Statistics 2000, Section 001, Final (300 Points)

May 4, 2000, Dr. Jürgen Symanzik

Your Name: _____

First look at all 6 questions. Then start with the question that looks easiest to you. Continue with a more difficult question. Try to answer as many questions as possible in these 110 minutes.

Note that you will obtain at least partial credit if you indicate a correct formula but your final result is incorrect. If you just rely on your calculator without indicating the formula that should be used and your result is incorrect, you will obtain no credit at all for this part of a question.

Question 1: Short Answers (**50 Points**)

1. For $x_1 = 5, x_2 = 4, x_3 = 10, x_4 = 0, x_5 = 20, x_6 = 14, x_7 = -5, x_8 = 4, x_9 = -2$, and $n = 9$, determine the following sums: (**12 Points**)

$$\sum_{i=1}^{n/3} x_i \cdot x_{i+1} =$$

$$\sum_{i=n-2}^{n} i^2 \cdot x_{(i)} =$$

$$\sum_{i=1}^{n} (i!) \cdot \frac{x_4}{100 + x_i} =$$

2. Let $z_i = \frac{x_i - \overline{x}}{s}$, $i = 1, \ldots, 8$, be the z–scores for the 9 values given in (1.) above. You do not even have to calculate the individual values for each z_i to indicate which of the following statements are correct: (**6 Points**)

   (a) The correlation coefficient r between x and z must be negative since I subtract $\overline{x}$ from each x_i.

   (b) The correlation coefficient r between x and z is somewhere between -0.99 and -0.01.

   (c) The correlation coefficient r between x and z is somewhere between 0.01 and 0.7.

   (d) The correlation coefficient r between x and z is somewhere between 0.7 and 0.99.

   (e) The correlation coefficient r between x and z is exactly -1.

   (f) The correlation coefficient r between x and z is exactly 0.

   (g) The correlation coefficient r between x and z is exactly 1.

3. Determine the slope and the y–intercept of the lines whose equations are given as: (**6 Points**)

   (a) $5x - 2y = -2$

   (b) $5y + 5x + 5 = 0$

4. A foreign lottery has a game called "6 out of 25". You make a selection of 6 numbers between 1 and 25 and you win the big prize if exactly these numbers are drawn in the weekly drawing. How many different combinations are possible to select 6 out of 25 numbers? Calculate this value! (**6 Points**)

5. What is wrong with this graphic from "Time" (April 9, 1979, p. 57), reprinted in Edward R. Tufte's book "The Visual Display of Quantitative Information". Provide a better graphical representation of the same data. (**8 Points**)

6. Let Z be a standard Normal variable, i.e., $Z \sim N(0,1)$, and X be a Normal variable with mean $\mu = -2$ and variance $\sigma^2 = 9$, i.e., $X \sim N(-2, 3^2)$. Determine the following: (**12 Points**)

   (a) $P(Z < 1.84) =$

   (b) $P(-2.0 < Z < 1.0) =$

   (c) $P(X < 1.84) =$

   (d) $P(-2.0 < X < 1.0) =$

   (e) Find a number # so that
   $P(Z < \#) = 0.20$

   (f) Find a number # so that
   $P(X < \#) = 0.20$

Question 2: Numbers and Graphs (**49 Points**)

The following Poverty Rates have been taken from the Web at
`http://www.census.gov/hhes/poverty/povanim/pvmaptxt.html`.
See the Web page (and linked pages) indicated above for an exact definition of "Poverty".

We are only interested in the Poverty Rates of **all** 12 states in the "Southwest". And here are the Poverty Rates of these states (averaged over 1996 to 1998):

State	Poverty Rate
Arizona	18.1
Arkansas	17.2
California	16.3
Colorado	9.3
Kansas	10.1
Louisiana	18.6
Missouri	10.4
Nevada	9.9
New Mexico	22.4
Oklahoma	14.8
Texas	16.1
Utah	8.5

1. Calculate the Arithmetic Mean, the Variance, and the Standard Deviation of the Poverty Rates of these 12 states. (**10 Points**)

2. Do you have to calculate the **Sample Variance** or **Population Variance** in (1.) above? Circle your answer. (**3 Points**)

3. Calculate the Median $\tilde{x}$ and the Range R of the Poverty Rates. (**6 Points**)

4. Draw a Histogram of the Poverty Rates (Start at 5.0, using a class width of 5.0). (**8 Points**)

5. Draw a meaningful (!) Stem–and–Leaf Plot of the Poverty Rates. (**8 Points**)

6. Draw a Boxplot of the Poverty Rates. (**8 Points**)

7. Are there any outliers in this data set? Circle your answer: **Yes** or **No**.

   If yes, then also circle the number(s)/state(s) you consider as outliers in the table on the previous page. (**3 Points**)

8. Based on your plots in (4.) through (6.) above, does your data appear to belong to a Normal distribution? Circle your answer: **Yes** or **No** (**3 Points**)

Question 3: Linear Regression & Correlation (**73 Points**)

The average prices (in dollars) per ounce of gold and silver for the years 1986 through 1994 are given below. (Source: U.S. Bureau of Mines.)

Year	Gold	Silver
1986	368	5.47
1987	478	7.01
1988	438	6.53
1989	383	5.50
1990	385	4.82
1991	363	4.04
1992	345	3.94
1993	361	4.30
1994	389	5.30

1. Draw a scatterplot where x is the silver price and y is the gold price. Does the appearance of the scatterplot suggest a linear relationship between x and y? (**10 Points**)

2. Fit a least squares (linear regression) line to the data. Indicate what your variables stand for. It might help to know that:

$$\sum_{i=1}^{n} x_i = 46.91, \sum_{i=1}^{n} x_i^2 = 253.6095, \sum_{i=1}^{n} y_i = 3,510, \sum_{i=1}^{n} y_i^2 = 1,383,102, \sum_{i=1}^{n} x_i y_i = 18,625.9$$

(**38 Points**)

3. Calculate Pearson's correlation coefficient r between x and y. How can we interpret this value for our given data set? (**10 Points**)

4. Based on your calculations in (2.), what is the predicted price of an ounce of gold when the price of an ounce of silver is \$4.00?

   Do you think it is safe to predict that an ounce of gold costs about \$237 if the price of an ounce of silver drops to \$1.00? Explain. (**15 Points**)

5. Indicate which residual plots should be plotted for this data set. You should describe **at least 2** different residual plots that might be useful here. You **do not** have to draw any such plot to get full credit. (**10 Points**)

Question 4: Probability Distributions (**43 Points**)

In our first quiz, the following grades (representing full–letter grades) have been reported for the 42 students that participated in this quiz:

# students	points	$p(y)$	$F(y)$
9	0.0		
14	1.0		
7	2.0		
8	3.0		
4	4.0		

We introduce a random variable Y that represents the points for Quiz #1.

1. Add the values for $p(y)$, i.e., the probability distribution for Y, to the table above and draw a spike graph of $p(y)$. (**10 Points**)

2. Add the values for $F(y)$, i.e., the cumulative probability function (or cumulative distribution function, cdf) for Y, to the table above and draw a graph of $F(y)$. (**10 Points**)

3. Indicate $P(Y \leq 3.0)$? (**3 Points**)

4. Calculate $\mu = E(Y)$, i.e., the mean (expected value) of Y. (**10 Points**)

5. Calculate $\sigma^2 = Var(Y)$, i.e., the variance of Y. (**10 Points**)

Question 5: Binomial Probability Distributions (**45 Points**)

In a previous quiz, only 5 out of 25 Stat 2000 students answered a particular question correctly. Let us assume that in the upcoming academic year all new Stat 2000 students have to answer this same question in one of their quizzes. Based on previous years, the upper enrollment limit of 200 students has been obtained each year so we can safely assume that there will be exactly 200 students again during the next year that will attend this class.

Let X be the random variable that describes the number of students that will correctly answer this question in the next year. It is safe to asssume that students' ability to answer a particular question does not change over time.

1. Indicate the probability distribution of X. Complete the following formula that relates to the probability distribution of X: (**5 Points**)

$$X \sim$$

2. Calculate the mean and the variance of X. (**10 Points**)

3. What is the **exact** probability that 40 students will answer this question correctly? Be as efficient in your calculations as possible. (**10 Points**)

4. What is the probability that 30 through 50 students will answer this question correctly? You may use an **approximation** if **appropriate**. (**20 Points**)

Question 6: Confidence Intervals & Tests of Significance (**40 Points**)

A friend who hears that you have been taking a statistics course asks you to help with a chemistry lab report. She has made four independent measurements of the specific gravity of a compound. The results are

$$3.82, 3.93, 3.67, 3.78.$$

The lab manual says that repeated measurements will vary according to a Normal distribution with standard deviation $\sigma = 0.15$. The true mean μ of the specific gravity of a compound depends on the temperature and several other factors.

1. The lab manual asks for a 95% confidence interval for the the true specific gravity. Your friend does not know how to calculate this. Do it for her. (**20 Points**)

2. The lab manual says that the true specific gravity at a temperature of 80F is 3.9. However, at a higher temperature, the specific gravity will be considerably less. Unfortunately, your friend did not record the exact temperature but assumes that it might have been about 80F. State a test of significance where the null hypothesis is that the true specific gravity is 3.9 and the alternative is that the true specific gravity is less than 3.9.

   Indicate the test statistic and calculate the p–value. Can you reject H_0 at the 5% level of significance? (**20 Points**)