HW4_solution

Problem 1 (a)

```
mileage = read.table("mileage.csv", header = TRUE, fill = TRUE, sep=",")
mileage$x1 = as.factor(mileage$x1)
summary(mileage)
```

```
##        y          x1          x2
##  Min.   :27.40   A:9   Min.   :0.000
##  1st Qu.:30.00   B:8   1st Qu.:1.000
##  Median :33.00   C:5   Median :1.000
##  Mean   :32.12         Mean   :1.364
##  3rd Qu.:33.88         3rd Qu.:2.000
##  Max.   :35.60         Max.   :3.000
```

```
fit_prob1 = lm(y ~ x1+x2, data = mileage)
summary(fit_prob1)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = mileage)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6171 -1.6321  0.5508  1.3756  4.0021
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  32.0171     1.0005  32.002   <2e-16 ***
## x1B           1.5218     1.2650   1.203    0.245
## x1C           0.5252     1.6194   0.324    0.749
## x2           -0.4192     0.6042  -0.694    0.497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.532 on 18 degrees of freedom
## Multiple R-squared:  0.09453,    Adjusted R-squared:  -0.05638
## F-statistic: 0.6264 on 3 and 18 DF,  p-value: 0.6072
```

(b)

```
confint(fit_prob1)
```

```
##                 2.5 %     97.5 %
## (Intercept) 29.915164 34.1189970
## x1B         -1.135886  4.1795680
## x1C         -2.877095  3.9274823
## x2          -1.688644  0.8502126
```

There is 95% chance that the true β_1 would be in between [-1.135886, 4.1795680].

(c) If we split the null into $\beta_1 = 0$ and $\beta_2 = 0$ and adjust for the Bonferroni correction, the result is still not significantly enough for rejecting the null hypothesis.

Problem 2

```
seeding = as.factor(rep(c(25,50,75,100,125,150),4))
filed = as.factor(c(rep(1,6),rep(2,6),rep(3,6),rep(4,6)))
yield = c(5.1,5.3,5.3,5.2,4.8,5.3,5.4,6.0,5.7,4.8,4.8,4.5,5.3,4.7,5.5,5.0,4.4,4.9,4.7,4.3,4.7,4.4,4.7,4
data_prob2 = data.frame(seeding,filed,yield)
summary(data_prob2)
```

```
##  seeding filed      yield
##  25 :4   1:6   Min.   :4.100
##  50 :4   2:6   1st Qu.:4.700
##  75 :4   3:6   Median :4.850
##  100:4   4:6   Mean   :4.954
##  125:4         3rd Qu.:5.300
##  150:4         Max.   :6.000
```

```
fit_prob2 = lm(yield ~ seeding+filed, data = data_prob2)
anova(fit_prob2)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df Sum Sq Mean Sq F value   Pr(>F)
## seeding    5 1.2671 0.25342  2.1261 0.118366
## filed      3 1.9646 0.65486  5.4941 0.009488 **
## Residuals 15 1.7879 0.11919
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is greater than 0.05, we cannot reject the null hypothesis.

Problem 3 (a)

```
block = as.factor(rep(c(1,2,3,4,5),4))
treatment = as.factor(c(rep(1,5),rep(2,5),rep(3,5),rep(4,5)))
speed = c(12,2,1,8,7,20,14,17,12,17,13,7,13,8,14,11,5,10,3,6)
data_prob3 = data.frame(block,treatment,speed)
summary(data_prob3)
```

```
##  block treatment     speed
##  1:4   1:5       Min.   : 1.00
##  2:4   2:5       1st Qu.: 6.75
##  3:4   3:5       Median :10.50
##  4:4   4:5       Mean   :10.00
##  5:4             3rd Qu.:13.25
##                  Max.   :20.00
```

```r
fit_prob3 = lm(speed ~ treatment+block, data = data_prob3)
anova(fit_prob3)
```

```
## Analysis of Variance Table
##
## Response: speed
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treatment  3  310.0 103.333 14.8503 0.0002421 ***
## block      4  124.5  31.125  4.4731 0.0192167 *
## Residuals 12   83.5   6.958
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value is small, we reject the null.

(b)

```r
pairwise.t.test(data_prob3$speed,data_prob3$treatment, p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  data_prob3$speed and data_prob3$treatment
##
##   1      2      3
## 2 0.0028 -      -
## 3 0.2608 0.2608 -
## 4 1.0000 0.0069 0.5912
##
## P value adjustment method: bonferroni
```

The result suggests there are differences bewteen treatment 1, 2 and treatment 2, 4.

Problem 4

```r
treatment = as.factor(rep(c(1,2,3,4),4))
period = as.factor(c(rep(1,4),rep(2,4),rep(3,4),rep(4,4)))
diet = as.factor(c(4,1,3,2,1,4,2,3,3,2,1,4,2,3,4,1))
milk = c(192,195,292,249,190,203,218,210,214,139,245,163,221,152,204,134)
data_prob4 = data.frame(treatment,period,diet,milk)
fit_prob4 = lm(milk ~ treatment+period+diet, data = data_prob4)
summary(fit_prob4)
```

```
##
## Call:
## lm(formula = milk ~ treatment + period + diet, data = data_prob4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -32.12 -14.56   0.00  10.38  37.62
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   224.62      27.81   8.078 0.000193 ***
## treatment2    -32.00      24.87  -1.287 0.245646
## treatment3     35.50      24.87   1.427 0.203398
## treatment4    -15.25      24.87  -0.613 0.562289
## period2       -26.75      24.87  -1.076 0.323482
## period3       -41.75      24.87  -1.679 0.144237
## period4       -54.25      24.87  -2.181 0.071945 .
## diet2          15.75      24.87   0.633 0.549934
## diet3          26.00      24.87   1.045 0.336134
## diet4          -0.50      24.87  -0.020 0.984613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.17 on 6 degrees of freedom
## Multiple R-squared:  0.7132, Adjusted R-squared:  0.2831
## F-statistic: 1.658 on 9 and 6 DF,  p-value: 0.277
```

```r
anova(fit_prob4)
```

```
## Analysis of Variance Table
##
## Response: milk
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatment  3 9929.2  3309.7  2.6751 0.1409
## period     3 6539.2  2179.7  1.7618 0.2540
## diet       3 1995.7   665.2  0.5377 0.6736
## Residuals  6 7423.4  1237.2
```

Since the p-value is large, we cannot reject the null hypothesis.