FINAL EXAM

STA 110A
Brani Vidakovic

Name ____

TA: ______________________ Section: ________

Notes:

1. This is an open book and open notes exam.
2. You must show your work and explain your answer in order to receive credit.
3. The exam carries 100 points.
4. The points assigned to each problem are indicated at the beginning of that problem. Use them to plan your time. You have 3 hours to finish.
5. The exam has 8 problems. Make sure your copy has all of them!

1. Sharp Willies.[15pts] After loosing the contract with Sherwood Rascals, Sharp Willie, an arrow producing enterprise, has sent a number of employees to four educational institutions for technical training. The company hoped that the training would improve employee productivity and product quality. At the end of program Sharp Willie tested 40 graduates. The scores are:

Program A:	95 88 90 99 89 93 95 97 85 90
Program B:	92 88 80 75 67 78 92 80 77 69
Program C:	85 81 86 91 78 81 86 90 75 83
Program D:	98 65 74 82 90 62 75 85 70 82

1. Fill-in the ANOVA table and test that all programs are the same at $\alpha = 1\%$.
2. What assumptions on data are needed to apply an ANOVA test.
3. Explain (in the terms of the problem) what is the error of first kind here.

```
Analysis of variance on score
Source      SS      df        MS        F
program                     382.3
error
total     3367.9
```

2. College Entrance Test [10pts]. Because of the role of college aptitude test scores in college entrance decision, there are minicourses that purport to teach students how to take these tests. A particular aptitude test has been found to produce scores that are normally distributed, with mean θ and standard deviation 80. If the minicourse directed at this test is effective (on average), the mean score θ of students who take the course is larger than 500; otherwise it is not. We want to test

$$H_0 : \theta \geq 500 \quad \text{versus} \quad H_1 : \theta < 500,$$

and our prior for θ is $N(480, 30^2)$.

1. Find the prior probabilities of hypotheses H_0 and H_1, π_0 and π_1.
2. If we observed 491, 487, 492, and 482 find the posterior probabilities of hypotheses H_0 and H_1, p_0 and p_1, and make the decision.
3. What is 95% credible set for θ.

3. Great white shark.[15pt] One of the most feared predators in the ocean is the great white shark *Carcharodon carcharias.* Although it is known that the white shark grows to a mean length of 21 feet (record: 39 feet), a marine biologist believes that the great white sharks off the Bermuda coast grow much longer due to unusual feeding habits. To test this claim, a number of full-grown great white sharks are captured off the Bermuda coast, measured and then set free. However, because the capture of sharks is difficult, costly, and very dangerous, only three are sampled. Their lengths are 24, 20, 22 feet.

1. Do the data provide sufficient evidence to support marine biologist's claim? Use $\alpha = 0.1$.
2. What assumptions must be made in order to carry out the test?

For 3 and 4 assume that the population standard deviation σ is known and is 4 feet.

3. What sample size is needed for a 99% confidence interval for the unknown mean length of a white shark if we want the length of the confidence interval to be 8 msec?

4. What is the power $(1-\beta)$ of the test $H_0 : \mu = 21$ against the one-sided alternative when $H_1 : \mu = 23$? (Level α is 0.05, $\bar{X} = 22, n = 3, \sigma = 4$.)

4. Multiple Choice.[10pts] A student answers a multiple choice examination question that has 5 possible answers. Suppose that the probability that the student knows the answer to the question is 0.75 and the probability that the student guesses is 0.25. If student guesses, probability of correct answer is 0.2.

(i) What is the probability that the fixed question is answered correctly?

(ii) If it is answered correctly what is the probability that the student really knew the correct answer.

5. Sales of air conditioners.[15pts] An appliance store dealer believes that the weekly sales of air conditioners are dependent upon the average outside temperature during the week. In support of this claim the dealer randomly selects eight weeks and records the average outdoor temperature and the sales of air conditioners during the week. This information is listed in the following chart:

Average Outside Temperature, x	Number of Air Conditioners Sold, y
75	4
83	7
67	2
89	9
95	12
79	6
81	6

The MINITAB output for the regression procedure is given below:

```
The regression equation is
sales = - 22.2 + 0.354 temp

Predictor        Coef       Stdev    t-ratio       p
Constant      -22.213       1.704     -13.04   0.000
temp          0.35412     0.02085      16.98   0.000

s = 0.4660      R-sq = 98.3%     R-sq(adj) = 98.0%

Analysis of Variance

SOURCE       DF          SS          MS         F        p
Regression    1      62.628      62.628    288.39    0.000
Error         5       1.086       0.217
Total         6      63.714



     Fit  Stdev.Fit         95% C.I.          95% P.I.
  13.199      0.428   ( 12.098, 14.299)  ( 11.571, 14.826)
```

1. Test the hypothesis $H_0 : \beta_1 = 0.3$ against two sided alternative.
2. Find 98% confidence interval for the unknown population intercept β_0.
3. How many sales do you predict when temperature is 100? Find 99% confidence interval for the mean response when temperature is 100.
4. Discuss the regression output (one paragraph).

Basic metabolic rate.[10 pts.] A group of researchers is studying basal metabolic rate, i.e. the number of calories your body consumes when you are in a reclined, resting state. They study a group of 15 subjects, both female and male, and measure their heights, weights, and metabolic rates with a gas exchange analysis. Their goal is to see if weight and height correlate with metabolic rate.

bmr	weight	height	bmr	weight	height
1800	70	170	1850	71	175
1725	72	169	1810	73	177
1799	74	172	1847	75	173
1850	76	178	1200	50	155
1344	54	160	1322	55	166
1400	56	161	1386	58	163
1375	63	168	1450	63	170
1433	64	170			

```
The regression equation is
bmr = 414 + 29.2 weight - 4.39 height

Predictor        Coef        Stdev    t-ratio        p
Constant        413.8        968.1       0.43    0.677
weight         29.241        5.745       5.09    0.000
height         -4.391        7.692      -0.57    0.579

s = 76.83       R-sq = _______     R-sq(adj) = 89.7%

Analysis of Variance

SOURCE       DF          SS           MS          F         p
Regression    2      732985       366493      62.09     0.000
Error        12       70828         5902
Total        14      803813

SOURCE       DF      SEQ SS
weight        1      731062
height        1        1924
```

1. What is $R - sq$?

2. Would you accept the hypothesis that the intercept of is equal to zero against two sided alternative? Why?

3. What `bmr` can be predicted for `weight=70` and `height=175`.

7. Alcohol and Marriage. [10 pts] A national survey was conducted to obtain info on the alcohol consumption patterns of American adults by marital status. A random sample of 1772 residents 18 years old and over, yielded the data below. Do the data suggest at 5% significance level that marital status and alcohol consumption patterns are statistically dependent?

	Abstain	1 - 60	over 60
Single	67	213	74
Widowed	85	633	129
Divorced	27	60	15

noindent **8. True-False . [15pts]**

Question	True	False
1. Whatever data set you have – at least 75 percent of the data is within 3 standard deviations from the mean.		
2. It may happen that the mean of a data set can fall outside interquartile range ($[Q_1, Q_3]$).		
3. In testing statistical hypotheses when you fix α then β is fixed for all possible alternatives.		
4. In testing equality of means by ANOVA an assumption on equality of population variances is needed.		
5. In doing contingency tables you got a zero count in a cell Thus, testing for independence is impossible since you have division by 0 in the process of calculating chi-square statistics.		
6. A part of the regression output is an ANOVA test that tests that all slopes are equal.		
7. In ANOVA, the sum of numerator and denominator degrees of freedom in the F statistics is equal to the total sample size minus one.		
8. When sample size is larger than 30 than any t-test can use z-cut-points.		
9. Bayesian statisticians can do inference even if they have no data.		
10. If p-value is 0.02, that means that the probability that H_0 is true is 0.02.		
11. It is impossible to use normal approximation on discrete distributions simply because the normal distribution is continuous.		
12. The mean θ in a population is the quantity of interest. generally, it is more likely that the average of two observations is closer to the unknown mean than a single observation.		
13. In testing the equality of means in two populations you did not assume that the variances are the same. Then in general, number of degrees of freedom for the **t** statistic may be different than $n_1 + n_2 - 2$.		
14. p-value for the test $H_0 : \mu = 2$ versus $H_1 : \mu > 2$ is 0.75. Then the p-value in testing H_0 against $H_1 : \mu \neq 2$ is 0.5.		
15. The probability of the event A is $P(A) = 0.7$. The probability of the event B is $P(B) = 0.5$ The probability of $A \cdot B$ is $P(A \cdot B) = 0.35$. Events A and B must be independent.		
16. The probability of the event A is $P(A) = 0.7$. The probability of the event B is $P(B) = 0.5$ The probability of $A \cdot B$ is $P(A \cdot B) = 0.35$. Then $P(A \cup B) = 0.85$.		
17. A possible estimator of the mean μ when the observations are normal is $\hat{\mu} = \bar{X}$. The estimator is a random variable.		
18. In testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$ the population is not normal. Than the only test applicable is **t** test.		
19. In doing regression with one variable the **t** test for testing that the slope is zero and the ANOVA F test coincide.		
20. If $P(A\|B) = P(A)$, then the events A and B are necessarily independent.		