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

Question 1a

Paired data *t*-test is the correct parametric analysis.
[11 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; try a stem and leaf plot or dot plot but can't really say); independence (can't test: sample is small and the observations are not in any particular order).

Conclusion: There is evidence from these data (P = 0.006) that the mean intake does not equal the mean expenditure in these players. Note: using the textbook tables we get 2-sided 0.1 > P > 0.05.

```
> soccer
  expen intake diff
1  14.4   14.4  0.0
2  12.1    9.2  2.9
3  14.3   11.8  2.5
4  14.2   11.6  2.6
5  15.2   12.7  2.5
6  15.5   15.0  0.5
7  17.8   16.3  1.5
> t.test(soccer$expen,soccer$intake,pair=T)

      Paired t-test

data:  soccer$expen and soccer$intake
t = 4.1309, df = 6, p-value = 0.006141
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.7279498 2.8434788
sample estimates:
mean of the differences
               1.785714

> stem(soccer$diff)

  The decimal point is at the |

  0 | 05
  1 | 5
  2 | 5569
```

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

Conclusion: Out of 6 non-zero differences, 0 were negative, so a 2-sided P-value is twice the left tail of Bin(6, 0.5). There is evidence from these data (P = 0.031) that the median intake does not equal the median expenditure in these players.

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. In this example, they both lead to the same conclusion but the *t*-test gives a smaller P-value.

```
> 2*pbinom(0,6,.5)
[1] 0.03125
```

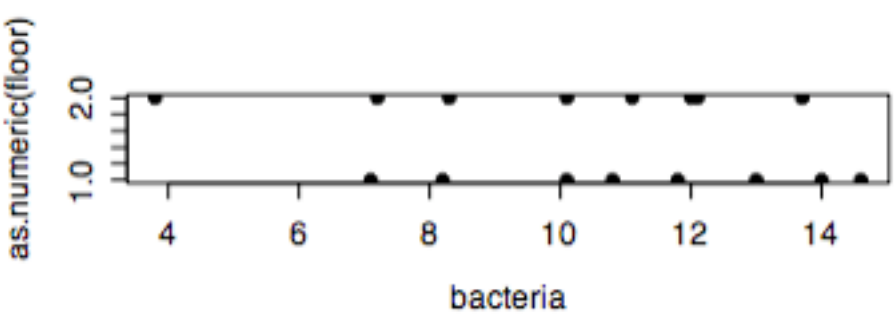
Question 1b

Independent-sample *t*-test is the correct parametric analysis.
[16 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 = 0.35) that the mean airborne bacteria is different in carpeted and uncarpeted rooms. Note: using the textbook tables we get 2-sided 0.5 > P > 0.2.

```
> airborne
   bacteria  floor
1      11.8   Carp
2       8.2   Carp
3       7.1   Carp
4      13.0   Carp
5      10.8   Carp
6      10.1   Carp
7      14.6   Carp
8      14.0   Carp
9      12.1 Uncarp
10      8.3 Uncarp
11      3.8 Uncarp
12      7.2 Uncarp
13     12.0 Uncarp
14     11.1 Uncarp
15     10.1 Uncarp
16     13.7 Uncarp
```




```
> t.test(bacteria~floor, airborne, var.equal=T)

      Two Sample t-test

data:  bacteria by floor
t = 0.9558, df = 14, p-value = 0.3554
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.757205  4.582205
sample estimates:
  mean in group Carp mean in group Uncarp
             11.2000               9.7875
```

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: $F_0 = 1.4374$, so there is no evidence from these data ($P = 0.64$) that the variance in airborne bacteria is different in carpeted and uncarpeted rooms. Note: using the textbook tables we get 2-sided $P > 0.5$.

```
> varbact <- sapply(split(airborne$bacteria,airborne$floor),var)
> varbact
     Carp    Uncarp
 7.168571 10.304107
> varbact[2]/varbact[1]
  Uncarp
1.437400
> 2*(1-pf(varbact[2]/varbact[1],7,7))
   Uncarp
0.6440893
```

Question 2
[5 marks.]

Here, $n_1 = 8$, $n_2 = 8$, $\alpha = 0.01$, $\delta = 2$, and we use $\sigma^2 = s_p^2 = 8.73634$. From tables, $z_{0.005} = 2.576$. We find that the chance of a Type II error is 89%.

```
> (7*varbact[1]+7*varbact[2])/14
   Carp
8.73634
> pnorm(qnorm(.995)-2/sqrt(mean(varbact)*2/8))-pnorm(-qnorm(.995)-
2/sqrt(mean(varbact)*2/8)
+ ))
Error: syntax error
> pnorm(qnorm(.995)-2/sqrt(mean(varbact)*2/8))-pnorm(-qnorm(.995)-
2/sqrt(mean(varbact)*2/8))
[1] 0.889203
```

Question 2
[3 marks.]

William Sealey Gosset + 3 interesting facts.