# STATISTICS 3N03/3J04 – TEST #3 SOLUTIONS

**Question 1a**

**Paired data *t*-test is the correct parametric analysis.**
***[12 marks if all of the following is given; maximum 8 marks for a wrong analysis.]***

**Assumptions:** Normality (can't test with such a small sample but it looks OK on a stem and leaf plot or dot plot); independence (can't test: sample is small and the observations are not in any particular order).

**Conclusion:** There is no evidence from these data (P = 0.61) that the mean noise level is different in acceleration and deceleration lanes in Bangkok. Note: using the textbook tables we get 2-sided 0.8 > P > 0.5.

```
> bangkok
    acc  dec diff
1  78.1 78.6 -0.5
2  78.1 80.0 -1.9
3  79.6 79.3  0.3
4  81.0 79.1  1.9
5  78.7 78.2  0.5
6  78.1 78.0  0.1
7  78.6 78.6  0.0
8  78.5 78.8 -0.3
9  78.4 78.0  0.4
10 79.6 78.4  1.2

> stem(bangkok$diff)

  The decimal point is at the |

  -1 | 9
  -0 | 53
   0 | 01345
   1 | 29

> t.test(bangkok$acc,bangkok$dec,pair=T)

      Paired t-test

data:  bangkok$acc and bangkok$dec
t = 0.5311, df = 9, p-value = 0.6082
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.5540866  0.8940866
sample estimates:
mean of the differences
                   0.17
```

**Sign test is the correct nonparametric analysis**
***[5 marks if all of the following is given.]***

**Conclusion:** Out of 9 non-zero differences, 3 were negative, so a 2-sided P-value is twice the left tail of Bin(9, 0.5). There is no evidence from these data (P = 0.51) that the mean noise level is different in acceleration and deceleration lanes in Bangkok.

The *t*-test is more powerful than the sign test. The sign test is more robust than the *t*-test because it does not assume normality.

```
> 2*pbinom(sum(bangkok$diff<0), sum(bangkok$diff!=0), 0.5)
[1] 0.5078125
```

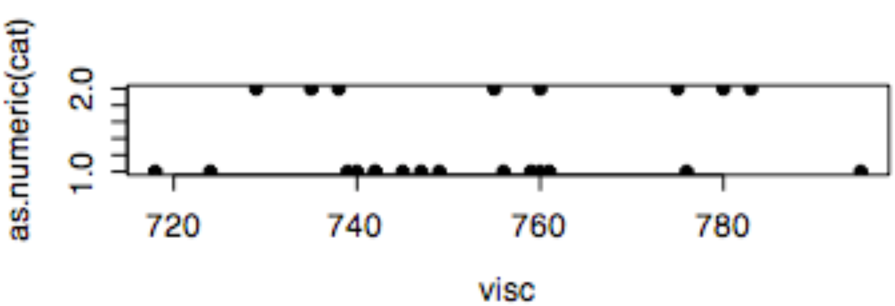
**Question 1b**

**Independent-sample *t*-test is the correct parametric analysis.**
***[18 marks if all of the following is given, including the F-test; maximum 8 marks for a wrong analysis.]***

**Assumptions:** Normality (can't test with such a small sample but it looks OK on comparative stem and leaf or dot plots); independence within and between samples (can't test: samples are small and the observations are not in any particular order); homoscedasticity (accepted by the *F*-test below).

**Conclusion:** There is no evidence from these data (P = 0.45) that the mean viscosity is different after changing the catalyst. Note: using the textbook tables we get 2-sided 0.5 > P > 0.2.

```
> catalyst
   visc cat
1   724   A
2   718   A
3   776   A
4   760   A
5   745   A
6   759   A
7   795   A
8   756   A
9   742   A
10  740   A
11  761   A
12  749   A
13  739   A
14  747   A
15  742   A
16  735   B
17  775   B
18  729   B
19  755   B
20  783   B
21  760   B
22  738   B
23  780   B
```




```
> t.test(visc~cat,catalyst,var.equal=T)
```

```
      Two Sample t-test

data:  visc by cat
t = -0.7672, df = 21, p-value = 0.4515
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -24.76795  11.41795
sample estimates:
mean in group A mean in group B
        750.200         756.875
```

**Two-sided *F*-test is the correct test for homoscedasticity.**

**Assumptions:** Same as for the *t*-test. Normality (can't test with such a small sample but it looks OK on comparative stem and leaf or dot plots); independence within and between samples (can't test: samples are small and the observations are not in any particular order).

**Conclusion:** $F_0 = 1.238$, so there is no evidence from these data ($P = 0.69$) that the variance in viscosity is different after changing the catalyst. Note: using the textbook tables we get 2-sided $P > 0.5$.

```
> catvar<-sapply(split(catalyst$visc,catalyst$cat),var)
> catvar
       A        B
365.8857 452.9821
> sqrt(catvar)
       A        B
19.12814 21.28338
> (14*catvar[1]+7*catvar[2])/21
       A
394.9179
> sqrt((14*catvar[1]+7*catvar[2])/21)
       A
19.87254
> catvar[2]/catvar[1]
       B
1.238043
> 2*(1-pf(catvar[2]/catvar[1],7,14))
        B
0.6926031
```

**Question 2**

***[5 marks.]***

Here, $\alpha = 0.05$, $\beta = 0.01$, $\delta = 0.2$, and we use $\sigma^2 = s_d^2 = 1.024$. From tables, $z_{0.025} = 1.960$ and $z_{0.01} = 2.326$. We see that 471 locations would be required.

```
> var(bangkok$diff)
[1] 1.024556
> var(bangkok$diff)*(qnorm(.975)+qnorm(.99))^2/(0.2^2)
[1] 470.5904
```