MAT 167: Statistics

Test I: Chapters 1-3

Instructor: Anthony Tanbakuchi

Fall 2007

Name: _____

Computer / Seat Number: _____

No books, notes, or friends. You may use the attached equation sheet, R, and a calculator. No other materials. If you choose to use R, copy and paste your work into a word document labeling the question number it corresponds to. When you are done with the test print out the document. Be sure to save often on a memory stick just in case. Using any other program or having any other documents open on the computer will constitute cheating.

You have until the end of class to finish the exam, manage your time wisely.
If something is unclear quietly come up and ask me.
If the question is legitimate I will inform the whole class.

Express all final answers to 3 significant digits. Probabilities should be given as a decimal number unless a percent is requested. Circle final answers, ambiguous or multiple answers will not be accepted. Show steps where appropriate.

The exam consists of 13 questions for a total of 40 points on 12 pages.

This Exam is being given under the guidelines of our institution's
Code of Academic Ethics. You are expected to respect those guidelines.

Points Earned: _____ **out of 40 total points**

Exam Score: _____

1. Given the following data collected from a random sample of individual's heights in cm:

   170 185 155 168 162 164 280

   (a) (3 points) If you were asked to calculate the *"average"* height for this data, what measure of center (mean, median, mode, midrange) would you use? Give a clear reason why you choose this measure in terms of the data and what the measure represents.

   (b) (2 points) Compute the measure of center you recommended above for the data.

   (c) (2 points) Using the Range Rule of Thumb, estimate the standard deviation for the data.

   (d) (2 points) Compute the standard deviation for the data.

2. A certain type of mouse has a mean weight of 500g and a standard deviation of 30g.
   (a) (2 points) Construct an interval using the Empirical Rule which you would expect 95% of the weights to fall within.

   (b) (2 points) Find the z-score for a mouse with a weight of 445g.

   (c) (1 point) Would you consider a mouse with a weight of 445g unusual?

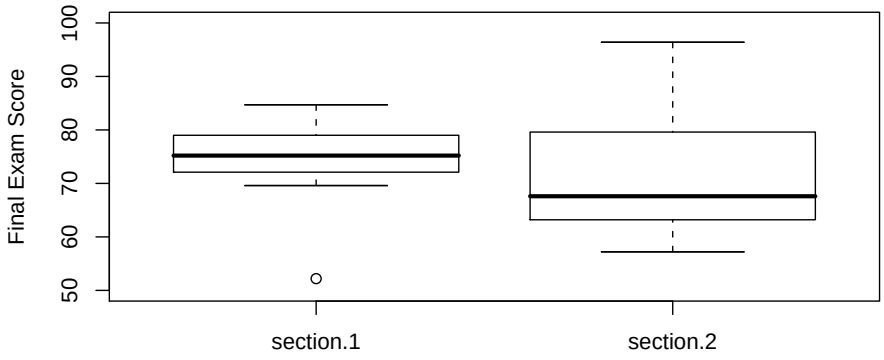
3. (2 points) "38% of adults in the United States regularly visit a doctor". This conclusion was reached by a college student after she had questioned 520 randomly selected members of her college. What is wrong with her survey?

4. (2 points) Give an example of sampling error.

5. (1 point) If the mean, median, and mode for a data set are all the same, what can you conclude about the data?

6. (2 points) Find ${}_{10}C_7$.

7. The following is a modified box plot of final exam scores for two different sections of a statistics class taking the same exam.



(a) (1 point) Which section had the highest score on the exam? Give an approximate value for the highest score.

(b) (1 point) Which section had the higher median score? Give the approximate median score for that section.

(c) (1 point) A useful measure of variation is the inter-quartile range which is calculated as $IQR = Q3 - Q1$. Find the approximate IQR for section 2.

(d) (1 point) What percent of data lies within the IQR?

8. (2 points) What is the probability of randomly selecting three people all born on the same day

of the year? (Ignore leap years).

9. If a couple has 5 children, what is the probability that:
   (a) (1 point) They have exactly 2 girls and 3 boys (in that order).

   (b) (2 points) They have at least one boy.

10. (3 points) In a blood testing procedure, blood samples from 5 people are combined into one mixture. The mixture will only test negative if all the individual samples are negative. If the probability that an individual sample tests positive is 0.20, what is the probability that the mixture will test positive?

11. (3 points) If a class consists of 12 freshmen and 8 sophomores, find the probability of randomly selecting three students in the following order: a sophomore then a sophomore then a freshman without replacement.

12. (2 points) A cruise ship has 1000 people on it. Of the 1000 people, 25 are crew members. Find the probability of randomly selecting 10 people without replacement and none of them are crew members.

13. Given the following frequency table summarizing data from a study:

age.years	frequency
0-9	5.00
10-19	8.00
20-29	12.00
30-39	2.00

(a) (1 point) Construct a relative frequency table.

(b) (1 point) What is the probability of randomly selecting someone from the study who is in the age range 10-29?

MAT67 Exam 1 Tanbakuchi

Name_____

Select the best answer from the choices provided.

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

Determine whether the given value is a statistic or a parameter.

1) A health and fitness club surveys 40 randomly selected members and found that the average weight of those questioned is 157 lb. 1) _____

A) Parameter B) Statistic

Determine which of the four levels of measurement (nominal, ordinal, interval, ratio) is most appropriate.

2) Nationalities of survey respondents. 2) _____

A) Ratio B) Nominal C) Interval D) Ordinal

Identify the number as either continuous or discrete.

3) The number of stories in a Manhattan building is 22. 3) _____

A) Discrete B) Continuous

Identify which of these types of sampling is used: random, stratified, systematic, cluster, convenience.

4) A tax auditor selects every 1000th income tax return that is received. 4) _____

A) Convenience

B) Random

C) Systematic

D) Cluster

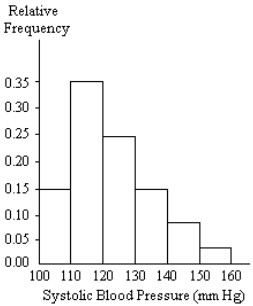
E) Stratified

Solve the problem.

5) On a test of 40 items, Ariel got 31 correct. What percent were correct? 5) _____

A) 129.032258% B) 77.5% C) 1.29032258% D) 0.775%

A nurse measured the blood pressure of each person who visited her clinic. Following is a relative–frequency histogram for the systolic blood pressure readings for those people aged between 25 and 40. Use the histogram to answer the question. The blood pressure readings were given to the nearest whole number.



6) Approximately what percentage of the people aged 25–40 had a systolic blood pressure reading between 110 and 139 inclusive? 6) _____

A) 89% B) 75% C) 39% D) 59%

Determine which score corresponds to the higher relative position.

7) Which score has the better relative position: a score of 51 on a test for which $\bar{x} = 48$ and $s = 5$, a score of 5.8 on a test for which $\bar{x} = 4.8$ and $s = 0.7$ or a score of 434.3 on a test for which $\bar{x} = 383$ and $s = 57$? 7) _____

A) A score of 434.3 B) A score of 5.8 C) A score of 51

Find the mode(s) for the given sample data.

8) 82, 39, 32, 39, 29, 82 8) _____

A) 39 B) 82 C) 82, 39 D) 50.5

Find the original data from the stem–and–leaf plot.

9) 9) _____

Stem	Leaves
52	2 6 7
53	2 4 8
54	1 7

A) 522, 526, 527, 532, 534, 538, 541, 547 B) 54, 58, 59, 55, 57, 61, 55, 61

C) 52267, 53248, 5417 D) 522, 526, 537, 532, 534, 538, 541, 557

Find the range for the given data.

10) To get the best deal on a microwave oven, Jeremy called six appliance stores and asked the cost of a specific model. The prices he was quoted are listed below: 10) _____

\$120 \$536 \$227 \$618 \$422 \$258

Compute the range.

A) \$120 B) \$31 C) \$498 D) \$536

Solve the problem.

11) If the standard deviation of a set of data is zero, what can you conclude about the set of values? 11) _____

A) The sum of the values is zero.

B) All values are equal to zero.

C) The sum of the deviations from the mean is zero.

D) All values are identical.

Use the empirical rule to solve the problem.

12) At one college, GPA's are normally distributed with a mean of 2.8 and a standard deviation of 0.6. What percentage of students at the college have a GPA between 2.2 and 3.4? 12) _____

A) 95.44% B) 99.74% C) 68.26% D) 84.13%

Answer the question, considering an event to be "unusual" if its probability is less than or equal to 0.05.

13) Assume that a study of 300 randomly selected school bus routes showed that 281 arrived on time. Is it "unusual" for a school bus to arrive late? 13) _____

A) Yes B) No

Answer the question.

14) Which of the following cannot be a probability? 14) _____

A) $\frac{1}{2}$ B) $\frac{3}{5}$ C) $\frac{5}{3}$ D) $\frac{2}{3}$

Estimate the probability of the event.

15) In a certain class of students, there are 12 boys from Wilmette, 6 girls from Kenilworth, 9 girls from Wilmette, 7 boys from Glencoe, 5 boys from Kenilworth and 3 girls from Glenoce. If the teacher calls upon a student to answer a question, what is the probability that the student will be from Kenilworth? 15) _____

A) 0.333 B) 0.143 C) 0.262 D) 0.208

Find the indicated probability.

16) A bag contains 4 red marbles, 3 blue marbles, and 5 green marbles. If a marble is randomly selected from the bag, what is the probability that it is blue? 16) _____

A) $\frac{1}{9}$ B) $\frac{1}{4}$ C) $\frac{1}{5}$ D) $\frac{1}{3}$

17) A sample of 100 wood and 100 graphite tennis rackets are taken from the warehouse. If 14 wood and 13 graphite are defective and one racket is randomly selected from the sample, find the probability that the racket is wood or defective. 17) _____

A) 0.565

B) 0.57

C) 0.135

D) There is insufficient information to answer the question.

18) The following table contains data from a study of two airlines which fly to Small Town, USA. 18) _____

	Number of flights which were on time	Number of flights which were late
Podunk Airlines	33	6
Upstate Airlines	43	5

If one of the 87 flights is randomly selected, find the probability that the flight selected arrived on time given that it was an Upstate Airlines flight.

A) $\frac{43}{87}$ B) $\frac{43}{48}$

C) $\frac{11}{76}$ D) None of the above is correct.

19) Based on meteorological records, the probability that it will snow in a certain town on January 1st is 0.269. Find the probability that in a given year it will not snow on January 1st in that town. 19) _____

A) 0.368 B) 3.717 C) 0.731 D) 1.269

Provide a written description of the complement of the given event.

20) When several textbooks are returned from editing, none of the books are found to be errorless. 20) _____

A) At least one of the textbooks is errorless.

B) None of the textbooks are errorless.

C) All of the textbooks are errorless.

Introductory Statistics
Quick Reference & R Commands

by Anthony Tanbakuchi. Version 1.1
http://www.tanbakuchi.com
ANTHONY@TANBAKUCHI·COM

Get R at: http://www.r-project.org
More R help & examples at:
http://tanbakuchi.com/Resources/R_Statistics/RBasics.html
R commands: **bold text**

1 Misc R

To make a vector / store data: **x=c(x1, x2, ...)**
Get help on function: **?functionName**
Get column of data from table: **tableName$columnName**
List all variables: **ls()**
Delete all variables: **rm(list=ls())**

$$\sqrt{x} = \texttt{sqrt(x)} \tag{1}$$
$$x^n = \texttt{x\^{}n} \tag{2}$$
$$n = \texttt{length(x)} \tag{3}$$
$$T = \texttt{table(x)} \tag{4}$$

2 Descriptive Statistics

2.1 Numerical

Let **x=c(x1, x2, x3, ...)**

$$\text{total} = \sum_{i=1}^{n} x_i = \texttt{sum(x)} \tag{5}$$
$$\min = \texttt{min(x)} \tag{6}$$
$$\max = \texttt{max(x)} \tag{7}$$
$$\text{six number summary} : \texttt{summary(x)} \tag{8}$$
$$\mu = \frac{\sum x_i}{N} = \texttt{mean(x)} \tag{9}$$
$$\bar{x} = \frac{\sum x_i}{n} = \texttt{mean(x)} \tag{10}$$
$$\tilde{x} = P_{50} = \texttt{median(x)} \tag{11}$$
$$\sigma = \sqrt{\frac{\sum (x_i-\mu)^2}{N}} \tag{12}$$
$$s = \sqrt{\frac{\sum (x_i-\bar{x})^2}{n-1}} = \texttt{sd(x)} \tag{13}$$
$$CV = \frac{\sigma}{\mu} = \frac{s}{\bar{x}} \tag{14}$$

2.2 Relative Standing

$$z = \frac{x-\mu}{\sigma} = \frac{x-\bar{x}}{s} \tag{15}$$

Percentiles

$$P_k = x_i, \qquad \text{(sorted } x\text{)}$$
$$k = \frac{i-0.5}{n}\cdot 100\% \tag{16}$$

To find x_i given P_k, i is:

$$1 : L = \frac{k}{100\%}\cdot n \tag{17}$$

2 : if L is an integer: $i = L + 0.5$; otherwise i=L and round up.

2.3 Visual

All plots have optional arguments:
main="" sets title
xlab="", ylab="" sets x/y-axis label
type="p" for **p**oint plot
type="l" for **l**ine plot
type="b" for **b**oth points and lines
Ex: `plot(x, y, type="b", main="My Plot")`
Histogram: **hist(x)**
Stem & leaf: **stem(x)**
Box plot: **boxplot(x)**
Barplot: **plot(T)** (where `T=table(x)`)
Scatter plot: **plot(x,y)** (where `x, y` are ordered vectors)
Time series plot: **plot(t,y)** (where `t, y` are ordered vectors)
Graph function: **curve(expr, xmin, xmax)** plot `expr` involving x

2.4 Assessing Normality

Q-Q plot: **qqnorm(x); qqline(x)**

3 Probability

Number of successes x with n possible outcomes.
(Don't double count!)

$$P(A) = \frac{x_A}{n} \tag{18}$$
$$P(\bar{A}) = 1 - P(A) \tag{19}$$
$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \tag{20}$$
$$P(A \text{ or } B) = P(A) + P(B) \qquad \text{if } A, B \text{ mutually exclusive} \tag{21}$$
$$P(A \text{ and } B) = P(A)\cdot P(B|A) \tag{22}$$
$$P(A \text{ and } B) = P(A)\cdot P(B) \qquad \text{if } A, B \text{ independent} \tag{23}$$
$$n! = n(n-1)(n-2)\cdots 2\cdot 1 = \texttt{factorial(n)} \tag{24}$$
$${}_nP_k = \frac{n!}{(n-k)!} \qquad \text{Perm. no elements alike} \tag{25}$$
$$= \frac{n!}{n_1!n_2!\cdots n_k!} \qquad \text{Perm. } n_1 \text{ alike, } \ldots \tag{26}$$
$${}_nC_k = \frac{n!}{(n-k)!k!} = \texttt{choose(n,k)} \tag{27}$$

4 Random Variables

4.1 Discrete Distributions

$$P(x_i) : \text{ probability distribution} \tag{28}$$
$$E = \mu = \sum x_i \cdot P(x_i) \tag{29}$$
$$\sigma = \sqrt{\sum (x_i-\mu)^2 \cdot P(x_i)} \tag{30}$$

4.2 Continuous Distributions

CDF $F(x)$ gives area to the left of x, $F^{-1}(p)$ expects p is area to the left.

$$f(x) : \text{ probability density} \tag{31}$$
$$E = \mu = \int_{-\infty}^{\infty} x\cdot f(x)\,dx \tag{32}$$
$$\sigma = \sqrt{\int_{-\infty}^{\infty} (x-\mu)^2\cdot f(x)\,dx} \tag{33}$$
$$F(x) : \text{ cumulative prob. density (CDF)} \tag{34}$$
$$F^{-1}(x) : \text{ inv. cumulative prob. density} \tag{35}$$
$$F(x') = \int_{-\infty}^{x'} f(x)\,dx \tag{36}$$
$$p = P(x < x') = F(x') \tag{37}$$
$$x' = F^{-1}(p) \tag{38}$$
$$p = P(x > a) = 1 - F(a) \tag{39}$$
$$p = P(a < x < b) = F(b) - F(a) \tag{40}$$

4.3 Sampling Distributions

$$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{41}$$
$$\mu_{\hat{p}} = p \qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \tag{42}$$

4.4 Binomial Distribution

$$\mu = n\cdot p \tag{43}$$
$$\sigma = \sqrt{n\cdot p\cdot q} \tag{44}$$
$$P(x) = {}_nC_x p^x q^{(n-x)} = \texttt{dbinom(x, n, p)} \tag{45}$$

4.5 Poisson Distribution

$$P(x) = \frac{\mu^x\cdot e^{-\mu}}{x!} = \texttt{dpois(x, } \mu\texttt{)} \tag{46}$$

4.6 Normal Distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{47}$$
$$p = P(z < z') = F(z') = \texttt{pnorm(z')} \tag{48}$$
$$z' = F^{-1}(p) = \texttt{qnorm(p)} \tag{49}$$
$$p = P(x < x') = F(x') = \texttt{pnorm(x', mean=}\mu\texttt{, sd=}\sigma\texttt{)} \tag{50}$$
$$x' = F^{-1}(p) = \texttt{qnorm(p, mean=}\mu\texttt{, sd=}\sigma\texttt{)} \tag{51}$$

4.7 t-DISTRIBUTION

$$p = P(t < t') = F(t') = \texttt{pt(t', df)} \tag{52}$$
$$t' = F^{-1}(p) = \texttt{qt(p, df)} \tag{53}$$

4.8 χ^2-DISTRIBUTION

$$p = P(\chi^2 < \chi^{2'}) = F(\chi^{2'}) = \texttt{pchisq}(X^{2'}\texttt{, df}) \tag{54}$$
$$\chi^{2'} = F^{-1}(p) = \texttt{qchisq(p, df)} \tag{55}$$

4.9 F-DISTRIBUTION

$$p = P(F < F') = F(F') = \texttt{pf(F', df1, df2)} \tag{56}$$
$$F' = F^{-1}(p) = \texttt{qf(p, df1, df2)} \tag{57}$$

5 Estimation

5.1 CONFIDENCE INTERVALS

$$\text{proportion: } \hat{p} \pm E, \qquad E = z_{\alpha/2} \cdot \sigma_{\hat{p}} \tag{58}$$
$$\text{mean } (\sigma \text{ known}): \bar{x} \pm E, \qquad E = z_{\alpha/2} \cdot \sigma_{\bar{x}} \tag{59}$$
$$\text{mean } (\sigma \text{ unknown, use } s): \bar{x} \pm E, \qquad E = t_{\alpha/2} \cdot \sigma_{\bar{x}}, \qquad df = n-1 \tag{60}$$
$$\text{variance: } \frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L}, \qquad df = n-1 \tag{61}$$

5.2 CRITICAL VALUES

$$z_{\alpha/2} = P(z > \alpha) = \texttt{qnorm(1-alpha/2)} \tag{62}$$
$$t_{\alpha/2} = P(t > \alpha) = \texttt{qt(1-alpha/2, df)} \tag{63}$$
$$\chi^2_L = P(\chi^2 < \alpha) = \texttt{qchisq(alpha/2, df)} \tag{64}$$
$$\chi^2_R = P(\chi^2 > \alpha) = \texttt{qchisq(1-alpha/2, df)} \tag{65}$$

5.3 REQUIRED SAMPLE SIZE

$$\text{proportion: } n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2 \quad (\hat{p} = \hat{q} = 0.5 \text{ if unknown}) \tag{66}$$
$$\text{mean: } n = \left(\frac{z_{\alpha/2} \cdot \hat{\sigma}}{E}\right)^2 \tag{67}$$

6 Hypothesis Tests

`alternative` can be:
`"two.sided"`, **`"less"`**, **`"greater"`**

6.1 1-SAMPLE PROPORTION

$H_0 : p = p_0$
`prop.test(x, n, p=`p_0**`, alternative="two.sided")`**

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \tag{68}$$

6.2 1-SAMPLE MEAN (σ KNOWN)

$H_0 : \mu = \mu_0$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{69}$$

6.3 1-SAMPLE MEAN (σ UNKNOWN)

$H_0 : \mu = \mu_0$
`t.test(x, mu=`μ_0**`, alternative="two.sided")`**
Where **`x`** is a vector of sample data.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}, \qquad df = n-1 \tag{70}$$