Name: _____

Computational Public Health Statistics Midterm, S04

- ✎ Hand in your completed **Lab Workbook** at the beginning of the exam and pick it up at the end.
- ✎ Write all answers in your **blue book**, **numbering** each response.
- ✎ **Time limit**: 2 hours.
- ✎ **Part A** is Close Book. **Part B** is Procedure Book. When you are done with Part A, turn in *this page* and pick up your Procedure Notebook. Please **do not change answers in Part A** after starting Part B.

Ch 11

1. Provide a synonym for "variance."
2. What is the square root of the variance called?
3. Provide two visual clues of distributional spread seen on a boxplot.
4. What *symbol* is used to denote the *population variance*?
5. Using statistical notation, provide the two-sided *alternative* hypotheses for the independent *t* test.

Ch 12

6. This statistic quantifies the variability of the group means around the grand mean. Choices: (a) F_{stat} (b) df_W (c) df_B (d) s^2_W (e) s^2_B
7. What symbol is used to denote the mean of population *i*?
8. What symbol is used to denote the (sample) grand mean?
9. Write the null tested by ANOVA (for 3 groups).
10. In an ANOVA comparing 4 groups of 8 people each (32 people total), df_B = ____ and df_w = _____.

Ch 13

11. What statistical test can be used to assess whether the variances of 2 *or more* groups differ significantly?
12. What statistical test is the non-parametric analogue of ANOVA?
13. Other than "independence," identify a *distributional* assumptions for ANOVA.
14. Identify a *validity* assumption.

Ch 14

15. This is the value of *r* when all points fall directly on a line with a downward slope.
16. What symbol is used to denote the correlation coefficient in the population? [Provide the Greek symbol, or if you are unable to properly draw the symbol, its name.]
17. Which correlation coefficient indicates the strongest correlation? Choices: (a) .43 (b) .22 (c) −.08 (d) −.22 (e) −.44
18. Other than checking for outliers, why is it important to scrutinize the scatter plot before calculating *r*?
19. True or false: If *r* is zero, you can be certain that X and Y are not related [in any way].

Ch 15

20. What does the slope of a simple regression model indicate?
21. What symbol is used to denote the slope *parameter*?
22. What is "squared" in a least squared regression line?
23. Provide one distributional assumptions needed for inference concerning regression analysis.

Name: _____

Part B (Procedure Notebook)

1. An $F_{stat} = 3.97$ with 2 and 20 degrees of freedom. Draw an *F* curve for the problem and shade the area corresponding to the *p* value region. Then, provide the 95th percentile landmark for the problem and determine whether *p* is greater or less than .05. [3 pts]

2. A $t_{stat} = 2.36$ with 25 degrees of freedom. Draw the *t* curve for this problem and shade the areas corresponding to the two-sided *p* value. Then, provide a landmark on the curve to determine the approximate *p* value. Is *p* less than .05? Is it less than .01? [3 pts]

3. **MIQ:** [10 pts]. Intelligence quotients in 6 monkeys are:

   88 100 102 106 106 116

   (A) Determine the 5 point summary for these data.
   (B) Determine the IQR.
   (C) Identify outside values, if any.
   (D) Identify the value of top whisker.
   (E) Identify the value of bottom whisker.

4. **MONKEY2:** Summary statistics for two independent groups are shown below. Use an *F* ratio statistic to test whether the variances differ significantly. Let $\alpha = .05$. Show all calculations and hypothesis testing steps, including statements of H_0 and H_1. Briefly discuss your results. [9 pts]

GROUP	*n*	Mean	Std. Deviation
1	6	2.500	3.521
2	6	2.000	1.871

– Exam continues on next page –

Name: _____

5. **MONKEY4:** A total of 12 monkeys are assigned four different interventions. (For now let us call the interventions, intervention 1, intervention 2, and so on.) We want to conduct an ANOVA to see whether effects of treatments differ. To save time I've calculated summary statistics (table below), and sums of squares: SS_B = 170.917 and SS_B = 40.000. (Assume ANOVA assumptions have been met.) [12 pts]

TABLE: Summary statistics for problem 5.

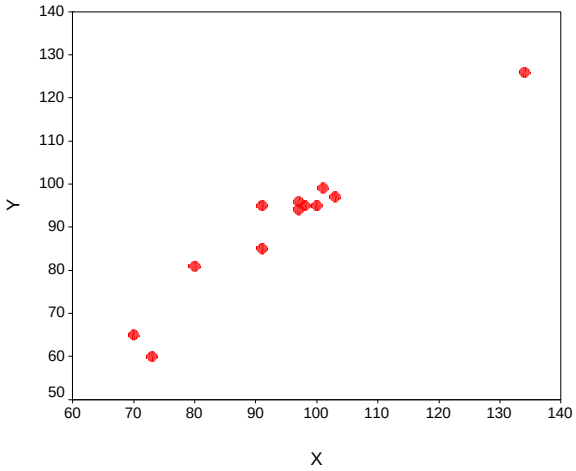
Intervention	*n*	Mean change	Std. Deviation
1	3	−1.33	2.517
2	3	9.00	3.606
3	3	5.33	0.577
4	3	2.67	0.577
Total	12	3.92	4.379

(A) List H_0 and H_1.
(B) df_B =
(C) df_W =
(D) Mean Square Between (s^2_B) =
(E) Mean Square Between (s^2_W) =
(F) F_{stat} =
(G) The *p* value for the ANOVA is .003. Do groups differ significantly?
(H) Conduct a post hoc test comparing group 1 and 2 (H_0: $\mu_1 = \mu_2$). In calculating results, show calculation of the standard error of the mean difference, t_{stat} and df.
(I) The two-sided *p* value for the post hoc test H_0: $\mu_1 = \mu_2$ without Bonferroni correction is .00048. What is the *p* value with Bonferroni correction?

– Exam continues on next page –

Name: _____

6. **GENERIC-SCATTER:** Twelve (12) bivariate observations are plotted below. [12 pts]



(A) Interpret this plot.
(B) r = .97. Interpret this statistic.

SPSS provides the following regression output:

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
Model	B	Std. Error	Beta		
(Constant)	9.089	7.317		1.242	.243
X	.943	.079	.966	11.871	.000

(C) Write down the regression model.
(D) Interpret the slope estimate.
(E) Calculate a 95% confidence interval for the slope parameter.
(F) Interpret the 95% confidence interval.

7. List three (3) methods for comparing averages even when groups have very different variances. (You need not limit your list to inferential methods.)

8. List three (3) methods of describing the relation seen in the graph below.