Patrick Fitzgerald
ADS 534 Statistical Modeling
Lab # 1

Numerical Summary of the Data

Figure 1: Summary of Gestational Age (weeks)

The UNIVARIATE Procedure
Variable: gestage (gestage)

Moments			
N	100	Sum Weights	100
Mean	28.89	Sum Observations	2889
Std Deviation	2.53419045	Variance	6.42212121
Skewness	-0.0564008	Kurtosis	-0.2953667
Uncorrected SS	84099	Corrected SS	635.79
Coeff Variation	8.77186032	Std Error Mean	0.25341904

Figure 2: Summary of Head Circumference (cm)

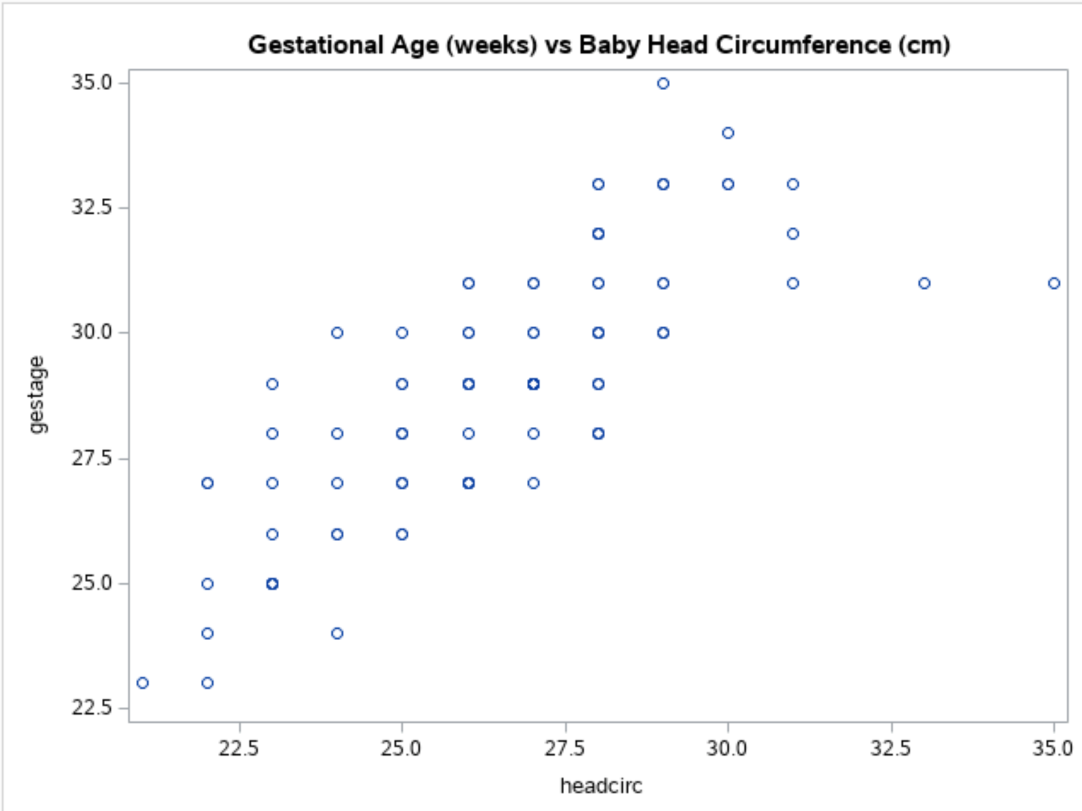
The UNIVARIATE Procedure
Variable: headcirc (headcirc)

Moments			
N	100	Sum Weights	100
Mean	26.45	Sum Observations	2645
Std Deviation	2.53211693	Variance	6.41161616
Skewness	0.25964754	Kurtosis	0.55232742
Uncorrected SS	70595	Corrected SS	634.75
Coeff Variation	9.57322092	Std Error Mean	0.25321169

Analyze the effect of **headcirc** on **gestage**

(a) Scatterplot of **gestage** versus **headcirc**

Figure 3: Scatterplot of Gestational Age vs Head Circumference



(b) What is the model we will fit?

Figure 4: Estimates of Fit-Model

Parameter Estimates						
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	8.22379	1.67862	4.90	<.0001
headcirc	headcirc	1	0.78133	0.06318	12.37	<.0001

(b) What is the estimate for the effect of **gestage** on **headcirc**? How do you interpret this?

Figure 5: ANOVA table for gestage vs headcirc

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	387.50122	387.50122	152.95	<.0001
Error	98	248.28878	2.53356		
Corrected Total	99	635.79000			

Due to a very small p-value we find that it is unlikely we would have obtained such a large F statistic (Figure 5) if β_1 were in fact equal to zero. Therefore, we conclude $\beta_1 \neq 0$ and there is significant evidence at the $\alpha = 0.05$ level to conclude that there is a linear relationship between gestational age and head circumference (see next section question **(a)**).

Inferences based on the model

(a) Perform the appropriate t-test to determine if there is a significant relationship between **gestage** and **headcirc**?

$$H_0: \beta_1 = 0 \text{ and } H_A: \beta_1 \neq 0$$

Referring to **Figure 5:** ANOVA table for gestage vs headcirc we can see that we have a small p-value (< 0.0001) which indicates that we reject the null hypothesis $H_0: \beta_1 = 0$ in favor of the alternative hypothesis $H_A: \beta_1 \neq 0$. There is significant evidence at the $\alpha = 0.05$ level to conclude that there is a linear relationship between gestational age and head circumference

(b) How does this compare to the F-test result given in the output?

The p-values are the same because of the relationship between a t random variable and an F random variable that has 1 numerator degrees of freedom, that is to say for all simple linear regression: $(t_{(n-2)})^2 = F_{(1,\, n-2)}$.
For this example, we can see that the t-value produced in Figure 4 is equal to 12.37 and $(12.37)^2 = \sim 152.95$.

(c) What if we were only interested in testing if increased **gestage** lead to an increase in **headcirc**? Perform this test.

(d) Find a two-sided 95% confidence interval for β1, the regression coefficient of **gestage**.

$\alpha = 1 - (95/100) = 0.05$
Degrees of Freedom (df) = n-2 = 98
Critical Value is t-statistic having (n-2) df and a cumulative probability (p*) = 0.975
$p^* = 1 - \alpha/2 = 0.05/2 = 0.975$
t = 1.66 (from Appendix B in text)
$\alpha = 0.05$, p* = 0.975, df = 98, t = 1.66
Margin of Error = critical value * standard error
ME = 1.66 * 0.06318 = 0.105
The 95% CI for β_1 (gestage) is 0.781 ± 0.105 which is
(0.676, 0.886)

(e) Find a two-sided 95% confidence interval of the mean value of **headcirc** for those with a **gestage** of 33 weeks.

Figure 6: Occurrences of gestage = 33

Gestage = 33 Weeks

Obs	headcirc	length	gestage	birthwt	momage	toxemia
1	30	38	33	1490	32	0
2	29	42	33	1480	30	0
3	28	39	33	1250	31	1
4	31	42	33	1480	30	0
5	30	38	33	1350	34	1
6	29	42	33	1420	28	1
7	29	38	33	1160	33	1
8	28	41	33	1320	36	1

Figure 7: Mean for headcirc where gestage = 33

The MEANS Procedure

Analysis Variable : headcirc headcirc				
N	Mean	Std Dev	Minimum	Maximum
8	29.2500000	1.0350983	28.0000000	31.0000000

From the data obtained in Figures 6 & 7, I created a table and using R was able to obtain a 95% CI based on the mean head circumference (cm) for a gestational age of 33 weeks.
R-code:

```
> headcirc <- c(30, 29, 28, 31, 30, 29, 29, 28 )
> a <- mean(headcirc)
> b <- sd(headcirc)
> n <- 8
> error <- qt(0.975, df=n-1)*b/sqrt(n)
> left <- a-error
> right <- a+error
> left
[1] 28.38464
> right
[1] 30.11536
```

So, the 95% CI for head circumference is 28.384 cm to 30.115 cm based on the acquired dataset.

(f) How do you interpret this interval?

We can say that we are 95% confident that the mean head circumference in centimeters will fall in the interval 28.384 cm and 30.115 cm for a gestational age of 33 weeks. There is 5% chance that the head circumference at 33-week gestational age will not fall into this interval.

(g) Calculate the prediction interval of **headcirc** for a future observation with **gestage** of 33 weeks.

Sample Estimate ± (t-multiplier * SE)

From our model (Figure X.X) we can see that the predicted value for a gestational age of 33 weeks corresponds to a head circumference of 34 cm.
We can see from the same table (Figure X.X) that the standard error (SE) is 0.06318.
Putting all of this together we see:

$$34 \pm (1.66 * 0.06318) = (33.895, 34.105)$$

(h) How do you interpret this prediction interval?

Our 95% prediction interval for a future observation with gestage of 33 weeks tells us that future head circumferences based on a 33 week gestational age will fall between 33.895 cm and 34.105 cm 95% of the time. There is 5% chance that the head circumference at 33-week gestational age will not fall into this interval.