Statistics 151 – Practice Final Exam 4

Instructor: Mike Kowalski

Instructions:

1. *Read all the instructions carefully.*
2. *This is a closed book exam.*
3. *You may use the STAT 151 formula sheets and tables provided and a non-programmable calculator only.*
4. *You are not to use any scrap paper of your own. There is plenty of blank space for scrap work.*
5. *You have 180 minutes to complete the exam.*
6. *The exam consists of 30 multiple choice questions worth 1 mark each, and 22 written answer marks. Your final percentage will be based out of a total of 50 marks.*
7. *Make sure your name and signature are on this page and that your ID number is on the back of this page.*

Make sure you mark all your multiple choice answers on the back of this page in the allotted time. These will be your official answers. You will not be given time to fill it out after time expires!

You must return ALL pages at the end of the exam. Missing pages will result in a score of 0 and further disciplinary actions may be taken.

Name: _____ Signature: _____

Component	Worth	Mark
Multiple Choice	30	
Written Answer		
Question 1	10	
Question 2	4	
Question 3	5	
Question 4	3	
Total	50	

ID:_____

Make sure you mark all your multiple choice answers on this page in the allotted time. These will be your official answers. You will not be given time to fill it out after time expires!

Multiple Choice Answer Sheet:

Question	Answer
1	D
2	A
3	D
4	A
5	B

Question	Answer
16	C
17	C
18	C
19	D
20	A

6	A
7	B
8	B
9	B
10	B

21	D
22	B
23	C
24	B
25	B

11	D
12	D
13	D
14	D
15	B

26	D
27	B
28	B
29	B
30	D

1. The nicotine content in cigarettes of a certain brand is normally distributed with mean (in milligrams) μ. The brand advertises that the mean nicotine content of their cigarettes is 1.5, but you believe that the mean nicotine content is actually higher than advertised. To explore this, you test the hypotheses H_0: $\mu = 1.5$, H_a: $\mu > 1.5$ and you obtain a P-value of 0.062. Which of the following is true?

A) At the $\alpha = 0.05$ significance level, you have proven that H_0 is true.
B) You have failed to obtain any evidence for H_a.
C) There is some evidence against H_a, but it is not significant at $\alpha = 0.05$.
D) There is some evidence against H_0, but it is not significant at $\alpha = 0.05$.

2. A consumer group surveyed the prices for a certain item in five different stores, and reported the average price as \$15. We visited four of the five stores, and found the prices to be \$10, \$15, \$15, and \$25. Assuming that the consumer group is correct, what is the price of the item at the store that we did not visit?

A) \$10 B) \$15 C) \$20 D) \$25

3. The stronger the effect of a factor on the response variable in a one-way ANOVA,

A) The smaller the between-groups variation and the larger the within-groups variation.
B) The smaller the between-groups variation and the smaller the within-groups variation.
C) The larger the between-groups variation and the larger the within-groups variation.
D) The larger the between-groups variation and the smaller the within-groups variation.

4. An automobile insurer has found that repair claims have a mean of \$920 and a standard deviation of \$370. Suppose that the next 9 claims can be regarded as a random sample from the long-run claims process. The probability that the average of the 9 claims is larger than \$1000 is about

A) 0.2578. B) 0.2162. C) 0.6486. D) 0.4168.

5. In a particular game, a fair die is tossed. If the number of spots showing is six you win \$6, if the number of spots showing is five you win \$3, and if the number of spots showing is four you win \$1. If the number of spots showing is one, two, or three you win nothing. You are going to play the game twice. The probability that you will win at least \$1 total on the two plays of the game is

A) 8/9. B) 3/4. C) 1/9. D) 1/4.

6. When a lower confidence level is used, all other parts of the confidence interval held constant,

A) The standard error will stay the same but the margin of error will be smaller.
B) The standard error will be smaller but the margin of error will stay the same.
C) The standard error and the margin of error will both be smaller.
D) The standard error and the margin of error will stay the same.

7. A radio talk show host with a large audience is interested in the proportion $\hat{p}$ of adults in his listening area that think the drinking age should be increased to 21. To find this out he poses the following question to his listeners: "Do you think that the drinking age should be increased to 21?" He asks listeners to phone in and vote "yes" if they agree the drinking age should be lowered and "no" if they do not. You are told that the proportion $\hat{p}$ of those who phoned in and answered yes is $\hat{p}$ = 0.25 and the standard error $S.E.(\hat{p})$ of the proportion is 0.04. The number of people who phoned in is

A) 5. B) 118. C) 100. D) 22.

8. Suppose you want to estimate the average height of 25-year-old males, in inches. From historical data you estimate standard deviation to be approximately 2.4. If you want to estimate average height to within 0.5 inches with 99% confidence, roughly how many observations should you select?

A) 89. B) 153. C) 52. D) 123.

9. A certain population follows a normal distribution with mean μ. You collect data and test the hypotheses H_0: $\mu = 1$, H_a: $\mu \neq 1$. You obtain a *P*-value of 0.052. Which of the following is true?

A) A 90% C.I. interval for μ will not include the value 1. A 95% C.I. interval for μ will not include the value 1.
B) A 90% C.I. interval for μ will not include the value 1. A 95% C.I. interval for μ will include the value 1.
C) A 90% C.I. interval for μ will include the value 1. A 95% C.I. interval for μ will not include the value 1.
D) A 90% C.I. interval for μ will include the value 1. A 95% C.I. interval for μ will include the value 1.

10. A statistic is said to be unbiased if
A) it is used for only honest purposes.
B) the mean of its sampling distribution is equal to the true value of the parameter being estimated.
C) the survey used to obtain the statistic was designed so as to avoid even the hint of racial or sexual prejudice.
D) both the person who calculated the statistic and the subjects whose responses make up the statistic were truthful.

11. A university administrator obtains a sample of the academic records of past and present scholarship athletes at the university. The administrator reports that no significant difference was found in the mean GPA (grade point average) for male and female scholarship athletes (*p-value* = 0.287). This means

A) the GPAs for male and female scholarship athletes are identical except for 28.7% of the athletes.
B) the maximum difference in GPAs between male and female scholarship athletes is 0.287.
C) the chance that a pair of randomly chosen male and female scholarship athletes would have a significant difference in GPAs is 0.287.
D) the chance of obtaining a difference in GPAs between male and female scholarship athletes as large as that observed in the sample if there is no difference in mean GPAs is 0.287.

12. The first baseman in a baseball game is thrown two balls during an inning. Define the events:
A - catches both balls, B - catches at least one ball, C - misses both balls
Which of the following is false?
A) P(B|A)=1
B) A and C are disjoint
C) P(C|B)=0
D) A and B are independent

13. At a small Midwestern college, 60% of the students are female, 70% of the female students are liberal arts majors, and 50% of the male students are liberal arts majors. A student is selected at random to represent the school at a conference. The probability that the selected student is a liberal arts major is

A) 1.20. B) 0.20. C) 0.42. D) 0.62.

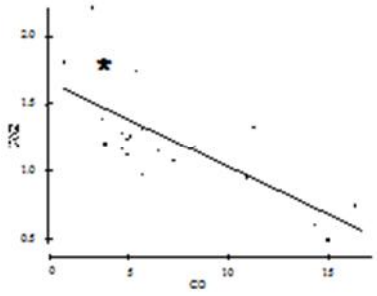
14. A level α two-sided significance test rejects the null hypothesis $H_0: \mu = \mu_0$ when

A) the t test statistic differs from μ_0 by a significant amount.
B) the t test statistic differs from μ_0 by at least α.
C) the value μ falls outside a level 1-α confidence interval for μ_0.
D) the value μ_0 falls outside a level 1-α confidence interval for μ.

15. Two years ago the true proportion of people interested in starting a small business was 20%. You wish to determine if that proportion has increased. Today you take a sample of 200 people and find that 35% are interested in starting a small business. What is the test statistic for your test?

A) 1.96
B) 5.30
C) 4.45
D) 17.68

16. Consider the following scatterplot of amounts of CO (carbon monoxide) and NOX (nitrogen oxide) in grams per mile driven in the exhausts of cars. The least-squares regression line has been drawn in the plot.



Which is true about the point (*):

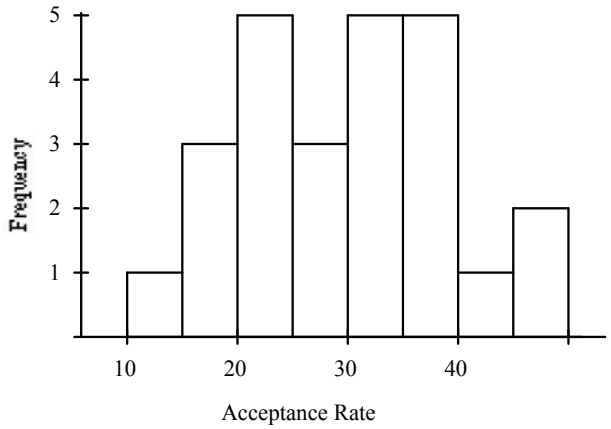
a) The product $z_x z_y$ is positive and the residual is positive.

b) The product $z_x z_y$ is positive and the residual is negative.

c) The product $z_x z_y$ is negative and the residual is positive.

d) The product $z_x z_y$ is negative and the residual is negative.

17. The following histogram represents the distribution of acceptance rates (percent accepted) among 25 business schools in 1995.



What is a possible value for the third quartile?

a) 32 **b)** 28 **c)** 36 **d)** 42

18. A SLR model is proposed to analyze the association between storage temperature and shelf life of a certain brand of cough syrup. A sample of 37 bottles were stored at a different temperatures and each was tested weekly (using some test) to determine if the syrup has expired or not. The response is the week the syrup was deemed to have expired. Using the information below answer the questions:

Parameter estimates:

Parameter	Estimate	Std. Err.	DF	T-Stat	P-Value
Intercept		1.430		90.24	<0.0001
Slope		0.061		-29.67	<0.0001

What is the estimated effect of increasing storage temperature by 5 degrees?
a) -1.81 b) 120 c) -9.05 d) 486.39

Use the following to answer the questions 19 - 20.
Does systolic blood pressure increase with age? A sample of 27 individuals was randomly selected from a population with ages ranging from 20 to 70 years. The response is measured as the average systolic blood pressure at a fixed time during the day over a 1-week period. Consider SLR model for average blood pressure given age. Use the following output to answer the questions:

Parameter estimates:

Parameter	Estimate	Std. Err.	DF	T-Stat	P-Value
Intercept	76.84	4.6288	25	16.60	<0.0001
Slope	1.41	0.1006	25	14.02	<0.0001

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	7745	7745	197.38	<0.0001
Error	25	981	39.24		
Total	26	8726			

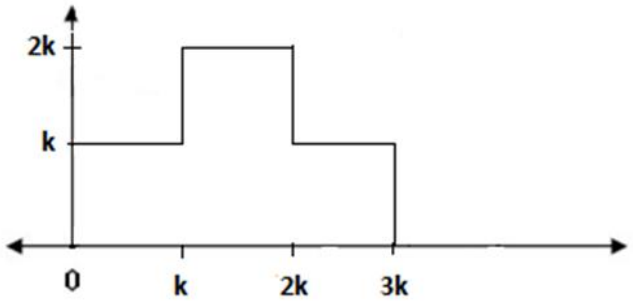
19. What is the predicted blood pressure of a 53-year old?
A) 82.17 B) 407.39 C) 322.16 D) 151.57

20. The percentage of variation in blood pressure explained by the simple linear regression on age is
A) 88.76% B) 11.24% C) 10.06% D) less than 0.01%

21. The *test-statistic* from a sample of $n = 5$ observations from a normal population for the two-sided test of H_0: $\mu = 6$, H_a: $\mu \neq 6$ has the value 1.5. Based on this information

A) P-value < 0.10
B) 0.10 < P-value < 0.15.
C) 0.15 < P-value < 0.20.
D) 0.20 < P-value.

22. Consider the random variable, X, with the following density curve:



What is k?

a) 0.41 b) 0.5 c) 0.25 d) 0.71

23. The ages of people in a class (to the nearest year) are as follows:

Age	18	19	20	21	22	23	24	25	32
Number of students	14	120	200	200	90	30	10	2	1

The median is

A) over 21 B) less than 20 C) 20 D) 21

24 I flip a special coin twice and count the number of heads. The flips are independent and always land either heads or tails. Which of the following is a valid assignment of probabilities for the number of heads observed in two flips? Note: There is NOT a typo in this question. Only one answer is correct.

A)	Number of heads	0	1	2
	Probability	1/4	1/3	1/4
B)	Number of heads	0	1	2
	Probability	4/9	4/9	1/9
C)	Number of heads	0	1	2
	Probability	1/2	1/2	1/2
D)	Number of heads	0	1	2
	Probability	1/4	1/4	1/2

Use the following to answer questions 25 – 27

A researcher is studying treatments for agoraphobia with panic disorder. The treatments are to be the drug Imipramine at the doses 1.5 mg per kg of body weight and 2.5 mg per kg of body weight. There will also be a control group given a placebo. Thirty patients were randomly divided into three groups of 10 each. One group was assigned to the control, and the other two groups were assigned to the two treatments. After 24 weeks on treatment, each of the subject's symptoms were evaluated through a battery of psychological tests, where high scores indicate a lessening of symptoms. Assume the data for the three groups are independent, and the data are approximately normal. The means and standard deviations of the test scores for the three groups are given below.

Mean test score	Std. dev. in score	Group
70	12	Control
85	18	Dose = 1.5
100	20	Dose = 2.5

An ANOVA *F* test was run on the data. Below are a portion of the results.

Source	df	SS	MS	F-Stat	P-value
Treatments					
Error			289.333		
Total		12312			

25. The value of the test statistic is
    A) 8.06
    B) 7.78
    C) 0.58
    D) 0.29

26. The best estimate for the common standard deviation is
    A) 17.290.
    B) 16.703.
    C) 16.667.
    D) 17.009.

27. The distribution of the test-statistic under the null hypothesis that all three groups have the same mean is:
    A) An F-distribution with 2 and 29 degrees of freedom.
    B) An F-distribution with 2 and 27 degrees of freedom.
    C) An F-distribution with 3 and 27 degrees of freedom.
    D) An F-distribution with 3 and 29 degrees of freedom.

28. An old saying in golf is "You drive for show and you putt for dough." The point is that good putting is more important than long driving for shooting low scores and hence winning money. To see if this is the case, data on the top 16 money winners on the PGA tour in 1997 are examined. The average number of putts per hole for each player is used to predict their total winnings using the simple linear regression model

$$\text{Average 1997 winnings} = \alpha + \beta\,(\text{average number of putts per hole})$$

This model was fit to the data using the method of least squares. The following results were obtained from statistical software.

Parameter estimates:

Parameter	Estimate	Std. Err.	DF	T-Stat	P-Value
Intercept	7897000	3023000		2.61	
Slope	-5026000	1698000		-2.96	

The p-value for test to determine if there is a negative linear association between average number of putts per hole and average winnings is:

A) less than 0.005.
B) between 0.005 and 0.01.
C) between 0.01 and 0.02.
D) greater than 0.02.

Use the following for 29-30: 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	df	SS	Mean Square	F-Statistic	p-value
Treatment	3	448.575	149.53	63.36	<0.0001
Error	36	84.852	2.36		
Total	39	533.427			

29. The estimated overall effect of computer labs on mean score is
    a) 13.2 b) 6.6 c) 7.4 d) 37.7

30. The sum-of-squared residuals for the model under the alternative hypothesis in the one-factor ANOVA F-test for any mean differences is:
    a) 63.36 b448.58 c) 533.43 d) 84.85

Part 2 - Long Answer (20 marks total): You must show your work to receive credit for your answers.

Problem 1 (10 marks total): In a study on first year science student expectations a survey was conducted on 60 randomly selected students (30 male and 30 female) just starting their science degree program at a particular university. All students are taking the exact same courses. Before starting any classes students were asked what their expected average mark (as a percentage) will be at the end of their first year. These students were followed and asked again after their midterms in their first semester. The summary statistics are displayed below.

"__ before" are the results before starting any classes.
"__ after" are the results after their midterms in the first semester.
The **diff**erence rows give the summary statistics on paired differences.

Note: Not all of the rows will be required. It is up to you to decide which summary statistics are appropriate to answer the questions.
Summary statistics:

Column	n	Mean	Std. Dev.
Male before	30	77.99	6.11
Female before	30	74.34	4.03
Male after	30	68.03	8.78
Female after	30	69.87	7.33
Diff: Male before - Male after	30	9.96	4.53
Diff: Female before - Female after	30	4.47	4.78
Diff: Male before - Female before	30	3.65	7.19
Diff: Male after - Female after	30	-1.84	10.81

a) **(3 marks)** In their report, they estimate that the change in expectations for males before starting classes vs. after their midterms was a drop between 9.34 and 10.58. What confidence level did they use? Estimate this within a range using the t-table.

b) **(2 marks)** You decided to estimate this change (before-after for males) yourself. Based on a sample of 125 males you estimated with 98% that the change was between 11.25 and 12.75. What was the sample standard deviation of the differences in your sample?

c) **(5 marks)** Do males and females have different expectations before starting classes? Carry out an appropriate test. State clearly the null and alternative hypothesis in terms of parameters you have defined. Calculate the test-statistic and show how it is calculated (state clearly the estimate and standard error). Give the p-value probability statement showing how the p-value is calculated with reference to its null distribution. Find the p-value range using the t-tables, and then give your conclusion in your own words.

H_0 : _____ H_a : _____

Test-Statistic=

$p-value=$

$Conclusion$:

Problem 2: (4) Vitamin C and the Common Cold (Randomized Experiment): Does vitamin C help prevent the occurrence of the common cold? A study was conducted on 550 volunteers. At the beginning of the winter the subjects were randomly assigned to one of two treatment groups, the vitamin C group and the placebo group. Both groups received an adequate supply of their respective drugs to last through the winter at 1,000 mg per day. At the end of the study, a physician who was unaware of the treatment to which the subject had been assigned interviewed each subject and determined whether the subject had suffered from a cold during the study period. The table below is a summary of the results:

		Outcome (Response)		
		Cold	No Cold	Total
Treatment (Explanatory)	Placebo	189	84	273
	Vitamin C	161	116	277
	Total	350	200	550

Calculate a 96% confidence interval for the difference in the proportion of individuals who suffered from the cold for the two groups. Clearly identify (in the blanks below) the estimate, the standard error of the estimate and the critical value. Then calculate the interval. Based on your result **comment** about whether or not you think vitamin C helped.

Estimate: _____

$S.E.(Estimate)$: _____

Critical Value:_____

Problem 3: (5 marks total) In order to apply for a particular $10,000 loan your annual salary must be less than $30,000. Out of the applications received, 30% are approved for the full amount ($10,000), 50% are approved for half ($5,000), and 20% are rejected. The resulting decisions are independent of an individual's salary. Salaries in this population are known to be approximately Normally distributed with mean $35,000 and standard deviation $8,000.

a) (2 marks) Define the probability distribution for the loan amount given to individuals in this population.

b) (1 mark) What is the average or expected amount loaned to any individual in this population?

c) (2 marks) Given that a randomly selected individual from the population didn't get a loan, what is the probability they earn less than $25,000?

Problem 4: (3 marks) A Cola company decides to only advertise their brand on a particular university campus if the level of consumer support they have is below 60%. They will estimate support and base their decision on a random sample of 81 students. Suppose at a particular campus the actual proportion of support is 70% (but they don't know this). What is the probability that they will (in error) decide to advertise on this campus?