***** PRACTICE VERSION *****

WAIVER EXAM
STATISTICAL METHODS, WAGNER SCHOOL

Your name (PRINT) Solutions

Your N number ____________________

In taking this exam,
I have adhered to the practices that are outlined in the Wagner Academic Code.
I will not disclose any information about this exam to my fellow Wagner students.

____
Your signature

1. You have 100 minutes to complete this exam.
2. You may use any regular calculator, but you may not use your phone as a calculator.
3. You may use a one-sided, handwritten 8 ½ by 11 inch "cheat sheet" of formulas, etc. – but no other books or other supplementary materials are permitted.
4. You have been given a set of statistical tables to use during the exam.
5. You will hand in this exam document at the end of the period. Write down and show all of your work on this document (not on the statistical tables, not on scrap paper)
6. You may only take the waiver exam once.
7. You do not need to get all of the questions right to get a waiver. If you see a question that you can't answer, just keep going.
8. Good luck!

Question 1

According to the IEA Clean Coal Centre, there are 7000 coal-fired power plants in the world. Among those plants, emissions are normally distributed, with a mean of 200 pounds of CO_2/million BTU, and a standard deviation of 50 pounds of CO_2/million BTU.

Part A

Worldwide, what is the median emission of a coal-fired power plant?

✓ The median is 200 pounds of CO_2/million BTU

___ The median is 200 divided by 50 pounds CO_2/million BTU, or 4 pounds CO_2/million BTU

___ Not enough information is given to answer this question

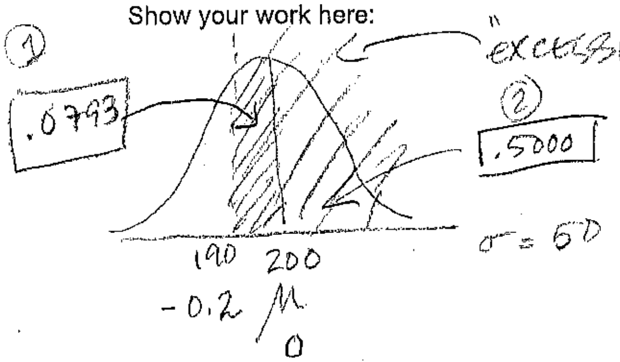
Part B

Coal-fired power plants are called "excessive polluters" if they emit > 190 pounds CO_2/million BTU. Worldwide, how many plants are "excessive polluters?" *(Fill in the blank and show your work).*

Worldwide, 4055 plants are "excessive polluters."

Show your work here:

"excessive" is all shaded; all shaded has two parts



$$Z_{190} = \frac{190-200}{50} = \frac{-10}{50} = -0.2 \Rightarrow$$

(1) .0793 on the negative side

(2) .5000 on the positive side

.5793

there are 7000 plants in total

$.5793 \times 7000 = 4055$ plants are "excessive polluters"

Question 2

You are studying attitudes towards solar power in the US. You ask 110 randomly-selected homeowners if they would be interested in installing residential solar power units, if the federal government subsidized 35% of the cost. Of those in your sample, 44 say that they would.

Part A

Provide a 95% confidence interval estimate of the proportion of homeowners in the population who would install solar power, with the 35% government subsidy. Fill in the blanks with the lower and upper bounds of your confidence interval. Show your calculations below.

Your answer: .31 <= Proportion of homeowners who would install <= .49

$N = 110$

$P_s = \frac{44}{110} = 0.4$

$\sigma_p = \sqrt{\frac{P_u(1-P_u)}{N}} = \sqrt{\frac{(.5)(.5)}{110}}$

$= .047$

Note: if you learned to solve this as $\sqrt{\frac{(.4)(.6)}{110}}$ That's OK, too.

$.04 - (1.96)(.047) < P_u < 0.4 + (1.96)(.047)$

$\boxed{.31 \le P_u \le .49}$ don't worry about rounding.

Part B

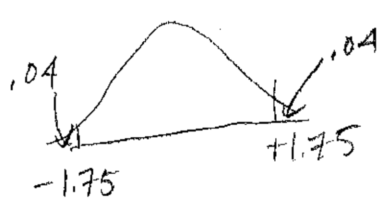
If you had been asked to provide a 92% confidence interval estimate above, what critical value would you have used (instead of 1.96)? Briefly explain how you arrived at your answer.

I would have used a critical value of 1.75, because:

$92 \rightarrow \alpha = .08$

therefore .04 on each tail

The corresponding z value is +/- 1.75



Question 3

Part A

A study finds that children who attend parochial schools (schools affiliated with religious organizations) perform better in college than those who attend regular public schools. In regression analysis, for example, the more years of parochial school a child attends, the higher his/her college GPA ($R^2 = .64$; highly statistically significant). A friend hears about this study, and argues that it demonstrates that parochial schools provide better preparation for college. From the perspective of what you learned in your basic statistics course, what is the single greatest problem with this reasoning?

The single greatest problem with this reasoning is:

correlation isn't causation; there may be other differences between children who attend parochial schools and those that don't. There may be other differences between long- and short- attenders of parochial schools.

Part B

In a sample of 101 young adults who attended Catholic schools, the mean college GPA was 3.4 (s.d = 1.0). In a sample of 101 young adults who attended public schools, the mean was 3.2 (s.d. = 1.0). Do young adults who attended Catholic schools have higher GPAs than those who attended public schools? Do a statistical test, using an alpha of .05. Be sure to show:

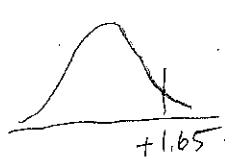
a) the null and alternative hypothesis;
b) the critical value for the test statistic that you're calculating (and how you got it) XX
c) your calculations of the test statistic; and
d) your conclusion: Is there an association between s~~upport for the plan and region of the~~ city?

You may use another blank sheet of this exam to continue, if you need more space.

$\bar{X}_{cath} = 3.4, S_c = 1$
$\bar{X}_{public} = 3.2, S_p = 1$

a) $H_0: \mu_{cath} = \mu_{public}$, or $\mu_{cath} - \mu_{pub} = 0$
$H_1: \mu_{cath} > \mu_{public}$, or $\mu_{cath} - \mu_{pub} > 0$

b) critical value
large sample → Z distribution
one tailed $\alpha = .05$
$Z_{critical} = +1.65$



c) test statistic = $Z_{obtained}$ =

$$= \frac{\bar{X}_{cath} - \bar{X}_{pub}}{\sigma_{\bar{X}_c - \bar{X}_p}} = \frac{3.4 - 3.2}{.14} = \frac{.2}{.14} = \boxed{+1.42}$$

$$\sigma_{\bar{X}_c - \bar{X}_p} = \sqrt{\frac{s_c^2}{N_c - 1} + \frac{s_p^2}{N_p - 1}} = \sqrt{\frac{1}{101-1} + \frac{1}{101-1}} = \sqrt{.02} = .14$$

d) Zobtained isn't in the critical region, can't reject Ho.
Insufficient evidence of assn betw. score & school type.

Question 4

You are studying the relationship between school expenditures and performance on standardized tests. Using data from 50 US states, you have data on yearly average per-student expenditures, and average test scores in each state. For example, your first 3 data points look like this:

State	Yearly average per-student expenditure (thousands of dollars) (X)	Average test score (Y)
New York	8.5	600
Montana	4.3	450
Florida	3.2	325

You run a linear regression to predict average test score, based on yearly average per-student expenditure. Your output indicates that the regression coefficient, b= +25, and the intercept a=300.

a. Write the regression equation that predicts average test score, based on yearly average per-student expenditure.

score' = 300 + 25 (yearly average expenditure in thousands of dollars)

b. How would you interpret the value of regression coefficient (b), in plain English? (Hint: your answer should begin with the phrase, "For every additional ...")

For every additional thousand dollars spent per pupil, test score increases 25 points, on average.

c. If a state spends an average of $4,000 per student every year, what is the predicted average test score? (Show your work and fill in the blank below)

Predicted average test score = 304.

Score = 300 + 25(4) (expenditure in $1000s of dollars)
= 400

Question 5

Part A

You are studying the factors that contribute to differences in test scores between school districts. Using a sample of 100 randomly selected school districts in New York State, you perform a regression analysis of the relationship between average household income in the district (X) and average student test scores (Y).

a. You want to cite a number that reflects *how accurately* you can predict average student test scores, given household income in the district. Which of the following would be the best choice?

___ SST, the total sum of squares for the regression
___ b, the regression coefficient on household income
___ $t_{obtained}$ for b, the regression coefficient on household income
_X_ r, the bivariate correlation coefficient for X and Y
___ the Y intercept (the "constant") for the regression

b. You want to cite a number that reflects *whether there is a statistically significant association* between average household income and average test scores. Which of the following would be the best choice?

___ SST, the total sum of squares for the regression
___ b, the regression coefficient on household income
_X_ $t_{obtained}$ for b, the regression coefficient on household income
___ r, the bivariate correlation coefficient for X and Y
___ the Y intercept (the "constant") for the regression

Part B

As manager of a hospital outpatient care system, you oversee two clinics, Clinic A and Clinic B. You have received many complaints about long waiting times in Clinic A, and you decide to investigate. You draw a random sample of 200 patients at each clinic, and calculate the average wait times at both clinics.

To compare wait times, you conduct a one tailed test of the difference of means, with an alpha level of .05. What is the appropriate value for $z_{critical}$ for that test?
___ z critical is -1.96
___ z critical is +/- 1.96 (that's "plus or minus")
_X_ z critical is -1.65
___ z critical is +/- 1.65 (that's "plus or minus")

Suppose you conduct your statistical test, doing everything correctly. You find that you cannot reject the null hypothesis (in other words, you can't conclude that Clinic A has longer waiting times). Which of the following are true (check all that apply)
___ It is possible that you made a Type 1 error
_X_ It is possible that you made a Type 2 error
___ It is possible that you made a Type 1 error and a Type 2 error, at the same time