Statistics 151 Final – Practice Final

Instructor: Mike Kowalski

Instructions:
1. *This is a closed book exam.*
2. *You may use the STAT 151 formula sheets and tables provided and a calculator only.*
3. *You have 120 minutes to complete the exam.*
4. *The exam is out of a total of 50 marks (20 multiple choice and six long answer problems worth a total of 30).*
5. *Make sure you mark all your multiple-choice responses on the answer page (page 2), as these will be taken as your answers.*
6. *Show all your work for the long answer section to receive full credit.*
7. *Read the instructions for the long answer problems.*
8. *Use the backs of the pages at the end for scrap work.*
9. *Make sure your name and signature are on the front and that your ID number is on page two.*
10. *Make sure to hand back ALL pages including the formula sheets and the tables. Missing pages will result in a mark of zero.*

Name: ______________________________________

Signature: ____________________________

Part		**Possible Marks**	
Multiple Choice	20 questions	20	
Written #1	2 parts	6	
Written #2	4 parts	8	
Written #3	1 part	4	
Written #4	1 part	4	
Written #5	1 part	4	
Written #6	1 part	4	
Total		**50**	

ID:________________________

PART 1 (MULTIPLE CHOICE QUESTIONS)

Response Section for Multiple Choice

1.	a	b	c	d	11.	a	b	c	d
2.	a	b	c	d	12.	a	b	c	d
3.	a	b	c	d	13.	a	b	c	d
4.	a	b	c	d	14.	a	b	c	d
5.	a	b	c	d	15.	a	b	c	d
6.	a	b	c	d	16.	a	b	c	d
7.	a	b	c	d	17.	a	b	c	d
8.	a	b	c	d	18.	a	b	c	d
9.	a	b	c	d	19.	a	b	c	d
10.	a	b	c	d	20.	a	b	c	d

Use the following to answer question 1:

You measure the weights of a random sample of twenty-five male runners. The sample mean is x-bar = 60 kilograms (kg). Suppose that the mean weights of male runners follow a normal distribution with unknown mean μ and standard deviation $\sigma = 5$ kg. You compute a 95% confidence interval for μ.

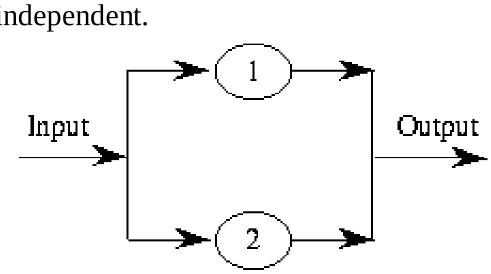
1. Suppose I had measured the weights of a random sample of 100 runners rather than 25 runners. Which of the following statements is true?
   A) The margin of error for our 95% confidence interval would increase
   B) The margin of error for our 95% confidence interval would decrease
   C) The margin of error for our 95% confidence interval would stay the same, since the level of confidence has not changed
   D) σ would decrease

2. Event *A* occurs with probability 0.2. Event *B* occurs with probability 0.8. If *A* and *B* are disjoint, then
   A) *P*(*A* and *B*) = 0.16.
   B) *P*(*A* or *B*) = 1.0.
   C) *P*(*A* and *B*) = 1.0.
   D) *P*(*A* or *B*) = 0.16.

3. In a large population of adults, the mean IQ is 112 with a standard deviation of 20. Suppose 200 adults are randomly selected for a market research campaign. The distribution of the sample mean IQ is
   A) exactly normal, mean 112, standard deviation 20.
   B) approximately normal, mean 112, standard deviation 0.1.
   C) approximately normal, mean 112 , standard deviation 1.414.
   D) approximately normal, mean 112, standard deviation 20.

Use the following to answer questions 4:

A system has two components that operate in parallel, as shown in the diagram below. Because the components operate in parallel, at least one of the components must function properly if the system is to function properly. The probabilities of *failures* for the components 1 and 2 during one period of operation are .20 and .03, respectively. Let *F* denote the event that the component 1 *fails* during one period of operation and *G* denote the event that component 2 *fails* during one period of operation. The component failures are independent.



4. The event corresponding to the above system failing during one period of operation is
   A) *F* and *G*. B) *F* or *G*. C) not *F* or not *G*. D) not *F* and not *G*.

5. Suppose that *A* and *B* are two independent events with *P*(*A*) = .2 and *P*(*B*) = .4. *P*(*A* and not *B*) is
   A) 0.08. B) 0.12. C) 0.52. D) 0.60.

6. Event A has probability 0.4. Event B has probability 0.5. If A and B are disjoint, then the probability that both events occur is
   A) 0.0. B) 0.1. C) 0.2. D) 0.9.

7. The heights (in inches) of males in the United States are believed to be normally distributed with mean μ. The average height of a random sample of twenty-five American adult males is found to be x-bar = 69.72 inches and the standard deviation of the twenty-five heights is found to be *s* = 4.15. The standard error of x-bar is
   A) 0.17. B) 0.69. C) 0.83. D) 2.04.

8. Suppose that *A* and *B* are two independent events with *P*(*A*) = 0.2 and *P*(*B*) = 0.4. *P*(*A* or *B*) is
   A) 0.08. B) 0.12. C) 0.52. D) 0.60.

Use the following to answer questions 9-10:

I wish to determine the correlation between the height (in inches) and weight (in pounds) of 21-year-old males. To do this, I measure the height and weight of two 21-year-old men. The measured values are

	Male #1	Male #2
Height	70	75
Weight	160	200

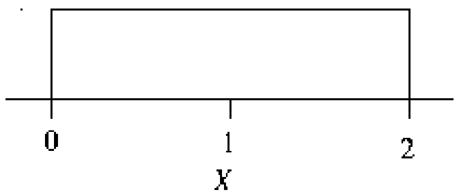
9. The correlation *r* computed from the measurements of these males is
   A) 1.0.
   B) positive and between 0.25 and 0.75.
   C) near 0, but could be either positive or negative.
   D) exactly 0.

10. The correlation *r* computed from the measurements on these males would have units
    A) inches. B) pounds. C) inches-pounds. D) no units.

11. We know the random variable x-bar has approximately a normal distribution because of
    A) the law of large numbers.
    B) the central limit theorem.
    C) the law of proportions.
    D) the fact that probability is the long run proportion of times an event occurs.

Use the following to answer question 12:

The probability density of a random variable *X* is given in the figure below.



12. The probability that $X = 1.5$ is
    A) 0. B) 1/4. C) 1/3. D) 1/2.

Use the following to answer questions 13-14:

Do children's fear levels change over time, and, if so, in what ways? Little research has been done on the prevalence and persistence of fears in children. In 1989 two researchers surveyed a group of ninety-four third and fourth grade children asking them to rate their level of fearfulness about a variety of situations. Two years later, the children again completed the same survey. The researchers computed the mean fear rating for each child in both years and were interested in the relation between these ratings. They then assumed that the true regression line was

$$\mu_{1991\ \text{Mean Rating}} = \alpha + \beta(1989\ \text{Mean Rating})$$

and that the assumptions for regression inference were satisfied. This model was fit to the data using least squares. The following results were obtained from statistical software.

Parameter estimates:

Parameter	**Estimate**	**Std. Err.**	**DF**	**T-Stat**	**P-Value**
Intercept	0.877917	0.1184			
Slope	0.397911	0.0676			

S = 0.2374 R-Sq = 0.274

13. The correlation between the 1989 and 1991 mean fear ratings is
    A) 0.274. B) 0.40. C) 0.523. D) 0.632.

14. Suppose the researchers test the following hypotheses about the slope β:

    $$H_0: \beta = 0,\ H_a: \beta \neq 0$$

    The value of the *t* statistic for this test is
    A) 0.27. B) 0.40. C) 5.89. D) 7.41.

Use the following to answer questions 15 - 17:

At what age do babies learn to crawl? Does it take longer to learn in the winter when babies are often bundled in clothes that restrict their movement? Data were collected from parents who brought their babies into the University of Denver Infant Study Center to participate in one of a number of experiments between 1988 and 1991. Parents reported the birth month and the age at which their child was first able to creep or crawl a distance of four feet within one minute. The resulting data were grouped by month of birth. The data are for January, May, and September.

Birth month	Average crawling age	SD	*n*
January	29.84	7.08	32
May	28.58	8.07	27
September	33.83	6.93	38

Crawling age is given in weeks. Assume the data are three independent SRSs, one from each of the three populations (babies born in a particular month) and that the populations of crawling ages have normal distributions.

A partial ANOVA table is given below.

Source of Variation	Sum of Squares	d.f.	Mean Square	F-Statistic	p-value
Treatment (Between)	505.26				
Error (Within)	5024.30		53.45		
Total	5529.56				

15. The null hypothesis for the ANOVA *F* test is that the population mean crawling ages are equal for all three birth months. The alternative hypothesis is
    A) that the population mean crawling age is larger for January than the other two months.
    B) that the population mean crawling age is larger for May than the other two months.
    C) that the population mean crawling age is larger for September than the other two months.
    D) None of the above.

16. The test-statistic is
    A) 4.73
    B) 3.15
    C) 252.63
    D) 0.101

17. The denominator degrees of freedom for the ANOVA *F* test are
    A) 2. B) 3. C) 94. D) 97.

18. In their advertisements, a new diet program would like to claim that their methods result in a mean weight loss of more than ten pounds in two weeks. In order to determine if this is a valid claim, they hire an independent testing agency that then selects twenty-five people to be placed on this diet. The agency should be testing the null hypothesis H_0: $\mu = 10$ and the alternative hypothesis
A) H_a: $\mu < 10$.
B) H_a: $\mu > 10$.
C) H_a: $\mu \neq 10$.
D)
$$H_a: \mu \neq 10 \pm \frac{\sigma}{\sqrt{n}}$$

Use the following to answer questions 19:

An SRS of size 100 is taken from a population having proportion 0.8 of successes. An independent SRS of size 400 is taken from a population having proportion 0.5 of successes.

19. The sampling distribution for the difference in the sample proportions $p_1 - p_2$ has standard deviation equal to
A) 1.3. B) 0.40. C) 0.047. D) 0.002.

20. In a test to compare the speeds of two types of computers, ten benchmark programs written in C+ were run on both types of computers. The CPU time in minutes was measured and recorded.

Program	1	2	3	4	5	6	7	8	9	10
Computer 1	14.2	7.1	8.4	10.8	10.5	8.4	10.2	22.9	10.1	11.3
Computer 2	15.0	9.0	9.4	9.9	13.8	7.4	10.9	20.1	14.1	13.5

	Computer 1	Computer 2
Mean	11.39	12.31
Variance	20.09877778	14.00988889

Select the appropriate test:
A) One sample t-test
B) Paired t-test
C) Two independent sample t-test (assuming equal variances)
D) Two independent sample t-test (not assuming equal variances)

Long Answer:
When asked to carry out an appropriate test:

1) set up the hypotheses
2) calculate the test statistic (state all important components of the test statistic)
3) state the distribution of the test statistic (i.e. t_9 or $F_{3, 46}$)
4) approximate the p-value
5) comment on the model assumptions (if needed)
6) and state your conclusion in plain English.

Question 1 (6 marks):

The Gallup Poll once asked a random sample of 1540 adults, "Do you jog?" Suppose that in fact it is known that 15% of all Americans jog.

a) (3 marks) What is the approximate sampling distribution (and the mean and standard deviation) of the sample proportion of joggers for samples of size 1540?

b) (3 marks) What sample size would be required to reduce the standard deviation of the sample proportion to one-third the original value (based on the sample of 1540)?

Question 2 (8 marks):

A study was conducted to determine whether there is a linear relationship between the breaking strength of wooden beams and the specific gravity of wood. The following summary statistics were observed from a sample of 5 observations.

Parameter estimates:

Parameter	Estimate	Std. Err.	DF	T-Stat	P-Value
Intercept	5.974	2.005			
Slope	11.701	3.759			

S = 0.47 $r^2 = 0.76$

a) (1 marks) Write down the estimated regression equation in terms of the variables.

b) (2 marks) State and interpret the correlation coefficient and the coefficient of determination.

c) (3 marks) Calculate a 95% confidence interval for the change in average breaking strength as gravity is increased by 1 unit?

d) (2 marks) What is the sum of the squared residuals for the regression model?

Question 3 (4 marks):
An SRS of 64 postal employees found that the average time these employees had worked for the postal service was 7 years with a variance of 4 years. Is there any evidence that the mean time spent in the postal service has declined from the value of 7.4, that it was 20 years ago? Assume that the observations are approximately normal. Carry out an appropriate test to support your answer.

Question 4: (4 marks)
Have attitudes toward investing in the stock market changed in recent years as the growth of stocks has slowed? In 1995, a random sample of 100 adults that had investments in the stock market found that only 20 said they were investing for the long haul rather than to become rich or make quick profits. A random sample of 100 adults that had investments in the stock market in 2002 found that 36 were investing for the long haul rather than to become rich or make quick profits. Let p_1 and p_2 be the population proportion of all adults with investments in the stock market in 1995 and in 2002, respectively, that were investing for the long haul.

Calculate a 95% confidence interval for the difference in the proportion of individuals that were investing for the ling haul in 2002 as opposed to 1995.

Question 5: (4 marks)
An article reported the results of a planned experiment contrasting four different teaching methods. ***n*** students were randomly allocated, an equal number to each method. After completing the experimental course, a 1-hour examination was administered. The table below summarizes the scores on a 10-minute retention test that was given 6 weeks later. Assume all populations are in fact normal with some common variance.

Group	Teaching Method	Sample Mean	Sample S.D.
1	Lectures only	30.25	1.29
2	Lectures and Assignments	31.82	1.43
3	Lectures and Computer Labs	37.61	1.74
4	Lectures, Assignments and Computer Labs	37.67	1.64

Source of Variation	Sum of Squares	d.f.	Mean Square	F-Statistic	p-value
Treatment (Between)	448.473				
Error (Within)					
Total	533.427	39			

Does there appear to any significant evidence of a difference among the four teaching methods? Carry out an appropriate test.

Question 6 (4 marks)

Two machines are used to fill plastic bottles with dishwashing detergent. Two random samples are taken from and the following results are obtained:

Machine	***n***	**Sample Mean**	**Sample Variance: s^2**
1	30	30.87	0.113
2	30	31.28	0.125

Is there a real difference between their average fills? Carry out an appropriate test to answer this question.