475.101/102/107/108 Semester 2 2000
Assignment 5 Solutions

Question 1.

(a) **(i)** Two independent samples, observational study.
(ii) Paired, experiment.
(iii) Two independent samples, experiment.

(b) We observed 8 positive signs and 2 negative signs. Assuming null hypothesis is true and ignoring hypothesized values of 0 in the data, use $Y \sim$ Binomial(n=10, p=0.5).

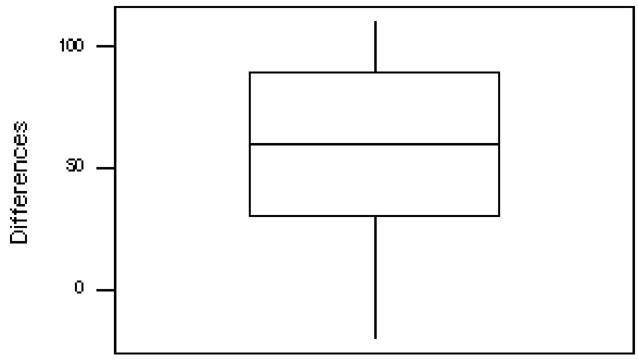
Let y = minimum of the number of +'s and the number of -'s = minimum of 2 and 8 = 2.
P-value $= 2\times\text{pr}(Y \le y) = 2\times\text{pr}(Y \le 2) = 2\times 0.055 = 0.11$.

(c) $\text{df}_1 = k - 1 = 4 - 1 = 3$.
$\text{df}_2 = n - k = (14+11+8+13) - 4 = 46 - 4 = 42$.
$\text{df}_{tot} = \text{df}_1 + \text{df}_2 = 3 + 42 = 45$.
$\text{f}_0 = \dfrac{191.82}{74.88} = 2.562$.

Question 2.

(a) Stem-and-leaf of Differences
N = 17
Leaf Unit = 10

```
  1    -0 2
  1    -0
  1     0
  6     0 22333
  8     0 45
 (2)    0 66
  7     0 8999
  3     1 111
```



The plot of the data show no major departures from Normality. There is one slightly unusual low value (-20) and possible signs of two modes, but no major problems. The small sample size means we cannot read too much into these features.

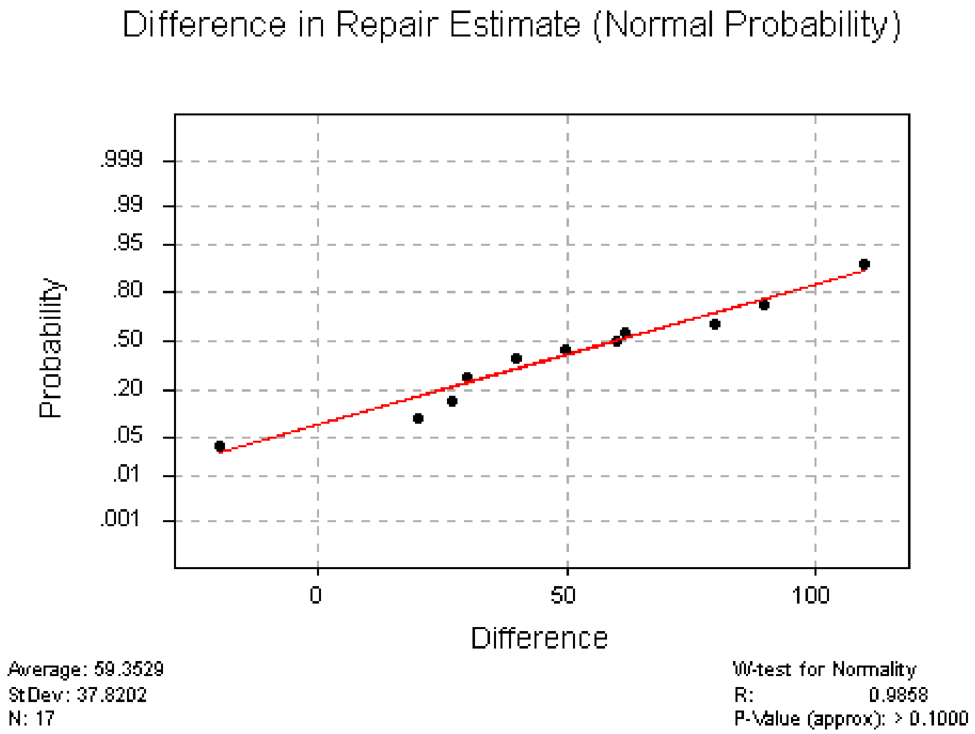
(b) The parameter of interest here is μ_{Diff}, the underlying mean of the differences in the repair estimates between the 2 panel beaters. We wish to test $H_0 : \mu_{Diff} = 0$ vs $H_1 : \mu_{Diff} \neq 0$.

```
Paired T-Test and Confidence Interval

Paired T for Panelbeater1 - Panelbeater2

                      N      Mean      StDev      SE Mean
Panelbeater 1        17      663.2     312.5      75.8
Panelbeater 2        17      603.9     286.6      69.5
Difference           17      59.35     37.82      9.17
```

```
95% CI for mean difference: (39.91, 78.80)
T-Test of mean difference = 0 (vs not = 0): T-Value = 6.47  P-Value =
0.000
```

The *P-value* of 0.000 provides very strong evidence against H_0. Thus we have very strong evidence that the average difference in repair estimates between the two panel beaters is not 0. With 95% confidence the repair estimate for Panel Beater 1 is, on average, between \$39.91 and \$78.80 more than that for Panel Beater 2.

(c)

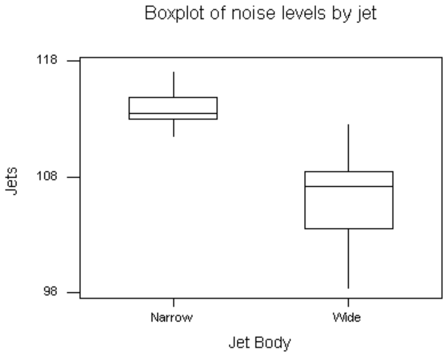


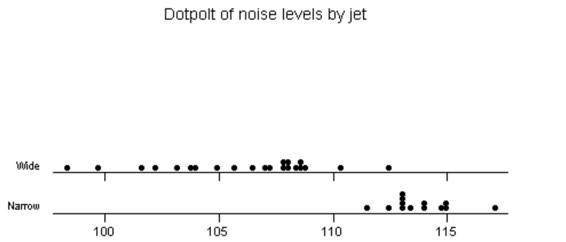
The null hypothesis being tested is that differences in repair estimate between the two panel beaters are Normally distributed versus the alternative hypothesis that they are not Normally distributed. The Normal probability plot and the *W*-test (*P-value* > 0.100) provide no evidence against the null hypothesis, thus indicating that the assumption that the underlying distribution of the differences in repair estimates between Panel Beater 1 and Panel Beater 2 is Normal is reasonable.

(d) From the plots and test in (a) and (c), we appear to be no major problems with assuming that the underlying distribution of the differences is Normal. As there are no major problems with this assumption, we have no reason to doubt the validity of the *t*-test carried out in (b).

Question 3.

(a) (i)





(ii) The plots show that the noise level from narrow-bodied jets is centred higher than the noise level from wide-bodied jets while the data for wide-bodied jets appears to be more spread out.

(iii) It appears that the noise level from wide-bodied jets is skewed to the left while that of narrow-bodied jets is very slightly skewed to the right. Neither data set looks badly skewed.

(b) (i) The parameter of interest here is $\mu_W - \mu_N$ where μ_W is the mean noise level (in decibels) of the wide-bodied jets and μ_N is the mean noise level (in decibels) of the narrow-bodied jets. $H_0: \mu_W - \mu_N = 0$ vs $H_1: \mu_W - \mu_N \neq 0$.

We have very strong evidence (*P-value* = 2.73E-10 ≈ 0.000) of a difference in the mean noise level between wide-bodied and narrow-bodied jets. With 95% confidence the mean noise level of the wide-bodied jets is between 6.04 and 9.46 decibels less than that of the narrow-bodied jets.

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Wide*	*Narrow*
4	Mean	106.022727	113.769231
5	Variance	12.1542208	1.93897436
6	Observations	22	13
7	Hypothesized Mean Difference	0	
8	df	30	
9	t Stat	-9.2481707	
10	P(T<=t) one-tail	1.3673E-10	
11	t Critical one-tail	1.69726036	
12	P(T<=t) two-tail	**2.73E-10**	
13	t Critical two-tail	2.04227035	
14			

(ii) The parameter of interest here is $\tilde{\mu}_W - \tilde{\mu}_N$ where $\tilde{\mu}_W$ is the median noise level (in decibels) of the wide-bodied jets and $\tilde{\mu}_N$ is the median noise level (in decibels) of the narrow-bodied jets. We wish to test $H_0: \tilde{\mu}_W - \tilde{\mu}_N = 0$ vs $H_1: \tilde{\mu}_W - \tilde{\mu}_N \neq 0$.

We have very strong evidence (*P-value* ≈ 0.0000) of a difference in the median noise level between wide-bodied and narrow-bodied jets. With 95% confidence the median noise level of the wide-bodied jets is between 5.4 and 9.4 decibels less than that of the narrow-bodied jets.

(c) Due to the fact that both data sets are not badly skewed and there are no severe signs of non-Normality a 2 independent sample *t*-test is most appropriate for this data.

Question 4.

(a) From the dotplots we can see that lubricant 2 appears to have lower average wear and tear scores than the other lubricants. There does not seem to be large differences between the average wear and tear scores for lubricants 1, 3 and 4. The wear and tear scores for lubricant 1 seem to be much less varied than that for the other lubricants. The wear and tear scores for lubricant 2 seem to be a little right skew.

(b) The outcome of one test should not have any effect on the outcome of another test. They are independent of each other. This satisfies the independence assumption for the *F*-test.

(c) The highest standard deviation is 6.66 for lubricant 2. The lowest is 2.09 for lubricant 1. The ratio of the highest standard deviation to the lowest is 6.66/2.09 = 3.2.

(d) There are doubts about the validity of the *F*-test. The difference in variabilities between the groups is greater than that which is acceptable, even with identical sample sizes. There are no major problems with assuming normality or the independence assumption.

(e) The non-parametric alternative to the *F*-test is the Kruskal-Wallis test.

(f) **(i)** Let μ_1 be the average wear and tear scores for lubricant 1, similarly define μ_2 , μ_3 and μ_4 for lubricants 2, 3 and 4. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_0 : The mean wear and tear score is the same for the four different lubricants.

(ii) H_1 The mean wear and tear score is different for at least two of the four lubricants.

(iii) As the *P-value* from the *F*-test is 0.000, we have extremely strong evidence against the null hypothesis. Thus, we have extremely strong evidence that the mean wear and tear score is different for at least two of the four different lubricants.

(g) **(i)** We can not determine which single lubricant has the lowest mean wear and tear score.

(ii) We can not determine which single lubricant has the highest mean wear and tear score.

(iii) We estimate, with 95% confidence, that the mean wear and tear score for lubricant 2 is between 5.3 and 16.2 units lower than the mean wear and tear score for lubricant 3.

(iv) There are significant differences in the average wear and tear scores at the 5% level between: lubricants 1 and 3; lubricants 2 and 3; lubricants 2 and 4.

Question 5.

(a) The most appropriate design for this experiment would be to use three samples, one for each of the two daily dosages of antibiotic and one control group (that received no antibiotic), and use one-way analysis of variance. There must be some form of blocking on the weight of the pigs. For example, the pigs could be formed into two blocks of 30: the heaviest 30 pigs and the lightest 30 pigs. Pigs from the heavier block are then randomly allocated to each treatment group and similarly for pigs from the lighter block. The hypotheses we would be interested in testing would be: H_0: The mean weight gain is the same for the three groups versus H_1: At least one of the mean weight gains is different from another.

(b) The most appropriate design for this experiment would be a paired design in order to cut down on the variability (which might otherwise overshadow any difference in means) of the bacteria counts between subjects. There are 32 subjects available. Each subject will wear a ring on one day and not wear a ring on the other. Randomly allocate 16 subjects to wear a ring on the first day. The hypotheses we would be interested in testing would be: H_0: $\mu_{\text{diff}} = 0$ versus H_1: $\mu_{\text{diff}} \neq 0$ where μ_{diff} is the mean difference in the bacteria counts.

Question 6. [6 marks]

(a) "It skips over the fact that the way most polls are done, one in 20 will be flat wrong – that is, outside the margin of error's seven point envelope."

The margin of error has been determined at a **95% confidence level**.

(b) **Selection bias:** These polls are telephone polls and miss the 6% of New Zealanders who have no phone and also those who are unlisted. **Nonresponse bias:** Some people are difficult to contact and some people don't want to give answers the questions. I.e., there are a group of people with whom the pollsters can't talk.

(c) **understates** the true margin of error. **(d)** **overstates** the true margin of error.