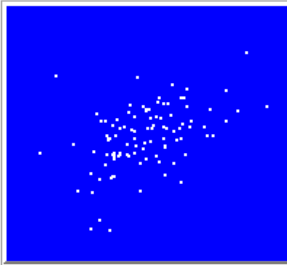
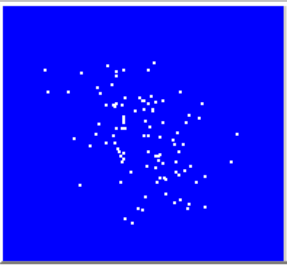
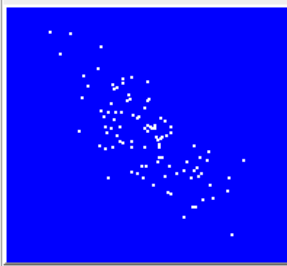
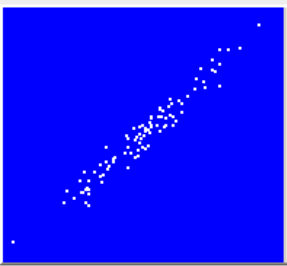


**INSTRUCTIONS:** Read the questions carefully and completely. Answer each question and show work in the space provided. Partial credit will not be given if work is not shown. When asked to explain, describe, or comment, do so within the context of the problem. Be sure to include units when dealing with quantitative variables.

1. **[20 pts]** Short answer.
  - a) [3] Statistics is about ... \_\_\_\_\_. (Fill in the blank with one word.)
  - b) [4] For each variable listed below indicate what type of variable it is, categorical/qualitative or numerical/quantitative. For categorical/qualitative variables give several categories. For numerical/quantitative variables indicate the units that would accompany the value of the variable.
    - Hair color
    - Circumference of a tree trunk
  - c) [4] A sample of 10 homes has a mean price of \$207,500. The most expensive home in the sample is \$500,000. If the most expensive home is removed from the sample, what is the mean price of the remaining 9 homes? In order to get credit you must show your work.
  - d) [2] On the first exam in Stat 104 in fall 2010, the mean score was 82.1 points and the standard deviation was 12.84 points. A student scored 60 points on the exam. What is that student's standardized score?
  - e) [2] Physical flexibility ratings are recorded for a sample of young adult males. A value of 1 indicates very low physical flexibility. A value of 10 indicates very high physical flexibility. The five-number summary is: minimum = 1,  $Q_1 = 5$ , median = 7.5,  $Q_3 = 9$ , maximum = 10.
    - What is the shape of the distribution? Explain how you determined the shape.

f) [3] Match the Plot with the correct value of the correlation coefficient,  $r$ .

	<input type="radio"/> 0.97 <input type="radio"/> 0.45 <input type="radio"/> -0.38 <input type="radio"/> -0.74		<input type="radio"/> 0.97 <input type="radio"/> 0.45 <input type="radio"/> -0.38 <input type="radio"/> -0.74
	<input type="radio"/> 0.97 <input type="radio"/> 0.45 <input type="radio"/> -0.38 <input type="radio"/> -0.74		<input type="radio"/> 0.97 <input type="radio"/> 0.45 <input type="radio"/> -0.38 <input type="radio"/> -0.74
Match the correlations with the plots.		Answers	

g) [2] Why is random assignment of treatments to subjects important in an experiment?

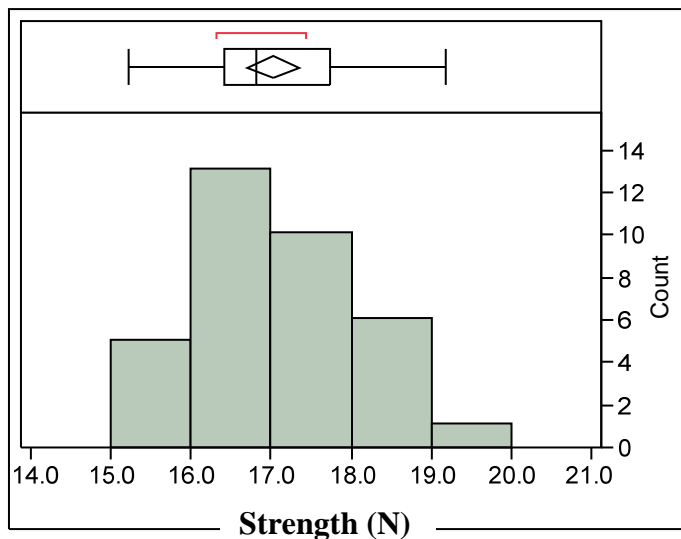
2. [18 pts] A study is performed on the effect of temperature on the strength of plastic welds. There are seven different temperatures (140, 150, 160, 170, 180, 190 and 200 C). At each temperature five specimens of the plastic material are welded. The welds are then subjected to a load until the weld breaks. The maximum load (in Newtons) before the weld breaks is the strength of the weld. The thirty five welds (5 specimens at each of 7 temperatures) are run in a random order. All welds are made with the same equipment by the same operator. All strengths are measured with the same equipment with the same operator.

a) [3] Why is this study an experiment and not an observational study?

b) [3] What is the response variable? Is this variable categorical or numerical? If categorical give the categories if numerical give units of measurement.

- c) [3] What is the explanatory variable? Is this variable categorical or numerical? If categorical give the categories if numerical give the units of measurement.
- d) [3] Identify an outside variable that is controlled in the study and indicate how that variable is controlled.
- e) [3] Is there a control group in the study? Explain briefly.
- f) [3] How is replication within this study accomplished?

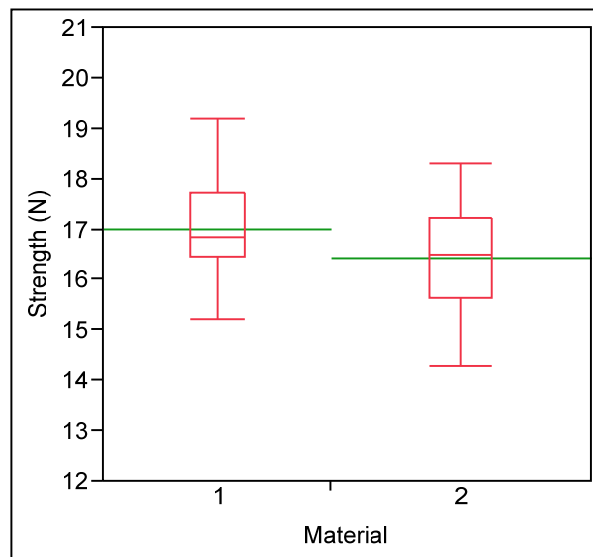
3. [17 pts] Data from the group of 35 weld strengths is displayed below.



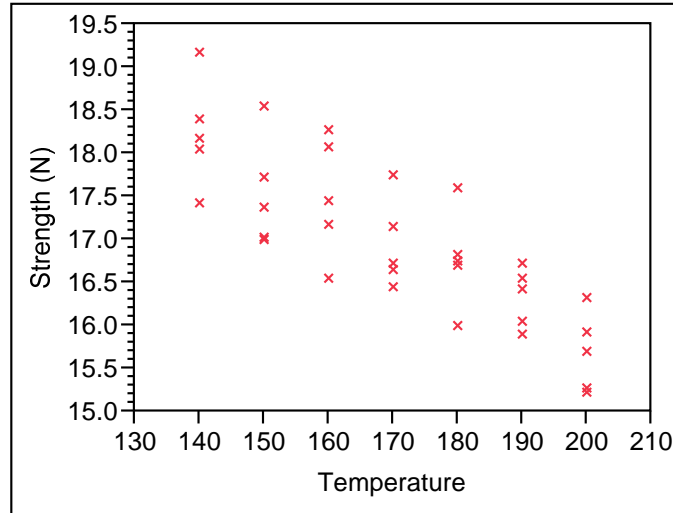
100.0%	maximum	19.18
75.0%	quartile	17.72
50.0%	median	16.82
25.0%	quartile	16.42
0.0%	minimum	15.22
	Mean	17.00
	Std Dev	0.946
	N	35

- a) [3] Looking at the histogram describe the shape of the distribution of strength.

- b) [3] What are the values of the sample mean strength and the sample median strength? Include appropriate units.
- c) [3] What are the values of the sample standard deviation and the sample IQR? Include appropriate units.
- d) [3] Which summaries of center and variability are most appropriate for these data? Explain briefly.
- e) [5] Below are box plots for the strengths of welds for two different plastic materials, Material 1 and Material 2. The longer horizontal lines through each box represent the mean weld strength. Use the information given in the box plot to compare the materials in terms of their effect on the mean strength and the variability in strengths.



4. [25 pts] The investigators that performed the study on plastic weld strength described in problem 2 thought that the temperature used to create the weld would have an effect on the strength. Below is information on the strength of welds using Material 1 and the temperature used to make those welds.



Temperature (C)		Strength (N)			
Mean	170	Mean	17.00		
Std Dev	20.292	Std Dev	0.946	Correlation	-0.828

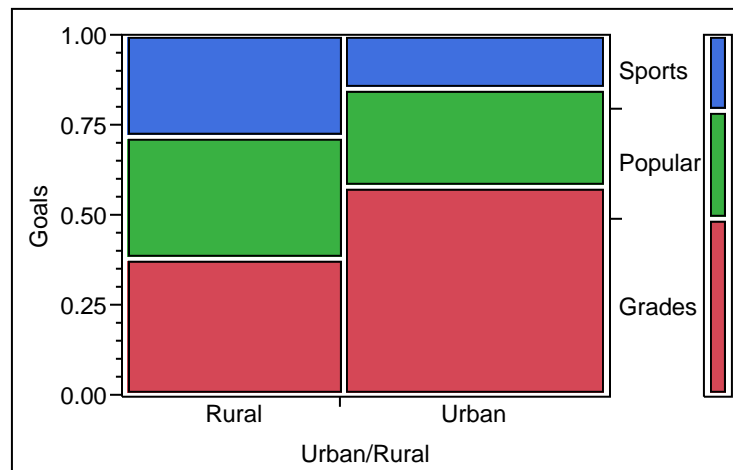
- a) [2] What is the explanatory variable? What is the response variable?
- b) [2] Describe the relationship between temperature and strength.
- c) [4] Compute the least squares estimate of the slope,  $b$ . Round your final answer to 4 decimal places.

- d) [4] Give an interpretation of the slope within the context of the problem.
- e) [3] Compute the least squares estimate of the y-intercept,  $a$ . Round your final answer to 2 decimal places.
- f) [3] Give the equation of the least squares regression line and use it to predict the strength for a weld that is welded at 200 C.
- g) [2] One of the welds made at 200 C has strength of 15.7 N. What is the residual for this weld?
- h) [3] Graph the least squares regression line on the plot on the previous page. In order to get full credit it must be obvious to me that you are using the equation to draw the line.
- i) [2] What percentage of variation in weld strength is explained by the linear relationship with temperature?

5. [10 pts] A sample of 327 students in grades 4 – 6 in selected schools in Michigan were asked the following question about their school goals, specifically “What would you most like to do at school?” The students could choose either “make good grades,” “be good at sports” or “be popular.” It was also noted whether the student was from a rural school or an urban school. A summary of how many students from each type of school and their answers to the question is given in the table below.

	Grades	Popular	Sports	
Rural	57	50	42	149
Urban	103	49	26	178
	160	99	68	327

- a) [1] Who was data collected on?
- b) [2] What data was collected?
- c) [2] Considering only students from urban schools, what proportion of students from urban schools answered that they want to make good grades?
- d) [5] Below is a mosaic plot of the row percentages. What does this plot tell you about the differences in school goals for students from rural and urban schools? Be sure to refer specifically to the mosaic plot, and the percentages this plot displays, in answering this question.



6. **[10 pts]** A study reported in *Alzheimer's and Dementia: the Journal of the Alzheimer's Association* in 2005 looked at the relationship between folate in the diet and Alzheimer's disease. The study used information gathered from the Baltimore Longitudinal Study of Aging. The study found that diets high in folate might help reduce the risk of Alzheimer's disease. Source: [www.cbsnews.com](http://www.cbsnews.com)

In that study the researchers enrolled 579 volunteers (359 men, 220 women) 60 and older without Alzheimer's disease and followed them for nine years. The researchers looked at the percentage of antioxidant vitamins (E, C, carotenoids) and B vitamins (folate, B-6, and B-12) contained in participants' diets.

During the nine year follow-up period 57 participants developed Alzheimer's disease.

The researchers then compared the nutrient intake of those who developed Alzheimer's disease with the nutrient intake of those who did not develop the disease. The researchers found that those with a higher dietary intake of folate had an almost 60 percent lower rate of Alzheimer's disease.

- a) [3] Why is this study an observational study and not an experiment? Explain briefly.
- b) [3] Was this a prospective or a retrospective study? Explain briefly.
- c) [2] What is the explanatory variable? Is it categorical or numerical?
- d) [2] What is the response variable? Is it categorical or numerical?

## Formulas

**five number summary**

**Min**       $Q_1$       **Median**       $Q_3$       **Max**

**Range = Max – Min**

**IQR =  $Q_3 - Q_1$**

**mean**      **standard deviation**      **standard score**

$$\bar{x} = \frac{\sum x}{n} \quad s_x = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad z_x = \frac{x - \bar{x}}{s_x}$$

$$\bar{y} = \frac{\sum y}{n} \quad s_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n - 1}} \quad z_y = \frac{y - \bar{y}}{s_y}$$

**correlation coefficient**

$$r = \frac{\sum z_x z_y}{n - 1} = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n - 1)s_x s_y}$$

$$R^2 = (r)^2$$

**slope**      **y - intercept**      **prediction equation**

$$b = r \frac{s_y}{s_x} \quad a = \bar{y} - b\bar{x} \quad \hat{y} = a + bx$$

**residual =  $y - \hat{y}$**

**approximate margin of error =  $\frac{1}{\sqrt{n}}$**