BIOS 600-001 (Bordonali)
Final Exam 1
12/9/10

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 will not discuss the contents of this exam with anyone in the class that yet to take the 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 40 who have been diagnosed with lung cancer. The true mean proportion of these individuals who survive five years after diagnosis is $p = 0.10$. 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 2 or fewer 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 20 or fewer patients survive five years?

(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 outcome of one observation has an effect on the outcome of other observations.

(3 pts) 7. What is the definition of a p-value?

(2 pts) 8. What is the definition of Type I error, α?

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

(2 pts) 9. What is the definition of Type II error, β?

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 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 t test
    B) ANOVA F test
    C) Two independent sample t test
    D) One sample Z test for proportions

_____ (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) One sample Z test for proportions
    B) Paired samples t test
    C) Two independent sample t test
    D) Two independent sample Z test for proportions.

_____ (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) ANOVA F test
    B) Paired samples t test
    C) One sample Z test for proportions.
    D) Chi-square test

(2 pts) 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 increase in BMD in elderly women taking a calcium supplement is the same as the increase in elderly women on placebo.

A) One sample t test

B) Two sample t test

C) Paired t test

D) Z test

Use the following information to answer questions 14-16

Consider a 1979 study examining the relationship between conjugated estrogen use and cervical cancer. The investigators enrolled 183 women with cervical cancer as well as 187 women without cervical cancer, who were selected randomly from voter registration records. Each woman was then asked about her history of estrogen use. In all, 74 women reported estrogen use, where 55 of these women had cervical cancer.

(2 pts) 14. What kind of sampling method was used in this study?

A) Naturalistic

B) Purposive Cohort

C) Case-Control

(2 pts) 15. What is the Odds Ratio describing the relationship between estrogen use and cervical cancer?

(3 pts) 16. Report a 90% confidence interval for the OR.

Use the following information to answer questions 17-18.

Consider the NIH-sponsored Women's Health Initiative experimental study that enrolled 16,608 postmenopausal women to investigate the relationship between estrogen exposure and the risk of adverse health events. About half of these women were randomly assigned to take estrogen ($n_1 = 8506$), while the remaining subjects received a placebo ($n_2 = 8102$). These women were followed for 5 years and the number of adverse health events were tabulated. Women from the treatment group experienced 751 adverse health events, while women from the control group experienced 623 adverse health events.

_____ (2 pts) 17. What kind of sampling method was used in this study?

A) Naturalistic

B) Purposive Cohort

C) Case-Control

(2 pts) 18. What is the Relative Risk of experiencing an adverse health event for this data?

_____ (3 pts) 19. 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 95% 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

_____ (3 pts) 20. If we prefer a 90% confidence interval instead, what would happen to the required sample size n?

A) n would decrease.

B) n would increase.

C) n would not change.

Use the following information to answer questions 21-25.

Cytomegalovirus (CMV) is a type of herpes virus commonly found in adults in the U.S. that often lies dormant in the body over long periods. Recent studies shown that CMV may be related to high blood pressure and may be a major cause of atherosclerosis and stenosis.

Stenosis refers to the narrowing of blood vessels, resulting in restricted blood flow, which can turn into a dangerous medical condition. One common cause is the buildup of plaque deposits along the walls of the blood vessel. This condition can be treated by an **atherectomy** procedure to remove the plaque burden in the vessel.

While this atherectomy procedure is generally thought to be effective long-term, it can occasionally lead to **Coronary Restenosis**, which is the renarrowing of the blood vessel after receiving treatment to clear the blockage. We would like to investigate whether or not CMV infection is associated with Coronary Restenosis.

The table below displays the observed number of patients experiencing Coronary Restenosis within 6 months of atherectomy, classified according to CMV infection status.

	Experienced Restenosis	No Restenosis	Total
CMV+	21	28	49
CMV-	2	24	26
Total	23	52	75

(2 pts) 21. Consider the proportions of Restenosis in the CMV+ and CMV- populations. What is the estimated difference between the two population proportions of Restenosis between the two groups?

(2 pts) 22. We would like to perform the χ^2 test of general association. What are the null and alternative hypotheses for this test?

_____ (1 pt) 23. We calculate the χ^2 test statistic to be $X^2_{\text{stat}} = 9.879$. What are the degrees of freedom associated with this test statistic?

A) 1

B) 2

C) 4

D) 74

(2 pts) 24. What can we say about the p-value associated with the test statistic from #23?

A) $0.001 < p < 0.005$

B) $0.002 < p < 0.01$

C) $0.005 < p < 0.01$

D) $p > 0.01$

(3 pts) 25. At the $\alpha = 0.05$ level of significance, what is our decision about H_0 and conclusion in context of the original populations?

A) Reject H_0 and conclude that CMV infection is independent of risk of Coronary Restenosis.

B) Reject H_0 and conclude that CMV infection is somehow associated with the risk of Coronary Restenosis.

C) Reject H_0 and conclude that CMV infection has a significant linear relationship with the risk of Coronary Restenosis.

D) Reject H_0 and conclude that the risk of Coronary Restenosis is significantly higher for people that do not have CMV infection than for people who do have CMV infection.

(1 pt) 26. What can we say about the direction of relationship between the risk of Coronary Restenosis between CMV+ and CMV- individuals?

A) $\hat{p}_{\text{CMV+}}$ is significantly greater than $\hat{p}_{\text{CMV-}}$

B) $\hat{p}_{\text{CMV+}}$ is significantly less than $\hat{p}_{\text{CMV-}}$

C) We can't say anything about the direction of the relationship between $\hat{p}_{\text{CMV+}}$ and $\hat{p}_{\text{CMV-}}$

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 understood that 32.8% have mothers who smoked during pregnancy. We would like to test whether the proportion of maternal smoking for children with an oral cleft is identical to the proportion 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 can we say about the p-value for this test?

(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.

_____ (2 pts) 35. Which type of sampling method is preferred in the study of a rare disease?

A) Naturalistic
B) Purposive Cohort
C) Case-Control

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

_____ (2 pts) 36. **True/False.** The error terms in MLR are normally distributed.

_____ (2 pts) 37. **True/False.** The variability of the residuals can change between observations.

Use the following information to answer questions 38-45

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 whether an expectant mother's 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. Therefore, we fit the following Multiple Linear Regression model using software. The output is displayed below.

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

where $x_1 =$ gestational age in weeks

and $x_2 =$ toxemia status $= \begin{cases} 1\ , & \text{if diagnosed with toxemia} \\ 0\ , & \text{if not diagnosed with toxemia} \end{cases}$

ANOVA Table

Source	df	SS	MS	F	p-value
Between	2	619.25	309.625	48.95	$< .0001$
Within	97	643.51	6.325		
Total	99	1262.76			

Parameter Estimate Table

Parameter	Estimate	Std Error	t Stat	p-value
Intercept	6.28	3.192	1.97	0.0517
GestAge	1.07	0.112	9.55	$< .0001$
Toxemia	-1.78	0.694	-2.56	0.0120

_____ (2 pts)

38. What percentage of the variation in infant length can be explained by the linear model with gestational age and toxemia?

A) 24.52%

B) 96.23%

C) 49.03%

D) 50.96%

_____ (2 pts) 39. What is the test statistic to test $H_0 : \beta_1 = \beta_2 = 0$ vs. H_A : at least one $\beta_i \neq 0$?

A) $F = 48.95$
B) $t = 1.97$
C) $t = 9.55$
D) $t = -2.56$

_____ (2 pts) 40. What is the test statistic to test whether Toxemia has a significant linear relationship with Length, after adjusting for Gestational Age?

A) $F = 48.95$
B) $t = 1.97$
C) $t = 9.55$
D) $t = -2.56$

_____ (3 pts) 41. Report the 90% confidence interval for β_1, the slope coefficient for gestational age.

A) 6.28 ± 5.3115
B) 1.07 ± 0.1864
C) -1.78 ± 1.1548
D) 1.07 ± 0.7072

_____ (2 pts) 42. **True/False.** Based on the fitted model, the infants of toxemic mothers are expected to be shorter on average than the infants of non-toxemic mothers.

(2 pts) 43. Based on the fitted model, what is the predicted length of infants born to toxemic mothers having gestational age $x_1 = 35$ weeks?

(3 pts) 44. Interpret the estimate of β_2, the slope coefficient for toxemia, in the context of the fitted model and the original population.

(3 pts) 45. **Bonus**: Suppose we remove the variable Gestational Age from our model, so that Toxemia is the only explanatory variable (i.e., Simple Linear Regression). If we then test whether the Toxemia term is a significant linear predictor of Length, this is equivalent to another test we learned about this semester. What is it?