

**Biostatistics 140.655, 2017-18**  
**QUIZ 3**

**Quiz Guidelines:**

Please read the following quiz guidelines carefully:

- For this quiz, you are to work ALONE. You may use your course notes and lab materials to help answer the questions.
- Submit your answers to Courseplus by 5pm Friday February 23rd.
- DO NOT discuss this quiz or your solution to this quiz with other students from the course on Wednesday Feb 21<sup>st</sup> through Friday Feb 23<sup>rd</sup>. The solution to the quiz will be available Saturday, Feb 24<sup>th</sup>.
- By submitting your answers to Courseplus, you are acknowledging that you have read the guidelines carefully and will adhere to these guidelines.

**Scientific Background:**

You will analyze data from the National Cooperative Gallstone Study (NCGS). The major interest of the study was to determine the safety of the drug, chenodiol, for the treatment of cholesterol gallstones. Patients were randomly assigned to high dose (750 mg per day), low dose (375 mg per day), or placebo. The available data set consists of 103 patients with floating gallstones who were assigned to the high dose ( $n = 62$ ) or placebo groups ( $n = 41$ ).

In the NCGS, it was suggested that chenodiol would dissolve gallstones, but in doing so, might increase levels of serum cholesterol. As a result, serum cholesterol (mg/dL) was measured at baseline, 6, 12, 20 and 24 months of follow-up. NOTE: there is missing data in the trial. The vast majority of the missing data is due to patient drop-out; i.e. the patients have data up to some follow-up and then do not return for any additional assessments. There are two treatment group patients that have “intermittent” missing data; that is, they have some measurements, but not all measurements.

**Objective:** To determine if the rate of high serum cholesterol, defined as a serum cholesterol level  $\geq 240$  mg/dL, differed in the treatment groups at both **12** and **24 months**.

**Variables:**

Trt: 1 = high dose, 2 = placebo

ID: subject ID

Month: 0, 6, 12, 20 and 24

R: serum cholesterol level (mg/dL)

Y\_binary: indicator of serum cholesterol level  $\geq 240$  mg/dL

Reference: Wei, L.J. and Lachin, J.M. (1984). Two-Sample Asymptotically Distribution-Free Tests for Incomplete Multivariate Observations. Journal of the American Statistical Association, 79, 653-661.

**Notation:** Let

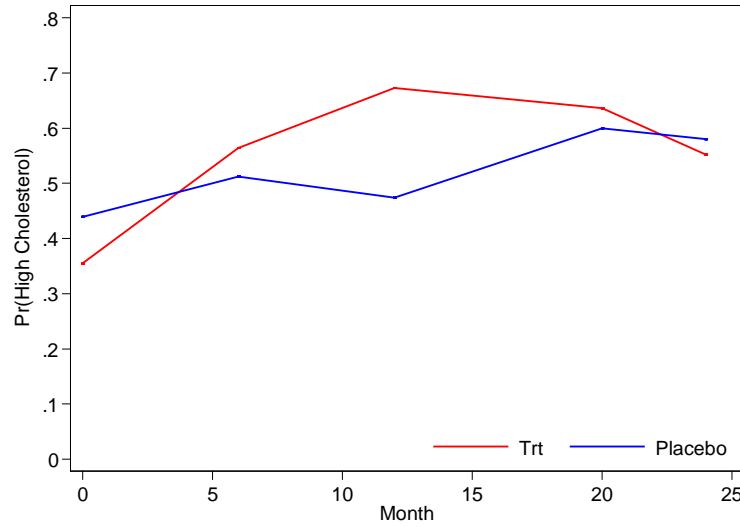
$Y_{ij}$  be the indicator for high serum cholesterol for subject  $i$  ( $i$  in 1, 2, ..., 103) at month  $j$  ( $j = 0, 6, 12, 20, 24$ ).

$month_{ij}$  be the month of follow-up, 0, 6, 12, 20, and 24

$drug_i$  be the indicator that subject  $i$  received the high dose relative to placebo

### **Exploratory analysis of the mean:**

The figure below displays the proportion of subjects with high serum cholesterol for each follow-up separately by treatment.



The following mean model was considered:

$$\begin{aligned} \text{Log} \left[ \frac{\Pr(Y_{ij} = 1 | drug_i, month_{ij})}{\Pr(Y_{ij} = 0 | drug_i, month_{ij})} \right] = & \beta_0 + \beta_1(month_{ij} = 6) + \beta_2(month_{ij} = 12) + \beta_3(month_{ij} = 20) + \beta_4(month_{ij} = 24) + \\ & \beta_5(month_{ij} = 6)drug_i + \beta_6(month_{ij} = 12)drug_i + \beta_7(month_{ij} = 20)drug_i + \\ & \beta_8(month_{ij} = 24)drug_i \end{aligned}$$

1. The interpretation of  $\beta_2$  is:
  - a. The log odds of high serum cholesterol at baseline in the placebo group.
  - b. The log odds ratio of high serum cholesterol comparing 12-months to baseline in the placebo group.
  - c. The log odds of high serum cholesterol at 12-months in the placebo group
  - d. The log odds ratio of high serum cholesterol comparing 12-months to baseline in the drug group.
  
2. Based on the model above and the objective of the analysis, the hypothesis test of interest is:
  - a.  $H_0: \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$  vs.  $H_1: \text{at least one } \beta \text{ is not } 0$
  - b.  $H_0: \beta_5 = \beta_7 = 0$  vs.  $H_1: \text{at least one } \beta \text{ is not } 0$
  - c.  $H_0: \beta_6 = \beta_8 = 0$  vs.  $H_1: \text{at least one } \beta \text{ is not } 0$
  - d.  $H_0: \beta_6 = \beta_7 = 0$  vs.  $H_1: \text{at least one } \beta \text{ is not } 0$

### **Exploratory analysis of the within-subject correlation:**

The table below summarizes patterns of responses within and between patients in the randomized trial.

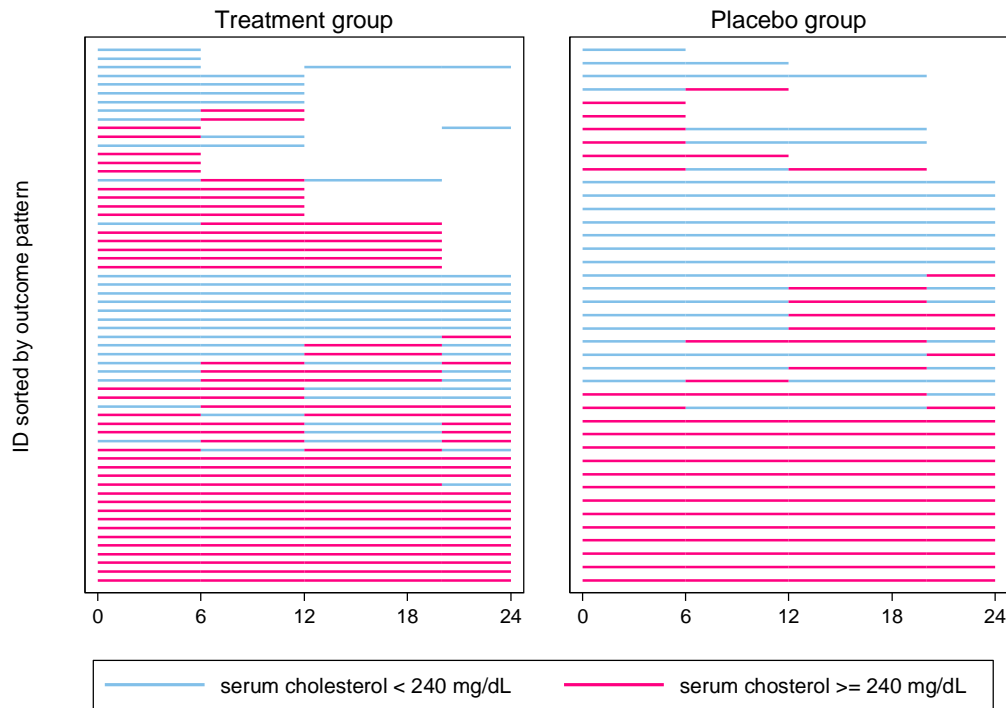
```
. xttab y_binary
```

y_binary	Overall		Between		Within
	Freq.	Percent	Freq.	Percent	Percent
0	208	46.53	72	69.90	67.25
1	239	53.47	79	76.70	69.09
Total	447	100.00	151	146.60	68.21

(n = 103)

3. Among the 447 observations,
  - a. Roughly 47 percent of the observations have serum cholesterol level  $\geq 240$  mg/dL
  - b. Roughly 47 percent of the observations have serum cholesterol level  $< 240$  mg/dL
  - c. Roughly 53 percent of the patients have serum cholesterol level  $\geq 240$  mg/dL
  - d. Roughly 53 percent of the patients have serum cholesterol level  $< 240$  mg/dL
  
4. Among patients with at least one serum cholesterol level  $\geq 240$  mg/dL,
  - a. The average number of follow-ups with serum cholesterol levels  $\geq 240$  mg/dL is 67 percent.
  - b. The average number of follow-ups with serum cholesterol levels  $\geq 240$  mg/dL is 77 percent.
  - c. The average number of follow-ups with serum cholesterol levels  $\geq 240$  mg/dL is 69 percent.
  - d. The average number of follow-ups with serum cholesterol levels  $< 240$  mg/dL is 67 percent.

The figure below displays the lasagna plot separately for each treatment group.



5. Comparing the treatment to placebo group,
  - a. There is a greater proportion of patients with missing data in the treatment group, as compared to the placebo group.
  - b. There is a greater proportion of patients with missing data in the placebo group, as compared to the treatment group.
6. Among patients with no missing data in the treatment group,
  - a. Roughly the same proportion of patients have serum cholesterol less than 240 mg/dL for all follow-ups and serum cholesterol  $\geq$  240 mg/dL for all follow-ups.
  - b. A larger proportion of patients have serum cholesterol less than 240 mg/dL for all follow-ups compared to serum cholesterol  $\geq$  240 mg/dL for all follow-ups.
  - c. A smaller proportion of patients have serum cholesterol less than 240 mg/dL for all follow-ups compared to serum cholesterol  $\geq$  240 mg/dL for all follow-ups.
  - d. We cannot determine this information from the lasagna plot.

The two tables below summarize results of the trial for the 103 patients.

TABLE 1:

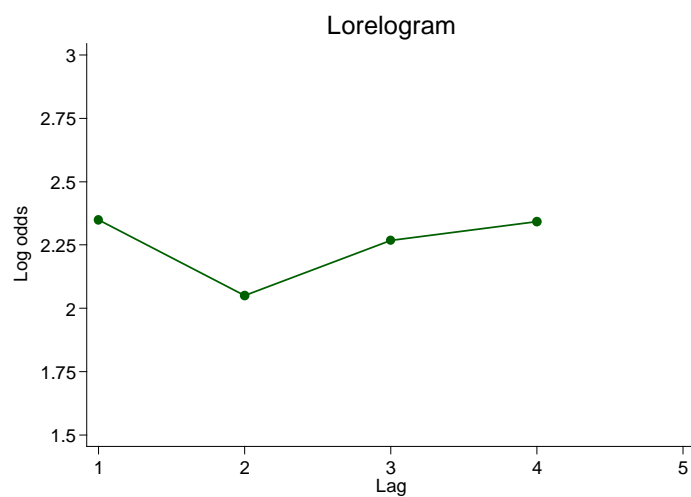
High Serum Cholesterol	Month		Total
	0	6	
No	63	47	110
Yes	40	56	96
Total	103	103	206

TABLE 2:

High Serum Cholesterol at baseline (month 0)	High Serum Cholesterol at 6 months		Total
	No	Yes	
No	42	21	63
Yes	5	35	40
Total	47	56	103

7. To understand the correlation in the binary response of high serum cholesterol within a subject comparing baseline to the 6-month follow-up, which of the two tables (Table 1 or Table 2) provide the relevant information?
- TABLE 1
  - TABLE 2

The figure below displays the estimated lorelogram for the binary outcome of high serum cholesterol.



8. Based on the estimated lorelogram, a reasonable working correlation model would be:
- Independence
  - Unstructured
  - Exchangeable
  - Autoregressive

The logistic model specified above was fit to the data assuming an unstructured working correlation model. The results of the “xtgee” command are presented below:

```
GEE population-averaged model
Group and time vars:      id month
Link:                     logit
Family:                   binomial
Correlation:              unstructured
Scale parameter:          1
Number of obs             =      447
Number of groups          =      103
Obs per group: min       =        2
                        avg       =      4.3
                        max       =        5
Wald chi2(8)              =     25.78
Prob > chi2               =     0.0011
```

y_binary	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
month						
6	1.478376	.414091	1.40	0.163	.8538169	2.559796
12	1.350557	.4335687	0.94	0.349	.7198663	2.533808
20	2.239639	.7396483	2.44	0.015	1.172373	4.278487
24	1.928654	.621925	2.04	0.042	1.025111	3.628589
drug_month6	1.488216	.5075608	1.17	0.244	.7627124	2.903828
drug_month12	2.493097	.9854039	2.31	0.021	1.148945	5.409773
drug_month20	1.193842	.4939527	0.43	0.668	.5305932	2.68616
drug_month24	1.069356	.4360159	0.16	0.869	.4809012	2.377873
_cons	.6349206	.1283624	-2.25	0.025	.4272001	.9436425

```
. test drug_month12 drug_month24
```

```
( 1) drug_month12 = 0
( 2) drug_month24 = 0
```

```
      chi2( 2) =      5.95
Prob > chi2 =      0.0510
```

9. The estimated relative odds of high serum cholesterol, comparing the 12-month follow-up to baseline, among subjects receiving the placebo are 1.35. The estimated relative odds of high serum cholesterol, comparing the 12-month follow-up to baseline, among subjects receiving the high dose treatment are:
  - a. 2.49
  - b.  $2.49 + 1.35$
  - c.  $2.49 \times 1.35$
  - d.  $0.63 \times 2.49 \times 1.35$