STATISTICS 401C

Fall 2008

Laboratory Assignment 5

1. In a study of environmental pollution due to industrialization, lead concentrations in suspended particulate matter and sediments were collected at 37 inshore stations in Mombasa, Kenya. These are (in mg/kg dry weight):

48	53	44	55	52	39	62	38	23	27	41	37	41	46	32	17	32	41	23
12	3	13	10	11	5	30	11	9	7	11	77	210	38	112	52	10	6	

Use hand computation for all parts of this problem. From these $\bar{y} = 37.24$ and $s = 37.12$ were calculated.

(a) Conduct a test to determine if there is sufficient evidence in the data to indicate that the mean lead concentration exceeds 30 mg/kg dry weight (Use $\alpha = .05$). State the hypotheses, the t-statistic, and the rejection region.

(b) Assuming that the population standard deviation of the lead concentration is $\sigma \approx 30$ mg/kg dry weight, use the charts in Table 3 to determine the power of the above test to detect mean lead concentrations of 36, 39, 42, and 45 (to the best accuracy possible).

μ_a	d	β	Power=1-β
36			
39			
42			
45			

(c) Explain why the following are reasonable assessments based on the solution to part (b).

i. There is a very good chance of falsely failing to reject the null hypothesis if the actual mean lead concentration is only 36 mg/kg dry weight.

ii. There is a very good chance of rejecting the null hypothesis if the actual mean lead concentration is as large as 45 mg/kg dry weight.

(d) Before the samples were to be taken, suppose that you were required to make a sample size determination to ensure that the probability of accepting H_0 is kept below 0.1 when the actual mean lead concentration exceeds 54 if $\alpha = .05$ is used. Give an approximate value for the minimum sample size necessary as close as you can read using the charts in Table 3). Also use the sample size table supplied to get this estimate. Use $\sigma \approx 30$ as in part (b).

2. Lead poisoning results from an unnatural craving for such substances as paint and affects a large number of children each year causing them to suffer severe retardation. But it has also been investigated whether such cravings have a nutritional explanation. In a study, each rat in a group of 20 were randomly assigned to a normal diet (control group) or to a calcium-deficient diet(experimental group). Each rat occupied a separate cage and was monitored to observe the quantity of a .15% lead acetate solution consumed. The data are:

Group	y, Quantity of .15% $Pb(C_2H_3O_2)_2$ consumed										n	$\sum y$	$\sum y^2$
Control	5.4	6.2	3.1	3.8	6.5	5.8	6.4	4.5	4.9	4.0	10	50.6	268.76
Experimental	8.8	9.5	10.6	9.6	7.5	6.9	7.4	6.5	10.5	8.3	10	85.6	752.22

`Note:` Use hand calculation for parts (a) and (b). Show work. Use a JMP analysis for answering parts (c) and (d). Turn in the JMP output.

a.) Test the research hypothesis that the mean quantity of lead acetate consumed by the rats in the experimental group is greater than that consumed by the rats in the control group. First, state H_0 and H_a in terms of appropriate parameters you define. Use a t-test with $\alpha = 0.05$ and assume equal variance for the two populations. Interpret the result.

b.) Compute a 90% confidence interval for the difference in population means in lead consumption between control and experimental groups. Assume equal variance for the two populations. Use this interval to test the hypothesis you stated in part a). State your conclusion.

Use the JMP data table `lead.JMP` to obtain answers to the following questions:

c.) Based on the point estimates of the variances s_1^2 and s_2^2, box plots, and the normal probability plots, does it appear that the population variances for the populations of rats fed control and experimental diets are equal is a plausible assumption? Explain.

d.) Obtain the JMP output necessary to answer the parts a.) and b.) above. `Circle` the t-statistic, the p-value, and the confidence interval in the JMP output. Use the p-value to test the hypothesis stated in part a.) using $\alpha = 0.05$

3. Following the March 24, 1989 grounding of the tanker *Exxon Valdez* in Alaska, approximately 35,500 tons of crude oil were released into Prince William Sound. The paper "The deep benthos of Prince William Sound, Alaska, 16 months after *Exxon Valdez* oil spill" [*Marine Pollution Bulletin(1998), 36:118-130*] reports on an evaluation of deep benthic infauna after the spill. The following data are population abundance (individuals per square meter) at 13 sites: 7 within the oil trajectory and 6 outside the oil trajectory, at a depth of 100 m.

Site	y, Population Abundance (count/meter^2)							n	$\sum y$	$\sum y^2$
Within	3,228	2,032	3,256	3,816	2,438	4,897	1,346	7	21,013	71,448,569
Outside	1,676	2,008	2,224	1,234	1,598	2,182		6	10,922	20,624,700

`Note:` Use hand calculation for parts (a) to (c) and show work. Use an appropriate JMP analysis of the data for answering parts (d),(e), and (f). `Turn in the marked JMP output` and show work.

(a) Calculate the sample means $\bar{y}_{within}$ and $\bar{y}_{outside}$ and sample variances s^2_{within} and $s^2_{outside}$, respectively, of the two samples. Is there evidence to believe that the *variances* of the population abundance for the two populations of sites may be different? Explain.

(b) Calculate, by hand, Welch's two-sample t-statistic and perform a test of the hypothesis that there is a difference in the mean population abundance between the sites within and outside the oil trajectory. Use $\alpha = 0.05$. State H_0 and H_a in terms of appropriate `population parameters you define and label clearly`. Interpret your findings and show your work.

(c) Compute a 90% confidence interval for the difference in population means you defined in part (b) for the two populations assuming unequal population variances.

(d) Report evidence you find from your JMP output (a) for supporting the normality assumption of the two samples, and (b) for supporting the conclusion that the population variances are not equal.

(e) The t-statistic for testing the hypothesis you stated in Part (b) under the unequal variances assumption can be computed in JMP. Circle it in the JMP output and copy here the statistic and the p-value associated with it extracted from the JMP output.

(f) Extract the 90% confidence interval for the difference in population means as in Part (c) computed under unequal variances assumption from the JMP output. Copy it here and circle it in the JMP output.

Due Thursday 23, October 2008 (turn-in during lab)