Hao-HW7

7.24

a) There is a positive relationship between calories and carbs

b) Explanatory = calories; response = carbs

c) We might want to fit a regression in order to estimate the amount of carbs based on caloric information

d) No, the fan-shape of the residual plot is evidence of heteroscedasticity

7.26

a) $y = 105.97 + 0.61x$

b) Slope: each additional cm of shoulder girth is associated with an additional 0.61 cm of height. Intercept: assuming zero cm of shoulder girth, height would be 106cm. Since zero cm of shoulder girth obviously wouldn't make sense, the intercept simply serves to adjust the height of the regression line

c) 0.45: 45% of the variation in height is explained by the variation in shoulder girth, or 45% of the variation in height is explained by this model

d) 166.76cm

e) $166.76 - 160 = 6.76$cm residual; the positive residual indicates that the model over-estimated the student's height

f) No, a shoulder girth of 56cm falls outside of the range of shoulder girths within the data set; as such, using this model in this case would be extrapolation and incorrect

```
x_bar = 107.2
x_sd = 10.37
y_bar = 171.14
y_sd = 9.41
r = 0.67

# slope
b1 = y_sd * r / x_sd
b1
```

```
## [1] 0.6079749
```

```
# intercept
b0 = y_bar - b1 * x_bar
b0
```

```
## [1] 105.9651
```

```
# r2
r2 = r^2
r2
```

```
## [1] 0.4489
```
```
# predict height of student with 100cm shoulder girth
x = 100
y = b0 + 100 * b1
y
```
```
## [1] 166.7626
```

7.30

a) y = -0.357 + 4.034x

b) The estimated heart weight for a cat that weighted 0kg would be -0.357g; this is obviously a meaningless measure, and the intercept simply adjusts the height of the regression line

c) For every additional kg of body weight, the estimated heart weight of a cat would increase by 4.034g

d) 64.41% of the variation in heart weight is explained by variation in body weight

e) 0.803

```
# correlation coefficient
r2 = 0.6441
r = sqrt(r2)
r
```
```
## [1] 0.8025584
```

7.40

a) 0.133

b) Yes, the t-stat of 4.13 indicates that the slope is positive at a 95% confidence level

c) Linear relationship (non-linear patterns in the data or residuals) - appears satisfied
No auto-correlation (e.g. seasonality or other trends in the residuals) - appears satisfied
Homoscedasticity (i.e. constant variance or no fan-shape in the residuals plot) - appears satisfied

```
x_bar = -0.0883
y_bar = 3.9983
b0 = 4.010
b1 = (y_bar - b0) / x_bar
b1
```
```
## [1] 0.1325028
```
```
0.0322 * 4.13
```
```
## [1] 0.132986
```