BIOS 662, Fall 2018
Solution to Midterm Exam

Question 1

Question 2

Investigation of potentially incorrect data values: Below is a list of reasons for setting various values to missing, with the IDs of the corresponding women given in parentheses.

- A gestational age of 75 is impossible (GA_ultra for M2349).
- Months are numbered 1 through 12, so rand_month values of 0 (M1190) and 15 (M1722) are errors.
- Number of previous pregnancies cannot be negative, so values of -9 are errors (M1190 and M1410).

(a) A histogram, stem-and-leaf plot or boxplot suggests substantial skewness in the data, with a long tail towards lower values of GA. A QQ plot also shows substantial departure from normality See, for instance, Figure 1. To test for normality, use the Kolmogorov-Smirnov test. We need the Lilliefors version to adjust for having to estimate the mean and variance.

H_0 : GA_est is normally distributed; H_A : GA_est is not normally distributed.

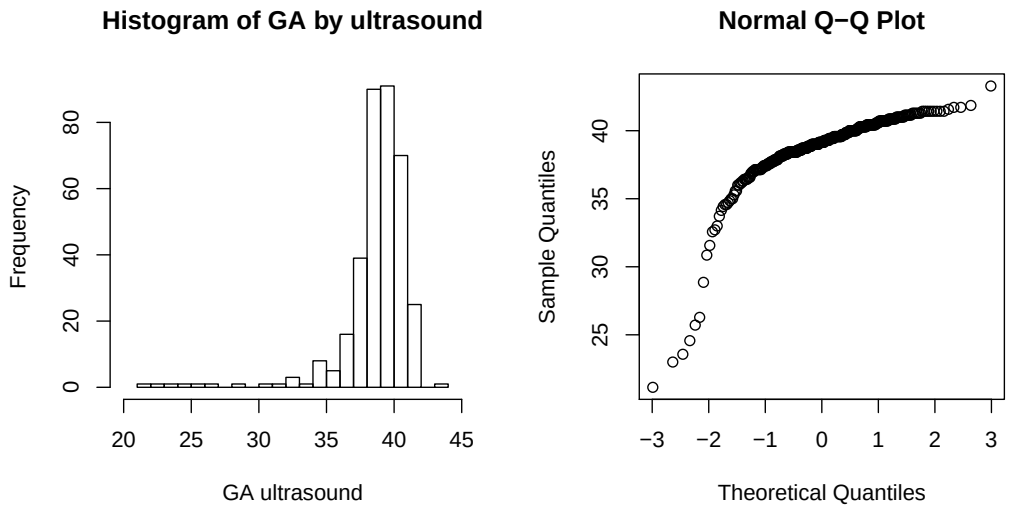
SAS gives the p-value for the Lilliefors version automatically; R needs the function lillie.test.

In the SAS output we see $D = 0.186$ with $p < 0.01$, so we reject H_0 and conclude that GA_est is not normally distributed.

```
proc univariate normal;
   var ga_ultra;
```

```
                  Tests for Normality

Test                   --Statistic---    -----p Value------

Shapiro-Wilk           W     0.698496    Pr < W     <0.0001
Kolmogorov-Smirnov     D     0.185894    Pr > D     <0.0100
Cramer-von Mises       W-Sq  3.729445    Pr > W-Sq  <0.0050
Anderson-Darling       A-Sq  22.17517    Pr > A-Sq  <0.0050
```

Figure 1: Histogram and normal QQ plot for GA by ultrasound



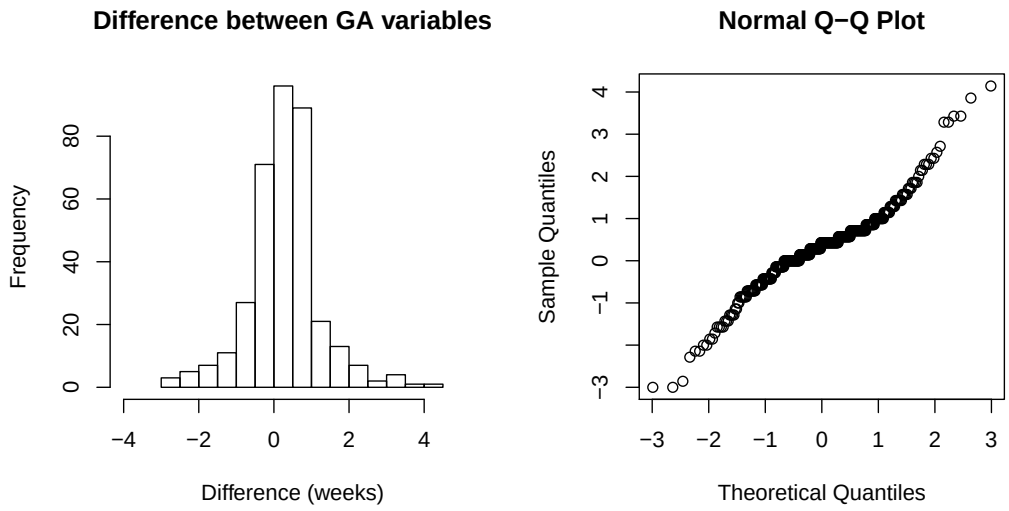
(b) Because the data are paired (each woman has gestational age estimated by each method), we don't have two independent samples. So let $Y_i = X_{1i} - X_{2i}$ where X_{1i} is the GA by ultrasound and X_{2i} the GA estimated at birth for woman i. Assume $E(Y_i) = \mu$ for all i. The observations on different women are independent. We want to test $H_0 : \mu_{\text{diff}} = 0$ versus $H_A : \mu_{\text{diff}} \neq 0$. The histogram in Figure 2 shows that the distribution of differences is reasonably symmetric but the QQ plot suggests that it may not be reasonable to assume normality. The sample size is large, so we will use the CLT / Slutsky's Theorem to give a test using the Z statistic. (Using a t-test would also be reasonable.) The critical region is $C_{0.05} = \{Z : |Z| > 1.96\}$.

$$Z = \frac{\bar{Y} - 0}{s/\sqrt{n}} = \frac{0.3180}{0.960/\sqrt{358}} = 6.268 > 1.96.$$

So we reject H_0 and conclude that the mean GA by ultrasound differs from the mean GA estimated at birth. The sample means are 38.73 and 38.41 months respectively, so the mean GA by ultrasound appears to be larger than the mean GA estimated at birth. A t-test gives similar results (see SAS output below).

```
The TTEST Procedure

Variable:  ga_diff

   N        Mean     Std Dev
 358      0.3180      0.9601

    DF    t Value    Pr > |t|
   357       6.27      <.0001
```

Figure 2: Histogram and normal plot for differences between GA variables



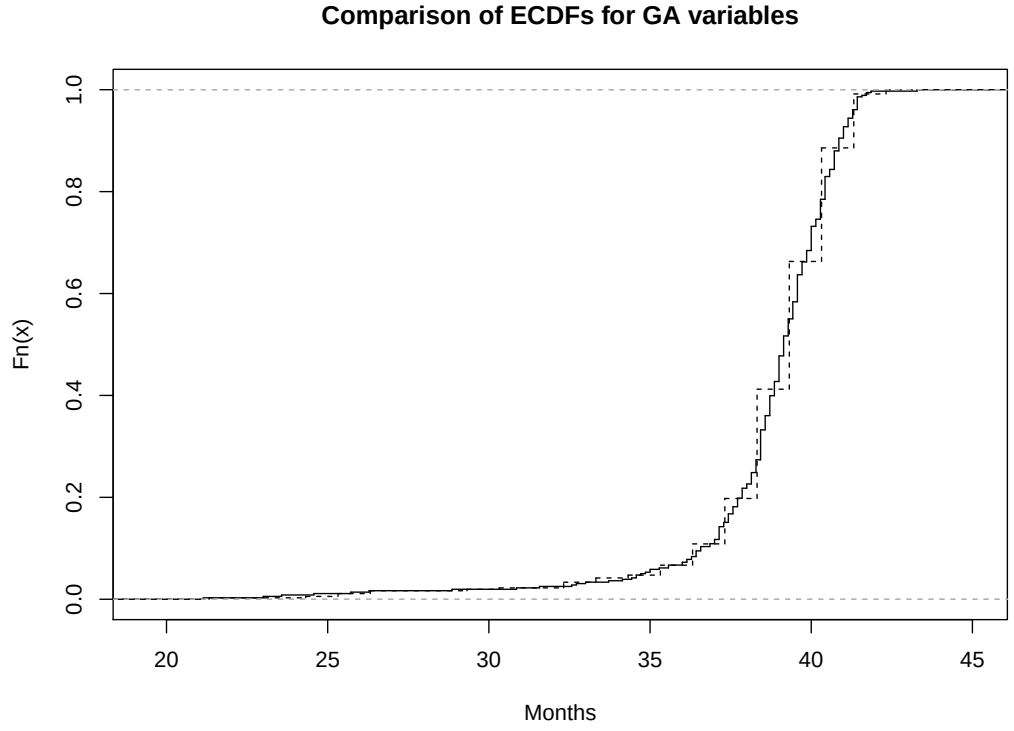
(c) Because the means differ by 0.3180 months, we add this to each GA_est observation. This eliminates the difference between the means. One way to test whether there are other differences between the distributions after eliminating the difference between the means is to use the Kolmogorov-Smirnov test. The KS test in this situation assumes independence of the two samples, but here the two measures are used for each infant and so the assumption of independence is violated. As we don't have a test that does not make the independence assumption we will use the KS test even though it is not ideal.

We want to test $H_0 : F_1(y) = F_2(y)$ for all y versus $H_A : F_1(y) \neq F_2(y)$ for at least one y, where F_1 and F_2 are the CDFs of the two GA variables. To run the KS test we first need to create a single variable that has the GA values of both types, along with an indicator of which are the GA_ultra and which the GA_est values. In the SAS code that follows these are the variables GA and GA_GROUP, respectively. The "mc" in the last line of the code is so that SAS doesn't take forever to run. The p-values for both the asymptotic and the exact versions of the test are < 0.05, so we reject the hypothesis that the two distributions have the same shape. However, looking at the plot in Figure 3, the only noticeable difference is that the step function for the GA_EST version has bigger steps because it is given in whole weeks whereas the other version is in weeks plus days. So, although the difference is statistically significant it isn't really a meaningful difference.

A test for equality of variances for the two variables does not reject the null hypothesis of equality (though here too the assumption of independence of the two samples is not met. The histogram in Figure 4 shows that the distribution of differences is reasonably symmetric.

```
proc npar1way;
   class ga_group;
   var ga;
   exact ks / mc;

  Kolmogorov-Smirnov Two-Sample Test

D = max |F1 - F2|                0.1397
Asymptotic Pr >  D               0.0019

Monte Carlo Estimate
Exact Pr >= D                    0.0010
```

Figure 3: ECDFs for the GA variables after eliminating the difference in means



(d) The following table has the two versions of GA classified as stated. The observed proportion of agreement is $p_a = (33 + 170 + 91)/359 = 0.82$. As seen from the SAS output, the chance-corrected measure of agreement is $\kappa = 0.68$ and the associated 95% CI for the true agreement is $(0.61, 0.75)$. The agreement is reasonable, though not great.

Figure 4: Histogram of the difference between the GA variables after eliminating the difference in means

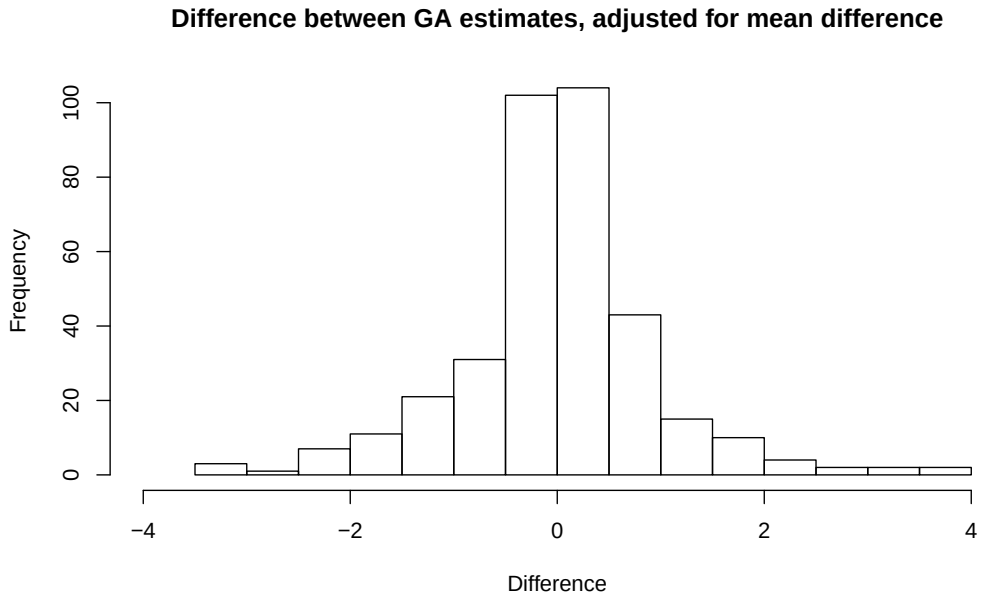


Table of ga_cat by ga_catp

ga_cat	ga_catp			
Frequency	1. LT 37	2. 37-40	3. GE 40	Total
1. LT 37	33	6	0	39
2. 37-40	6	170	30	206
3. GE 40	0	22	91	113
Total	39	198	121	358

Frequency Missing = 1

Kappa Statistics

Statistic	Value	ASE	95% Confidence Limits	
Simple Kappa	0.6826	0.0366	0.6108	0.7544
Weighted Kappa	0.7182	0.0335	0.6526	0.7838

Effective Sample Size = 358
Frequency Missing = 1

(e) If randomization is spread evenly across the calendar year, we would expect the proportion of women randomized in January to be 31/365, the proportion in February to be 28/365, etc. We need to conduct a goodness of fit test to see whether the distribution of the number of women randomized is consistent with this. This is similar to the example starting on page 33 of the overheads on "Categorical Data: Contingency Tables" with 12 probabilities rather than just 3.

$$H_0 : \pi_{\text{Jan}} = 31/365 = 0.084932, \pi_{\text{Feb}} = 28/365 = 0.076712, \ldots \pi_{\text{Dec}} = 31/365 = 0.084932$$

$$H_A : \text{at least one of the equalities is false.}$$

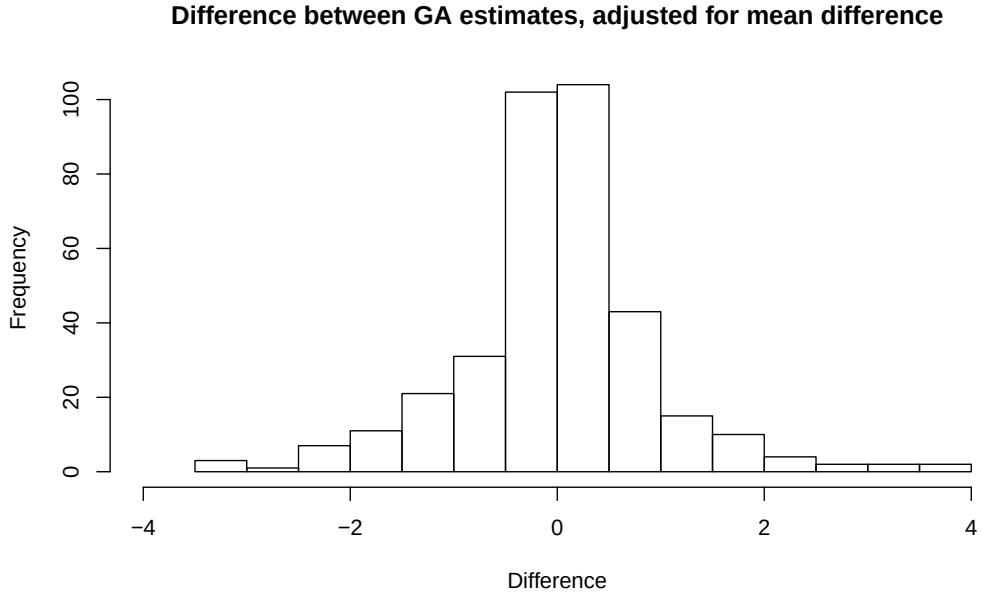
Using the SAS code and output below, $p = 0.0004$, so we reject H_0 and conclude that the proportion of women randomized each month is not consistent with the number of days in the month. Some slight variation from expected is likely to be because of the number of weekend days in a particular month. But we see that December, in particular, has a much lower percent randomized than expected just by the length of the month. Because of holidays in December, which means fewer working days and also potential participants having other priorities than being in a clinical trial, trials often struggle to randomize many participants in that month.

```
proc freq data=bw;
   table rand_month / testp=(8.4932 7.6712 8.4932 8.2192 8.4932 8.2192
        8.4932 8.4932 8.2192 8.4932 8.2192 8.4932);

The FREQ Procedure
                                              Test
rand_month    Frequency     Percent     Percent
-------------------------------------------------
         1           22        6.16        8.49
         2           17        4.76        7.67
         3           42       11.76        8.49
         4           24        6.72        8.22
         5           41       11.48        8.49
         6           35        9.80        8.22
         7           40       11.20        8.49
         8           22        6.16        8.49
         9           32        8.96        8.22
        10           39       10.92        8.49
        11           29        8.12        8.22
        12           14        3.92        8.49

     Chi-Square Test
for Specified Proportions
-------------------------
Chi-Square        33.4334
DF                     11
Pr > ChiSq         0.0004

Effective Sample Size = 357
Frequency Missing = 2
```

(f) Table 1 gives percentages of women randomized by month and live births by month. The percentage of birth each month is much more even than that of the number of women randomized. (I said not to do a test, but if one does the same test as in (e) for the births it yields $p = 0.5$.) That could be because the number of births in a month is not influenced substantially by the number of holidays in the month. (It is probably also because I had an error when I created the dataset – I used the month from each woman's date of birth rather than from her infant's date of birth.)

Month	1	2	3	4	5	6	7	8	9	10	11	12
Randomized	6.2	4.8	11.8	6.7	11.5	9.8	11.2	6.2	9.0	10.9	8.1	3.9
Births	6.1	6.4	8.6	7.5	7.5	9.7	9.5	9.7	8.6	11.4	7.2	7.5

Table 1: Percentages of women randomized each month and births each month



(g) We use a χ^2 test of trend. Let ρ_i denote the probability of preterm delivery in previous pregnancy category i. We want to test $H_0 : \rho_0 = \rho_1 = \rho_2 = \rho_{3+}$ against $H_A : \rho_0 \leq \rho_1 \leq \rho_2 \leq \rho_{3+}$ or $\rho_0 \geq \rho_1 \geq \rho_2 \geq \rho_{3+}$ with at least one of the inequalities being strict. From the SAS output, $X^2_{\text{trend}} = (-0.0335)^2 = 0.0011$ with $p = 0.9733$. So we do not reject H_0 and conclude there isn't much evidence for a monotonic trend in the risk of preterm delivery with the number of previous pregnancies.

```
Table of preterm by ppnum

preterm     ppnum

Frequency|
Col Pct  | 0      | 1      | 2      |3+      |  Total
---------+--------+--------+--------+--------+
       0 |     57 |    117 |     72 |     28 |    274
         |  85.07 |  92.13 |  88.89 |  84.85 |
---------+--------+--------+--------+--------+
       1 |     10 |     10 |      9 |      5 |     34
         |  14.93 |   7.87 |  11.11 |  15.15 |
---------+--------+--------+--------+--------+
Total          67      127       81       33      308

Frequency Missing = 51

Cochran-Armitage Trend Test
---------------------------
Statistic (Z)       -0.0335
One-sided Pr <  Z    0.4867
Two-sided Pr > |Z|   0.9733
```

(h) Now we want to test whether there is an association between treatment group and preterm delivery. We use a χ^2 test of association or, equivalently, test whether the proportion of preterm deliveries is the same in the two treatment groups. We want to test H_0 : preterm delivery is independent of treatment group, versus H_A : preterm delivery is associated with treatment group. The χ^2 test yields $p = 0.14$, hence we do not reject the null hypothesis. From the SAS output we see that the preterm percentages in the two groups are 13.5 and 8.6, that is, the percentage is actually nominally higher in the prenatal treatment group. So there is no evidence that treatment of periodontal disease reduces the risk of preterm delivery.

```
group     preterm

Frequency|
Row Pct  |       0|       1|  Total
---------+--------+--------+
       1 |    148 |     23 |    171
         |  86.55 |  13.45 |
---------+--------+--------+
       2 |    171 |     16 |    187
         |  91.44 |   8.56 |
---------+--------+--------+
Total         319       39      358

Statistics for Table of group by preterm

Statistic                     DF       Value      Prob
------------------------------------------------------
Chi-Square                     1      2.2040    0.1376
```

(i) We have pocket depth measurements on each woman at two time points (apart from some missing values, which we exclude from this analysis). Because the two measurements on each woman are not independent, we cannot use a two-sample test. Instead, we calculate the difference and test whether the mean of the differences is zero, that is, $H_0 : \mu_{\text{diff}} = 0$ versus $H_A : \mu_{\text{diff}} \neq 0$. Taking the difference as the pocket depth after delivery minus the pocket depth at baseline, the difference will be positive if pocket depth has increased (that is, periodontal disease has progressed) and negative if it has declined (that is, if there has been an improvement).

Even after omitting missing values, the sample size is large ($n = 286$), so we can rely on the CLT and Slutsky's Theorem to give a test using the Z statistic.

The critical region is $C_{0.05} = \{Z : |Z| > 1.96\}$.

Using the SAS output below, the test statistic is $z = (-0.0835 - 0)/\sqrt{0.2173/286} = -3.03 > 1.96$. The associated p-value is 0.002. (Using a one-sample t-test gives a very similar p-value – see the SAS output.)

So we reject H_0 and conclude that the mean average pocket depth changed from baseline to delivery. The estimate of mean change is -0.08mm. That is average pocket depth became slightly worse, even though about half of the participants had their periodontal disease treated in the interim.

```
N                          286    Sum Weights               286
Mean                -0.0835129    Sum Observations    -23.884703
Std Deviation       0.46619515    Variance            0.21733791

           Tests for Location: Mu0=0

Test           -Statistic-    -----p Value------

Student's t    t  -3.02949    Pr > |t|    0.0027
```

(j) Now we do have two independent sets of measurements – the data from the two treatment groups (with the data within each group being the difference between the pocket depth measurements at the two time points, as in part (e)). The sample sizes in the two groups are still large ($n_1 = 61$ in the prenatal group and $n_2 = 62$ in the post-partum group), so we can again rely on the CLT and Slutsky's Theorem to give a test using the Z statistic. Here $H_0 : \mu_{\text{diff},1} = \mu_{\text{diff},2}$ versus $H_A : \mu_{\text{diff},1} \neq \mu_{\text{diff},2}$.

The critical region is again $C_{0.05} = \{Z : |Z| > 1.96\}$.

$$Z = \frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(0.0130 - (-0.1662)) - 0}{\sqrt{\frac{0.4910^2}{132} + \frac{0.4284^2}{154}}} = 3.26$$

As $3.26 > 1.96$ we reject the null hypothesis and conclude there is a difference between the change in pocket depth in the prenatal treatment group compared with the post-partum treatment group. The associated p-value is 0.0011. Using a t-test yields similar results

```
The TTEST Procedure

Variable:  pd_chng

group            N        Mean     Std Dev

1              132      0.0130      0.4910
2              154     -0.1662      0.4284
Diff (1-2)              0.1792      0.4583


group         Method               Mean       95% CL Mean

1                                0.0130     -0.0716   0.0975
2                               -0.1662     -0.2344  -0.0980
Diff (1-2)    Pooled             0.1792      0.0721   0.2862
Diff (1-2)    Satterthwaite      0.1792      0.0710   0.2873

Method           Variances        DF    t Value    Pr > |t|

Pooled           Equal           284       3.30      0.0011
Satterthwaite    Unequal      262.16       3.26      0.0013
```

(k) In part (i) we saw that overall there is a very marginal improvement in the mean pocket depth. From part (j) we see that the change in pocket depth differs significantly between the two treatment groups. Looking at the point estimates and confidence intervals for the change within each group (in the SAS output in part (j), we see that pocket depth tended to get worse in the post-partum treatment group. For the prenatal treatment group the point estimate shows a small improvement in mean pocket depth, but the associated 95% confidence interval includes 0, so there is not a statistically significant (or clinically meaningful improvement. So, even if effective periodontal therapy does have the potential to reduce the risk of premature birth or low birthweight, the periodontal therapy given in this study does not appear to have had much effect on periodontal disease, but is better than leaving the periodontal disease untreated and so may have an effect on birth outcomes.

Question 3

(a) Below is a 2×2 table of exposure by case status along with the column percentages, that is the percentage in each exposure category among cases and among controls.

	Control	Case	Total
Exposed = 0	69	51	120
	79.3	58.6	
Exposed = 1	18	36	54
	20.7	41.4	
Total	87	87	174

Let π_1 be the probability of being exposed in the control group and π_2 the corresponding probability in the case group. We want to test $H_0 : \pi_1 = \pi_2$ versus $H_A : \pi_1 \neq \pi_2$. The sample size is large enough to use a χ^2 test of association. The critical region is $C_{0.05} = \{X^2 : X^2 > \chi^2_{1,0.95} = 3.84\}$. From the SAS output below we see that $X^2 = 8.7 > 3.84$ and the corresponding p-value is 0.003. Thus we reject the null hypothesis and conclude that moderate to severe periodontal disease is associated with premature birth.

```
Statistic                     DF       Value      Prob
------------------------------------------------------
Chi-Square                     1      8.7000    0.0032
```

(b) Because this is a case-control study, the appropriate measure of association is the odds ratio. From the 2×2 table we obtain

$$\widehat{\text{OR}} = \frac{n_{11}n_{22}}{n_{21}n_{12}} = \frac{69 \times 36}{51 \times 18} = 2.71.$$

This is confirmed by the SA output below, which gives the 95% CI as $(1.38, 5.30)$.

```
            Estimates of the Common Relative Risk (Row1/Row2)

Type of Study     Method                  Value     95% Confidence Limits
-----------------------------------------------------------------------
Case-Control      Mantel-Haenszel        2.7059       1.3823       5.2967
  (Odds Ratio)    Logit                  2.7059       1.3823       5.2967
```

(c) Using the Mantel-Haenszel method, from the SAS code and output below we obtain an estimated odds ratio (adjusted for age) of 3.05, with 95% CI $(1.48, 6.28)$.

```
proc freq;
   table age_group*exposed*case / norow nopercent cmh;
```

```
          Estimates of the Common Relative Risk (Row1/Row2)

Type of Study     Method                  Value     95% Confidence Limits
-------------------------------------------------------------------------
Case-Control      Mantel-Haenszel        3.0506       1.4819       6.2802
  (Odds Ratio)
```

(d) The adjusted odds ratio of 3.05 is reasonably similar to the unadjusted one of 2.71, so age group does not appear to be a substantial confounder of the association between periodontal disease and premature birth. Investigating whether age group is associated with both the exposure and the outcome will show that numbers of cases and controls are equal within each age group (because of the matching on age), so age and case status are not associated in this dataset. From separate 2×2 tables for the three age groups, we obtain estimated odds ratios of 2.23, 4.86 and 9.63, respectively. This suggests that the odds ratios are not homogeneous across the age groups and so it may not be appropriate to pool them using the Mantel-Haenszel estimator. (We have not covered a test of homogeneity of the odds ratios and I did not expect you to look for such a test. The Breslow-Day test for homogeneity does not reject the null hypothesis of homogeneity. This test is part of the output from the SAS code above.)

```
     Breslow-Day Test for
Homogeneity of the Odds Ratios
------------------------------
Chi-Square              1.8405
DF                           2
Pr > ChiSq              0.3984
```

(e) For this part we need the 2×2 table in a different form. To obtain this in SAS, we need each pair to be a single observation in the SAS dataset. One way to do this is to rename the exposure variables so that those for cases and controls are distinct, split the dataset into two, one consisting of cases, the other of controls, and then merge the two on the part of the ID that is common to the members of a pair. This yields the following table.

		Controls		
		$E = 0$	$E = 1$	Total
Cases	$E = 0$	44	7	51
	$E = 1$	25	11	36
		69	18	87

We want to test $H_0 : \pi_1 = \pi_2$ versus $H_A : \pi_1 \neq \pi_2$. We do so using McNemar's test statistic. Here $n_{12} + n_{21} = 7 + 25 = 32 > 30$, so we can use the χ^2 approximation. The critical region is $C_\alpha = \{\mathrm{M} : \mathrm{M} > \chi^2_{1,0.95}\} = \{\mathrm{M} : \mathrm{M} > 3.84\}$.

$$\mathrm{M} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} = \frac{(7-25)^2}{7+25} = 10.1 > 3.84.$$

So we reject the null hypothesis and conclude that moderate to severe periodontal disease is associated with premature birth. From the SAS output we see that the associated p-value is 0.0015.

```
Statistic (S)         10.1250
DF                          1
Asymptotic Pr >  S     0.0015
Exact      Pr >= S     0.0021
```

To estimate the odds ratio, whether one uses n_{12}/n_{21} or n_{21}/n_{12} depends on how the table is set up. Here $\widehat{\mathrm{OR}}_\mathrm{M} = n_{21}/n_{12} = 25/7 = 3.57$.

$$\widehat{\mathrm{Var}}(\ln(\widehat{\mathrm{OR}}_\mathrm{M})) \approx \frac{1}{n_{12}} + \frac{1}{n_{21}} = \frac{1}{25} + \frac{1}{7} = 0.183$$

So an approximate 95% CI on the log scale is $\log(3.57) \pm 1.96 \cdot \sqrt{0.183}$, that is $(0.435, 2.111)$. On the original scale this becomes $(e^{0.435}, e^{2.111}) = (1.54, 8.23)$.

(f) The estimate in (d) is most appropriate because it takes into account the matched case-control design. Although part (c) takes into account age group, it assumes frequency matching rather than individual matching and doesn't take into account the second matching factor (number of previous pregnancies).

(g) The estimated odds ratio of 3.57 in part (d) is fairly large, the lower end of the 95% confidence interval is well above 1 and the p-value from McNemar's test is substantially below 0.05, so the evidence from the matched case control study is strong. Evidence from a case-control study is relatively weak, for several reasons including concerns about the representativeness of the controls, potential differences in recall of exposure by cases versus controls, assessment of timing of exposure relative to becoming a case and, as in any observational study, the possibility that there may be unmeasured confounders. In this particular study it should be relatively easy to recruit appropriate controls (women giving birth in the same facility as the cases), exposure recall is not an issue for the main exposure (measured periodontal disease) but the timing of the exposure measurement is (after giving birth, so it is not known when the periodontal disease developed). Also, there may have been unmeasured confounders, such as smoking.