BIOS 600-001 (Bordonali)
Final Exam 2
12/17/10

Name: ____________________________________
(Please print)

Exam Instructions

1. Answer ONLY what the question asks! This is not a homework, so there is no need to write a formal conclusion with every answer. If the question asks for a value, just calculate that value and provide it.
2. For multiple choice questions, pick ONE answer and write the letter of the answer choice on the line next to the problem number. If you circle an answer that does not match your written choice, I will grade your written answer choice.
3. For non-MC questions, circle/underline/box your final answer.
4. You are not required to show your work. However, partial credit for non-MC problems can only be given when work is provided. No partial credit will be given for MC problems.

Quantities you may find useful:

$$z_{0.9} = 1.28; \qquad z_{0.95} = 1.645; \qquad z_{0.975} = 1.96; \qquad z_{0.99} = 2.33; \qquad z_{0.995} = 2.576$$

Honor Statement. I pledge that I have not used any notes, text, or any other reference materials other than the "Formulas and Tables for Gerstman" (with personal additions) during this examination. I pledge that I have neither given nor received any aid from any other person during this examination, and that the work presented here is entirely my own. I furthermore pledge that I will not reveal any of the material on this examination, either in the form of the exact question or the topics covered, to any person for any reason. **I also pledge that I have not discussed the contents of the previous final exam with anyone in the class that took the early exam.** I pledge that I will report all Honor Code violations observed by me. I understand that if I have committed any of the above, I have violated the UNC Honor Code.

Signature and date:

____

Use the following information to answer questions 1-4.

Consider the population of individuals under 45 who have been diagnosed with Hodgkin's lymphoma. The true mean proportion of these individuals who survive five years after diagnosis is $p = 0.90$. Suppose we take a sample of 10 patients from this population, and evaluate X, the number of patients surviving five years.

(2 pts) 1. What is the expected value, μ, of X?

(2 pts) 2. What is the standard deviation, σ, of X?

(4 pts) 3. What is the exact probability of seeing 8 or more patients survive five years?

(3 pts) 4. Now suppose we take a sample of $n = 100$ patients from this population. What is the **approximate** probability of seeing 80 or fewer patients survive five years? (Hint: Don't use the binomial formula!)

______ (2 pts)

5. **True/False.** For a binomial random variable with n trials, each observation is categorized as being either a failure or success.

______ (2 pts)

6. **True/False.** For a binomial random variable, the probability of success for one observation affects the probability of success for the next observation.

______ (2 pts)

7. In a test of statistical hypotheses, what does the p-value tell us?

   A) The probability that the null hypothesis is true.

   B) The probability that the alternative hypothesis is true.

   C) The probability that we observe data as or more extreme than our study, assuming the null hypothesis is true.

   D) The probability that the null hypothesis is true, assuming we observe data as or more extreme than our study.

(3 pts)

8. What is the definition of Type I error, α?

______ (2 pts)

9. What is the definition of power, $1 - \beta$?

   A) The probability of rejecting the null hypothesis given the null hypothesis is true

   B) The probability of rejecting the null hypothesis given the alternative hypothesis is true

   C) The probability of failing to reject the null hypothesis given the null hypothesis is true

   D) The probability of concluding that the alternative hypothesis is true when it is not

   E) None of the above

For questions 10-13, read the presented information carefully. Then decide what type of test would be most appropriate for each scenario.

_________ (2 pts) 10. A study of vector control in an African village found that the mean sprayable surface area was 249 ft^2 with sample standard deviation 39.82 ft^2 in a SRS of $n = 40$ homes. We wish to compare the true mean sprayable surface area in this village to the country-wide mean sprayable surface area reported by WHO and widely accepted as true.

A) One sample Z test

B) Two independent sample t test

C) One sample Z test for proportions

D) One sample t test

_________ (2 pts) 11. Benign prostatic hyperplasia is an enlargement of the prostate that affects older men and causes bladder problems like painful urination. A study of a minimally invasive procedure for the treatment of BPH enrolled $n = 150$ men who rated their pain at baseline, and again 3 months after the procedure. The continuous variable "Pain" is measured on a scale from 1 to 10. We would like to determine whether the post-treatment pain is significantly reduced from the pre-treatment pain, on average.

A) Paired samples t test

B) One sample Z test for proportions

C) Two independent sample Z test for proportions.

D) Two independent sample t test

_________ (2 pts) 12. Recently, standards for levels of carbon monoxide (CO) in the workplace were relaxed, so that workers were now allowed to be subjected to environments containing more CO than before. Prior to this, 300 workers were tested, and 24 showed signs of respiratory illness. One year after the standards were relaxed, another 300 workers were tested, and 35 showed signs of respiratory illness. We want to test the null hypothesis that the population proportion of respiratory illness after the standards were lowered is the same as before the standards were lowered.

A) One sample Z test for proportions.

B) Chi-square test

C) ANOVA F test

D) Paired samples t test

________ 13. We want to know whether taking a calcium supplement increases bone density (BMD) in elderly women with osteoporosis. We randomize 26 women to take a calcium supplement and 26 women to take placebo, and measure the increase in bone density over 6 months. Assume BMD is normally distributed. In the calcium supplement group, the mean increase in BMD is 50 g/cm^2, with standard deviation 10 g/cm^2. In the placebo group, the mean increase in BMD is 20 g/cm^2, with standard deviation 12 g/cm^2. We will test the null hypothesis that the average increase in BMD in elderly women taking a calcium supplement is the same as the average increase in elderly women on placebo.

(2 pts)

A) Two sample t test

B) Z test

C) Paired t test

D) One sample t test

Use the following information to answer questions 14-15

An investigator wants to determine whether vitamin deficiency is associated with birth defects. By reviewing birth certificates during a single year in a large U.S. county, 189 infants born with Neural Tube Defects (NTDs) are enrolled as well as 600 other infants, selected at random from the certificates. A dietary questionnaire was then administered to the mothers. In all, 221 mothers reported no use of supplementary vitamins, with 84 of these mothers giving birth to infants with an NTD.

________ 14. What kind of sampling method was used in this study?

(2 pts)

A) Purposive Cohort

B) Case-Control

C) Naturalistic

(2 pts) 15. What is the Odds Ratio describing the relationship between vitamin use and NTDs?

Use the following information to answer questions 17-23.

In sub-saharan African, malaria is a major cause of anemia during pregnancy. Several studies have explored whether Insecticide Treated Nets (ITNs) could reduce malaria prevalence and anemia prevalence among pregnant women. An observational study (Marchant, 2002) also investigated the association between ITN and malaria. In this study, 505 pregnant women were enrolled from six villages in the Kilombero valley area of Tanzania. These women were then asked about ITN use and blood samples were taken to determine malarial status. There were $n_1 = 266$ women reporting ITN use, 68 of whom had malaria. There were an additional 79 women with malaria that did not use ITNs.

__________ (2 pts) 16. What kind of sampling method was used in this study?

A) Purposive Cohort

B) Case-Control

C) Naturalistic

(2 pts) 17. What is the Relative Risk of having malaria for this data?

(3 pts) 18. Give a 95% CI for the RR.

(2 pts) 19. What is the estimated difference in risk of having malaria between ITN users and non-ITN users?

(2 pts) 20. What are the hypotheses we would use to test whether the risk of malaria is the same between the two groups?

(2 pts) 21. What is the test statistic?

(2 pts) 22. What is the p-value? Either exact values or ranges are acceptable.

________ (2 pts) 23. If we were to perform the χ^2 test instead, how would the p-value compare to the one we calculated in #22? Note: There is **no need** to compute the χ^2 statistic.

A) The χ^2 p-value would be larger than the p-value in #22.
B) The χ^2 p-value would be smaller than the p-value in #22.
C) The χ^2 p-value would be equal to the p-value in #22.
D) We don't know the relationship unless we actually perform the χ^2 test.

________ (2 pts) 24. If we were to perform the χ^2 test instead, what degrees of freedom would we use?

A) 1
B) 2
C) 4
D) 504

________ (3 pts) 25. The rate of HIV/AIDS infection is at epidemic levels in Zambia, with a hypothesized national prevalence of 17% among adults aged 15-49. Our goal is to create a 90% confidence interval for p with a margin of error $m = 0.05$. Using the hypothesized value of 0.17, what sample size is necessary to achieve this margin of error?

A) 45
B) 153
C) 217
D) 262

________ (2 pts) 26. If we prefer a 95% confidence interval instead, what would happen to the margin of error m? Assume that we will use the sample size calculated in #25.

A) m would decrease.
B) m would increase.
C) m would not change.

Use the following information to answer questions 27-32.

A study was conducted to investigate the relationship between maternal smoking during pregnancy and oral cleft, a specific type congenital malformation. In a random sample of 27 children with oral cleft, 15 have mothers who smoked during pregnancy.

(5 pts) 27. Construct a 95% confidence interval for the population proportion of maternal smoking among children with oral cleft.

Among children who suffer from a congenital malformation **other than** oral cleft, it is assumed (or, generally believed) that the constant percentage of maternal smoking is 32.8%. Using the data from above, we would like to test whether the proportion of maternal smoking for children with an oral cleft is the same as the known percentage of maternal smoking for children with other types of malformations.

(2 pts) 28. What are the null and alternative hypotheses for this test?

(2 pts) 29. Calculate the standard error SE_p under H_0. Use the formula corresponding to the test statistic for the test in #28.

(2 pts) 30. What is the value of the test statistic corresponding to the test in #28?

(3 pts) 31. What is the p-value for this test? Either exact values or ranges are acceptable.

(2 pts) 32. Suppose the true population proportion of children with oral cleft whose mothers smoked is actually $p_1 = 0.25$. To conduct the two-sided hypothesis test at $\alpha = 0.01$ with 90% power, how large a sample would be required?

__________ (2 pts) 33. If we were to perform the previous test at $\alpha = 0.05$, using the sample size calculated above, what would happen to the power $1 - \beta$?

A) Power would decrease.

B) Power would increase.

C) Power would stay the same.

__________ (2 pts) 34. In a χ^2 test, we calculate "expected" values in order to calculate the test statistic X^2_{stat}. These expected values are calculated assuming ________.

A) H_0 is true.

B) H_A is true.

C) nothing about the hypotheses.

Please answer the following questions regarding assumptions of Multiple Linear Regression (MLR).

__________ (2 pts) 35. **True/False.** The error terms in MLR are normally distributed.

__________ (2 pts) 36. **True/False.** The variability of the residuals can change between observations.

Use the following information to answer questions 39-45

In the population of low birth weight infants, it has been shown that gestational age (weeks) is a significant linear predictor of infant length (cm). In particular, length increases as gestational age increases. This was discovered via a Simple Linear Regression model.

$$y = \alpha + \beta_1 x_1 + \epsilon, \quad \text{where } x_1 = \text{gestational age}$$

We now wish to investigate how an expectant mother's age and diagnosis of toxemia during pregnancy affects the length of her child. Toxemia, or preeclampsia, is a pregnancy condition in which high blood pressure and protein in the urine develop after the 20th week (late 2nd or 3rd trimester) of pregnancy. The only cure for toxemia is delivery of the baby, so we might expect toxemia and gestational age to be related. Similarly, older mothers may be more likely to have high-risk pregnancies. Therefore, we fit the following Multiple Linear Regression model using software. The output is displayed below.

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon \quad ,$$

$$\begin{aligned} \text{where } x_1 = &\quad \text{gestational age in weeks} \\ \text{and } x_2 = &\quad \text{mother's age in years} \\ \text{and } x_3 = &\quad \text{toxemia status} = \begin{cases} 1 \ , & \text{if diagnosed with toxemia} \\ 0 \ , & \text{if not diagnosed with toxemia} \end{cases} \end{aligned}$$

ANOVA Table

Source	df	SS	MS	F	p-value
Between	3	621.37	207.123	31.00	$< .0001$
Within	96	641.39	6.681		
Total	99	1262.76			

Parameter Estimate Table

Parameter	Estimate	Std Error	t Stat	p-value
Intercept	6.04	3.233	1.87	0.0649
GestAge	1.05	0.116	9.05	$< .0001$
MomAge	0.03	0.045	0.67	0.5752
Toxemia	-1.78	0.696	-2.56	0.0122

_________ (2 pts) 37. **True/False.** Based on the model results, we can say that mother's age is a significant linear predictor of infant length.

_________ (2 pts) 38. **True/False.** Based on the fitted model, the infants of toxemic mothers are expected to be longer on average than the infants of non-toxemic mothers.

________ (2 pts) 39. What percentage of the variation in infant length remains unexplained by the linear model with gestational age, mother's age and toxemia?

A) 16.40%
B) 83.60%
C) 49.20%
D) 50.79%

________ (2 pts) 40. What is the test statistic to test $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ vs. H_A : at least one $\beta_i \neq 0$?

A) $F = 31.00$
B) $t = 1.87$
C) $t = 9.05$
D) $t = 0.67$
E) $t = -2.56$

________ (2 pts) 41. What is the test statistic to test whether Toxemia has a significant linear relationship with Length, after adjusting for Gestational Age and Mother's Age?

A) $F = 31.00$
B) $t = 1.87$
C) $t = 9.05$
D) $t = 0.67$
E) $t = -2.56$

________ (3 pts) 42. Report the 95% confidence interval for β_1, the slope coefficient for gestational age.

A) 6.04 ± 5.3797
B) 1.05 ± 0.2308
C) -1.78 ± 1.3850
D) 1.05 ± 0.1930
E) -1.78 ± 1.1581

(2 pts) 43. Based on the fitted model, what is the predicted length of infants born to $x_2 = 30$ year-old toxemic mothers having gestational age $x_1 = 25$ weeks?

(3 pts) 44. Interpret the estimate of β_3, the slope coefficient for toxemia, in the context of the fitted model and the original population.

(2 pts) 45. **Bonus**: In class we discussed the relationship between ANOVA and the multiple linear regression model in depth. We also stated that a similar relationship exists between the pooled 2-sample t test and the multiple linear regression model.

Suppose we have data comparing blood pressure for two groups, Treatment and Control, and we want to compare the mean blood pressures for the two groups using a linear model. What is the population linear regression model that is equivalent to this two sample t test? Be as specific as possible.