Howard University
Department of Mathematics

Comprehensive Final (Fall 2005) *Tuesday, December 13th, 2005.*

Introduction to Statistics

Name:____
(*Please* ***PRINT*** *your name*)

Signature:______________________________

I.D. # __________________________

Show all work otherwise NO POINTS WILL BE AWARDED. All work must be neat and legible OTHERWISE POINTS WILL BE DEDUCTED. Partial credits will be given for work which demonstrates a working knowledge of the concepts.

Answer all questions. Each question is worth 25 points.
Begin each question on a new page.

Do not write in the columns below.

Question 1			***Question 7***	
Question 2			***Question 8***	
Question 3			***Question 9***	
Question 4				
Question 5			***Comprehensive Final Exam Total***	
Question 6			***Comprehensive Final Exam Grade***	
			Course grade	

Q1)

(a) Using the data set: 25, 32, 35, 18, 42, 39, 71, 29, 40, 19, 54, and 68:

(i) find the value that corresponds to the 60^{th} percentile.

(ii) find the percentile rank for the value 39.

(b) A list of the 50 most powerful women in America was published. The average age of these women was 49.74 years old with a standard deviation of 6.12 years. Using Chebyshev's Theorem, what is the range of ages in which at least 78% of the data lie?

Q2) The distribution on a Mathematics Test was obtained and the results are shown on the table below. Complete the table and find the mean, median, modal class, variance, and standard deviation for the data:

Class Limits	Frequency f	Class Boundaries	Midpoints X_m	$f.X_m$	Cummulative Frequency	$f.X_m^2$
46 - 56	4					
57 - 67	8					
68 - 78	16					
79 - 89	9					
90 - 100	6					

Q3)

(a) The table below shows the average grades and degrees for graduates.

	GRADES			
DEGREE	**C**	**B**	**A**	**TOTAL**
B.S	10	18	15	
B.A	16	12	11	
TOTAL				

Complete the table. If a graduate is selected at random, find the probability that:

(i) The graduate has a B.A. degree given that he or she has a B average.

(ii) Given that the graduate has a B.S. degree, he or she has an A average.

(b) Find the 95% confidence interval for the variance and standard deviation of the nicotine content of cigarettes manufactured if a sample of 20 cigarettes has a standard deviation of 1.6 milligrams. (*Use the Chi-Square Distribution)*.

Q4) A sales person from an auto dealership knows from past experience that, on the average, she will make a sale to 22% of her customers. What is the probability that, in 9 random selected presentations, she makes a sale to:

(i) Exactly 6 customers.

(ii) At most one customer.

(iii) At least one customer.

(Hint: Use the Binomial Distribution)

Q5)

(a) Of the members of a bowling league, 10% are widowed. If 200 bowling league members are selected at random, find the probability that 16 or more will be widowed. [*Use the normal approximation to the binomial distribution by first showing the test that this approximation can be used*]

(b) The average weight of young adult males is 160 pounds. The standard deviation is 10 pounds. If a sample of 28 people is selected from the population of 300, find the probability that the mean for the sample will be more than 156 pounds. *[Determine whether the correction factor is to be used]*.

Q6)

(a) A publisher wants to publish home improvement books. After a survey of the market, the publisher finds that the average price for this type of book is \$35. with a standard deviation of \$0.80. The publisher wants to target the middle 30% of the market. What should be minimum and maximum prices for the book assuming the variable is normally distributed?

(b) In an automobile service shop, the supervisor timed 8 employees and found that the average time it took them to change a tire was 24 minutes. The standard deviation of the sample was 3 minutes. Using the t-distribution, find the 99% confidence interval of the true mean.

Q7) Research at Florida State University was carried out to determine the effects of alcohol on the reactions of people to threat of electric shock. The "startle" response time for each subject was measured in milliseconds (ms) from which the mean was 37.9 ms and the standard deviation was 12.4 ms. Assuming the variable (blood alcohol level) is normally distributed:

i. Find the probability that the response time is between 40 ms and 50 ms.

ii. Find the probability that the response time is less than 35ms.

iii. Find the interval centered about the mean so that the probability that the response time falls in the interval is 0.94.

iv. What is the cut off response time for the highest 10% of the subjects?

Q8)

(a) In a survey of 32 adults, it was found that the mean age of a person's primary vehicle is 5.6 years. Assuming the standard deviation of the population is 0.8 year, find the 92% confidence interval of the population mean.

(b) A restaurant owner wishes to find the 90% confidence interval of the true mean cost of red wine. How large should the sample be if the owner wishes to be accurate within $0.06? A previous study showed that the standard deviation of the price was $0.10.

Q9)

(a) In a poll of 1000 likely voters, 550 say that the United States spends too little on fighting hunger at home. Find a 95% confidence interval for the true proportion of voters who feel this way. (*Hint: Use the confidence intervals for proportions*).

(b) A recent study indicates that 18% of the 100 women over age 62 were widows.

(i)How large a sample must be taken to be 90% confident that the estimate is within 0.05 of the true proportion of women over 62 who are widows?

(ii) If no estimate of the sample proportion is available, how large should the sample be?

(Hint: Use Sample Size for Proportions)

Table E The Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
[illegible]	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141

Table F The *t* Distribution

	Confidence intervals	50%	80%	90%	95%	98%	99%
	One tail, α	0.25	0.10	0.05	0.025	0.01	0.005
d.f.	Two tails, α	0.50	0.20	0.10	0.05	0.02	0.01
1		1.000	3.078	6.314	12.706	31.821	63.657
2		.816	1.886	2.920	4.303	6.965	9.925
3		.765	[illegible]				

Table G The Chi-Square Distribution

Degrees of freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
[illegible]	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
[illegible]						[illegible]	15.507	17.535	20.090	21.955

The three tables above were provided with the exam, as well as the following formulas:

Important Formulas

Range = highest value − lowest value

Midpoint (X_m) = $\dfrac{\text{lower boundary} + \text{higher boundary}}{2}$

Class width = upper boundary − lower boundary

Degrees for each section of the pie graph:

$$\text{degrees} = \frac{f}{n} \cdot 360°$$

f = frequency n = total number of subjects

Percent for each section of the pie graph: $\% = \dfrac{f}{n} \cdot 100$

Mean for individual data: $\bar{X} = \dfrac{\sum X}{n}$ or $\mu = \dfrac{\sum X}{N}$

Mean for grouped data: $\bar{X} = \dfrac{\sum f \cdot X_m}{n}$

Median for grouped data: $\text{MD} = \dfrac{(n/2) - \text{cf}}{f}(w) + L_m$

cf = cumulative frequency

w = upper boundary − lower boundary

L_m = lower boundary of the median class

Weighted mean: $\bar{X} = \dfrac{\sum w \cdot X}{\sum w}$

Midrange: $\text{MR} = \dfrac{\text{lower value} + \text{highest value}}{2}$

Variance for a population: $\sigma^2 = \dfrac{\sum (X - \mu)^2}{N}$

Variance for a sample: $s^2 = \dfrac{\sum X^2 - [(\sum X)^2/n]}{n - 1}$

Variance for grouped data:

$$s^2 = \frac{\sum f \cdot X_m^2 - [(\sum f \cdot X_m)^2/n]}{n - 1}$$

Standard deviation for a population: $\sigma = \sqrt{\dfrac{\sum (X - \mu)^2}{N}}$

Standard deviation for a sample:

$$s = \sqrt{\frac{\sum X^2 - [(\sum X)^2/n]}{n - 1}}$$

Standard deviation for grouped data:

$$s = \sqrt{\frac{\sum f \cdot X_m^2 - [(\sum f \cdot X_m)^2/n]}{n - 1}}$$

Coefficient of variation: CVat $= \dfrac{s}{\bar{X}} \cdot 100\%$ or $\dfrac{\sigma}{\mu} \cdot 100\%$

z score: $z = \dfrac{X - \mu}{\sigma}$ or $z = \dfrac{X - \bar{X}}{s}$

Percentile rank of a value x:

$$\text{Percentile} = \frac{\text{number of values below} + 0.5}{\text{total number of values}} \cdot 100\%$$

Value corresponding to a given percentile: $c = \dfrac{n \cdot p}{100}$

Interquartile range $= Q_3 - Q_1$

Multiplication rule 1: Total number of outcomes of a sequence when each event has k possibilities $= k^n$

Multiplication rule 2: Total number of outcomes of a sequence when each event has a different number of possibilities $k_1 \cdot k_2 \cdot k_3 \cdot \ldots \cdot k_n$

Permutation rule 1: Number of permutations of n objects is $n!$

Permutation rule 2: Number of permutations of n objects taken r at a time is ${}_nP_r = \dfrac{n!}{(n - r)!}$

Permutation rule 3: Number of permutations of n objects in which k_1 are alike, k_2 are alike, etc., is $\dfrac{n!}{k_1! \cdot k_2! \cdot \ldots \cdot k_m!}$

Combination rule: Number of combinations of r objects selected from n objects is

$${}_nC_r = \frac{n!}{(n - r)!r!}$$

Classical probability:

$$P(E) = \frac{\text{number of ways } E \text{ can occur}}{\text{total number of outcomes in the sample space}} = \frac{n(E)}{n(S)}$$

Empirical probability:

$$P(E) = \frac{\text{frequency of class}}{\text{total frequencies in the distribution}} = \frac{f}{n}$$

Addition rule 1 (mutually exclusive events):
$P(A \text{ or } B) = P(A) + P(B)$

Addition rule 2 (events not mutually exclusive):
$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Multiplication rule 1 (independent events):
$P(A \text{ and } B) = P(A) \cdot P(B)$

Multiplication rule 2 (dependent events):
$P(A \text{ and } B) = P(A) \cdot P(B \mid A)$

Conditional probability: $P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$

Complementary events: $P(\overline{E}) = 1 - P(E)$
$P(E) = 1 - P(\overline{E})$

Bayes's theorem: $P(A_1 \mid B_1) =$

$$\frac{P(A_1) \cdot P(B_1 \mid A_1)}{P(A_1) \cdot P(B_1 \mid A_1) + P(A_2) \cdot P(B_1 \mid A_2) + \cdots + P(A_n) \cdot P(B_1 \mid A_n)}$$

Mean for a probability distribution: $\mu = \sum X \cdot P(X)$

Variance for a probability distribution:
$\sigma^2 = \sum [X^2 \cdot P(X)] - \mu^2$

Expectation: $E(X) = \sum X \cdot P(X)$

Binomial probability: $P(X) = \frac{n!}{(n-X)!X!} \cdot p^X \cdot q^{n-X}$

Mean for binomial distribution: $\mu = n \cdot p$

Variance and standard deviation for the binomial distribution: $\sigma^2 = n \cdot p \cdot q \qquad \sigma = \sqrt{n \cdot p \cdot q}$

Multinomial probability:

$$P(M) = \frac{n!}{X_1!X_2!X_3! \ldots X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdot p_3^{X_3} \cdots p_k^{X_k}$$

Poisson probability: $P(X; \lambda) = \frac{e^{-\lambda}\lambda^X}{X!}$ where $X = 0, 1, 2, \ldots$

Hypergeometric probability: $P(A) = \frac{{}_aC_X \cdot {}_bC_{n-X}}{{}_{(a+b)}C_n}$

Mean of sample means: $\mu_{\bar{X}} = \mu$

Standard error of the mean: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Central limit theorem formula: $z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$

Central limit theorem formula when $n > 0.05N$:

$$z = \frac{\overline{X} - \mu}{(\sigma/\sqrt{n})\sqrt{(N-n)/(N-1)}}$$

Mean for binomial variable:
$\mu = n \cdot p$

Standard deviation for a binomial variable:
$\sigma = \sqrt{n \cdot p \cdot q}$

Finding a specific data value: $X = z \cdot \sigma + \mu$

z confidence interval for means:

$$\overline{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \overline{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

t confidence interval for means:

$$\overline{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Sample size for means: $n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$

where E is the maximum error of estimate

Confidence interval for a proportion:

$$\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + (z_{\alpha/2})\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Sample size for a proportion: $n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$

where $\hat{p} = \frac{X}{n}$ and $\hat{q} = 1 - \hat{p}$

Confidence interval for variance:

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

Confidence interval for standard deviation:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}$$