PLEASE PRINT THE FOLLOWING INFORMATION:

Name: _____   Instructor: _____

Student ID #: _____   Section/Time: _____

THIS EXAM HAS TWO PARTS.

PART I.

Part I consists of 30 multiple choice questions. Each correct answer is scored 2 points; each incorrect (or blank) answer is scored 0, so there is no penalty for guessing. You may do calculations on the test paper, but your answers must be marked on the OPSCAN sheet with a soft lead pencil (HB or No. 2 lead). Any question with more than one choice marked will be counted as incorrect. If more than one choice seems correct, chose the one that is most complete or most accurate. Make sure that your name and ID number are written and correctly bubbled on the OPSCAN sheet.

PART II.

Part II consists of 3 free response questions, with values as indicated. You must show all work in the space provided or elsewhere on the exam paper in a place that you clearly indicate. Work on loose sheets will not be graded.

FOR DEPARTMENT USE ONLY:

Part II.

Question	1	2	3
Score			

Part I	Part II	TOTAL

Part I

Problems 1 and 2 pertain to the following sample data:

$$4, \quad 2, \quad 0, \quad 4, \quad -1, \quad 1, \quad -2, \quad 0, \quad -3$$

1. The sample mean of this data set is about

    (a) 1.9   (b) 0.6   (c) 0.7   (d) 1.0   (e) 1.4

2. The sample standard deviation of this data set is about

    (a) 4.2   (b) 6.3   (c) 2.3   (d) 5.3   (e) 2.5

Problems 3 and 4 pertain to the data set of 75 measurements represented by the following stem and leaf diagram:

```
6 | 0 0 0 1 3 6 7 8
5 | 0 0 0 1 1 1 2 2 4 5 5 5 7 8 8 9 9
4 | 0 0 0 1 1 1 2 2 3 3 3 4 4 5 5 6 7 7 7 8 8 9 9
3 | 0 1 1 2 2 2 3 3 4 6 6 7 7 8 9
2 | 2 2 4 5 6 7 7 8
1 | 7 8 8 9
```

3. The sample median is about

    (a) 44   (b) 38   (c) 43.5   (d) 43   (e) 46

4. The percentile rank of the measurement 27 is about

    (a) 9   (b) 15   (c) 27   (d) 18   (e) 12

---

5. The distribution of the lengths of pregnancies in human beings is roughly bell-shaped with mean 268 days and standard deviation 7 days. The proportion of pregnancies that last at least one week, i.e., seven days, longer than the average is about

    (a) 0.16   (b) 0.68   (c) 0.84   (d) 0.50   (e) 0.32

---

6. Francesca is allergic to food additive A and to food additive B. If 22% of all restaurant dishes contain additive A, 17% contain additive B, and 8% contain both, then the probability that a randomly selected dish will contain at least one of the two additives, hence be inedible for Francesca, is about

    (a) 0.39   (b) 0.04   (c) 0.47   (d) 0.31   (e) 0.13

Problems 7–9 pertain to the information in the following two-way classification table, based on a random sample of 420 adults between 25 and 65 years of age. Assume that the table accurately reflects the composition of the whole population of such adults.

Highest level of Education	Single	Married	Divorced	Widow(er)	Total
Less than HS diploma	13	48	11	17	89
HS diploma or GDE	50	119	40	17	226
Bachelor's degree or higher	21	70	11	3	105
Total	84	237	62	37	420

7. The probability that a randomly selected adult between 25 and 65 is single is about

   (a) 0.29   (b) 0.38   (c) 0.84   (d) 0.03   (e) 0.20

8. The probability that a randomly selected adult between 25 and 65 is single, given that he has at least a bachelor's degree, is about

   (a) 0.24   (b) 0.20   (c) 0.04   (d) 0.38   (e) 0.12

9. The events S: the person selected is single and B: the person selected has at least a bachelor's degree are:

   (a) independent because $P(S \cap B) = P(S) \cdot P(B)$.
   (b) independent because $P(S \cup B) = P(S) + P(B)$.
   (c) independent because $P(S|B) = P(B)$.
   (d) dependent because $P(S \cap B) = P(S) \cdot P(B)$.
   (e) dependent because $P(S \cup B) \neq P(S) + P(B)$.

Problems 10 and 11 pertain to the probability distribution of the number X of times a randomly selected student visited the student health center in a recent semester.

x	0	1	2	3	4
$P(x)$	0.64	0.23	0.09	0.02	0.02

10. The probability that a randomly selected student visited the health center at least once is about:

    (a) 0.36   (b) 0.26   (c) 0.39   (d) 0.09   (e) 0.13

11. The average number of visits per student per semester is about:

    (a) 2.50   (b) 2.00   (c) 0.50   (d) 0.55   (e) 1.27

12. The yearly economics forecast by a university professor significantly underestimates the next year's economic growth with probability 0.20. The yearly economics forecast by an economist at an investment firm, made independently using different methods, significantly underestimates the next year's economic growth with probability 0.15. The probability that both forecasts will significantly underestimate the next year's economic growth is about

   (a) 0.05   (b) 0.35   (c) 0.32   (d) 0.03   (e) impossible to tell from the information given

13. Forty-three percent of adults in the U. S. know the number of acres in a square mile. In a trivia contest a team with four members is asked how many acres are in a square mile. The probability that none of them will know the answer is:

   (a) 0.43   (b) 0.28   (c) 0.03   (d) 0.11   (e) 0.00 (zero to two decimal places)

14. For the standard normal variable Z, $P(-0.83 < Z < 4.11)$ is about

   (a) 0.2033   (b) 0.7033   (c) 0.7967   (d) 0.5211   (e) impossible to tell

15. The temperature at 6:00 pm for days in May in the town in which a couple are planning an outdoor wedding at 6:00 pm is normally distributed with mean 70° F and standard deviation 2° F. The probability that the temperature at 6:00 pm on the day they have picked for their wedding will be between 66° F and 76° F is about:

   (a) 0.872   (b) 0.684   (c) 0.999   (d) 0.943   (e) 0.976

16. A company claims that 67% of adults can distinguish its bottled water from tap water. In a test of 250 randomly selected adults, 153 could do so. The probability of obtaining a sample proportion as low as this or lower, if the company's claim is true, is about:

   (a) 0.22   (b) 0.03   (c) 0.09   (d) 0.01   (e) 0.16

17. Household incomes in a certain region are skewed right with mean \$57,550 and standard deviation \$16,200. The probability that the mean of a random sample of 4400 household incomes would differ from the population mean by \$500 or more is about:

   (a) 0.12   (b) 0.09   (c) 0.04   (d) 0.21   (e) 0.18

18. The mean floor to knee height of adult American males is normally distributed with mean 23.2 inches and standard deviation 0.9 inch. If an engineer wishes to design an adjustable chair that will accommodate the middle 96% of all mean, then the minimum middle range of floor to knee heights that the chair must fit is

   (a) [22.3, 24.1]   (b) [20.0, 25.4]   (c) [21.6, 24.8]   (d) [21.7, 24.7]   (e) [21.4, 25.0]

19. In a sample of 25 three year old children whose mothers took a certain drug for epilepsy during pregnancy, the mean IQ score was 92 with standard deviation 2.25. Given that the population of IQ scores of all three year old children whose mothers took this drug during pregnancy is normally distributed, a 95% confidence interval for its mean is about:

   (a) $92 \pm 2.060 \frac{2.25}{\sqrt{25}}$ (b) $92 \pm 1.960 \frac{2.25}{\sqrt{25}}$ (c) $92 \pm 2.064 \frac{2.25}{\sqrt{25}}$ (d) $92 \pm 1.645 \frac{2.25}{\sqrt{25}}$

   (e) $92 \pm 1.711 \frac{2.25}{\sqrt{25}}$

20. In a random sample of 325 former prisoners in North Carolina, 29% were arrested again within one year. A 90% confidence interval for the proportion of all North Carolina prisoners who are arrested again within one year is about:

   (a) 0.29 ± 0.05 (b) 0.29 ± 0.06 (c) 0.29 ± 0.03 (d) 0.29 ± 0.07 (e) 0.29 ± 0.04

21. In the context of the previous problem, suppose that the confidence interval obtained is too wide. Making use of the information provided in that problem, the minimum sample size required in order to estimate the same population proportion to within two percentage points, still at 90% confidence, is about:

   (a) 684   (b) 1393   (c) 481   (d) 894   (e) 1108

22. A sample of the ages of 100 owners of small businesses had mean 43.6 years and standard deviation 4.5 years. A 95% confidence interval for the mean age of all owners of small businesses is about:

    (a) $[42.8, 44.4]$ (b) $[43.0, 44.2]$ (c) $[42.7, 44.5]$ (d) $[42.9, 44.3]$ (e) $[42.6, 44.6]$

Problems 23 and 24 pertain to the following information. In order to test whether the mean age of first onset of anorexia nervosa in girls has declined from the historic value of 15 years a researcher examines the records of 20 girls with the condition and obtains mean age at onset of 14.4 years with standard deviation 1.6 years. He will perform the test $H_0 : \mu = 15$ vs. $H_a : \mu < 15$ at the 5% level of significance.

23. The value of the test statistic is about:

    (a) 1.677   (b) -1.635 (c) -1.729 (d) 1.729   (e) -1.677

24. The rejection region is

    (a) $[1.729, \infty)$ (b) $(-\infty, -1.677]$ (c) $(-\infty, -0.050]$ (d) $(-\infty, -1.729]$ (e) $[1.725, \infty)$

---

25. In a test of hypotheses $H_0 : p = 0.42$ versus $H_a : p \neq 0.42$ at the 1% level of significance a sample of size 1250 produced $\hat{p} = 0.39$ and test statistic $z = -2.149$. The p-value (the observed significance) of the test is about:

    (a) 0.39   (b) 0.03   (c) -2.15 (d) -0.03 (e) 0.42

---

26. Knowing that drowsiness on the job significantly affects productivity, the management of a large manufacturing firm contemplates instituting an expensive program to encourage employees to get sufficient sleep at night. To test the efficacy of the program before a full-scale launch, they count the mean number of quality control rejections per day of manufactured goods on 60 randomly selected days before the inception of a pilot program, then count the mean number of quality control rejections per day of goods on 30 days after its inception. The appropriate formula for the test statistic in the relevant test of hypotheses is:

    (a) $T = \dfrac{\bar{d}}{s_d/\sqrt{n}}$ (b) $Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ (c) $T = \dfrac{(\bar{x}_1 - \bar{x}_2) - D_0}{\sqrt{s_p^2(\frac{1}{n_1} + \frac{1}{n_2})}}$, $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

    (d) $Z = \dfrac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}}$ (e) $T = \dfrac{\hat{\beta}_1 - B_0}{s_\varepsilon/\sqrt{SS_{xx}}}$

27. A corporation's internet provider is contracted to provide an internet connection with a mean speed of 55 megabits per second. The corporation's IT department periodically tests the mean speed to verify that it is not less than 55 megabits per second. The mean and standard deviation of the timing of 30 packets will be computed. The setup of the null and alternative hypotheses for the relevant test of hypotheses is:

   (a) $H_0 : \bar{x} = 55$ vs. $H_a : \bar{x} < 55$

   (b) $H_0 : \mu = 55$ vs. $H_a : \mu > 55$

   (c) $H_0 : \mu = 55$ vs. $H_a : \mu \neq 55$

   (d) $H_0 : \mu = 55$ vs. $H_a : \mu < 55$

   (e) $H_0 : z = 55$ vs. $H_a : z < 55$

Problems 28–30 pertain to the following information. The health department of a combined city and county government seeks to determine the relationship between an air pollution index (on a scale of 0 to 10 from best to worst) and the number of admissions to area hospitals for acute respiratory distress. It collects data from 75 randomly selected days (the air pollution index x and the number of admissions y for each day) and computes:

$$3.0 \leqslant x \leqslant 8.2 \qquad 27 \leqslant y \leqslant 102 \qquad \hat{y} = 13.6x - 16.2$$

$$r = 0.884 \qquad r^2 = 0.782 \qquad s_\varepsilon = 10.409$$

28. For each unit increase in the air pollution index, on average the number of admissions for respiratory distress

   (a) increases by about 0.88

   (b) increases by about 13.6

   (c) increases by about 10.4

   (d) decreases by about 16.2

   (e) changes by an indeterminate amount

29. For days on which the air pollution index is 5.5, the average number of admissions for respiratory distress is predicted to be about:

   (a) 58.6   (b) 91   (c) 74.8   (d) 10.4   (e) 31.2

30. The coefficient of determination, which gives proportion of the variability in the number of admissions for respiratory distress that is accounted for by the level of air pollution, is about:

   (a) 0.10   (b) 0.88   (c) 0.78   (d) 0.14   (e) 0.93

Part II

1. To test whether the mean resting pulse rates of men and women differ, the pulse rates of random samples of men and women were measured, with the following results:

$$\text{men} \quad n_1 = 360 \quad \bar{x}_1 = 73.5 \quad s_1 = 13.0$$
$$\text{women} \quad n_2 = 380 \quad \bar{x}_2 = 76.3 \quad s_2 = 12.8$$

The test is performed at the 1% level of significance.

(a) State the null and alternative hypotheses for the test. [2 points]

(b) State the formula for the test statistic and compute its value. Justify your answer. [4 points]

(c) Construct the rejection region and make a decision. [4 points]

(d) State a conclusion about the mean resting pulse rates of men and women, based on the test you performed. [2 points]

2. Two different treatments for chronic lymphocytic leukemia were independently given to randomly selected individuals with the disease, with the following results. The sample proportions are the proportion of patients in the group who survived one year.

$$\text{Treatment } A \quad n_1 = 110 \quad \widehat{p}_1 = 0.92$$
$$\text{Treatment } B \quad n_2 = 110 \quad \widehat{p}_2 = 0.80$$

Test whether the proportion of one-year survivors in the population all sufferers given treatment A is greater than the corresponding proportion in the population of all sufferers given treatment B, against the default that the proportions are the same. Test at the 1% level of significance in the following series of steps. (The sample are sufficiently large for the procedure to be valid.)

(a) State the null and alternative hypotheses for the test. [2 points]

(b) State the formula for the test statistic and compute its value. Justify your answer. [4 points]

(c) Construct the rejection region and make a decision. [4 points]

(d) State a conclusion in the context of the problem, based on the test you performed. [2 points]

(e) Compute the p-value (the observed significance) of the test and state what it means in the context of this problem. [2 points]

3. In an attempt to relate stride length to height, a police department measures the length of stride x and the height y (both in inches) of 90 men and women. The scatter diagram showed a linear trend. Summary information is:

$$23.2 \leqslant x \leqslant 31.7 \qquad 60.6 \leqslant y \leqslant 72.4 \qquad \bar{x} = 26.5 \qquad \bar{y} = 66.3$$

$$SS_{xx} = 433.6 \qquad SS_{xy} = 526.5 \qquad SS_{yy} = 1470.9 \qquad s_\varepsilon = 3.276$$

(a) Find the regression line for predicting y from x. [4 points]

(b) Use the answer to (a) to predict the height of the person whose footprints at a crime scene were spaced 25.7 inches apart. [2 points]

(c) Construct a 90% confidence or prediction interval, whichever is appropriate, for the height of the person who left the footprints described in part (b). [4 points]

(d) Find the linear correlation coefficient for stride length and height. [2 points]

(e) Heights of men are normally distributed with mean 70.5 and standard deviation 1.5. Using this fact and your answer to (c) (with your answer to (d) if you like) comment on how useful the stride length of the footprints are in identifying the height of the person who left them. [2 points]