Solutions for Exam 1 Math 263, Statistics and Biostatistics
September 13, 2005

1. *The table below shows the results of measurements that were done on rats which had been given a treatment to induce heart attacks. The first column shows the measured level of a certain enzyme in the rats blood after the heart attack, and the second column shows the fraction of pumping capability remaining for each rat after the attack.*

Enzyme level	Pumping fraction
0.0	0.9
0.1	1.0
0.4	0.8
1.8	0.5
2.4	0.4
2.5	0.5
2.7	0.6
3.3	0.3
5.4	0.1

(a) (15 pts) *Give the mean, the standard deviation and the five-number summary for both Enzyme level and Pumping fraction.*

$\overline{E} = 2.07$, $S_E = 1.74$, Min=0, Q1=.25, median=2.4, Q3=3, max=5.4. (Using TI-83+ calculator, you may get slightly different Q1 and Q3 values with the book definition or with other calculators)

$\overline{P} = .57$, $S_P = .29$, Min=.1, Q1=.35, median=.5, Q3=.85, max=1

(b) (5 pts) *Based on the 1.5 IQR rule find any outliers in the enzyme levels. Justify your answer.*

Using the answer to part a, we have that IQR=Q3-Q1=2.75, so 1.5×IQR=4.125. Outliers must either be larger than Q3+1.5×IQR=7.125, or smaller than Q1-1.5×IQR=-3.875. None of the date for enzyme level is an outlier.

(c) (5 pts) *Sketch a scatter plot of the data.*
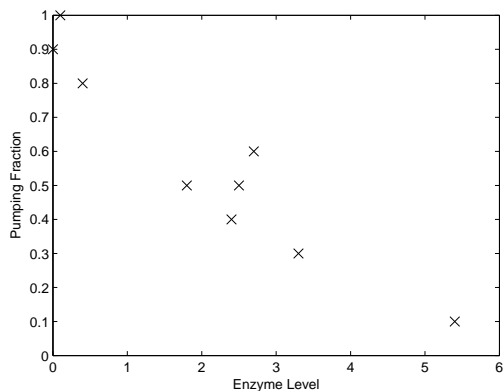
(d) (5 pts) *What is the correlation between the two observations?*

The correlation is $r = -0.95$

(e) (15 pts) *If we want to predict Pumping fraction based on Enzyme level, what (linear) formula should we use? What pumping level does it predict for an enzyme level of 2.0? Is this likely to be an accurate prediction? Justify your answer.*

We should use linear regression. Using a TI-83+ I got $\widehat{P} = .895 - .159E$. This formula predicts a value of .577 for the pumping fraction if the enzyme level is 2. The value of $r^2 = .901$ is quite close to 1, so the since 2 is in the midst of the data that we have for E, we can use the predicted value with some expectation that it will be close.

(f) (5 pts) *What pumping fraction would your formula predict for an enzyme level of 9.0? Is this likely to be an accurate prediction? Justify your answer.*

The value predicted is -0.534. This makes no sense, since pumping fraction should be non-negative. So we can see it must be unreliable. This is not surprising since the value 9 is well outside the range of E values that we had. This is extrapolation, which is unreliable.

2. *A group of nurses at UMC theorize that patients without health insurance tend to spend a shorter time as inpatients in the hospital than those with insurance. They analyze hospital records and report "There is a moderate correlation between having health insurance and the number of days that a patient stays in the hospital".*

   (a) (5 pts) *Is this an observational study or an experiment? Explain.*

   This is an observational study, using existing data. It doesn't not involve any treatment that is imposed on the patients.

   (b) (5 pts) *Identify the response variable and the explanatory variable.*

   The response variable is the time that each patient spends in the hospital, the explanatory variable is the insurance status of the patients.

   (c) (10 pts) *Why is the statement as reported wrong? How should they report their result?*

   Correlation is only defined between quantitative variables. The explanatory variable here is a categorical variable. They could have used the word association instead of correlation.

3. *A population of 999 students enrolled in a statistics class are randomly divided into three equal sized groups. One group is given additional intensive training in the use of TI-83+ calculators, the second group is given additional training in the use of minitab, and the third group is given no additional training. At the end of the semester, the score on the statistics final is collected for all students.*

   (a) (8 pts) *Draw a diagram outlining the design of this experiment.*

   (b) (5 pts) *Identify the experimental units and the treatment(s) in this experiment.*

   Experimental units, or subjects were the students. There were three treatments, TI83 training, Minitab training, and no additional training.

   (c) (7 pts) *What type of probability sample is used in this study?*

   This was a simple random sample (S.R.S.)

   (d) (10 pts) *Briefly describe how you might design a stratified random sample based on student GPA.*

   Divide the students into groups (Strata) according to their GPA, with one group in the 0-1 range, another 1-2, 2-3 and 3-4. Within each group randomly subdivide into three roughly equal subgroups one of which is trained in the TI83, one trained with Minitab, and the third receiving no additional training. Collect the data on exam scores for each of the subgroups and compare.