These data were collected from undergraduate students at New York University. During the week of midterms of the Fall 1996 semester, students were polled about their perceptions of stress due to various factors. These factors included academic components such as midterms, finals, and grades; social factors such as parents, partners, and friends; and economic factors such as housing and finances. Each student rated the stress due to each factor on a scale of 1 to 5 with 1 = “no stress”, 2 = “little stress”, 3 = “moderate stress”, 4 = “substantial stress”, and 5 = “extreme stress”. Demographic variables such as gender, number of credits enrolled, and academic year were also collected. The data for this sample were gathered so that there were an equal number of males and females, and an equal number of students in each of the four academic years.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Label</th>
<th>Additional Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMTWORK</td>
<td>Perceived stress due to amount of work</td>
<td>1 = no stress, 2 = little stress, 3 = moderate stress, 4 = substantial stress, and 5 = extreme stress</td>
</tr>
<tr>
<td>CREDITS</td>
<td>Number of credits taken in the current semester</td>
<td></td>
</tr>
<tr>
<td>EMPLOY</td>
<td>Perceived stress due to employment</td>
<td>1 = no stress, 2 = little stress, 3 = moderate stress, 4 = substantial stress, and 5 = extreme stress</td>
</tr>
<tr>
<td>GENDER</td>
<td>1 = Male; 2 = Female</td>
<td></td>
</tr>
<tr>
<td>SCHOOL</td>
<td>1 = Education; 4 = Arts &amp; Sciences</td>
<td></td>
</tr>
<tr>
<td>SLEEP</td>
<td>Average hours of sleep on school nights</td>
<td></td>
</tr>
</tbody>
</table>

Directions:
1. Ask if you have any questions about the exam.
2. Be concise in your answers. Do not add irrelevant or incorrect information to a response.
3. Cases of suspected cheating will be referred to the dean.
4. You will need to use your computer and the NELS data set to answer questions 4 and 5.
5. A formula sheet – one 8.5 by 11 inch paper is allowed. Calculator use is also permitted.
6. A copy of Table 1 is attached at the back of the exam.
7. There are 24 questions on this exam. Each is worth 4 points except 5a)-d) which are worth 5 points.
1. Students sampled attended either the College of Arts and Sciences or the School of Education at NYU. The bar graph depicts the relationship between gender and school attended. Use it to answer the following questions.

![Bar Graph](image)

a) Describe the nature of the relationship, or indicate that there is little or none, between SCHOOL and GENDER.

b) Is the Pearson correlation coefficient between SCHOOL and GENDER positive, negative, or near zero? Explain.

c) What is the probability that a student chosen at random from this data set is an education major?
2. Use the SPSS scatterplot of the relationship between the stress due to amount of school-work (AMTWORK) and the average hours of sleep on school nights (SLEEP) to answer the following questions about scatterplots and residuals.

   ![Scatterplot Image]

   a) Explain how the scatterplot indicates that linear regression is appropriate in this case.

   b) What was the stress level due to amount of school-work of the person with the largest negative residual?

   c) Describe the nature of the relationship between hours of sleep on school nights from the stress due to amount of school-work, using language that someone who has not taken statistics would understand.

   d) Is the slope of the regression line for predicting average hours of sleep on school nights from the stress due to amount of school-work positive, negative, or zero? Explain.

   e) What is the y-intercept of the regression equation? Explain its meaning or indicate why it is not meaningful.

   f) What is the predicted number of hours of sleep on school nights of someone who is extremely worried about his or her school-work (AMTWORK = 5)?
3. The following histogram depicts the distribution of SLEEP scores for the students questioned.

- **a)** Explain why it is reasonable to use the normal curve to approximate the distribution of SLEEP scores.

- **b)** According to the histogram, approximately what proportion (or percentage) of these students get 5 hours or less of sleep, on average, on school nights? Please show enough work to indicate your method.

- **c)** According to the normal distribution, approximately what proportion (or percentage) of students get 5 hours or less of sleep, on average, on school nights? Please show enough work to indicate your method.

- **d)** According to the normal distribution, what sleep score cuts off the bottom 30% of sleep scores? Please show enough work to indicate your method.
4. In the following problems, you are asked to describe the relationship between pairs of variables in the NELS data set. When describing the relationship, please use terms that someone who has not taken statistics could understand. For example, positive correlation is not sufficient. Also, select and report the value of a correlation coefficient to support your answer. If you cannot obtain the necessary SPSS output, you may request it from me. That will cost you 10 points on the exam.

a) Describe the relationship, or indicate that one does not exist, between years of math taken in high school (UNITMATH) and math achievement in twelfth grade (ACHMAT12).

b) Describe the relationship, or indicate that one does not exist, between the number of times the student was late to school in twelfth grade (LATE12) and the number of times the student skipped/cut classes in twelfth grade (CUTS12).

c) Describe the relationship, or indicate that one does not exist, between the school type attended in eighth grade (SCHTYP) and whether or not the student took any Advanced Placement classes in high school (APPROG).

d) Describe the relationship, or indicate that one does not exist, between family size (FAMSIZE) and self-concept in twelfth grade (SLFCNC12).
5. Can twelfth grade math achievement (ACHMAT12) be predicted by whether or not a student took advanced math in eighth grade (ADVMATH8) for students in the NELS data set? Perform a regression analysis using SPSS and use the results to answer the following questions. If you cannot obtain the necessary SPSS output, you may request it from me. That will cost you 2 points on the exam.

   a) What is the regression equation for predicting ACHMAT12 from ADVMATH8?

   b) Interpret the slope of the regression equation, in this case.

   c) Interpret the y-intercept of the regression equation, in this case.
a) Describe how one obtains the sampling distribution of means for samples of size $N = 8$.

b) If a certain data set is normally distributed with mean 70 and standard deviation of 2.5, give the mean and the standard deviation of the sampling distribution of means of size $N = 8$.

c) Describe the effect on your answer to part b) of increasing the sample size.

d) If we were to select a single sample of size $N = 8$, what is the probability that the mean of this sample will be two points higher than the actual population mean?
Exami 2 Solution

(a) Female students tended to be in school of education while males tended to be in Arts and Sciences

(b) Negative since females (high score) tended to be in school of education (low score)

\[
\frac{36+44}{36+44+47+33} = \frac{83}{160}
\]

2.) The data points cluster about a line

(a) 3

(b) The higher the stress level due to work, the fewer average hours of sleep a person gets.

(c) Negative: Increases in x-values correspond with decreases in y.

(d) y-intercept is 8.2. But stress due to work has no meaning at zero. (see front page.)

(e) 5.5 hours

3.) The distribution is reasonably symmetric about mean.

(a) Approx. \( \frac{39+71}{100} = 0.29 \)

(b) Z-score of 5: 5-6.31 = -0.99 From table, area is 0.1611

(c) Z-score with 3 as area to right is approx. \( z = -1.2 \). Change -1.2 to raw score: \( 613 + (1.31)(-1.2) = 586 \)

4.) Pearson's \( r = .423 \) (Spearman also OK). Kids who took more math in HS tended to have higher math achievement.

5.) \( r = .429 \) Kids who were late frequently in 12th grade also missed class frequently.

(c) The majority (86%) of public school students do not take AP classes. The majority of private school students (91.8% for religious schools, 70% for other private schools) do take AP classes.

(d) \( r = .022 \) There's no relationship between family size and self concept.

5a.) ACTMAT12 = 4.159(AADVANCEDMATH8) + 55,106

(b) Students who took advanced math in 8th grade tend to score about 4.2 pts higher on math achievement in 12th grade

(c) Students with no advanced math in 8th grade are predicted to score about 55 on 12th grade math achievement tests.

(a) Create all possible samples of size 8 and take the mean of all these samples.

(b) Mean to stand. dev. \( \frac{2.5}{\sqrt{8}} \approx 0.8839 \)

(c) Stand. dev. decreases, (Mean stays same.)

(d) Z-score \( \frac{72-70}{(2.5/\sqrt{8})} \approx 2.26 \) Area .0119