

# BIOS767\_FINAL

*Cheyenna Crowley*

4/29/2019

## Question 1A:

Variable: Birth Weight (g)	N	Mean	SD	Median	Range
Birth Order: 1	198	3028.02	558.56	3076.00	(920,4508)
Birth Order: 2	198	3102.87	552.25	3128.50	(1021,4508)
Birth Order: 3	198	3158.15	568.92	3201.50	(1380,4960)
Birth Order: 4	198	3194.53	566.33	3246.00	(910,4780)
Birth Order: 5	198	3208.24	581.58	3218.00	(1210,5018)
Low Birth Weight	366	2588.52	421.25	2726.00	(910,2990)
Not Low Birth Weight	624	3460.87	356.38	3374.00	(3000,5018)
Overall	990	3138.36	568.34	3175.00	(910,5018)

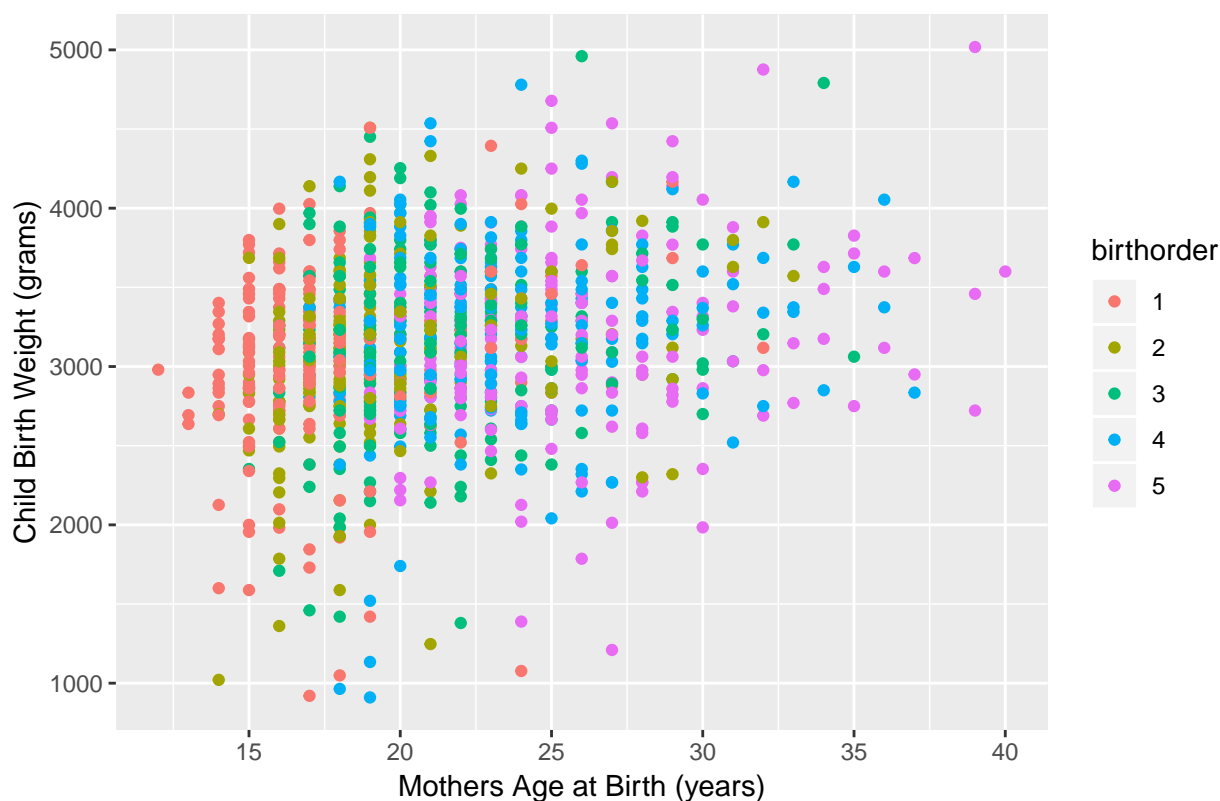


Figure 1: Child Birth Weight (grams) by Mothers Age (years)

## Question 1B:

Model M1:  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$   $i=1,\dots,K$   $j=1,\dots,5$

Parameter and Standard Error Estimates:

Parameter	Estimate	Standard Error
$\alpha_1$	3088.53	49.35
$\alpha_2$	3120.15	38.00
$\alpha_3$	3117.21	41.95
$\alpha_4$	3138.67	50.77
$\alpha_5$	3121.90	62.57
$\beta_1$	23.82	11.89
$\beta_2$	38.42	10.13
$\beta_3$	28.55	9.94
$\beta_4$	16.41	9.30
$\beta_5$	16.13	8.85

Our model  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$  is modeling the mean child birth weight determined by the child birth order and mothers age. Age is offset in this model by 20 years. The model has a varying intercept and slope for each child birth order. For a specific subject, we can expect that the birth order and the mothers age is related since a women can't have her second birth before her first thus implying the mother has to be older for the second birth than the first. There is no clear pattern for the estimates in relation to the child birth order but it can be noted that the estimates are all positive so with every one year increase in mothers age, the mean child birth weight increases.

#### Question 1C:

Model M1:  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$   $i=1,\dots,K$   $j=1,\dots,5$

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$

$H_1$ : Otherwise

Method: Likelihood Ratio Test

$\alpha = \text{Type 1 error} = 0.05$

Test Statistic:

$H_0$ :  $-2\log L = 15059.9$

$H_1$ :  $-2\log L = 15054.1$

$\text{LRT} = (15059.9 - 15054.1) = 5.8$

DF:  $10 - 6 = 4$

P-Value: 0.2146

Decision: Fail to reject  $H_0$

Conclusion/Summary:

Our null hypothesis was that our slopes for the model were equal for each birth order which corresponds to the model  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$  and our alternative hypothesis model was  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$  for  $i=1,\dots,K$   $j=1,\dots,5$ . Since these models are nested, we can fit these models and use their respective likelihoods to conduct a likelihood ratio test ( $-2\log(\text{Likelihood of } H_0 - \text{Likelihood of } H_A) = 15059.9 - 15054.1$ ) and achieve a Likelihood Ratio Test Statistic of 5.8. Our null model contains 6 parameters, and our alternative model contains 10, thus our test has 4 degrees of freedom. Using a chi-square distribution of 4 degrees of freedom and a test statistic of 5.8, we achieve a p-value of 0.2146. With a type 1 error,  $\alpha = 0.05$  which is less than our p-value of 0.2146 so we fail to reject the null and conclude that to model the mean birth weight the better model is  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$  compared to  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$  for  $i=1,\dots,K$   $j=1,\dots,5$ . Some inferences about the null model are that the slope does not vary for each of the mothers children

$(\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5)$  or that the estimated effect for every one year change in mothers age from 20 years old does not vary depending on the birthing order of the child.

#### Question 1D:

Model M1:  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$   $i=1,\dots,K$   $j=1,\dots,5$

$H_0 : \beta_j = 0, j=1,2,3,4,5$

$H_1$ : Otherwise

Method: Likelihood Ratio Test

Test Statistic:

$H_0$ :  $-2\text{Log } L=15070.7$

$H_1$ :  $-2\text{Log } L=15054.1$

$\text{LRT}=(15070.7-15054.1)=16.6$

DF:  $10-5=5$

P-Value: 0.0053

Decision: Reject  $H_0$

Conclusion/Summary:

Our null hypothesis was that mothers age was not significant in our model for each level of child birth order. Our null hypothesis model was  $E[Y_{ij}] = \alpha_j$  and our alternative hypothesis model was  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$  for  $i=1,\dots,K$   $j=1,\dots,5$ . Since these models are nested, we can fit these models and use their respective likelihoods to conduct a likelihood ratio test ( $-2\text{Log}(\text{Likelihood of } H_0 - \text{Likelihood of } H_A)=15070.7-15054.1$ ) and achieve a Likelihood Ratio Test Statistic of 16.6. Our null model contains 5 parameters, and our alternative model contains 10, thus our test has 5 degrees of freedom. Using a chi-square distribution of 5 degrees of freedom and a test statistic of 16.6, we achieve a p-value of 0.0053. With a type 1 error,  $\alpha=0.05$  which is greater than our p-value of 0.0053 we reject the null and conclude that to model the mean birth weight the better model is  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$  compared to  $E[Y_{ij}] = \alpha_j$  for  $i=1,\dots,K$   $j=1,\dots,5$ . Some inferences about the alternative model is that there is at least one slope ( $\beta_j$ ) that does not equal zero, thus mothers age at delivery is in the model.

#### Question 1E:

Model M1:  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$   $i=1,\dots,K$   $j=1,\dots,5$

$H_0 : \beta_1 = 0$  where  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$

$H_1$ : Otherwise

Method: Likelihood Ratio Test

Test Statistic:

$H_0$ :  $-2\text{Log } L=15070.7$

$H_1$ :  $-2\text{Log } L=15059.9$

$\text{LRT}=(15070.7-15059.9)=10.8$

DF:  $6-5=1$

P-Value: 0.0010

Decision: Reject  $H_0$

### Conclusion/Summary:

Our null hypothesis was that the estimated effect of mothers age was not significant in our model when there was only one slope regardless of child birthing order. Our null hypothesis model was  $E[Y_{ij}] = \alpha_j$  where  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$  and our alternative hypothesis model was  $E[Y_{ij}] = \alpha_j + \beta(\text{age}_{ij} - 20)$  (our null hypothesis in question b) for  $i=1, \dots, K$   $j=1, \dots, 5$ . Since these models are nested, we can fit these models and use their respective likelihoods to conduct a likelihood ratio test ( $-2\log(\text{Likelihood of } H_0 - \text{Likelihood of } H_A) = 15070.7 - 15059.9$ ) and achieve a Likelihood Ratio Test Statistic of 10.8. Our null model contains 5 parameters, and our alternative model contains 6, thus our test has 1 degrees of freedom. Using a chi-square distribution of 1 degrees of freedom and a test statistic of 10.8, we achieve a p-value of 0.0010. With a type 1 error,  $\alpha=0.05$  which is greater than our p-value of 0.0010 so we reject the null and conclude that to model the mean birth weight the better model is  $E[Y_{ij}] = \alpha_j + \beta(\text{age}_{ij} - 20)$  compared to  $E[Y_{ij}] = \alpha_j$  where  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$  for  $i=1, \dots, K$   $j=1, \dots, 5$ . Some inferences about the alternative model is that there is one slope ( $\beta$ ) that does not equal zero, thus mothers age at delivery is in the model or, the estimated effect of every 1 year change in mothers age at the time of birth from 20 years old is needed in the model and it does not vary by child birthing order.

### Question 1F:

Model M1:  $E[Y_{ij}] = \alpha_j + \beta_j(\text{age}_{ij} - 20)$   $i=1, \dots, K$   $j=1, \dots, 5$

$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$  where  $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$

$H_1$ : Otherwise

Method: Likelihood Ratio Test

Test Statistic:

$H_0$ :  $-2\log L = 15061.0$

$H_1$ :  $-2\log L = 15059.9$

$(15061.0 - 15059.9) = 1.10$

DF:  $6 - 2 = 4$

P-Value: 0.8943

Decision: Fail to reject  $H_0$

Summary/Conclusion:

Our null hypothesis was an intercept for each birthing order of the child was not needed in our model of mean birth weight. Our null hypothesis model was  $E[Y_{ij}] = \alpha + \beta(\text{age}_{ij} - 20)$  where  $\beta = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$  and  $\alpha = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$ . Our alternative hypothesis model was  $E[Y_{ij}] = \alpha_j + \beta(\text{age}_{ij} - 20)$  where  $\beta = \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5$  (our null hypothesis in question b) for  $i=1, \dots, K$   $j=1, \dots, 5$ . Since these models are nested, we can fit these models and use their respective likelihoods to conduct a likelihood ratio test ( $-2\log(\text{Likelihood of } H_0 - \text{Likelihood of } H_A) = 15061.0 - 15059.9$ ) and achieve a Likelihood Ratio Test Statistic of 1.1. Our null model contains 2 parameters, and our alternative model contains 6, thus our test has 4 degrees of freedom. Using a chi-square distribution of 4 degrees of freedom and a test statistic of 1.1, we achieve a p-value of 0.8943. With a type 1 error,  $\alpha=0.05$  which is less than our p-value of 0.8943 so we fail to reject the null and conclude that to model the mean birth weight the better model is  $E[Y_{ij}] = \alpha + \beta(\text{age}_{ij} - 20)$  compared to  $E[Y_{ij}] = \alpha_j + \beta(\text{age}_{ij} - 20)$  for  $i=1, \dots, K$   $j=1, \dots, 5$ . Some inferences about the null model is that there is one slope ( $\beta$ ) and one intercept ( $\alpha$ ) regardless of the birthing order.

### Question 1G:

Model M2:  $E[Y_{ij}] = \alpha + \beta(\text{age}_{ij} - 20)$   $i=1, \dots, K$   $j=1, \dots, 5$

Parameter	Estimates	Standard Error
$\alpha$	3103.25	28.3086
$\beta$	22.7397	4.0260

Estimates and 95% CI for probability of a low-birth weight for the first birth if a mothers age was 15, 20,25,30,35,40

$$E[Y_{i1}] = 3103.25 + 22.7397(age_{i1} - 20) \quad i=1,\dots,K$$

$$\Pr(\text{low child birth weight}) = \Pr(\text{birthwt} < 3000 \text{ grams})$$

Mothers Age at First Birth	Pr(low child birth weight)	95% CI
15	0.613	(0.0471, 0.988)
20	$1.33 * 10^{-4}$	$(1.013 * 10^{-8}, 0.0460)$
25	$3.44 * 10^{-11}$	$(1.071 * 10^{-17}, 2.544 * 10^{-6})$
30	$1.14 * 10^{-12}$	$(1.375 * 10^{-19}, 2.168 * 10^{-7})$
35	$2.59 * 10^{-12}$	$(3.901 * 10^{-19}, 3.945 * 10^{-7})$
40	$8.68 * 10^{-12}$	$(1.828 * 10^{-18}, 9.478 * 10^{-7})$

Using the predicted estimates and standard errors from fitting model M2 we can calculate the probability of low birth weight for a mother first child at ages 15,20,25,30,35, and 40. First we calculate the predicted estimates of birth weight by fitting the model to our data, then use the normal distribution to determine the probability the estimated mean is less than 3000. From the table above, we can see that the lower the mothers age the more likely the mother is to have a child with low birth weight. Mothers at age 15 were predicted to have a probability of 0.613 of having a child with low birth weight, while after age 15, the probability decreases immensely.

#### Question 1H:

$$\text{Model M2: } E[Y_{ij}] = \alpha + \beta(age_{ij} - 20) \quad i=1,\dots,K \quad j=1,\dots,5$$

$$\text{Model M3: } \Phi^{-1}(E[Y_{ij}^*]) = \theta + \psi(age_{ij} - 20) \quad i=1,\dots,K \quad j=1,\dots,5$$

Parameter	Estimates	Standard Error
$\theta$	-0.2917	0.0607
$\psi$	-0.0303	0.0101

In Model 3, we are estimating the mean of the indicator of low birth weight for the child ( $I(\text{birthwt} < 3000)$ ) as our response while in model 2 we are using the measurement in grams of the child birth weight. In part G, when we are calculating the predicted probability of a mother having a low birth weight child we are using the continuous measurement of birth weight in grams and then using the normal distribution to determine a probability. In theory, we should expect similar trends of estimated probabilities in part I, but it should be noted that model 2 can be more influenced by outliers and extremes than model 3 thus potentially skewing the results.

#### Question 1I:

$$\text{Model M3: } \Phi^{-1}(E[Y_{ij}^*]) = \theta + \psi(age_{ij} - 20) \quad i=1,\dots,K \quad j=1,\dots,5$$

Estimates and 95% CI for probability of a low-birth weight for the first birth if a mothers age was 15,20,25,30,35,40.

Mothers Age at First Birth	Pr(low child birth weight)	95% CI
15	0.4443	(0.3802,0.5099)
20	0.3853	(0.3406,0.4314)
25	0.3288	(0.2787,0.3821)
30	0.2760	(0.2089,0.3521)
35	0.2277	(0.1471,0.3286)
40	0.1846	(0.0980,0.3075)

As mothers age of delivery increases, the probability of a low birth weight decreases.

#### Question 1J:

Model M3:  $\Phi^{-1}(E[Y_{ij}^*]) = \theta + \psi(\text{age}_{ij} - 20)$   $i=1,\dots,K$   $j=1,\dots,5$

Model M4:  $\Phi^{-1}(E[Y_{ij}^*|U_i]) = \gamma + \delta(\text{age}_{ij} - 20) + U_i$   $i=1,\dots,K$   $j=1,\dots,5$   $U_i \sim N(0, \sigma_u^2)$

Parameter	Estimates	Standard Error
$\gamma$	-0.4063	0.08648
$\delta$	-0.0434	0.01318
$\sigma^2$	0.8933	0.1796

In Model 3,  $\Phi^{-1}(E[Y_{ij}^*]) = \theta + \psi(\text{age}_{ij} - 20)$   $i=1,\dots,K$   $j=1,\dots,5$ , which is a marginal model, we are assuming that the covariance matrix is the same for all subjects. In Model 4,  $\Phi^{-1}(E[Y_{ij}^*|U_i]) = \gamma + \delta(\text{age}_{ij} - 20) + U_i$   $i=1,\dots,K$   $j=1,\dots,5$   $U_i \sim N(0, \sigma_u^2)$ , we are assuming in this mixed model the responses are conditionally independent on some random variable  $U_i$  which allows us to model the covariance between and within subjects. We should expect the estimates in the models to be similar as both models are using a probit link to model the effect of mothers age on child low birth weight status.

#### Question 1K:

Model M4:  $\Phi^{-1}(E[Y_{ij}^*|U_i]) = \gamma + \delta(\text{age}_{ij} - 20) + U_i$   $i=1,\dots,K$   $j=1,\dots,5$   $U_i \sim N(0, \sigma_u^2)$

Variance Decomposition of  $Y_{i3}^*$  assuming  $\text{age}_{i3}=25$

	Variance
Between	0.0362
Within	0.1974
Total	0.2336
ICC	0.1551

For mothers delivery age of 25 the estimated variance for low birth weight for the mothers third child is 15.5% (0.0362/0.2336) due to variability between mothers and the estimated variance for low birth weight for the mothers third child is 84.5% due to the variability within mothers. The ICC is 0.1551 which shows the the proportion of the total variance in  $Y_{ij}$  that is accounted for by the mothers (“between mothers”). These values were calculated using 10000 simulations (iml) specific for a binary response and probit link.

### Question 1L:

The motivation for this study was to investigate how low birth weight is impacted by mothers age and birthing order. To investigate this, an indicator variable was created to define that a birth weight less than 3000 grams is considered a low birth weight. Different models can answer various forms of our question, but finding the correct model produces the most accurate results and applicable interpretation. One aspect of this study are the covariates involved, mothers age and birthing order. These two variables are related as the mothers age at delivery will be older for each subsequent birth. To account for this, a mixed model may be more appropriate. Determining the best model a series of diagnostics and models of best fit were conducted. First, we investigated a model that estimated the mean birth weight in grams and we determined whether if for each birthing order, the model would need a separate slope and intercept to accurately estimate the mean. After a set of likelihood ratio tests, it was deemed that a different slope and intercept for each birthing order was not necessary thus for further investigation an overall effect would be appropriate. Since our question of interest is more focused on a specified “low birth weight” category, we also investigated the probability associated with having a low birth weight by mothers age. It was found that as mothers age of delivery increases, the probability of a low birth weight decreases. When fitting a probit model of the binary variable “low birth weight”, we found that for mothers delivery age of 25 for their third child the estimated variance is 15.5% due to variability between mothers implying that the variance effect within mothers is substantial.