Answer 11.13

We perform regression in stata.

```
. regress sbp age
```

Source	SS	df	MS
Model	1502.66912	1	1502.66912
Residual	22.2720588	15	1.48480392
Total	1524.94118	16	95.3088235

Number of obs	=	17
F(1, 15)	=	1012.03
Prob > F	=	0.0000
R-squared	=	0.9854
Adj R-squared	=	0.9844
Root MSE	=	1.2185

sbp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
age	1.919118	.060326	31.81	0.000	1.790536	2.047699
_cons	97.78676	.6181575	158.19	0.000	96.46919	99.10434

The least square line is given by: $y = 97.8 + 1.92x$.

Answer 11.18

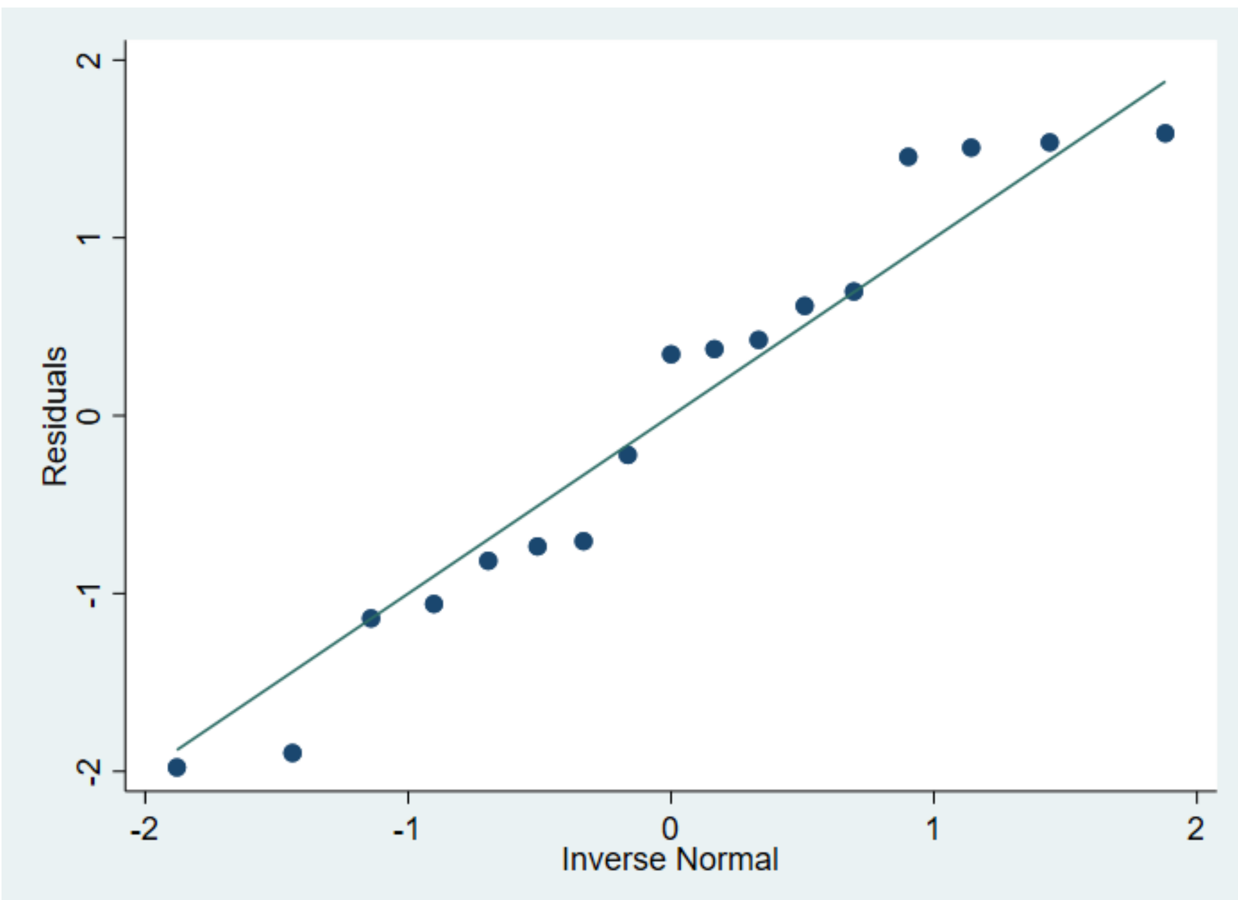

Figure 1. QQ plots for raw residuals.

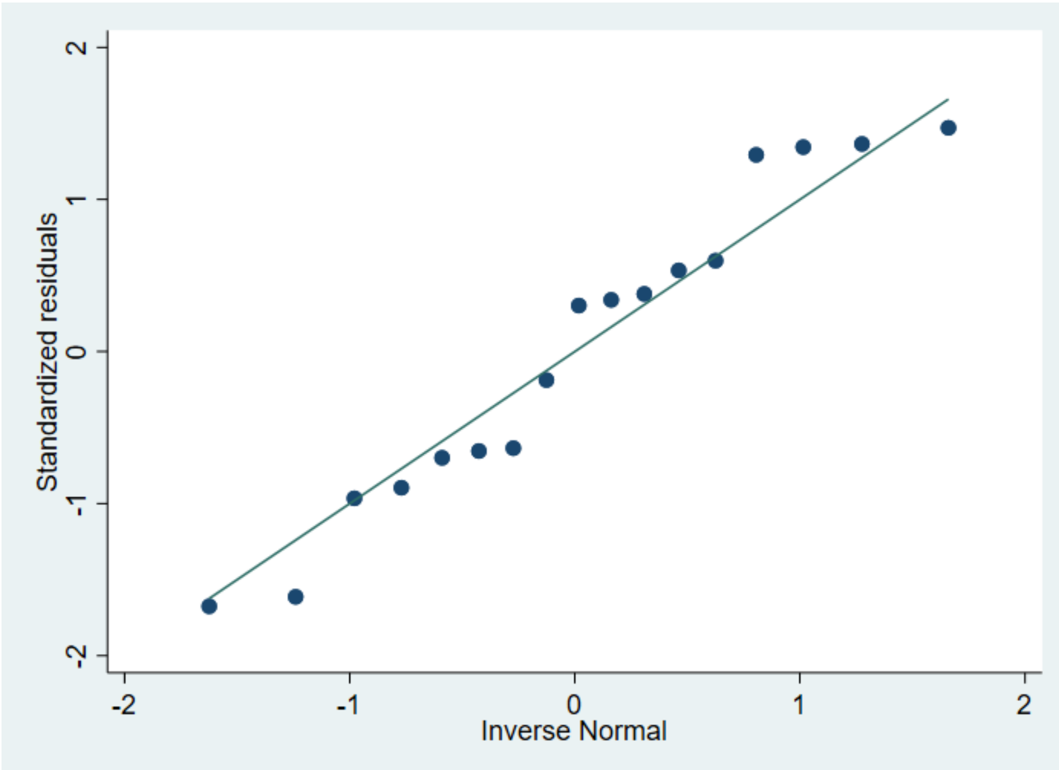

Figure 2. QQ plot for standardized residuals.

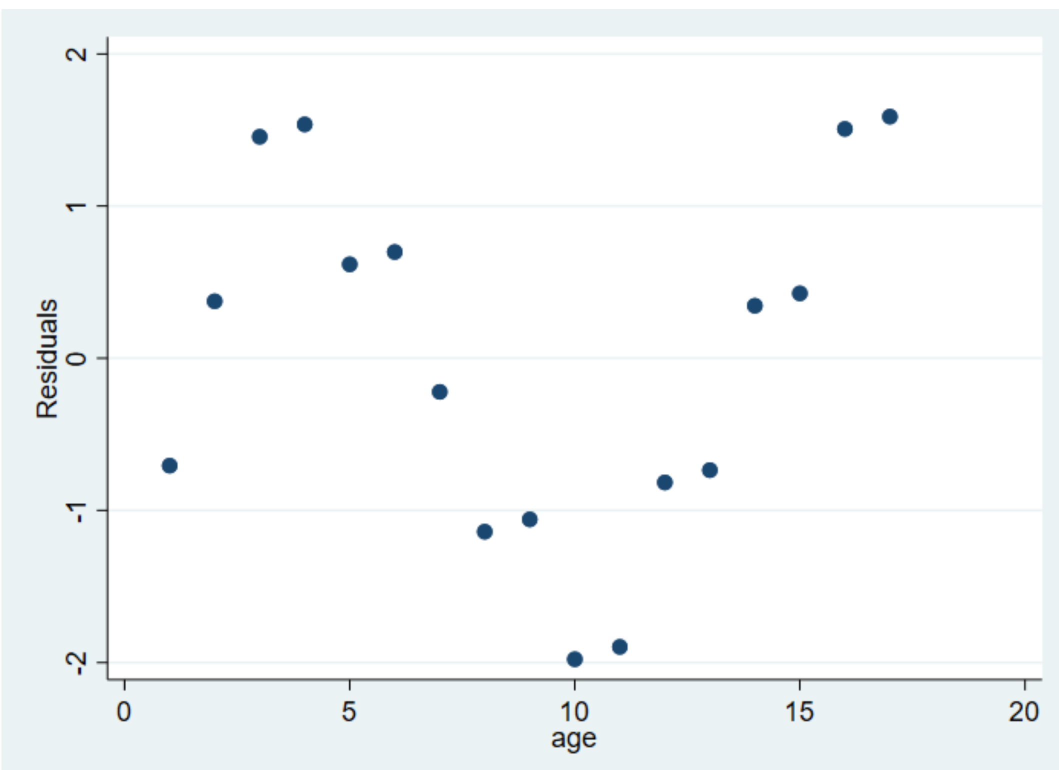

Figure 3. Scatter plot for raw residuals.

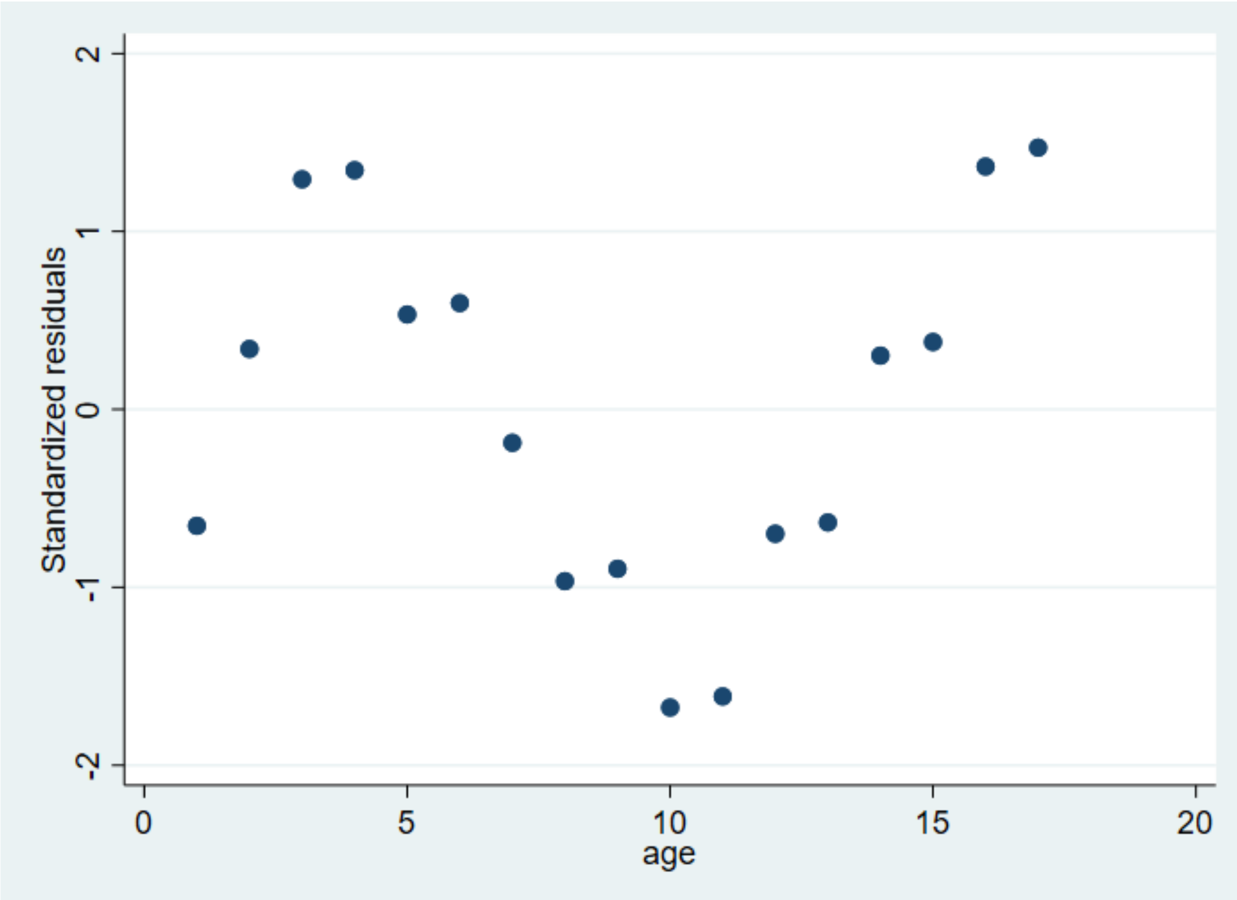

Figure 4. Scatter plot for standardized residuals.

The QQ plots are approximately linear. There is little deviation from the diagonal in the QQ plots and no indication of heteroscedasticity, nonlinearity, outliers or influential points in the scatterplots. Thus, we can proceed with our analysis of the model.

Answer 11.14
From the stata output, we can find that the 95% confidence interval of α is (96.5, 99.1), and the 95% confidence interval of β is (1.79 2,05).

Answer 11.15
We use the liner model to calculate the predicted blood pressure.

$$\hat{y} = 97.8 + 1.92 \times 13 = 122.8\, mm\ Hg$$

The predicted value is close to the observed value 122 mm Hg.

Answer 11.16

$$L_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{17} = 1785 - \frac{153^2}{17} = 408$$

$$L_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{17} = 226580 - \frac{1956^2}{17} = 1524.9$$

$$L_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{17} = 18387 - \frac{153 \times 1956}{17} = 783$$

$$s_{yx}^2 = \frac{L_{yy} - (L_{xy}^2/L_{xx})}{n-2} = \frac{1524.9 - (783^2/408)}{15} = 1.48$$

$$s_{yx} = 1.22$$

$$se = s_{yx}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{L_{xx}}} = 1.22\sqrt{\frac{1}{17} + \frac{(13-9)^2}{408}} = 0.382\ mm\ Hg$$

Answer 11.17

The predicted value for a 17-year-old is:

$$\hat{y} = 97.8 + 1.92 \times 17 = 130.44\ mm\ Hg$$

While the observed value is 132 mm Hg.

The standard error is:

$$se = s_{yx}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{L_{xx}}} = 1.22\sqrt{\frac{1}{17} + \frac{(17-9)^2}{408}} = 0.567\ mm\ Hg$$

Answer 11.25

A R- squared of 0.27 implies that about 27% of the variance for y can be explained by the linear regression model.

Answer 11.49

```
. regress thyroxine age
```

Source	SS	df	MS
Model	11.1283654	1	11.1283654
Residual	2.09563691	8	.261954614
Total	13.2240023	9	1.46933358

Number of obs	=	10
F(1, 8)	=	42.48
Prob > F	=	0.0002
R-squared	=	0.8415
Adj R-squared	=	0.8217
Root MSE	=	.51182

thyroxine	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
age	.3672728	.056349	6.52	0.000	.2373318	.4972137
_cons	-2.627274	1.614081	-1.63	0.142	-6.349351	1.094804

The least square line is given by: $y = -2.63 + 0.367x$.

Answer 11.51

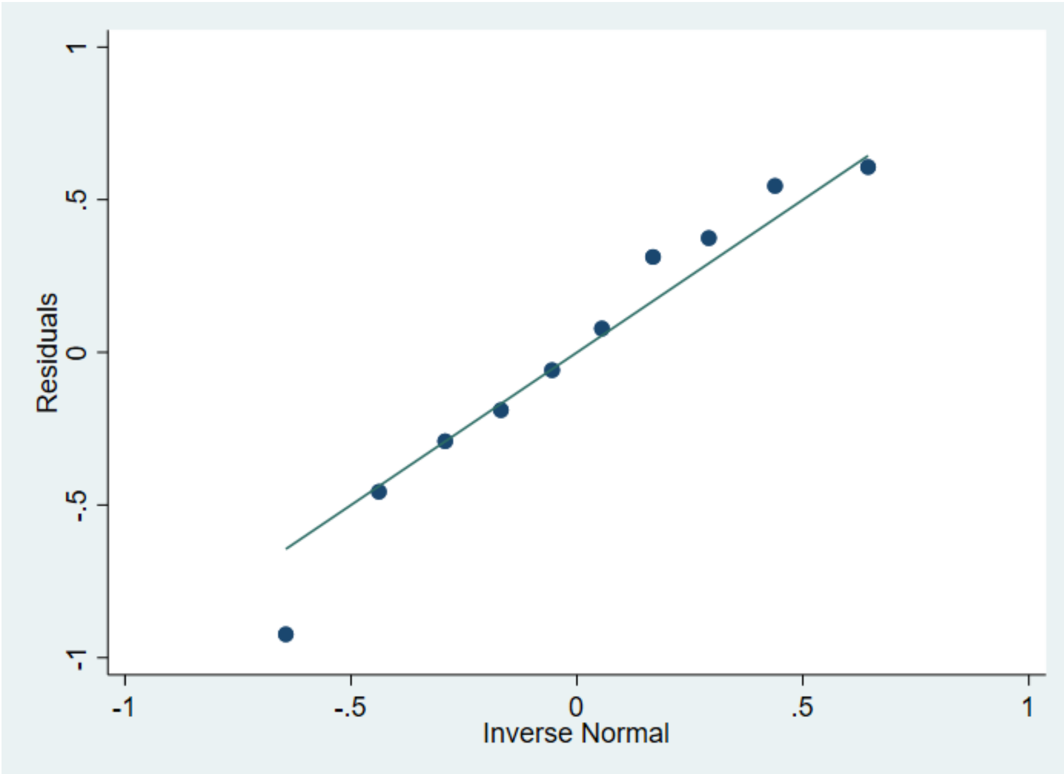

Figure 1. QQ plot for raw residuals.

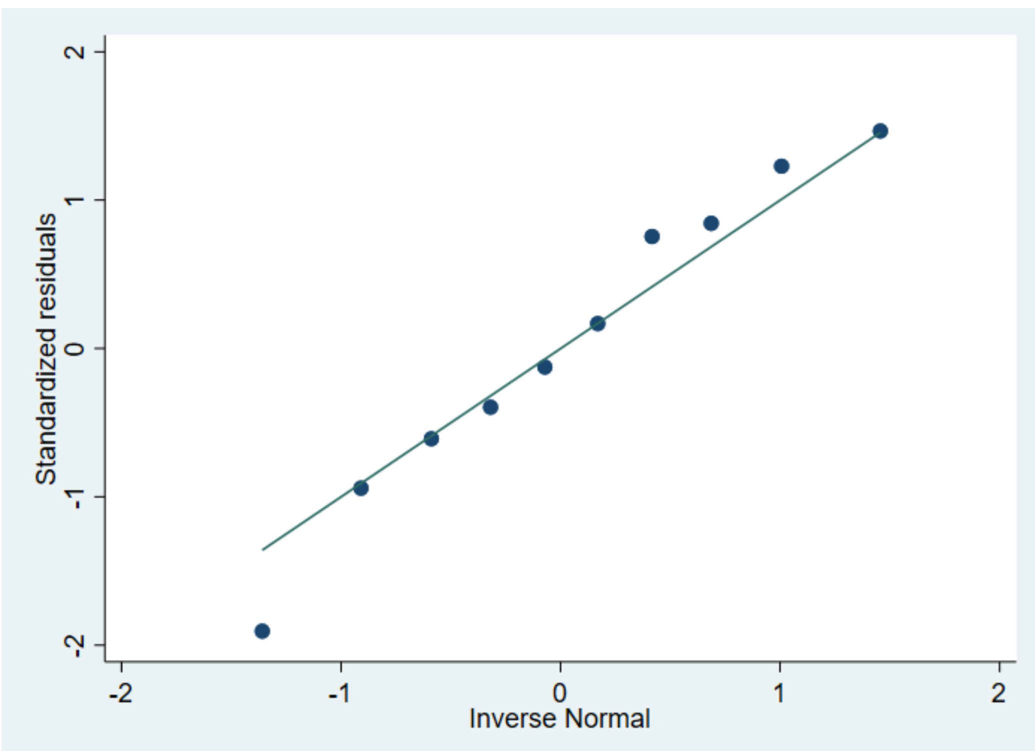

Figure 2. QQ plot for standardized residuals.

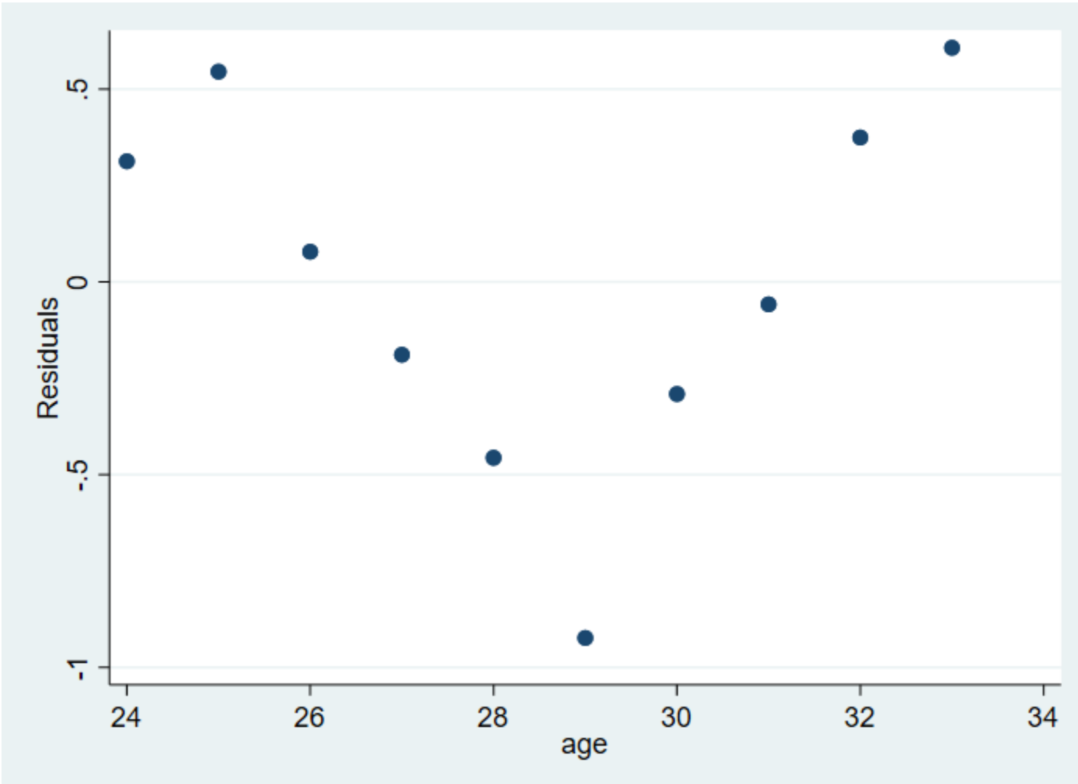

Figure 3. Scatter plot for raw residuals.

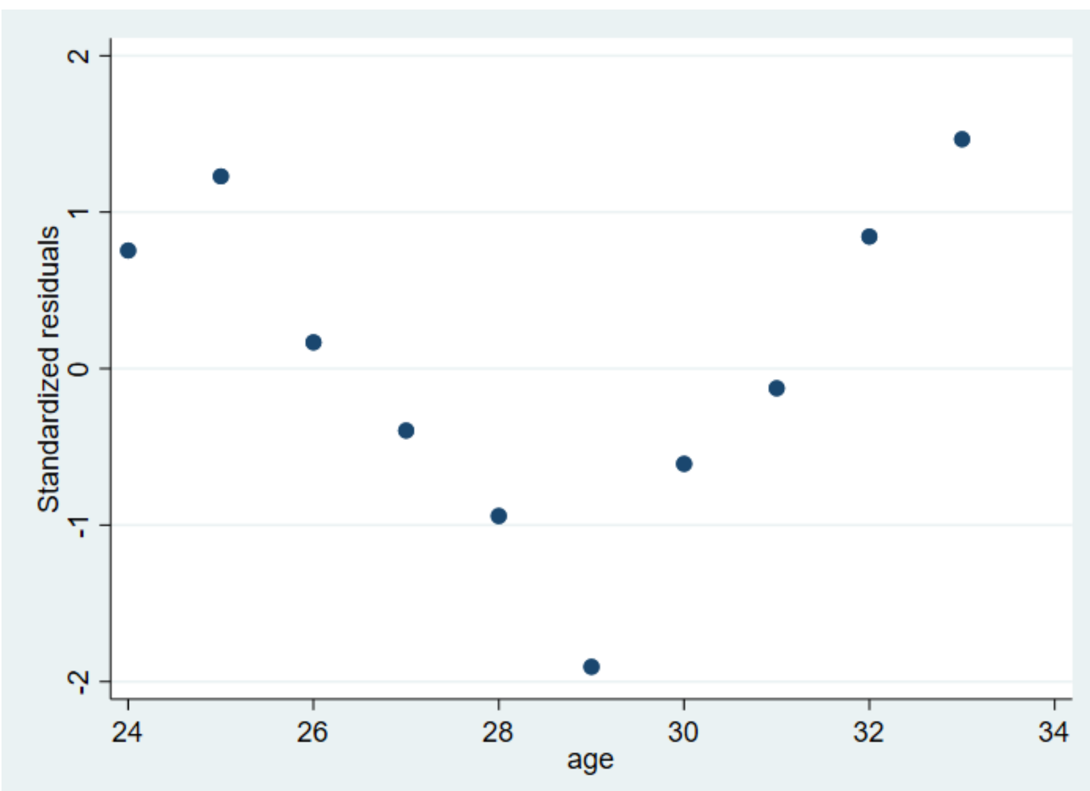

Figure 4. Scatter plot for standardized residuals.

There are some deviations from the diagonal in the QQ plots. There are nonlinearity in the scatterplots. Therefore, we may consider to use non-linear models instead of linear models.

Answer 11.50

The hypothesis is given by: $H_0: \beta = 0, H_1: \beta \neq 0$.

$$L_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{17} = 8205 - \frac{285^2}{10} = 82.5$$

$$L_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{17} = 627.88 - \frac{78.4^2}{10} = 13.224$$

$$L_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{17} = 2264.7 - \frac{285 \times 78.4}{10} = 783$$

$$Regr\ SS = \frac{L_{xy}^2}{L_{xx}} = \frac{30.3^2}{82.5} = 11.128$$

$$Res\ SS = 13.224 - 11.128 = 2.096$$

$$Res\ MS = \frac{2.096}{8} = 0.262$$

$$F = \frac{11.128}{0.262} = 42.48$$

Since $F_{1,8,0.999} = 25.42 < 42.48$, we have $p < 0.001$. We can conclude that there is a significant association between mean thyroxine level and gestational age.