**Homework 9 Solutions**

**Problem 1**

**# Problem 1(a)**

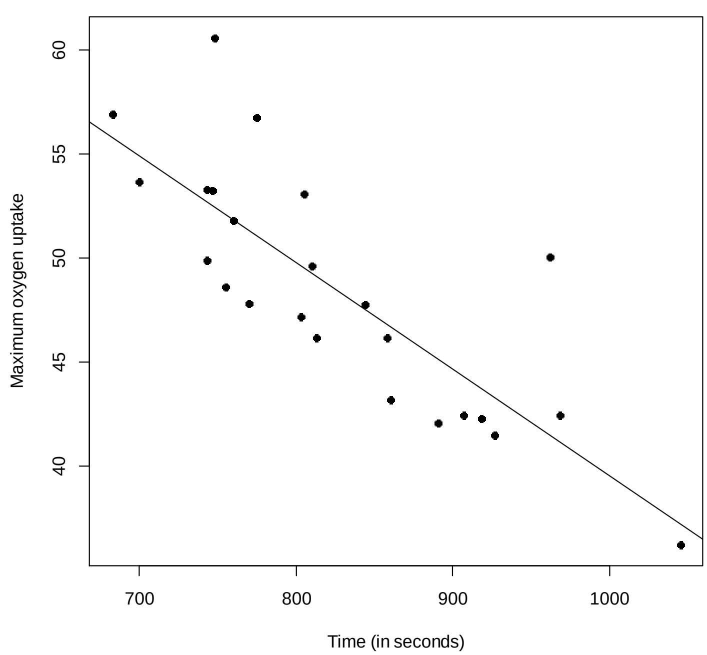
Here is the R code I used to estimate the population model:

```
time = c(918,805,891,962,968,907,770,743,1045,810,927,813,858,860,760,747,
       743,803,683,844,755,700,748,775)

max.O2 = c(42.33,53.10,42.08,50.06,42.45,42.46,47.82,49.92,36.23,49.66,41.49,
       46.17,46.18,43.21,51.81,53.28,53.29,47.18,56.91,47.80,48.65,53.67,60.62,56.76)
```

```
# Estimate the model
fit = lm(max.O2~time)
> fit
Coefficients:
(Intercept)          time
   90.89720      -0.05134
```

The estimate of $\beta_0$ is $b_0$ = 90.8972 and the estimate of $\beta_1$ is $b_1$ = -0.05134. Here is the scatterplot with the estimated model superimposed. The straight-line regression model seems to be plausible.

```
> plot(time,max.O2,xlab="Time (in seconds)",ylab="Maximum oxygen uptake",pch=16)
> abline(fit)
```




**# Problem 1(b)**

Here is the R code I used to perform population level inference (for the slope parameter):

```
fit = lm(max.O2~time)
> summary(fit)
```

```
Call:
lm(formula = max.O2 ~ time)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5425 -2.5733 -0.8386  0.8226  8.5555 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 90.897200   6.542737  13.893 2.27e-12 ***
time        -0.051344   0.007869  -6.525 1.46e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.497 on 22 degrees of freedom
Multiple R-squared:  0.6593,    Adjusted R-squared:  0.6438 
F-statistic: 42.57 on 1 and 22 DF,  p-value: 1.458e-06

> confint(fit,level=0.95)
                  2.5 %       97.5 %
(Intercept) 77.32839389 104.46600554
time        -0.06766371  -0.03502374
```

The p-value for the test that $\beta_1=0$ is 1.46e-06, which is incredibly small. Equivalently, note that the 95 percent interval for $\beta_1$ does not include 0 (and includes only negative values; i.e., we are 95 percent confident that $\beta_1$ is between -0.067 and -0.035). Thus, the time it takes to run two miles has a significant influence on maximum oxygen uptake in this population of runners.

**# Problem 1(c)**

Here is the R code I used to calculate confidence and prediction intervals:

```
fit = lm(max.O2~time)
> predict(fit,data.frame(time=900),level=0.90,interval="confidence")
       fit      lwr      upr
1 44.68785 43.10976 46.26593
```

When running two miles in 900 seconds, we are 90% confident that the population mean maximum oxygen uptake is between 43.10976 and 46.26593.

**# Problem 1(d)**

```
> predict(fit,data.frame(time=900),level=0.90,interval="prediction")
       fit      lwr      upr
1 44.68785 38.47834 50.89735
```

When running two miles in 900 seconds, we are 90% confident that the maximum oxygen uptake for an individual middle-aged man is between 38.47834 and 50.89735.

Difference: Part(c) is estimating the mean maximum oxygen uptake for a population of men. Part(d) is a statement about an individual man.

**Problem 2**

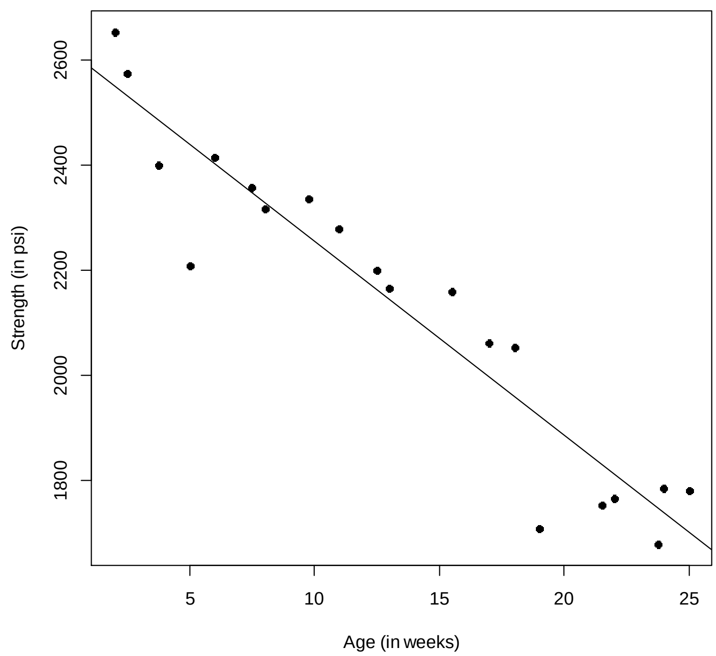
**# Problem 2(a)**

Here is the R code I used to estimate the population model:

```
age = c(15.50,23.75,8.00,17.00,5.00,19.00,24.00,2.50,7.50,11.00,13.00,3.75,
       25.00,9.75,22.00,18.00,6.00,12.50,2.00,21.50)

strength = c(2158.70,1678.15,2316.00,2061.30,2207.50,1708.30,1784.70,2575.00,
       2357.90,2277.70,2165.20,2399.55,1779.80,2336.75,1765.30,2053.50,2414.40,
       2200.50,2654.20,1753.70)

# Estimate the model
fit = lm(strength~age)
> fit

Coefficients:
(Intercept)          age
    2625.39       -36.96
```

The estimate of $\beta_0$ is $b_0$ = 2625.39 and the estimate of $\beta_1$ is $b_1$ = -36.96. Here is the scatterplot with the estimated model superimposed. The straight-line regression model seems to be plausible.

```
> plot(age,strength,xlab="Age (in weeks)",ylab="Strength (in psi)",pch=16)
> abline(fit)
```




**# Problem 2(b)**

Here is the R code I used to perform population level inference (for the slope parameter):

```
fit = lm(strength~age)
> summary(fit)
```

```
Call:
lm(formula = strength ~ age)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5425 -2.5733 -0.8386  0.8226  8.5555

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2625.385     45.347   57.90  < 2e-16 ***
age          -36.962      2.967  -12.46 2.75e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 99.05 on 18 degrees of freedom
Multiple R-squared:  0.8961,    Adjusted R-squared:  0.8903
F-statistic: 155.2 on 1 and 18 DF,  p-value: 2.753e-10

> confint(fit,level=0.95)
                 2.5 %     97.5 %
(Intercept) 2530.11529 2720.65563
age          -43.19484  -30.72875
```

The p-value for the test that $\beta_1=0$ is 2.75e-10, which is incredibly small. Equivalently, note that the 95 percent interval for $\beta_1$ does not include 0 (and includes only negative values; i.e., we are 95 percent confident that $\beta_1$ is between -43.19 and -30.72). Thus, there is strong evidence that the strength of the bond and age of the propellant are negatively linearly related in the population of motors.

**# Problem 2(c)**

Here is the R code I used to calculate confidence and prediction intervals:

```
fit = lm(strength~age)
> predict(fit,data.frame(age=20),level=0.95,interval="confidence")
      fit      lwr      upr
1 1886.15 1823.781 1948.518
```

When made from propellant that is 20 weeks old, we are 95% confident that the population mean strength is between 1823.781 psi and 1948.518 psi.

**# Problem 2(d)**

```
> predict(fit,data.frame(age=20),level=0.95,interval="prediction")
      fit      lwr      upr
1 1886.15 1668.905 2103.394
```

When made from propellant that is 20 weeks old, we are 95% confident that the strength of a single bond (i.e., for a single motor) is between 1668.905 psi and 2103.394 psi.