**Statistics 300 : Summer 2013**

**Instructor : L. C. Larsen**

**Student name & ID#:** _____

**Student signature:** _____

**Exam : Unit 1**

**Time allowed : 2 hours and 5 minutes**

**Resources allowed:**

== >   Textbook (Author: Triola)
== >   Notes/helps written by the student
== >   Quiz and exam solutions written by instructor
== >   Quiz and exam solutions written by the student
== >   Calculator/laptop of choice (no outside messages)
== >   Instructor at 916-346-6324

**Resources not allowed:**

== >   Consultants other than the instructor
== >   No phones, unless used as a calculator only

1.   **Use the data in the following table to answer parts (a) through (e).**

(1 point, 1 minute)
**(a)  What is the probability that a randomly selected person from this sample will have college-level education?**

| Education Level | Employment Type for People in Sample | | | | Total |
|---|---|---|---|---|---|
| | Farm | Factory | Health | Financial | |
| No Schooling | 256 | 267 | 212 | 292 | 1027 |
| Elementary | 226 | 245 | 288 | 294 | 1053 |
| Secondary | 267 | 228 | 257 | 291 | 1043 |
| College | 266 | 208 | 219 | 250 | 943 |
| Graduate | 207 | 293 | 263 | 233 | 996 |
| Total | 1222 | 1241 | 1239 | 1360 | 5062 |

(3 points, 3 minutes)
**(b)  What is the probability that a person randomly selected from this sample will have no schooling given that the person works in the "Farm" sector of the economy?**

(3 points, 3 minutes)
**(c)  What is the probability that a person randomly selected from this sample will have no schooling given that the person works in the "Financial" sector of the economy?**

(3 points, 3 minutes)
**(d)  Based on your answers to parts (b) and (c), say whether "no schooling" is independent or dependent on the sector of the economy, and explain why.**

(3 points, 3 minutes)
**(e)  What is the probability that a person randomly selected from this sample will have "Secondary" education or work in the "Health" sector of the economy?**

(5 points; 5 minutes)

**2. The scores of two different tests have bell-shaped distributions. The scores for Test A have a mean of 110 points and a standard deviation of 12 points. The scores for Test B have a mean of 330 and a standard deviation of 60 points. Use this information to answer the two questions below.**

**What percent of the scores for Test A should be found between 86 and 134? Show why!**

**Would a score of 220 on Test B be considered unusual? Show why!**

(10 points; 8 minutes)

**3.**

**Complete the columns in the "Frequency Distribution" table using the data values given below, and answer the two questions beoow the table.**

**Frequency Distribution**

| Class Limits Lower | Class Limits Upper | Tally | Frequency | Relative Frequency | Cumulative Frequency | Cumulative Relative Frequency | Class Midpoint | Class Boundary |
|---|---|---|---|---|---|---|---|---|
| 40 | 80 | | | | | | | |
| 90 | 130 | | | | | | | |
| 140 | 180 | | | | | | | |

| Class Width |
|---|
| |

| Data: | 75.9 | 91.4 | 146.7 | 101.1 | 150.1 |
|---|---|---|---|---|---|
| | 61.0 | 78.7 | 86.8 | 134.2 | |

**What is the frequency for Class #2 ?**          _____

**What is the uppoer limit for Class #1 ?**          _____

**4. Give a short definition of statistics** (1 point; 1 minute)**:**

_____

(5 points; 5 minutes)

   **5. For each situation below, select the appropriate statistical term from the list provided
   and write it in the blank next to the description or situation. Choose the term that is best
   connected to the <u>underlined text</u> in the description or situation.**

| Terms: | 1. randomization | 4. blinding | 7. experimental unit |
|--------|------------------|-------------|----------------------|
|        | 2. replication   | 5. placebo / control | 8. treatment |
|        | 3. confounding   | 6. block    |                      |

_____  Educators tested the effect of three different breakfasts on students
diagnosed with Attention Deficit Hyperactivity Disorder (ADHD).  A
stratified random sample of 2400 such kids included 300 boys and 300
girls in each of four age groups.  Critics say race and culture should have
been included in the structure of the stratified sampling.  In each stratum,
each child was assigned a normal, high fiber, or high protein breakfast
with equal and independent probability, and <u>the kids did not know which one</u>.

_____  Educators tested the effect of three different breakfasts on students
diagnosed with Attention Deficit Hyperactivity Disorder (ADHD).  A
stratified random sample of 2400 such kids included 300 boys and 300
girls in each of four age groups.  Critics say race and culture should have
been included in the structure of the stratified sampling.  In each stratum,
<u>each child</u> was assigned a normal, high fiber, or high protein breakfast
with equal and independent probability, and the kids did not know which one.

_____  Educators tested the effect of three different breakfasts on students
diagnosed with Attention Deficit Hyperactivity Disorder (ADHD).  A
stratified random sample of 2400 such kids included 300 boys and 300
girls in each of four age groups.  <u>Critics say race and culture should have
been included in the structure of the stratified sampling</u>.  In each stratum,
each child was assigned a normal, high fiber, or high protein breakfast
with equal and independent probability, and the kids did not know which one.

_____  Educators tested the effect of three different breakfasts on students
diagnosed with Attention Deficit Hyperactivity Disorder (ADHD).  A
stratified random sample of 2400 such kids included 300 boys and 300
girls in each of four age groups.  Critics say race and culture should have
been included in the structure of the stratified sampling.  In each stratum,
each child was assigned <u>a normal, high fiber, or high protein breakfast</u>
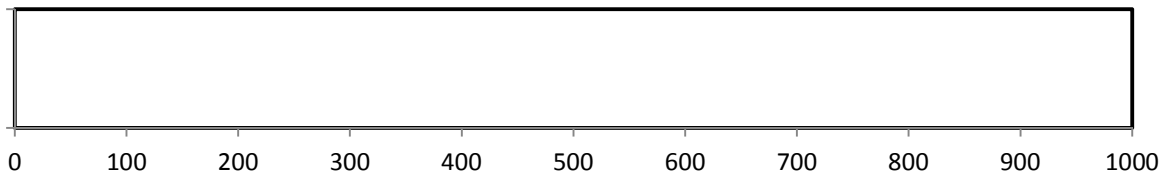with equal and independent probability, and the kids did not know which one.

_____  Educators tested the effect of three different breakfasts on students
diagnosed with Attention Deficit Hyperactivity Disorder (ADHD).  A
stratified random sample of 2400 such kids included 300 boys and 300
girls in each of four age groups.  Critics say race and culture should have
been included in the structure of the stratified sampling.  In each stratum,
each child was <u>assigned</u> a normal, high fiber, or high protein breakfast
<u>with equal and independent probability</u>, and the kids did not know which one.

(4 points; 4 minutes)

**6. A subway train has six safety systems that will stop the train in an emergency. Each system is independent of the others and each system will work with probability 0.6 each time an emergency happens.  What is the probability that exactly 5 of the systems will work the next time an emergency happens?**

(5 points and 4 points; 8 minutes)

**7. (a) For the set of 77 values shown below in sorted order, prepare a Boxplot inside the rectangle that is above the number line.**

| 0 | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 5 | 19 | 23 | 23 | 27 | 27 | 27 | 32 | 42 |
| 47 | 47 | 47 | 52 | 58 | 58 | 70 | 70 | 70 | 83 |
| 83 | 96 | 110 | 118 | 118 | 118 | 140 | 148 | 148 | 156 |
| 190 | 207 | 207 | 253 | 263 | 282 | 292 | 302 | 302 | 364 |
| 364 | 408 | 419 | 430 | 430 | 442 | 442 | 465 | 465 | 465 |
| 476 | 488 | 488 | 512 | 512 | 512 | 524 | 524 | 536 | 548 |
| 561 | 561 | 561 | 586 | 611 | 637 | 650 | 650 | 689 | 702 |
| 716 | 716 | 743 | 770 | 770 | 840 | 854 | | | |

**(b) For the set of 77 values shown above in sorted order, what percentile is represented by the value 70?**

(8 points; 10 minutes)
  **8. Answer parts (a), (b), and (c).**
     **Use the columns in the table in any way you wish to use them.**

     **(a)**    **Is this distribution "proper" (circle "YES" or "NO")?**     **YES**     **NO**

     **Why?** _____

| x | P(x) | | | | |
|-----|------|---|---|---|---|
| 146 | 0.26 | | | | |
| 163 | 0.16 | | | | |
| 183 | 0.22 | | | | |
| 215 | 0.36 | | | | |

     **(b)**    **Write the <u>formulas</u> for the mean, the variance, and the standard deviation of a discrete probability distribution.**

$$\mu = \underline{\hspace{8cm}}$$

$$\sigma^2 = \underline{\hspace{8cm}}$$

$$\sigma = \underline{\hspace{8cm}}$$

     **(c)**    **Write the <u>values</u> for the mean, the variance, and the standard deviation of <u>this</u> discrete probability distribution.**

$$\mu = \underline{\hspace{4cm}}$$

$$\sigma^2 = \underline{\hspace{4cm}}$$

$$\sigma = \underline{\hspace{4cm}}$$

(4 points and 2 points; 5 minutes)

**9. A quiz has two"true" (T) or "false" (F) questions and one "Pass" or "Fail" essay question. What is the sample space of possible answers to the set of three questions?**

**If an outcome is selected at random (all equally likely) from your sample space, what is the probability that your selection is "False" for the first problem and "Fail" for the essay?**

(4 points; 4 minutes)

**10. A lock uses a touch pad with the letters A B and C and the digits 1 2 3 4 5 6 7 and 8. The unlock code begins with 2 of the letters and ends with 4 of the digits. Each letter and each digit can be used each time a letter or a digit is needed. Every possible sequence of letters and digits is a different unlock code. How many different unlock codes are possible?**

(3 points; 3 minutes)

**11. You are one of twenty people that work in a company. Four of the 20 workers will be sent to New York for special training. How many sets of four workers are possible to form from the twenty workers, and what is the probability that you will not get to go to New York for the training?**

(4 points; 6 minutes)

**12. The Medicare administration wants to evaluate the standard deviation of the costs for a specific medical test this year to see if the standard deviation is acceptable under the Medicare rules. Because it is believed that variability is the same throughout each year, the Medicare staff uses the charges for the 229 times the medical test has been done so far this year and calculates the standard deviation of the 229 values. The result is that the calculated standard deviation is not acceptable.**

**Use the information in the "story" to answer the following:**

**(a) What is the population of interest to the Medicare administration?**

_____

_____

_____

**(b) What parameter was important for the Medicare administration to know?**

_____

_____

**(c) What statistic did the Medicare staff use instead?**

_____

_____

(14 points; 8 minutes)

**12. Use the data below to determine the value of each statistic. Write an expression for each statistic or describe how it is calculated in principle** (do NOT describe how to use the calculator to determine the result)**.**

| Data | | **Expression or Description** | **Value of statistic** |
|------|--------------|-------------------------------|------------------------|
| 40 | | | |
| 60 | **median** | | |
| 41 | | | |
| 56 | | | |
| 17 | | | |
| 59 | **mode** | | |
| 46 | | | |
| 24 | | | |
| 50 | **variance** | | |
| | **range** | | |
| | **standard deviation** | | |
| | **midrange** | | |
| | **mean** | | |

(6 points; 5 minutes)

13. **Last year, the dropout rate for High School students was 35%.  Using the "relative frequency" approach to probability, what is the probability that a random selection of five high school students at the beginning of this year will end up in the exact sequence shown below?**
(Show your work.)

| Student | Dropout |
|---------|---------|
| 1 | Yes |
| 2 | Yes |
| 3 | No |
| 4 | No |
| 5 | Yes |

(5 points; 6 minutes)

14. **Last year, the dropout rate for High School students was 35%.  Using the "relative frequency" approach to probability and assuming that decisions to drop out of school are independent, what are the mean and standard deviation for the number of students that will drop out this year at a high school with 3500 students?**

(5 points; 5 minutes)

15. **A gambler knows that playing "black" in a particular Roulette game will win with probability 0.475.  The gambler bets $10 on "black" for each spin of the wheel.  The gambler will get $10 if he wins and will lose $10 if he loses.  What is the expected value of the next two bets together?**

**Expected value = _____**

**Why?**

(6 points; 6 minutes)

**16. Circle the best answer for each situation.**

**An MTV program shows music videos and asks all viewers to call a free 800 number to rate each video on a scale of 1 to 10.**

| Simple Random | Systemmatic |
|---|---|
| Stratified Random | Cluster |
| Convenience | Census |

**A cable TV company rates the popularity of TV shows among its customers by constantly tracking the channel to which each of its cable connect boxes is tuned.**

| Simple Random | Systemmatic |
|---|---|
| Stratified Random | Cluster |
| Convenience | Census |

**A research company rates the popularity of TV shows by taking random samples of all Americans in each of 5 income groups within each of 5 age groups.**

| Simple Random | Systemmatic |
|---|---|
| Stratified Random | Cluster |
| Convenience | Census |

**The Department of Corrections (Prisons) selects a group of 5000 prisoners released in 2001 and studies key characteristics of their past lives to find out what types of decisions decrease the percent that return to prison at a later time.**

| retrospective | observational study |
|---|---|
| cross-sectional | |
| | experiment |
| prospective | |

**The Department of Corrections releases a group of 400 prisoners who share many key characteristics (race, education, family history, type of crime, etc.). Half of the 400 go into the army and the others do not, so the effect of military service can be studied.**

| retrospective | observational study |
|---|---|
| cross-sectional | |
| | experiment |
| prospective | |

**The Department of Corrections randomly selects 5000 prisoners in 2013 and conducts a detailed health exam on each one in order to study the present state of health in the prison population at that time.**

| retrospective | observational study |
|---|---|
| cross-sectional | |
| | experiment |
| prospective | |