Questions from*:* <http://www.math.wustl.edu/~jmding/math3200/chw/hw6.html>

1. *SAS code under “Problem 1” Heading. See Output pg. 1 for Q-Q Plot of Data. See Output pg. 2 for means and standard deviations of each sample. For (b), data input as a single sample since a Q-Q plots of two independent samples with the same sample size requires plotting the order statistics x(i) and y(i) against each other.*

(a) The samples are independent because 52 different clouds were randomly seeded to either a control (unseeded) or treatment (seeded) group which are not matched to clouds in the other group by some characteristic.

(b) Since most i/(n+1)th quartiles on the Q-Q plot are closer to the Seeded Cloud Rainfall axis, we conclude that seeded cloud rainfalls tend to be larger than unseeded cloud rainfall.

(c) For this test, let sample 1 be the seeded group let sample 2 be the unseeded group. Assume the samples are large enough to use CLT. The difference in the means of each sample is normally distributed with mean (mu1-mu2) and variance sigma12/n + sigma22/n. For this test, H0: mu1=mu2 and H1: mu1≠mu2. Our Z statistic is (442.1- 164.7)/sqrt (650.8^2/26 + 278.4^2/26)= 1.99827 which is greater than z0.025=1.9599. Thus, we reject the null hypothesis and conclude that there is significant difference between the means at the alpha= 0.05 level.

1. *SAS code under “Problem 2” heading. See Output pg. 3 for a scatter plot of the number of lesions overlaid by a 45 degree line through the origin. See output pg. 4 for the paired t-test statistics.*

(a) These samples are matched pairs since each virus treatment is rubbed onto each of 8 tobacco leaves. In this case, the leaves serve as blocks to naturally pair treatments 1 and 2 for comparison.

(b) The pairs tend to lie below the 45 degree line through the origin; for this scatter plot, 1 point lies above the line, and 7 lie below it. Thus, we conclude that virus 1 tends to produce more lesions than virus 2.

(c) Like in question 1, H0: mu1=mu2 and H1: mu1≠mu2. Our T statistic is 4/ (4.3095/sqrt (8)) = 2.6259 which has a p-value of 0.0341 as reported on the Output page.

1. *SAS code under “Problem 3” heading. See Output pg. 5 for the pooled and unspooled hypothesis test statistics and confidence interval limits for (a) and (b). Let method 1 denote sample 1 and method 2 denote sample 2.*
2. H0: mu1=mu2 and H1: mu1≠mu2. Since we assume that sigma12=sigma22, we use the pooled test statistics from the t-test output. Thus, we have 10+5-2=13 degrees of freedom and a t-value of -2.16 which has a p-value of 0.0499. The p-value is very slightly lower than alpha = 0.05 so we reject the null hypothesis and conclude that there is a significant difference between the means at the alpha=0.05 level.
3. H0: mu1=mu2 and H1: mu1≠mu2. Since we assume that sigma12≠sigma22, we use the unspooled or “Satterthwaite” test statistics from the output. Thus, we have 5.4808 degrees of freedom (calculated from formula 8.11 on page 280) and a t-value of -1.82 which has a p-value of 0.1238. The p-value is greater than alpha = 0.05 so we do not reject the null hypothesis at alpha = 0.05. This is the opposite conclusion as obtained after assuming equal variances; furthermore the p-values differ by 0.0739.

(c) For (a), the confidence interval would be calculated as follows: (12.0026-12.0143) – (t13,.025) [which is 2.160](0.00987)sqrt(1/10 + 1/5) < mu1-mu2 < (12.0026-12.0143) + (t13,.025) [which is 2.160](0.00987)sqrt(1/10 + 1/5). Thus, the 95% confidence interval is [-.0234, -6.59 E-6]. Thus, we are 95% confident that the difference of means is within this interval as calculated from (a). For (b), I will write the confidence interval from the t-test output since the calculation is difficult to write. The 95% confidence interval for the difference of means is [-.0278, 0.00442]. Note, the confidence interval from (b) clearly contains 0 while the interval from (a) barely excludes it. This is consistent with the results of our hypothesis test.

1. *SAS code under “Problem 4” heading. See Output pg.6 for the hypothesis test statistics and the confidence interval limits.*

(a) Let P1 be the probability that a birth is in the 1st month (January) H0: P1= P2=P3=P4=P5=P6=P7 = P8=P9=P10=P11=P12= 1/12= 0.083333. H1: There exists month *i* such that Pi ≠ Pj for some month , 1< j < 12. Thus, our chi square test statistic = sum (observed number – 0.083333\*200)2/ (0.083333\*200) = 19.7257 as reported on the output page, and there are 12-1= 11 degrees of freedom. Also reported, the p-value = 0.0492 which is less than alpha =0.05 (Note: the Chi-square value for this sample is less than X13,.05=22.362). Thus, we reject the null hypothesis and conclude that births are not spread uniformly through the year.

(b) Report p-value of test. As discussed above, the p-value is 0.0492 which corresponds to a chi-square statistic of 19.7257 with 11 degrees of freedom.

**PROBLEM 1:**

TITLE "Homework 6 Question 1";

**data** hw6q1;

input CityType $ Action $ Wallets @@;

datalines;

Big Returned 21 Big Kept 9

Suburbs Returned 18 Suburbs Kept 12

Medium Returned 17 Medium Kept 13

Small Returned 24 Small Kept 6

;

**run**;

**proc** **freq** data=hw6q1 order=data;

table CityType\*Action / chisq;

weight Wallets;

**run**;

**PROBLEM 2:**

TITLE "Homework 6 Question 2";

**data** hw6q2;

input LAST NEXT @@;

datalines;

2.0 50

1.8 57

3.7 55

2.2 47

2.1 53

2.4 50

2.6 62

2.8 57

3.3 72

3.5 62

3.7 63

3.8 70

4.5 85

4.7 75

4.0 77

4.0 70

1.7 43

1.8 48

4.9 70

4.2 79

4.3 72

;

**run**;

**proc** **gplot** data=hw6q2;

plot NEXT\*LAST;

**run**;

**proc** **reg** lineprinter;

model NEXT=LAST / r;

plot NEXT\*LAST= 'Y' predicted.\*LAST'P'/ overlay;

plot residual.\*LAST = '\*';

**run**;

**PROBLEM 3:**

TITLE "Homework 6 Question 3";

**data** hw6q3;

input h l @@;

x=log(l);

y=log(h);

\*For hospital cost h and length of stay l\*;

\*Data entered in single column in SAS. Listed as two columns to save paper\*;

datalines;

13728 13

8062 8

4805 13

5099 6

14963 33

4295 2

4046 9

3193 13

15486 16

9413 11

9034 19

8939 20

17596 26

1884 3

1763 5

1233 1

6286 30

2849 4

2818 4

2265 2

1652 9

1846 4

25460 18

4570 16

12213 10

5870 12

24484 52

4735 19

13334 9

35381 85

5681 8

7161 20

10592 41

;

**proc** **gplot** data=hw6q3;

plot y\*x;

**run**;

**proc** **reg** lineprinter;

model y=x / r clb corrb;

**run**;

**quit**;

Title 'Calculation and Test of Correlations, 95% CI';

ods output FisherPearsonCorr=corr;

**proc** **corr** data=hw6q3 fisher ( biasadj=no );

var x y;

**run**;

**PROBLEM 4:**

TITLE "Homework 6 Question 4";

**data** hw6q4;

input y x1 x2 x3 @@;

label y="Performance IQ" x1="Brain Size" x2="Height" x3= "Weight";

datalines;

124 81.69 64.50 118

150 103.84 73.30 143

128 96.54 68.80 172

134 95.15 65.00 147

110 92.88 69.00 146

131 99.13 64.50 138

98 85.43 66.00 175

84 90.49 66.30 134

147 95.55 68.80 172

124 83.39 64.50 118

128 107.95 70.00 151

124 92.41 69.00 155

147 85.65 70.50 155

90 87.89 66.00 146

96 86.54 68.00 135

120 85.22 68.50 127

102 94.51 73.50 178

84 80.80 66.30 136

74 93.00 74.00 148

86 88.91 70.00 180

84 90.59 76.50 186

134 79.06 62.00 122

128 95.50 68.00 132

102 83.18 63.00 114

131 93.55 72.00 171

84 79.86 68.00 140

110 106.25 77.00 187

72 79.35 63.00 106

124 86.67 66.50 159

132 85.78 62.50 127

137 94.96 67.00 191

110 99.79 75.50 192

86 88.00 69.00 181

81 83.43 66.50 143

128 94.81 66.50 153

124 94.94 70.50 144

94 89.40 64.50 139

89 93.59 75.50 179

;

**proc** **reg** lineprinter;

model y=x1 x2 x3 / r clb corrb;

**run**;

**quit**;

Title 'Calculation and Test of Correlations, 95% CI';

ods output FisherPearsonCorr=corr;

**proc** **corr** data=hw6q3 fisher ( biasadj=no );

var x y;

**run**;