This is a closed-book exam so do not refer to your notes, the text, or any other books (please put them on the floor). You may use a single sheet of notes or formulas and a calculator, but materials may not be shared. Normal $Z$, student $t$, Chi-Squared distribution, formula sheets, and four blank work sheets are attached to the exam.

You must show your work to get partial credit. Even correct answers will not receive full credit without justification. Please give all numerical answers to at least four correct digits or as exact fractions reduced to lowest terms. Write your solutions as clearly as possible and make sure it’s easy to find your answers (circle them if necessary), since you will not receive credit for work that I cannot understand or find. Good Luck!

If you find a question confusing please ask me, Huiyan or Floyd to clarify it.

Cheating on exams is a breach of trust with classmates and faculty, and will not be tolerated. After completing the exam please acknowledge the Duke Honor Code with your signature below:

I have neither given nor received unauthorized aid on this exam.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>/8</td>
</tr>
<tr>
<td>2.</td>
<td>/62</td>
</tr>
<tr>
<td>3.</td>
<td>/5</td>
</tr>
<tr>
<td>4.</td>
<td>/10</td>
</tr>
<tr>
<td>5.</td>
<td>/15</td>
</tr>
<tr>
<td>Total:</td>
<td>/100</td>
</tr>
</tbody>
</table>
Problem 1: True or false? Circle one and no explanations. (1pt each)

1. T  F If events $A$ and $B$ are independent, $P(A) = 0.3$ and $P(B) = 0.2$, then $A$ and $B$ are not mutually exclusive.

2. T  F Events $A$, $B$, $C$ are independent, if every pair is independent.

3. T  F For any two random variables $X$ and $Y$,

\[ E(X - Y) = E(X) - E(Y). \]

4. T  F For independent random variables $X$ and $Y$,

\[ \text{Var}(X - Y) = \text{Var}(X) - \text{Var}(Y). \]

5. T  F The rejection region is controlled by the $\alpha$ level and the alternate hypothesis.

6. T  F The p-value measures the probability that the hypothesis is true.

7. T  F If we keep the sample size fixed, the confidence interval gets wider as we increase the confidence coefficient.

8. T  F If the population std. deviation increases, the confidence interval decreases in width.
Problem 2: Circle the single correct answer. (2pt each)

1. The chances that you will be ticketed for illegal parking on campus are about 1/3. During the last nine days, you have illegally parked every day and have NOT been ticketed (you lucky person)! Today, on the 10th day, you again decide to park illegally. The chances that you will be caught are:
   a. greater than 1/3 since you were not caught in the last nine days.
   b. less than 1/3 since you were not caught in the last nine days.
   c. still equal to 1/3 since the last nine days do not affect the probability.
   d. equal to 1/10 since you were not caught in the last nine days.

2. A random variable $X$ has a probability distribution as follows:

<table>
<thead>
<tr>
<th>$X$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = x)$</td>
<td>$2k$</td>
<td>$3k$</td>
<td>$13k$</td>
<td>$2k$</td>
</tr>
</tbody>
</table>

Then the probability that $Pr(X < 2.0)$ is equal to
   a. 0.90  b. 0.25  c. 0.65  d. 0.15

3. Suppose that the allele for tallness (T) is dominant over shortness (t); that for Yellow (Y) is dominant over green (y); and that for roundness (W) is dominant over wrinkled(w). Suppose we cross two plants with genotypes TTYyWw and TtYyWw. The probability of a Tall, Yellow, Round plant is:
   a. $\frac{9}{16}$  b. $\frac{3}{32}$  c. $\frac{1}{16}$  d. $\frac{9}{32}$  e. $\frac{3}{16}$

4. If A and B are independent events with $P(A) = .6$ and $P(B) = .7$, then the probability that A occurs or B occurs or both occur is:
   a. 1.30  b. 0.88  c. 0.42  d. 0.10

5. Event A and B are disjoint. The probability that B will not happen is 0.7, and $P(A \text{ or } B)=0.5$. What is $P(A)$?
   a. 0.1  b. 0.2  c. 0.3  d. 0.4

6. Cans of soft drinks cost $.30 in a certain vending machine. What is the expected value and variance of daily revenue ($Y$) from the machine, if $X$, the number of cans sold per day has $E(X) = 125$, and $Var(X) = 50$ ?
   a. $E(Y) = 37.5$ $Var(Y) = 50$  b. $E(Y) = 37.5$ $Var(Y) = 4.5$
   c. $E(Y) = 37.5$ $Var(Y) = 15$  d. $E(Y) = 125$ $Var(Y) = 4.5$

7. Event A occurs with probability 0.4. The conditional probability that A occurs given that B occurs is 0.7, while the conditional probability that A occurs given that B does not occur is 0.2. What is the conditional probability that B occurs given that A occurs?
   a. 0  b. $\frac{4}{7}$  c. $\frac{5}{6}$  d. $\frac{7}{10}$
8. A college basketball player makes 80% of his free throws. At the end of a game, his team is losing by two points. He is fouled attempting a three-point shot and is awarded three free throws. Assuming each free throw is independent, what is the probability that he ties or wins the game by making some of the free throws?
   a. 0.896  b. 0.80  c. 0.64  d. 0.384

9. Which of the following is NOT an assumption of the Binomial distribution?
   a. All trials must be identical.
   b. All trials must be independent.
   c. Each trial must be classified as a success or a failure.
   d. The probability of success is equal to .5 in all trials.

10. The marks on a statistics test are normally distributed with a mean of 62 and a variance of 225. If the instructor wishes to assign A’s or higher to the top 30% of the students in the class, what mark is required to get a A or higher?
   a. 68.7  b. 71.5  c. 73.2  d. 74.6  e. 69.9

11. Suppose the test scores of 600 students are normally distributed with a mean of 76 and standard deviation of 8. The number of students scoring between 70 and 82 is:
   a. 272  b. 164  c. 260  d. 136  e. 328

12. The weight of medium-size tomatoes selected at random from a bin at the local supermarket is a random variable with mean $\mu = 10$ oz. and standard deviation $\sigma = 1$ oz. The weight of the tomatoes in pounds (1 pound = 16 oz.) is a random variable with standard deviation (in pounds)
   a. $\frac{1}{16}$  b. 1  c. 16  d. 256

13. The diameter of ball bearings are known to be normally distributed with unknown mean and variance. A random sample of size 25 gave a mean 2.5 cm. The 95% confidence interval (based on $t$ distribution) had length 4 cm. Then
   a. The sample variance is 4.84.
   b. The sample variance is 23.47.
   c. The sample variance is 26.03.
   d. The population variance is 4.84.
   e. The population variance is 26.03.

14. Researchers are studying yield of a crop in two locations. The researchers are going to compute independent 90% confidence intervals for the mean yield at each location. The probability that at least one of the intervals will cover the true mean yields at their locations is
   a. 0.81  b. 0.19  c. 0.99  d. 0.95
15. Suppose that the population of the scores of all high school seniors that took the SAT-M (SAT math) test this year follows a normal distribution with mean $\mu$ and standard deviation $s = 100$. You read a report that says, “On the basis of a simple random sample of 100 high school seniors that took the SAT-M test this year, a confidence interval for $\mu$ is 512.00 ± 25.76.” The confidence level for this interval is

a. 90%  b. 95%  c. 99%  d. more than 99.9%

16. In hypothesis testing, $\beta$ is the probability of committing an error of Type II. Then $1 - \beta$ is:

a. the probability of rejecting $H_0$ when $H_a$ is true
b. the probability of failing to reject $H_0$ when $H_a$ is true
c. the probability of failing to reject $H_0$ when $H_0$ is true
d. the probability of rejecting $H_0$ when $H_0$ is true

17. A certain population follows a normal distribution with mean $\mu$ and standard deviation $\sigma = 2.5$. You collect data and test the hypotheses

$$H_0 : \mu = 1, \quad H_a : \mu \neq 1.$$ 

You obtain a P-value of 0.022. Which of the following is true?

a. A 95% confidence interval for $\mu$ will include the value 1.
b. A 95% confidence interval for $\mu$ will include the value 0.
c. A 99% confidence interval for $\mu$ will include the value 1.
d. A 99% confidence interval for $\mu$ will include the value 0.
e. None of the above.

18. We wish to test if a new feed increases the mean weight gain compared to an old feed. At the conclusion of the experiment it was found that the new feed gave a 10 kg bigger gain than the old feed. A two-sample t-test with the proper one-sided alternative was done and the resulting p-value was .082. This means:

a. There is an 8.2% chance the null hypothesis is true.
b. There was only a 8.2% chance of observing an increase greater than 10 kg (assuming the null hypothesis was true).
c. There was only an 8.2% chance of observing an increase greater than 10 kg (assuming the null hypothesis was false).
d. There is an 8.2% chance the alternate hypothesis is true.
19. If the 95% confidence interval for a population mean \( \mu \) is (-1,1) and you are testing \( H_0: \mu = 1.3 \) against \( H_a: \mu \neq 1.3 \),
   a. you fail to reject \( H_0 \) at 1% level.
   b. you fail to reject \( H_0 \) at 5% level.
   c. you reject \( H_0 \) at 5% level.
   d. you reject \( H_0 \) at 1% level.
   e. There is insufficient information to make a decision.

20. A two sample t-test was performed using independent samples from two populations to test the null hypothesis \( \mu_1 - \mu_2 = 0 \). What conclusion can be made if a very small P-value (< 0.001) is obtained?
   a. We have evidence that \( \bar{X}_1 \) and \( \bar{X}_2 \) are the same.
   b. We have evidence that \( \mu_1 \) and \( \mu_2 \) are the same.
   c. We have evidence that \( \bar{X}_1 \) and \( \bar{X}_2 \) are not the same.
   d. We have evidence that \( \mu_1 \) and \( \mu_2 \) are not the same.

21. The infamous researcher, Dr. Gnirips, claims to have found a drug that causes people to grow taller. The coach of the Basketball team at Brandon University has expressed interest but demands evidence. Ten people are randomly selected from students at Brandon, their heights measured, the drug administered, and 2 hours later their heights remeasured. The results were as follows:

<table>
<thead>
<tr>
<th>Person</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Drug</td>
<td>68</td>
<td>69</td>
<td>74</td>
<td>78</td>
<td>70</td>
<td>66</td>
<td>71</td>
<td>70</td>
<td>71</td>
<td>65</td>
</tr>
<tr>
<td>Post-Drug</td>
<td>70</td>
<td>69</td>
<td>75</td>
<td>78</td>
<td>73</td>
<td>69</td>
<td>72</td>
<td>73</td>
<td>72</td>
<td>66</td>
</tr>
</tbody>
</table>

Making the appropriate normal assumptions, the decision rule for the hypotheses \( H_0: \) Drug has no effect versus \( H_a: \) Drug increases height at \( \alpha = .05 \) will be
   a. Reject \( H_0 \) if the test statistic is > 1.96
   b. Reject \( H_0 \) if the test statistic is > 1.645
   c. Reject \( H_0 \) if the test statistic is > 1.83
   d. Reject \( H_0 \) if the test statistic is > 1.73
   e. Reject \( H_0 \) if the test statistic is > 2.10
22. Suppose that 100 persons were selected at random in a certain city, and that each person was asked whether she/he thought the service provided by the fire department in the city was satisfactory. Shortly after this survey was carried out, a large fire occurred in the city. Suppose that after this fire, the same 100 persons were again asked whether they thought that the service provided by the fire department was satisfactory. The results are presented below:

<table>
<thead>
<tr>
<th></th>
<th>Satisfactory</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before the fire</td>
<td>80</td>
<td>20</td>
</tr>
<tr>
<td>After the fire</td>
<td>72</td>
<td>28</td>
</tr>
</tbody>
</table>

To test whether people's opinions are changed by the fire,

a. We should do a $\chi^2$ test for homogeneity.
b. We should do a $\chi^2$ test for independence.
c. We should not use a $\chi^2$ test.

23. A statistics professor would like to determine whether students in his class showed improved performance on the final examination as compared to the mid-term examination. A random sample of 4 students selected from a large class revealed the following mid-term and final scores:

<table>
<thead>
<tr>
<th>Student #</th>
<th>Mid-term</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>#2</td>
<td>62</td>
<td>79</td>
</tr>
<tr>
<td>#3</td>
<td>57</td>
<td>87</td>
</tr>
<tr>
<td>#4</td>
<td>68</td>
<td>88</td>
</tr>
</tbody>
</table>

Making the appropriate normal assumptions, the value of the test statistic is:

a. 19.25/8.30
b. 19.25/(8.30/2)
c. 19.25/\sqrt{(28.295/4 + 28.295/4)}
d. 19.25/\sqrt{(34.92/4 + 21.67/4)}

24. The fraction of the variation in the values of $y$ that is explained by the least-squares regression of $y$ on $x$ is

a. the correlation coefficient.
b. the slope of the least-squares regression line.
c. the square of the correlation coefficient.
d. the intercept of the least-squares regression line.
The next three questions refer to the following situation: The Excellent Drug Company claims its aspirin tablets will relieve headaches faster than any other aspirin on the market. To determine whether Excellent’s claim is valid, random samples of size 15 are chosen from aspirins made by Excellent and the Simple Drug Company. An aspirin is given to each of the 30 randomly selected persons suffering from headaches and the number of minutes required for each to recover from the headache is recorded. The sample results are:

<table>
<thead>
<tr>
<th></th>
<th>$X$</th>
<th>$s^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent (E)</td>
<td>8.4</td>
<td>4.2</td>
</tr>
<tr>
<td>Simple (S)</td>
<td>8.9</td>
<td>4.6</td>
</tr>
</tbody>
</table>

A 5% significance level of two-sample z-Test is performed to determine whether Excellent’s aspirin cures headaches significantly faster than Simple’s aspirin.

25. The appropriate hypothesis to be tested is:
   a. $H_0 : \mu_E - \mu_S = 0$  $H_a : \mu_E - \mu_S > 0$
   b. $H_0 : \mu_E - \mu_S = 0$  $H_a : \mu_E - \mu_S < 0$
   c. $H_0 : \mu_E - \mu_S = 0$  $H_a : \mu_E - \mu_S \neq 0$

26. Absolute value of the calculated value of the appropriate test statistic is:
   a. 1.61  b. 2.33  c. 0.65  d. 1.24  e. 0.85

27. Absolute value of the critical value for this test is:
   a. 1.960  b. 1.701  c. 2.048  d. 2.145  e. 1.645

The next set of questions refer to the following situation: A survey was conducted to investigate the severity of rodent problems in egg and poultry operations. A random sample of operators was selected, and the operators were classified according to the type of operation and the extent of the rodent population. A total of 78 egg operators and 53 turkey operators were classified and the summary information is:

<table>
<thead>
<tr>
<th>Count</th>
<th>egg</th>
<th>turkey</th>
</tr>
</thead>
<tbody>
<tr>
<td>mild</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>moderate</td>
<td>37</td>
<td>22</td>
</tr>
<tr>
<td>server</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>78</td>
<td>53</td>
</tr>
</tbody>
</table>

The corresponding $\chi^2$ test statistic is equal to 5.635.

28. The expected count in the (egg, mild infestation) cell is:
   a. 26.00  b. 33.33  c. 53.00  d. 31.55  e. 78.00
29. The approximate p-value is found to be:
   a. about .060  b. about .014  c. about .032  d. about .05

30. Which of the following is NOT correct?
   a. Operators who had both operations could not be used because this type of analysis requires each unit to be counted in one and only one cell.
   b. The null hypothesis is that the severity of the rodent problem is independent of the type of operator.
   c. The alternate hypothesis is that the proportion of turkey operators with mild, moderate, and severe rodent problems is different from the proportion of egg operators with mild, moderate, and severe rodent problems.
   d. A Type I error would be to conclude that the severity of rodent problems is dependent upon the type of operator while, in fact, the proportion of turkey operators with mild, moderate, and severe rodent problems is the same as the proportion of egg operators with mild, moderate, and severe rodent problems.
   e. A Type II error would be to conclude that the proportion of egg operators with mild, moderate, or severe rodent problems is the same as the proportion of turkey operators with mild, moderate, or severe rodent problems when in fact they are independent.

31. One reviewer of the study suggested that there may be a problem with the study because results from small operators were pooled with the results from large operators. Which of the following is NOT correct?
   1. Simpson’s paradox occurs when conclusions from a pooled table differ from the individual tables.
   2. Tables can be pooled when the underlying rates are equal among tables.
   3. Simpson’s paradox occurs when tables with unequal row totals are pooled.
   4. Inspection of the row or column percents will give a good clue if Simpson’s paradox is likely to occur.
**Problem 3:** Consider a system consisting of two components as in the figure below.

![Diagram of a system with two components](image)

The system works if either 1 or 2 works. Suppose the lifetime of each component, measured in hours, is a random variable $X_i$, $i = 1, 2$, with density function

$$f_{X_i}(x) = \begin{cases} \lambda_i e^{-\lambda_i x}, & \text{if } x \geq 0; \\ 0, & \text{if } x < 0. \end{cases}$$

where $\lambda_1 = 2/3$ and $\lambda_2 = 1/2$. Let $X$ denote the life time of the whole system (measured in hours). What is the density function of $X$? (5pt)
**Problem 4:** An investigator wishes to estimate the proportion of students at Duke University who have violated the honor code. A random sample of 100 Duke students is obtained. Instead of asking each, “Have you violated the honor code?”, the investigator makes up a deck of 100 cards, of which 50 are of type I and 50 are type II.

Type I: Have you violated the honor code (yes or no)?
Type II: Are you a Duke student? (yes or no)?

Each student in the random sample is asked to mix the deck, draw a card, and answer the resulting question truthfully. Let \( p \) denote the proportion of honor-code violators and let \( Y \) denote the number of yes responses.

Find an unbiased estimator for \( p \), that is, find a function of \( Y \) as your estimator for \( p \) and show that your estimator is unbiased. (10pt)

(*Hint:* let \( \lambda = P(\text{yes response}) \), so \( Y \sim Bin(100, \lambda) \). Thus \( Y/100 \) is an unbiased estimator of \( \lambda \). What is the relationship between \( \lambda \) and \( p \)?)
Problem 5: How does lateral acceleration - side forces experienced in turns that are largely under driver control - affect nausea as perceived by bus passengers? The article "Motion Sickness in Public Road Transport: The Effect of driver, Route, and Vehicle" (*Ergonomics*, 1999: 1646-1664) reported data on $x =$ motion sickness does (calculated in accordance with a British standard for evaluation similar motion at sea) and $y =$ reported nausea (%). Relevant summary quantities are

$$n = 17, \sum x_i = 222.1, \sum y_i = 193, \sum x_i^2 = 3056.69$$

a. Assume that the simple linear regression model is valid for relating these two variables, that is, we have in mind a model of the form: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for unknown parameters $\beta_0$ and $\beta_1$. The least square estimates $\hat{\beta}_0, \hat{\beta}_1$ are those values that “fit” the data the best. How are the estimates $\hat{\beta}_0, \hat{\beta}_1$ determined? That is, exactly what measure of “goodness of fit” do they optimize? (5pt)

b. Suppose $\hat{\beta}_0 = -8.7144$. Then $\hat{\beta}_1 =$? (5pt) (*Hint: if you know a point the regression line passes, you will be able to find the value of $\hat{\beta}_1$.*)

c. Suppose $SSE = \sum (y_i - \hat{y}_i)^2 = 418.2494$. What is the length of the 95% confidence interval for $\beta_1$? (5pt)
(Nearly) Blank Work-Sheet #1 (of 4)
(Nearly) Blank Work-Sheet #2 (of 4)
(Nearly) Blank Work-Sheet #3 (of 4)
(Nearly) Blank Work-Sheet #4 (of 4)