HOMEWORK 1 Solutions

BIOSTATISTICS 755

This question will use data on Orthodontic Measurements on Children. The data are from a study of dental growth measurements of the distance (mm) from the center of the pituitary gland to the pteryomaxillary fissure. Measurements were obtained on 11 girls and 16 boys at ages 8, 10, 12, and 14. Scientific goals of the study are to:

• determine whether distances over time are larger for boys than for girls, and

• whether the rate of change of distance over time is similar for boys and girls.

The data is available on the website in the file dental.txt, which contains the data and information about the study and variables.

Please use this dataset to answer the following questions. Note that dental.txt is given in the *wide format.* This format is good for questions 1, 4 and 5. For question 2 the *long format* will be needed. Please see the example from class on how to go from the wide to the long format.

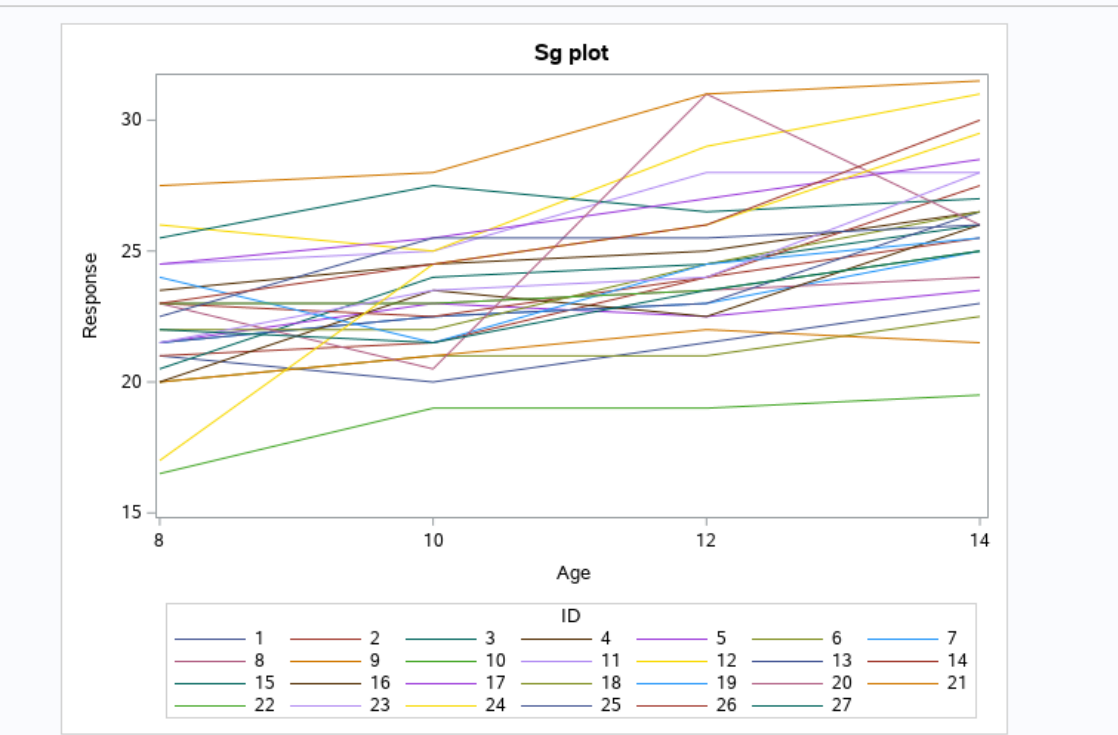
1. **(10 points)** Read the data into SAS and calculate sample means, standard deviations and variances of the distance measurements at each occasion.

Variable	Mean	Std Dev	Variance
Age8	22.1851852	2.4343225	5.9259259
Age10	23.1666667	2.1572775	4.6538462
Age12	24.6481481	2.8175781	7.9387464
Age14	26.0925926	2.7666873	7.6545584

2. Construct three time plots for data. For these plots:

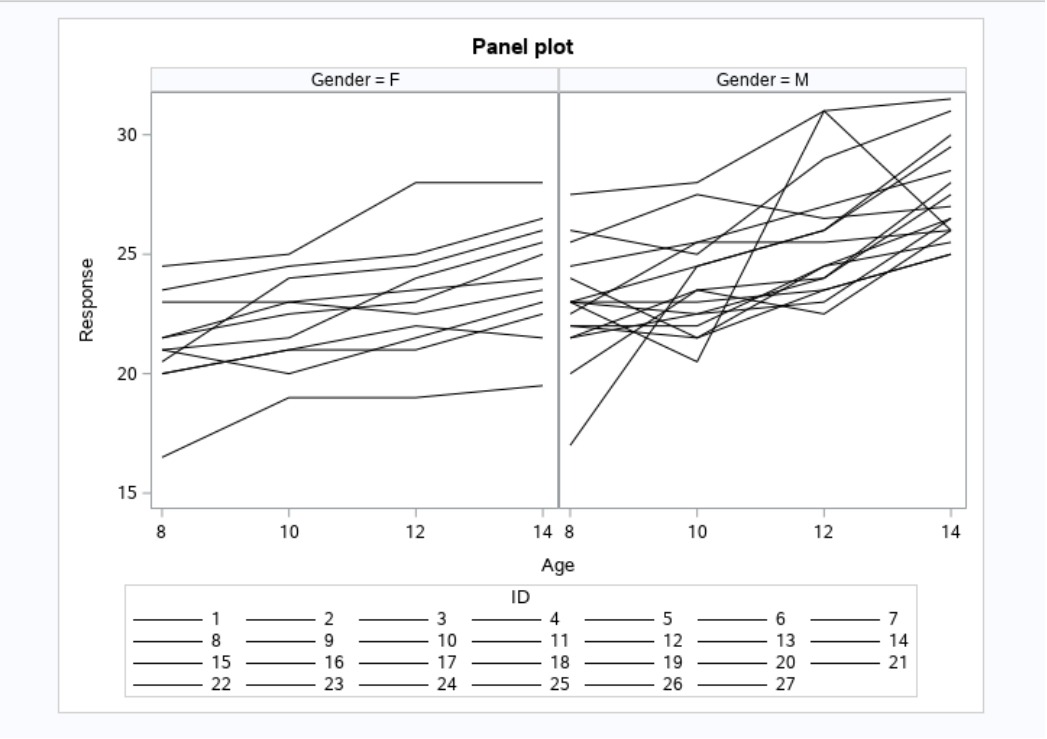
   (a) **(10 points)** Make all the colors differ by subject.

```
Proc SGplot data = dent_long;
series x=age y=response / group =ID LineAttrs= (pattern=1);
run;
```



(b) **(10 points)** Make a panel plot which separate panels for boys and girls (make all lines black).

```
Proc SGpanel data = dent_long;
PanelBy gender / columns=2;
series x=age y=response / group =ID LineAttrs= (pattern=1
color="black");
run;
```

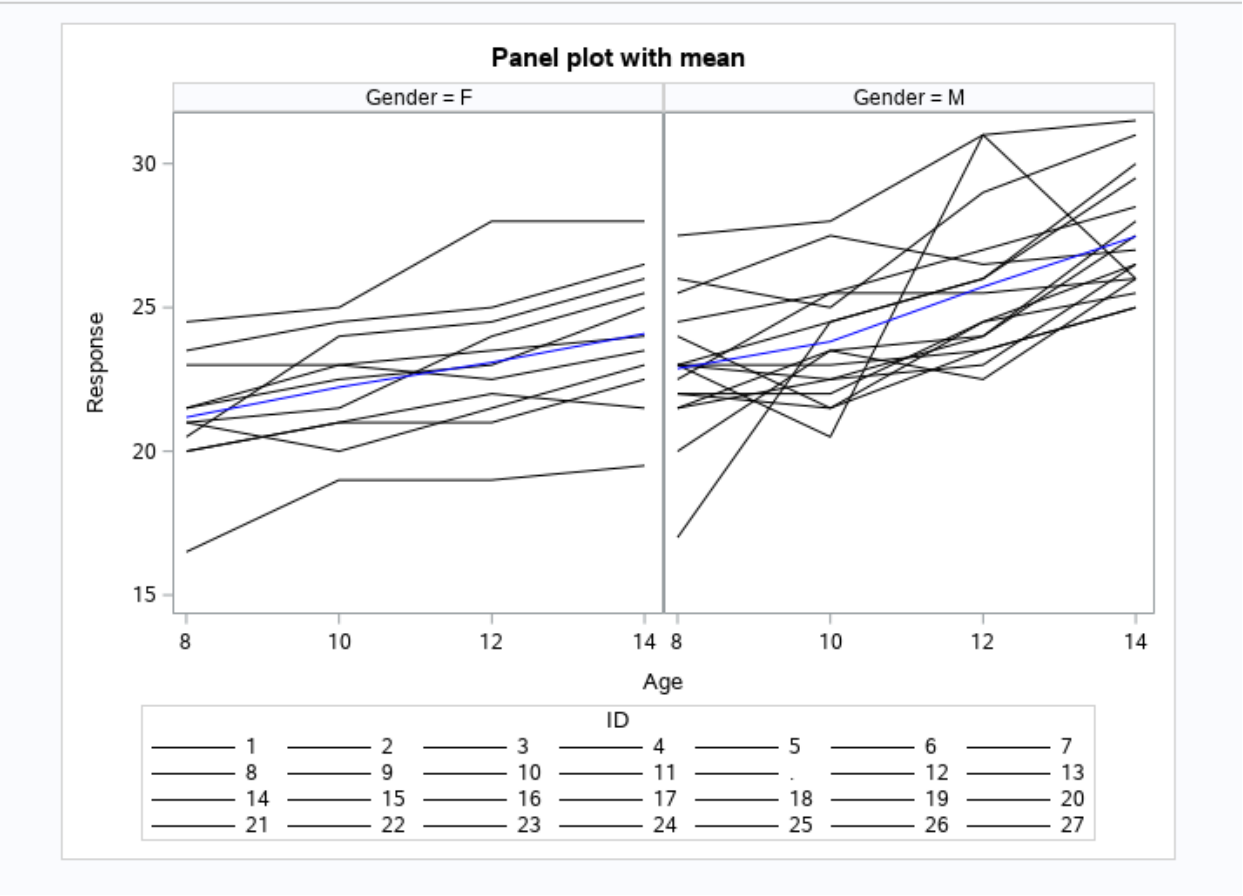


(c) **(10 points)** Repeat this plot adding a blue mean line to each panel.

```
proc sort data=dent_long;
by age gender;
run;
proc means mean data=dent_long;
by age gender;
var response;
output out=dent_long_mn mean=age_response_mn;
run;


data stacked_dent2;
set dent_long dent_long_mn;
run;
proc sort data=stacked_dent2;
by gender;
run;
proc sgpanel data= stacked_dent2;
panelby gender / columns=2;
series x=age y=response / group= ID lineattrs=(pattern=1
color='black');
series x=age y=age_response_mn / lineattrs=(pattern=1
color="blue");
```

```
run;
```



3. Using the plot in 2c:

(a) **(10 points)** discuss the differences in the means for boys and girls, and

The means for both boys and girls increase over time. The boys appear to have a slightly higher starting point, and a slope that is larger. This results in what appears to be a bigger difference in the means at age 14 versus age 8.

(b) **(10 points)** discuss the pattern in variation over time.

For both boys and girls, I would say the spread of the data is relatively consistent overtime and similar to each other. The within subject variation for the boys is markedly larger than the within subject variation for the girls. There appears to be more measurement error for the boys versus the girls. The between subject variation is similar for boys and girls.

4. **(10 points)** Calculate the 4 x 4 covariance and correlation matrices for the four repeated measures. Does correlation appear to vary by the time difference between measurements?

Covariance Matrix, DF = 26				
	Age8	Age10	Age12	Age14
Age8	5.925925926	3.285256410	4.875356125	4.039886040
Age10	3.285256410	4.653846154	3.858974359	4.532051282
Age12	4.875356125	3.858974359	7.938746439	6.197293447
Age14	4.039886040	4.532051282	6.197293447	7.654558405

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Age8	27	22.18519	2.43432	599.00000	16.50000	27.50000
Age10	27	23.16667	2.15728	625.50000	19.00000	28.00000
Age12	27	24.64815	2.81758	665.50000	19.00000	31.00000
Age14	27	26.09259	2.76669	704.50000	19.50000	31.50000

Pearson Correlation Coefficients, N = 27 Prob > |r| under H0: Rho=0				
	Age8	Age10	Age12	Age14
Age8	1.00000	0.62558 0.0005	0.71081 <.0001	0.59983 0.0009
Age10	0.62558 0.0005	1.00000	0.63488 0.0004	0.75933 <.0001
Age12	0.71081 <.0001	0.63488 0.0004	1.00000	0.79500 <.0001
Age14	0.59983 0.0009	0.75933 <.0001	0.79500 <.0001	1.00000

Covariance Matrix, DF = 10				
	Age8	Age10	Age12	Age14
Age8	4.513636364	3.354545455	4.331818182	4.356818182
Age10	3.354545455	3.618181818	4.027272727	4.077272727
Age12	4.331818182	4.027272727	5.590909091	5.465909091
Age14	4.356818182	4.077272727	5.465909091	5.940909091

Overall, the correlation appears to be relatively consistent over time.

5. **(10 points)** Calculate the 4 x 4 covariance and correlation matrices for the four repeated measures separately for boys and girls. Comment on the differences in the variance and correlation between boys and girls.

For females:

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**
Age8	11	21.18182	2.12453	233.00000	16.50000	24.50000
Age10	11	22.22727	1.90215	244.50000	19.00000	25.00000
Age12	11	23.09091	2.36451	254.00000	19.00000	28.00000
Age14	11	24.09091	2.43740	265.00000	19.50000	28.00000

Pearson Correlation Coefficients, N = 11 Prob > |r| under H0: Rho=0				
	Age8	**Age10**	**Age12**	**Age14**
Age8	1.00000	0.83009 0.0016	0.86231 0.0006	0.84136 0.0012
Age10	0.83009 0.0016	1.00000	0.89542 0.0002	0.87942 0.0004
Age12	0.86231 0.0006	0.89542 0.0002	1.00000	0.94841 <.0001
Age14	0.84136 0.0012	0.87942 0.0004	0.94841 <.0001	1.00000

For Males:

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**
Age8	16	22.87500	2.45289	366.00000	17.00000	27.50000
Age10	16	23.81250	2.13600	381.00000	20.50000	28.00000
Age12	16	25.71875	2.65185	411.50000	22.50000	31.00000
Age14	16	27.46875	2.08542	439.50000	25.00000	31.50000

Pearson Correlation Coefficients, N = 16 Prob > |r| under H0: Rho=0				
	Age8	**Age10**	**Age12**	**Age14**
Age8	1.00000	0.43739 0.0902	0.55793 0.0247	0.31523 0.2343
Age10	0.43739 0.0902	1.00000	0.38729 0.1383	0.63092 0.0088
Age12	0.55793 0.0247	0.38729 0.1383	1.00000	0.58599 0.0171
Age14	0.31523 0.2343	0.63092 0.0088	0.58599 0.0171	1.00000

Overall, the within subject correlation for the girls is higher than it is for the boys. The correlation is also slightly more consistent for the girls. The variation is similar for girls and boys.

6. **(10 points)** Recall the 4 correlations "truths" from the second set of slides. Describe on what these truths mean in the context of this example. In your description, try to be more specific than restating each "truth" in context. For example, don't use the word 'correlation' in your description, think about what the "truth" tells you about the data if taken as fact.

The four correlation truths in this context are:

1. the correlations are positive:

   Measurements from the same child are likely to be more alike than measurements from different children. That is, if a child's 1st measurement is higher than average, we expect their remaining measurements to be higher than average.

2. the correlations often decrease with increasing time separation

   If a child's 1st measurement is higher than average, we are more confident in their 2nd measurement being higher than average than their 3rd or 4th measurement being higher than average.

3. the correlations between repeated measures rarely ever approach zero

   If a child's 1st measurement is higher than average, than a later measurement is more likely to be higher than average than lower than average no matter how long between the measurements.

4. the correlation between a pair of repeated measures taken very closely rarely approaches one.

   No matter how close in time two measurements are taken, there will be some difference between them.

7. **(10 points)** State three potential sources of variability in the data in context of the problem. Using the time plot you constructed in 2(c) give your hypothesized ordering of which source of variability is the largest.

Three sources of variability:

1. differences between the kids

2. day to day or month to month differences in the true dental differences,

3. error when measuring the distances.

For girls, 1 is likely the largest, followed by 3 and 2.

For the boys, it's a little harder to tell. It might be the same as the girls, but it might be that 3 is the largest. Either way, 2 is probably the smallest.