 ounces (i.e. the total weight of the 100 selected bolts was 6.2 ounces). Which of the following is/are true?

   (a) The population consists of the 100,000 bolts in the warehouse

   (b) The sample is the average weight of the selected bolts (i.e. 0.062 ounces)

   (c) The parameter is the average weight of the selected bolts (i.e. 0.062 ounces)

   (d) Both (a) and (c)

   (e) Both (b) and (c)

7. A chocolate manufacturer produces boxes of chocolates whose weights X follow a normal distribution with mean $\mu = 16.22$ ounces and standard deviation $\sigma = 0.1$ ounces. Suppose the weight printed on the boxes of chocolates is 16 ounces. If we wish to determine the proportion of boxes that contain less than 16 ounces of chocolates, what is the proper way to mathematically begin this problem?

   (a) $Z > 16 = \cdots$

   (b) $X < 16 = \cdots$

   (c) $P(Z < 16) = \cdots$

   (d) $P(X < 16) = \cdots$

   (e) None of the above

8. In reference to question (7), find the probability that a randomly selected box contains less than 16 ounces of chocolates.

   $P(X<16) = P(Z<-2.2) = 0.0139$

   (a) 0.9861

   (b) 0.0139

   (c) 0.4129

   (d) 0.9750

   (e) 0.0446

9. In reference to question (7), suppose only the heaviest 61.79% of boxes are sent to retailers, while the lightest 38.21% of boxes are sent to discounters. To be sent to a discounter, how light must a box of chocolate be?

   $z = \frac{x-\mu}{\sigma} \rightarrow x = \mu + z\sigma = 16.22 + (-0.3)(0.1) = 16.19$

   (a) 15.68 ounces

   (b) 15.97 ounces

   (c) 16.19 ounces

   (d) 16.03 ounces

   (e) 16.25 ounces

10. In reference to question (7), what is the inter-quartile range (IQR) of the weights?

   $Q_1 = \mu - 0.67\sigma = 16.22 - 0.67(0.1) = 16.153$

   $Q_3 = \mu + 0.67\sigma = 16.22 + 0.67(0.1) = 16.287$

   $IQR = Q_3 - Q_1 = 0.134$

   (a) 0.067 ounces

   (b) 0.670 ounces

   (c) 0.200 ounces

   (d) 1.340 ounces

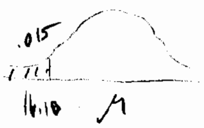
   (e) 0.134 ounces

11. An inexperienced researcher wants to test a new cold pill. Treatment and control groups were formed as follows: coffee drinkers were placed in the treatment group, non-coffee drinkers were placed in the control group. Which of the following is/are true?

   (a) If the treatment group has a lower incidence of colds, it must be due to the cold pill
   (b) The effect of coffee consumption and type of treatment are confounded
   (c) With the exception of coffee consumption and treatment type, the treatment and control groups are guaranteed to be "equivalent" or "similar"
   (d) This study is an observational study
   (e) None of the above

12. Suppose the weight of bags of tortilla chips follow a normal distribution with mean μ ounces and standard deviation $\sigma = 0.2$ ounces. If 1.5% of the bags weigh less than 16.10 ounces, what is the mean μ?

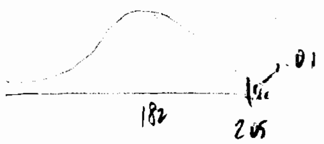
   (a) $\mu = 15.67$
   (b) $\mu = 16.53$
   (c) $\mu = 15.90$
   (d) $\mu = 16.44$
   (e) $\mu = 16.76$



$z = \frac{x-\mu}{\sigma} \to \mu = x - z\sigma$
$= 16.10 - (-2.17)(0.2)$
$= 16.53$

13. Cholesterol levels of Canadian women are normally distributed with mean $\mu = 182$ and standard deviation σ. If 1% of Canadian women have a cholesterol level more than 205, what is the standard deviation σ?

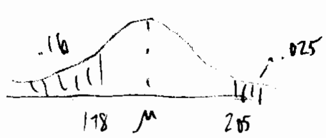
   (a) $\sigma = 9.87$
   (b) $\sigma = 11.50$
   (c) $\sigma = 4.94$
   (d) $\sigma = 8.65$
   (e) $\sigma = 5.21$



$z = \frac{x-\mu}{\sigma} \to \sigma = \frac{x-\mu}{z} = \frac{205-182}{2.33} = 9.87$

14. Cholesterol levels in Mexican men are normally distributed with mean μ and standard deviation σ. Out of 1000 randomly chosen Mexican males, 25 had a cholesterol level more than 205, while 160 had a cholesterol level less than 178. Use the empirical rule to find μ and σ.

   (a) $\mu = 196, \sigma = 9$
   (b) $\mu = 196, \sigma = 4.5$
   (c) $\mu = 187, \sigma = 9$
   (d) $\mu = 187, \sigma = 4.5$
   (e) Impossible to determine with the given information



$205 = \mu + 2\sigma$
$178 = \mu - \sigma$
$\Rightarrow 3\sigma = 27$
$\Rightarrow \sigma = 9 \Rightarrow \mu = 187$

15. Scores on the SAT exam follow a normal with mean $\mu = 500$ and standard deviation $\sigma = 100$. Scores on the ACT exam follow a normal distribution with mean $\mu = 18$ and standard deviation $\sigma = 6$. Suppose Hank got a 600 on the SAT exam, and Mario got a 25 on the ACT exam. Relatively speaking, who's academic knowledge is better? Why? *Assume both tests measure academic knowledge similarly.*

   (a) Hank's academic knowledge is better since his standard score ($z-$score) is higher than Mario's
   (b) Hank's academic knowledge is better since his percentile is higher than Mario's
   (c) Mario's academic knowledge is better since his standard score ($z-$score) is higher than Hank's
   (d) Mario's academic knowledge is better since his percentile is higher than Hank's
   (e) Both (c) and (d)

Hank → $z = 1$
Mario → $z = 1.167$

16. Which of the following is/are true?

   (a) The standard score (or z−score) describes how many standard deviations an observation lies above or below the mean

   (b) For any events A and B, we know that $P(A \text{ and } B) = P(A)P(B)$

   (c) Two events A and B are independent if $P(B|A) = P(A)$

   (d) Two events A and B are mutually exclusive if $P(A \text{ and } B) > 0$

   (e) Both (a) and (d)

17. It is known that 14% of all adults have diabetes. Given an adult has diabetes, the probability that he/she has high blood pressure is 0.30. Find the probability that a randomly selected adult has diabetes *and* high blood pressure.

   (a) 0.028

   (b) 0.084

   (c) 0.140

   (d) 0.042

   (e) 0.467

$P(D) = 0.14 \quad P(H|D) = 0.30$

$P(D \text{ and } H) = P(D)P(H|D) = 0.14(0.30) = 0.042$

18. A box contains 2 black, 3 red, and 5 green balls. Suppose two balls are randomly chosen from the box *without* replacement. What is the probability that the second ball is black given that the first ball was green (i.e. find $P(B_2|G_1)$)?

   (a) 0.100

   (b) 0.111

   (c) 0.200

   (d) 0.222

   (e) 0.350

$P(B_2|G_1) = \frac{2}{9} = 0.222$

19. In the U.S., 60% of new restaurants go out of business within the first year. If 3 new restaurants are randomly selected, find the probability that *none* remain in business after the first year. *Assume independence.*

   (a) 0.064

   (b) 0.936

   (c) 0.216

   (d) 0.784

   (e) 1.200

let D = out of biz.

$P(D_1 \text{ and } D_2 \text{ and } D_3) \overset{indep}{=} P(D_1)P(D_2)P(D_3)$

$= 0.60^3$

$= 0.216$

20. In reference to question (19), suppose 3 new restaurants are randomly selected. Find the probability that *1 or more* of the restaurants remain in business after the first year. *Assume independence.*

   (a) 0.064

   (b) 0.936

   (c) 0.216

   (d) 0.784

   (e) 1.200

$P(1 \text{ or more in biz}) = 1 - P(\text{none in biz})$

$= 1 - 0.216$

$= 0.784$

$0.064 = 0.4^3 = P(\text{all in biz after } 1^{st} \text{ year})$

21. As motivated in class, the ability to read mathematical equations is very important. One may suggest that variability could be measured by the following equation:

$$s_{abs} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \cdots + |x_n - \bar{x}|}{n}$$

Compute s_{abs} for the following dataset: 3.3, 2.8, 2.9. Note that $|\cdot|$ denotes absolute value.

(a) $s_{abs} = 0.30$

(b) $s_{abs} = 0.20$

(c) $s_{abs} = 0.265$

(d) $s_{abs} = 0.216$

(e) $s_{abs} = 0.447$

$n = 3$
$\bar{x} = 3$
$s_{abs} = 0.20$

22. Suppose a bowl has four chips: two of the chips are black, and two of the chips are red. One of the black chips has the number 1 written on it, while the other black chip has the number 2 written on it. One of the red chips has the number 2 written on it, while the other red chip has the number 3 written on it. Suppose one chip is randomly chosen from the bowl. Consider the following events:

$$A = \text{a black chip is chosen}$$
$$B = \text{the number on the chip is 2}$$

$B_1\ B_2\ R_2\ R_3$

What is $P(A|B)$?

(a) 0.50

(b) 0.25

(c) 0.75

(d) 0.375

(e) 0.825

$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{1/4}{2/4} = \frac{1}{2} = 0.50$

23. In reference to question (22), are A and B independent? Why?

(a) Not independent, since $P(B|A) \neq P(A)$

(b) Not independent, since $P(A|B) \neq P(B)$

* (c) Not independent, since $P(A \text{ and } B) \neq P(A)P(B)$

(d) Independent, since $P(A|B) = P(A)/P(B)$

(e) Independent, since $P(A|B) = P(A)$

$P(A) = 2/4 = 0.50$
Since $P(A|B) = P(A)$ then indp!

24. In reference to question (22), what is $P(A \text{ or } B)$?

(a) 1.00

(b) 0.50

(c) 0.25

(d) 0.75

(e) 0.60

$P(A \text{ or } B) \overset{\text{not mut}}{=} P(A) + P(B) - P(A \text{ and } B)$
$= \frac{2}{4} + \frac{2}{4} - \frac{1}{4} = 0.75$

25. Which of the following is/are true?

(a) I love Statistics for Business

(b) I think Matt's dog Hank (a.k.a. Hankenstein) is cute

(c) I know my TA's name

(d) I look forward to learning more about Statistics

(e) All of the above