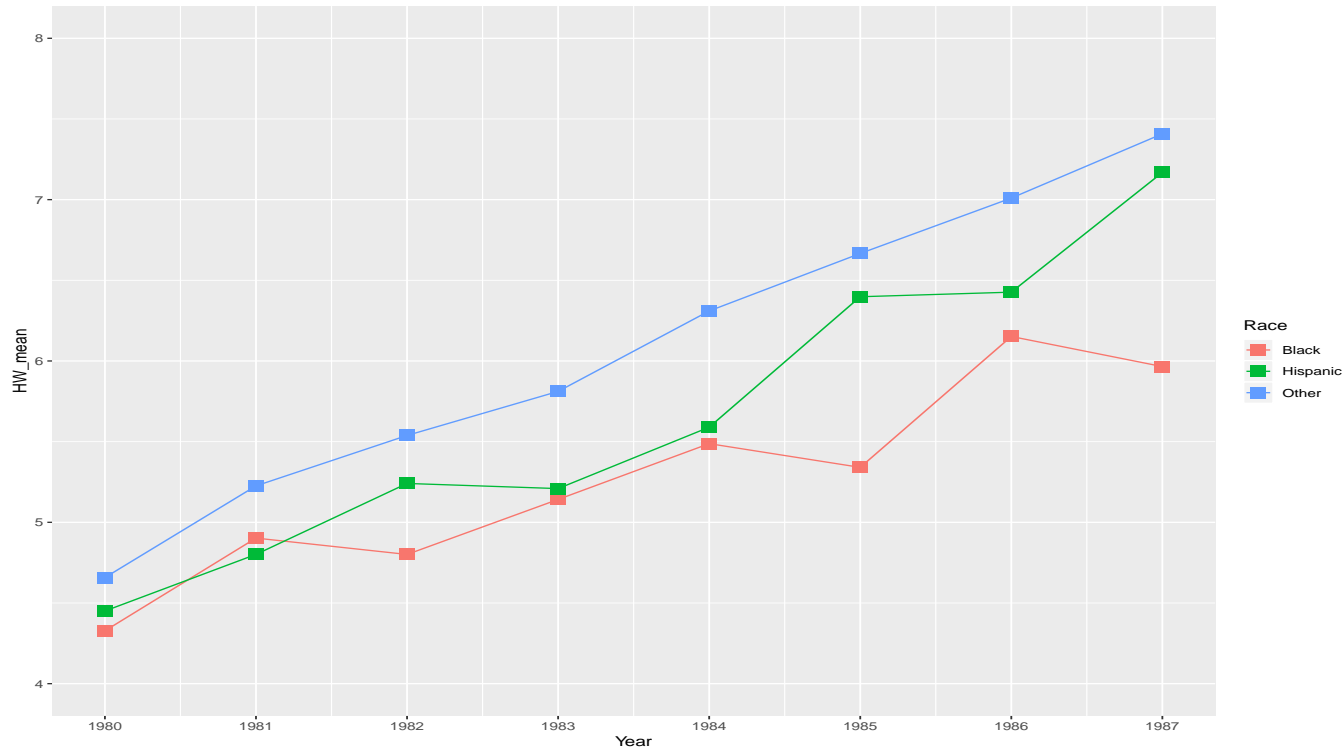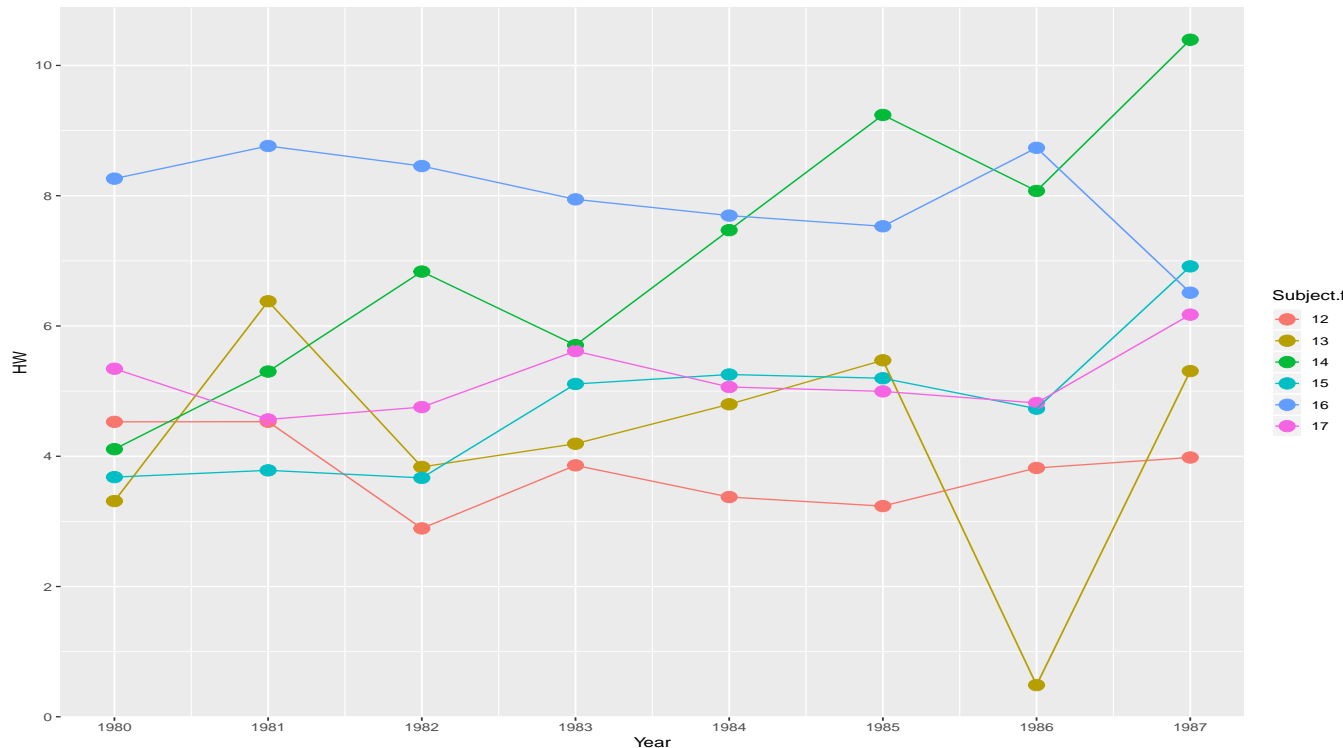# 1 Graphical analysis

1. **(2 marks)**

Figure 1: Plot of mean HW vs $Year$ grouped by $Race$

Yes since the trajectory of mean HW over time for the $Black$ group is different to the $Other$ and $Hispanic$ groups. **(2 marks)**

2. **(2 marks)**

Figure 2: Plot of HW vs $Year$ for first six subjects

Yes as it seems that the trajectory of HW over time is generally different from subject to subject. (**2 marks**). Yes since HW at 1980 differs substantially between subjects. (**2 marks**)

3.

2 Describing the model

4. (a) (**2 marks**)

$$Y_i = \begin{bmatrix} HW_{1i} \\ HW_{2i} \\ HW_{3i} \\ HW_{4i} \\ HW_{5i} \\ HW_{6i} \\ HW_{7i} \\ HW_{8i} \end{bmatrix}.$$

(b) (**4 marks**)

$$X_i = \begin{bmatrix} 1 & 0 & B_i & H_i & 0 & 0 \\ 1 & 1 & B_i & H_i & B_i & H_i \\ 1 & 2 & B_i & H_i & 2B_i & 2H_i \\ 1 & 3 & B_i & H_i & 3B_i & 3H_i \\ 1 & 4 & B_i & H_i & 4B_i & 4H_i \\ 1 & 5 & B_i & H_i & 5B_i & 5H_i \\ 1 & 6 & B_i & H_i & 6B_i & 6H_i \\ 1 & 7 & B_i & H_i & 7B_i & 7H_i \end{bmatrix}.$$

(c) (**2 marks**)

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}.$$

(d) (**2 marks**)

$$Z_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \end{bmatrix}.$$

(e) (**2 marks**)

$$\boldsymbol{\mu}_i = \begin{bmatrix} \mu_{0i} \\ \mu_{1i} \end{bmatrix}.$$

(f) (**2 marks**)

$$\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ \varepsilon_{5i} \\ \varepsilon_{6i} \\ \varepsilon_{7i} \\ \varepsilon_{8i} \end{bmatrix}.$$

3 Testing for random effects

5. (a) **(1 mark)**

$$HW_{ti} = \beta_0 + \beta_1 Y_{ti} + \beta_2 B_i + \beta_3 H_i + \beta_4 Y_{ti} \times B_i + \beta_5 Y_{ti} \times H_i + \mu_{0i} + \mu_{1i} Y_{ti} + \varepsilon_{ti}, \tag{1}$$

(b) **(1 mark)**

$$HW_{ti} = \beta_0 + \beta_1 Y_{ti} + \beta_2 B_i + \beta_3 H_i + \beta_4 Y_{ti} \times B_i + \beta_5 Y_{ti} \times H_i + \mu_{0i} + \varepsilon_{ti}, \tag{2}$$

(c) p-value $= 1.057962 \times 10^{-40}$ **(1 mark)**

(d) The reference model since $1.057962 \times 10^{-40} < 0.05$. **(2 marks)**

6. (a) Model (1) **(1 mark)**
   (b) **(1 mark)**

$$HW_{ti} = \beta_0 + \beta_1 Y_{ti} + \beta_2 B_i + \beta_3 H_i + \beta_4 Y_{ti} \times B_i + \beta_5 Y_{ti} \times H_i + \mu_{1i} Y_{ti} + \varepsilon_{ti}, \tag{3}$$

(c) p-value $= 4.456031 \times 10^{-38}$ **(1 mark)**

(d) The reference model since $4.456031 \times 10^{-38} < 0.05$. **(2 marks)**

4 Testing for fixed effects

7. (a) Model (1) **(1 mark)**
   (b) **(1 mark)**

$$HW_{ti} = \beta_0 + \beta_1 Y_{ti} + \beta_2 B_i + \beta_3 H_i + \mu_{0i} + \mu_{1i} Y_{ti} + \varepsilon_{ti}, \tag{4}$$
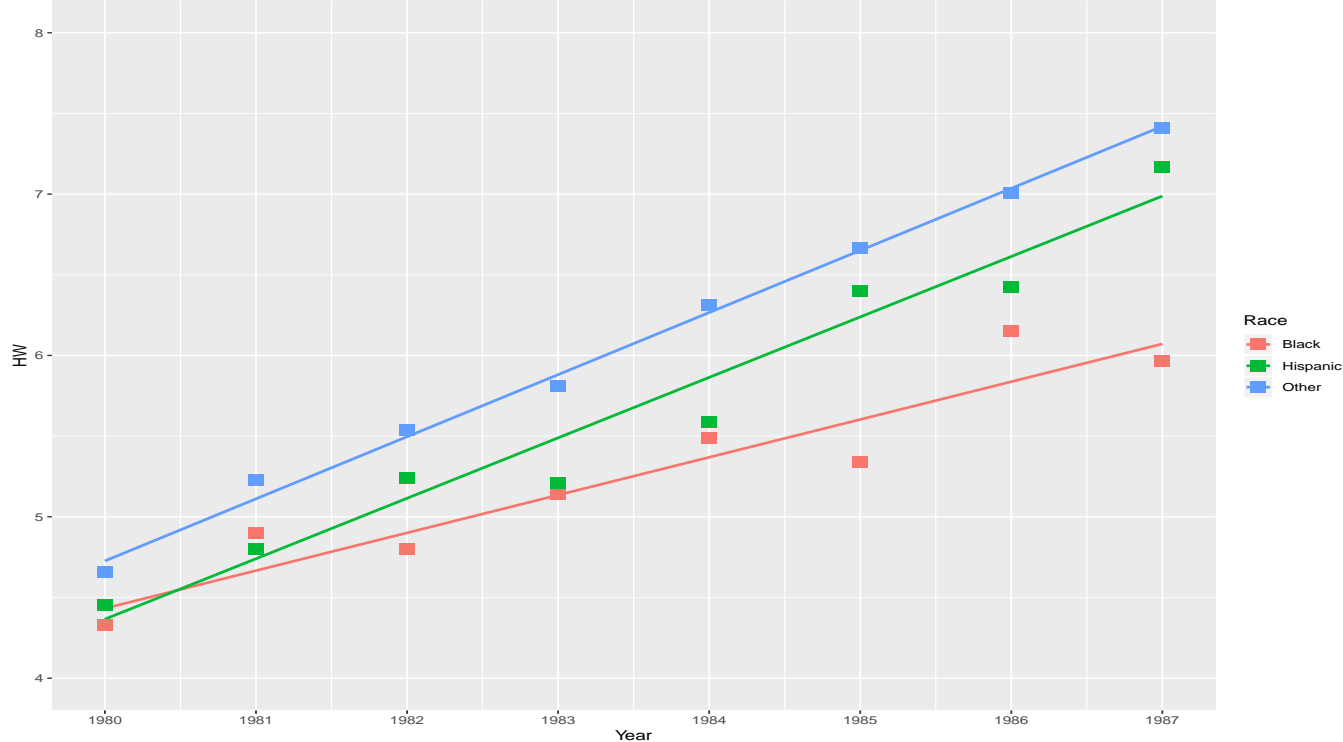
(c) p-value $= 0.0665$ **(1 mark)**

(d) The reference model since $0.0665 < 0.10$. **(2 marks)**
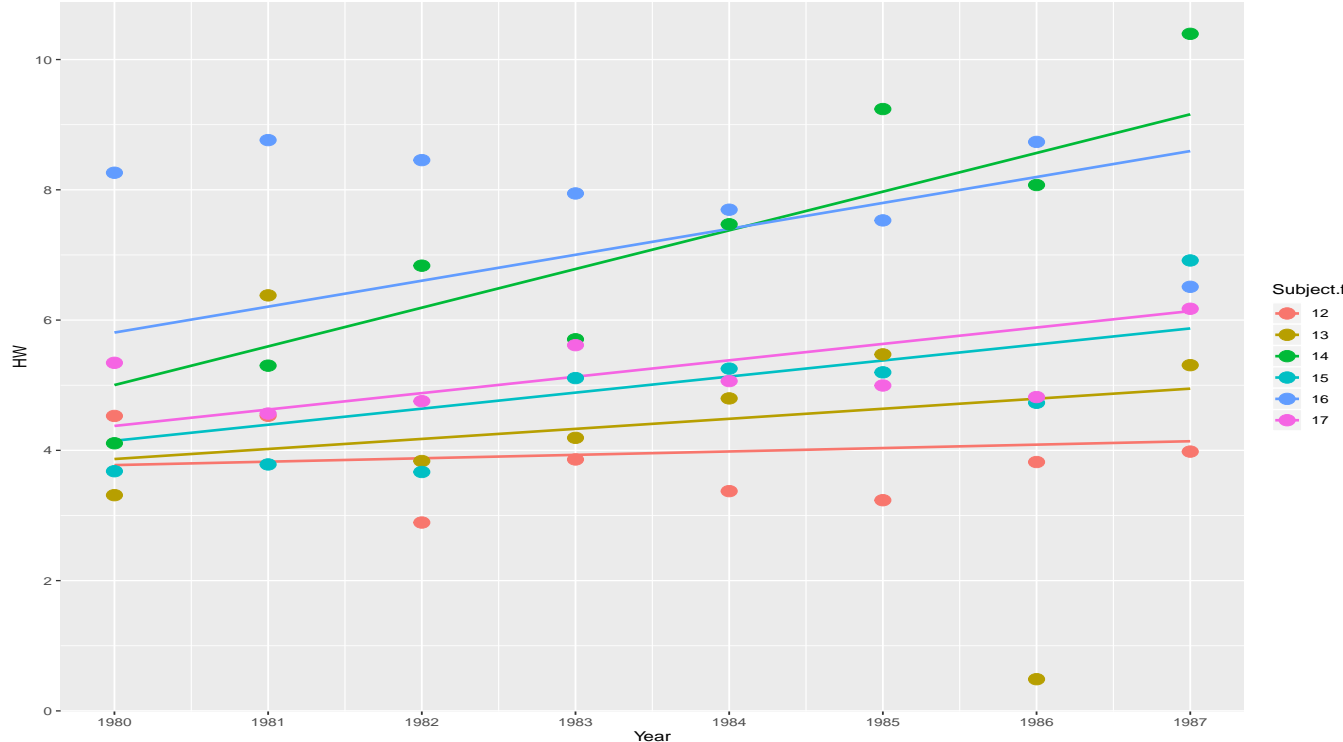
5 Diagnostics of your final linear mixed model

8. **(2 marks)**

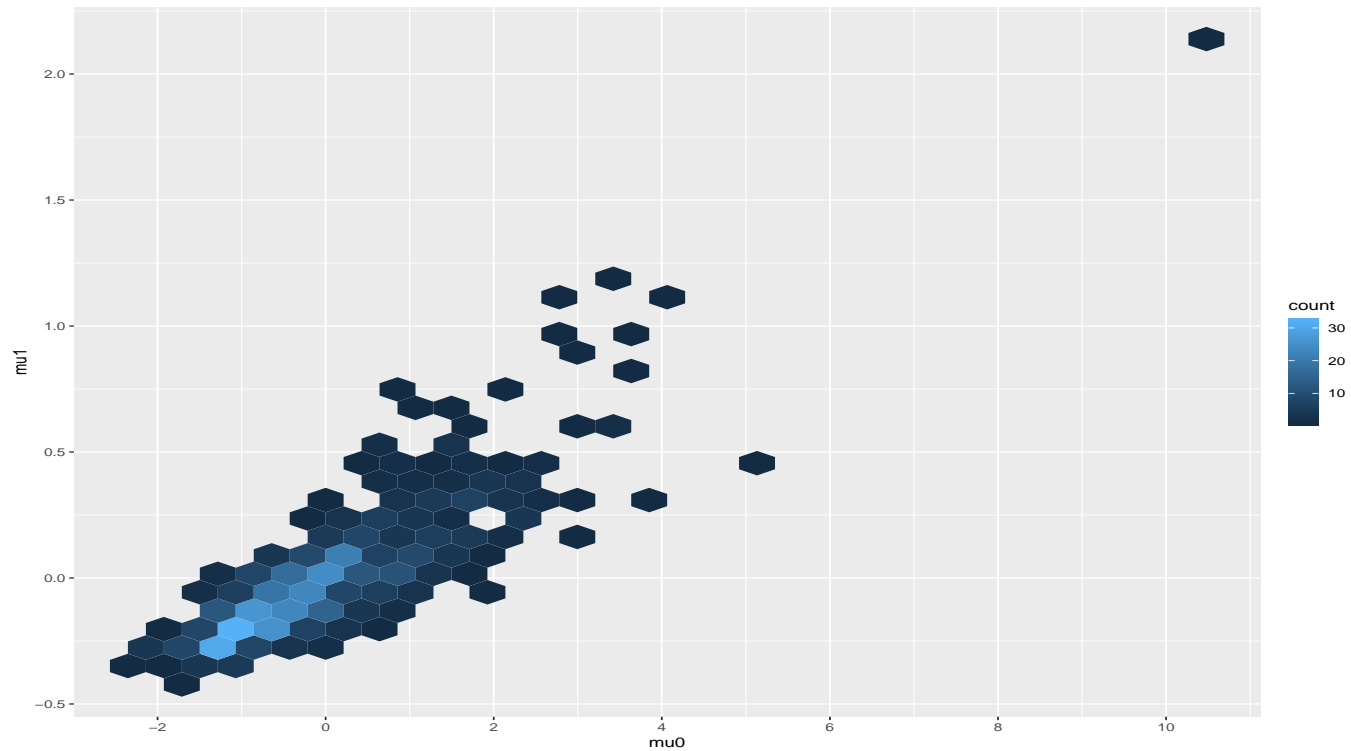Figure 3: Predicted marginal and observed mean values of HW as a function of $Year$

9. **(2 marks)**

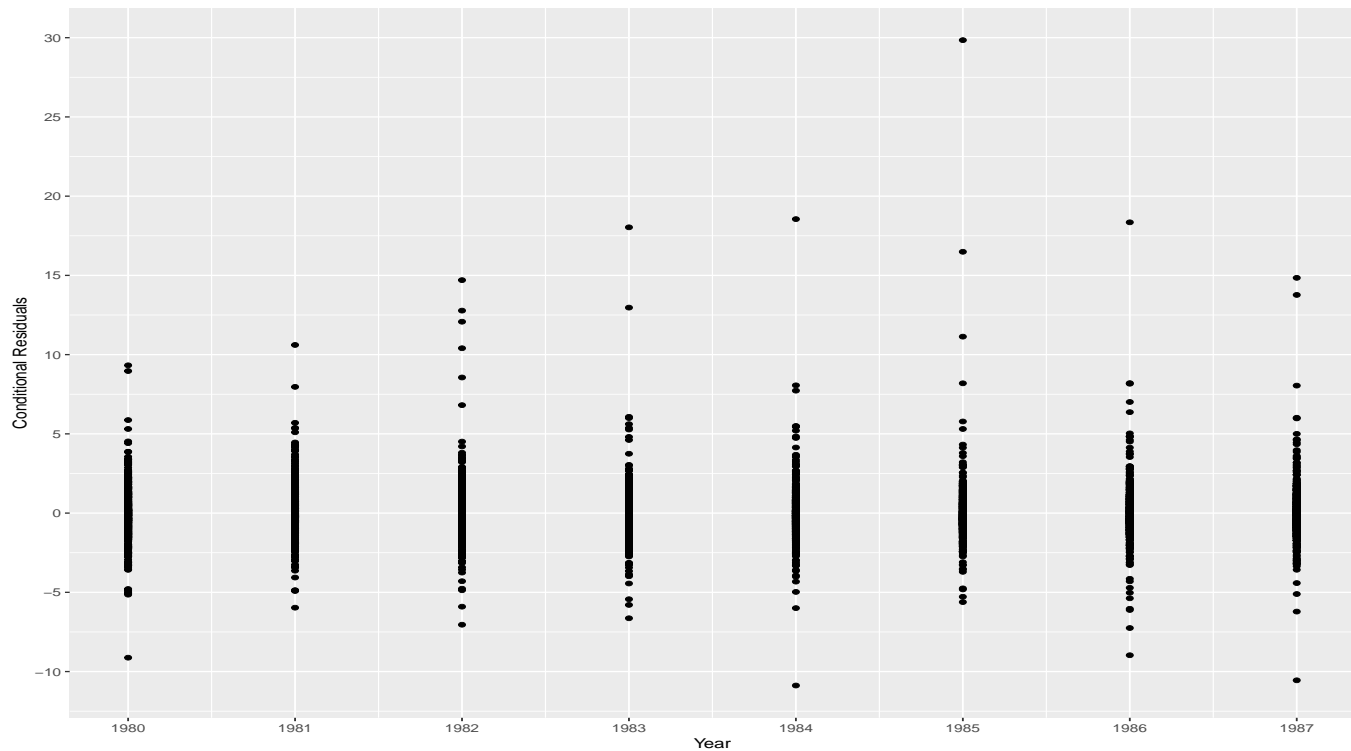Figure 4: Predicted conditional and observed values of HW as a function of $Year$ for first 6 subjects

10. **(2 marks)**

Figure 5: Hexagonal 2-D plot of the predicted random effect vector $\hat{\boldsymbol{\mu}}_i$

11. **(2 marks)**

Figure 6: Scatter plot of conditional residuals vs $Year$

6 Fixed effect estimates for your final linear mixed model

12. **(1 mark)**

Figure 7: Table 1

	Value	Std.Error	DF	t-value	p-value
(Intercept)	4.7265	0.1082	3812	43.6872	0.0000
Y	0.3848	0.0240	3812	16.0135	0.0000
B	-0.2942	0.2923	542	-1.0064	0.3147
H	-0.3611	0.2576	542	-1.4015	0.1616
Y:B	-0.1507	0.0649	3812	-2.3202	0.0204
Y:H	-0.0102	0.0572	3812	-0.1782	0.8586

13. **(1 mark)**

Figure 8: Table 2

	lower	est.	upper
(Intercept)	4.5144290	4.72654586	4.93866272
Y	0.3377151	0.38483145	0.43194779
B	-0.8684847	-0.29421375	0.28005716
H	-0.8671599	-0.36107629	0.14500728
Y:B	-0.2779812	-0.15066602	-0.02335084
Y:H	-0.1223955	-0.01019742	0.10200070

14. • The estimate of β_0 is 4.7265. We estimate that the mean hourly wage for *Other* working males in 1980 is \$4.7265. **(2 marks)**

• The estimate of β_1 is 0.3848. We estimate that the mean hourly wage for *Other* working males increases by \$0.3848 for each 1 year increase. or We estimate that the linear effect of *Year* on mean hourly wage for *Other* working males is \$0.3848. **(2 marks)**

• The estimate of β_2 is -0.2942. We estimate that the difference in mean hourly wage, in 1980, between *Black* and *Other* working males is -\$0.2942. or We estimate that the mean hourly wage for *Black* male workers in 1980 is \$0.2942 lower than the mean hourly wage for *Other* male workers in 1980. **(2 marks)**

• The estimate of β_3 is -0.3611. We estimate that the difference in mean hourly wage, in 1980, between *Hispanic* and *Other* working males is -\$0.3611. or We estimate that the mean hourly wage for *Hispanic* male workers in 1980 is \$0.3611 lower than the mean hourly wage for *Other* male workers in 1980. **(2 marks)**

• The estimate of β_4 is -0.1507. We estimate that the difference in the linear effect of *Year* on mean hourly wage between *Black* and *Other* male workers is -\$0.1507. or We estimate that the linear effect of *Year* on mean hourly wage for *Black* male workers is \$0.1507 lower than the linear effect of *Year* on mean hourly wage for *Other* male workers. **(2 marks)**

• The estimate of β_5 is -0.0102. We estimate that the difference in the linear effect of *Year* on mean hourly wage between *Hispanic* and *Other* male workers is -\$0.0102. or We estimate that the linear effect of *Year* on mean hourly wage for *Hispanic* male workers is \$0.0102 lower than the linear effect of *Year* on mean hourly wage for *Other* male workers. **(2 marks)**

15. Yes since the p-value $= 0.0204 < 0.05$ and the 95% CI $= $ [-0.278, -0.023] contains only negative values. **(3 marks)**

16. 1.3489 **(1 mark)**

17. The estimate of 1.3489 suggests that the mean hourly wage for *Other* male workers in 1987 is $1.3489 higher than the mean hourly wage for *Black* male workers in 1987, and this estimate is significant (p-value $= 0.00498 < 0.05$). Therefore there is sufficient statistical evidence to suggest that the mean hourly wage for *Other* male workers in 1987 is higher than the mean hourly wage for *Black* male workers in 1987. **(3 marks)**

18. 0.9164 **(1 mark)**

19. The estimate of 0.9164 suggests that the mean hourly wage for *Hispanic* male workers in 1987 is $0.9164 higher than the mean hourly wage for *Black* male workers in 1987, however this estimate is insignificant (p-value $= 0.12 > 0.05$). Therefore there is insufficient statistical evidence to suggest that the mean hourly wage for *Hispanic* male workers in 1987 is higher than the mean hourly wage for *Black* male workers in 1987. **(3 marks)**