Stats 252, Winter 2001

Midterm

Instructions

You may ask me any question you wish, but do not discuss with anyone else. In fact, you are encouraged to discuss it with me, via email or phone or in person. You may borrow any books I, or the libraries, have. The midterm is due **Tuesday, 9am, February 20**. Please slip it under my office door: MS 6151. Note that the 19^{th} is a holiday. You may use any resource materials you wish, as long as they are not human. For the sake of this midterm, listservers, bulletin boards, etc., will be considered "human". Attach extra pages, if you wish.

I. A marine biologist was interested in studying how sea slug larvae are able to find food. Once, researchers thought that the larvae simply drifted with currents, but soon came to believe that they are able to direct themselves to some extent. This theory would be strengthened considerably if it could be shown that the larvae can sense when food is nearby. One theory is that the algae that the larvae feed on secretes a chemical which the larvae can "smell". When near the seaweed, the larvae will change into sea slugs, and then feed on the algae. (I'm not making this up.)

To test this theory, the researcher collected samples of water from a tide pool in which the algae was present. He collected water at various times as the tide came in (and thus naturally diluted the concentration of the chemical.) Back in the lab, each sample had 15 larvae put into it. After a fixed amount of time, he recorded the percentage of these 15 larvae that had converted to sea slugs.

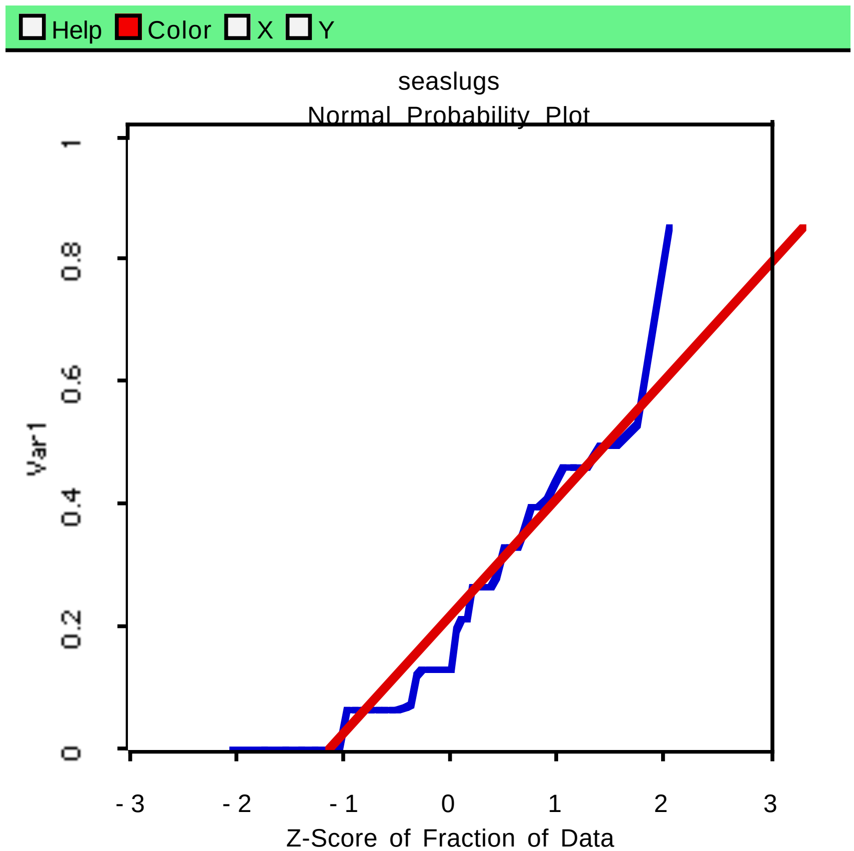
(If you want to see pictures of sea slugs, visit http://www.oz.net/~miranda/illustrat.html)

There are two variables: **time** (measured in minutes since low tide) and **percent** (percentage of larvae that converted). We will focus only on percent.

1. (25) Below is the normal probability plot for the percentage of sea slugs that converted. Note that it is very important that you carefully explain your reasoning when answering these quesitons and demonstrate that you know how to read this plot. (Note: "var1" is "percent").

a) What is the median of **percent**?

b) How does the distribution of percent differ from a normal distribution?

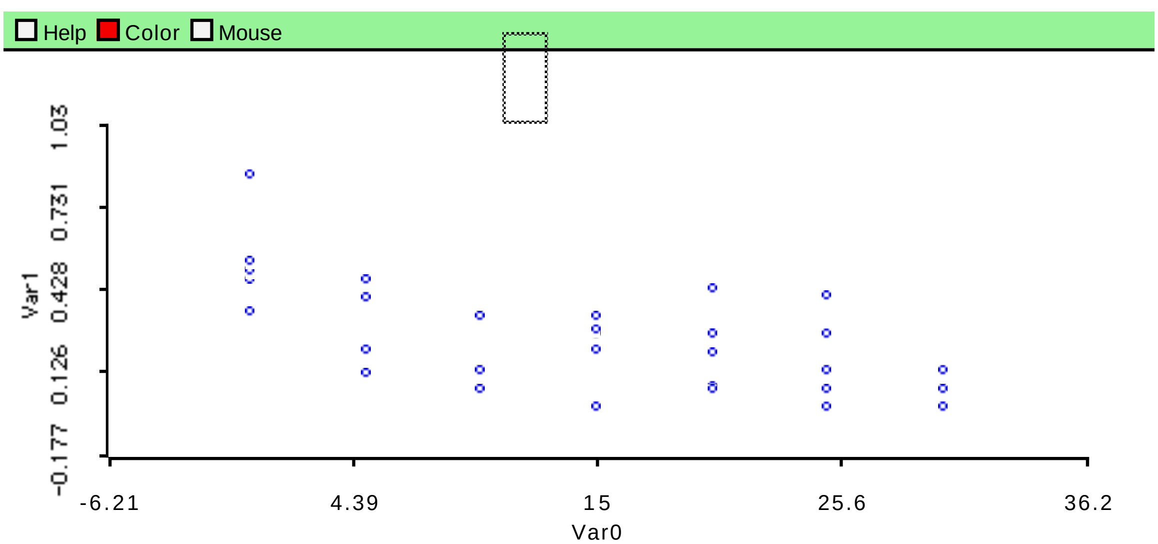

Help
Color
X
Y
seaslugs
Normal Probability Plot
Var1
0
0.2
0.4
0.6
0.8
1
- 3
- 2
- 1
0
1
2
3
Z-Score of Fraction of Data

Summary Statistics

VARIABLS	MEAN	StDv	variance	skewness	KURTOSIS	N
Time	15.00	10.12	102.44	0.00	-1.25	42.0
Percent	0.24	0.19	0.04	0.84	0.83	42.0

2. (15) Find a 90% confidence interval for the mean percent. State any assumptions you need.

3. (40) One hypothesis of the sea slug data was that at 15 minutes and beyond, the **median** percent would be less than it was for the measurements before 15 minutes. A visual inspection seems to agree:



a) Design a bootstrap routine to help you test this hypothesis.

- Give the algorithm for this routine. (You don't have to write code, (in fact don't bother), but you should explain step-by-step how the program would work.)
- Explain what the input would be.
- Explain what the output would be.
- Explain how to interpret the output to make a decision as to whether to reject the null hypothesis.

Assume a significance level of 5%. Assume that the data have been divided into two groups named "before" and "after". "Before" contains the percents for up to (but not including 15 minutes) and has 18 observations. "After" contains 25 observations: those for 15 minutes and after.

Hint: The test statistic is the difference of the two medians. The observed value here is Median(Before 15) - Median(15 and beyond) = 0.38 - .13 = 0.25. The null hypothesis is that the true value of this difference is 0, and this is the condition under which you perform the bootstrap.

4. (20) A researcher needs to measure the mass of ice crystals (nanograms). She believes her measurement error is close to .2 nanograms. She will estimate the true mass of each crystal by taking repeated independent measurements and computing the average. How many measurements should she take so that her estimate of the true value will be within .01, with 95% confidence?

5. In the blood pressure data set examined earlier, each of 15 patients' blood pressure was measured before and after receiving a drug. Let's focus on the diastolic:

dbefore	**dafter**	**ddif**
130	125	-5
122	121	-1
124	21	-3
104	106	2
112	101	-11
101	85	-16
121	98	-23
124	105	-19
115	103	-12
102	98	-4
98	90	-8
119	98	-21
106	110	4
107	103	-4
100	82	-18

The traditional test to test whether there was an "effect" is to look at the difference of the two variables and test whether or not the mean is lower than 0 (the "paired t-test"). A criticism of this is that the test assumes the data are normally distributed. Usually this is not too troublesome an assumption because even if the population is not normally distributed, the sample average will be approximately normally distributed, and hence the test statistic (the "t statistic") will be approximately t-distributed. But because the sample size is small here (15), there is reason to be cautious.

An alternative approach is to use a non-parametric test. Non-parametric tests are a broad class of tests that make no distributional assumptions on the data. The test we'll consider is somewhat crude, but has the benefit of requiring very few assumptions.

For each patient, record a "1" if the blood pressure went down (the drug was successful) and a "0" if it did not.

a) What's the observed probability that a patient's blood pressure will decrease?
b) Under the null hypothesis (i.e. the drug is ineffective), what is the probability that a patient's blood pressure will decrease? (Call this "pi".) (Careful: "ineffective" does NOT mean the drug will NEVER lower blood pressure.)
c) What is the p-value for this sample?
d) Would you reject the null hypothesis, based on the p-value? Use a significance level of 0.05.
e) With the significance level of 0.05, what is the smallest value for the observed "pi" for which you would reject the null hypothesis. (For example, if you saw 1 success out of 15, you would not. And probably for 2 out of 15 you would not. So how many successes out of 15 before you would reject?) (Hint: some of the commands in xlispstat might

prove useful. You don't need to load data in to use them, so it should work on every computer.)

f) With the significance level at 0.05, if the true probability of lowering a patient's blood pressure was 0.55, what is the power of this test?