⋆ Answer and mark clearly the questions in the provided answer sheets on each of these you should write down your name.
∗ **Note:** No points will be given if no arguments are provided for an answer. Using a hand calculator is allowed.
*Good Luck!* ∼∼ *Yuling* ☺

1. Consider the general linear regression model : $\underline{Y} = D\underline{\beta} + \underline{\epsilon}$, where $E(\underline{\epsilon}) = \underline{0}$, $\sigma^2\{\underline{\epsilon}\} = \sigma^2 \cdot I_{n\times n}$, $\underline{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^t$, $p = k + 1 < n$.
   (a) (10 points) Show that $\underline{b}$ is a least squares estimate of $\underline{\beta}$ if and only if $\underline{b}$ satisfies the normal equations:
   $$D^t\underline{Y} = D^tD\underline{b}.$$
   (b) (10 points) Now, assume $D$ is of full rank. Show that SSTO= $\underline{Y}^tP_1\underline{Y}$, SSR= $\underline{Y}^tP_2\underline{Y}$ and SSE= $\underline{Y}^tP_3\underline{Y}$ with each $P_j$, $j = 1, 2, 3$ be a $n \times n$, symmetric and idempotent matrix. Find $P_j$ and compute rank($P_j$), $j = 1, 2, 3$.
   (c) (10 points) If, furthermore, assume each $\epsilon_i$, $i = 1, \ldots, n$, distributes normally. Show the independence between SSR and SSE, and derive the distribution of $SSR/\sigma^2$ and that of $SSE/\sigma^2$, respectively.

2. (10 points) Show that the coefficient of multiple determination $R^2$ can be viewed as the coefficient of simple determination between the responses $Y_i$'s and the fitted values $\hat{Y}_i$'s.

3. (10 points) A student fitted a linear regression function for a class assignment. Show that: when the student regressed the residuals $e_i$'s against $Y_i$'s, the resulted fitted line has a positive slope; however, when the residuals were regressed against, the fitted values $\hat{Y}_i$'s, the resulted fitted line has a zero slope. Why?

4. (10 points) Let $E(Y_i) = \mu_1$, $i = 1, 2, 3$, $E(Y_i) = \mu_2$, $i = 4, 5, 6, 7$, and $E(Y_i) = \mu_3$, $i = 8, 9, 10, 11, 12$ where $\mu_1$, $\mu_2$ $\mu_3$ are unknown parameters. Also, assume that $Y_1$, $Y_2, \ldots$ $Y_{12}$ are independent normally distributed with common unknown variance $\sigma^2$.
   Write this in a linear regression model form: $\mathbf{Y_{12\times 1}} = \mathbf{D}\beta + \epsilon$, with suitable design matrix $D$ and unknown regression coefficients vector $\beta$.

   Find the best linear unbiased estimator (BLUE) for $\eta$, where $\eta = \mu_1 - (\mu_2 + \mu_3)/2$, and show how to utilize this for testing the hypothesis
   $H_0 : \mu_1 = (\mu_2 + \mu_3)/2$ v.s. $H_1 : \mu_1 \neq (\mu_2 + \mu_3)/2$.

   What if for one-sided $H_1 : \mu_1 > (\mu_2 + \mu_3)/2$?

5. A hospital surgical unit was interested in predicting survival in patients undergoing a particular type of liver operation. A random selection of 54 patients was available for analysis. From each patient record, the following information was extracted from the preopration evaluation:
$X_1$ : blood clotting score (BCS)
$X_2$ : prognostic index (PI)
$X_3$ : enzyme test (ET)
$X_4$ : liver function test score (LT).
$X_5$ : age (Age)
These constitute the pool of potential predictor variables and $Y$ : log of survival time (LST), is the response variable. A multiple linear regression model is fitted to the data.

Attached below is a part of the data analysis output by $R$, with (1), (2), (3), (4) and (5) missing, for the data set.
∗ Indicate where the numbers come from the given output for computing your answer to each of the following questions, otherwise no point will be given.

(a) (10 points) Utilize the available information to calculate (1), (2), (3) (4) and (5); explain your calculation. What is the fitted regression function?

(b) (10 points) Set up the ANOVA table for testing whether there is a regression relation, using $\alpha = 0.01$. Please state the null and alternative, also the decision rule and conclusion.

(c) (10 points) Obtain the Bonferroni joint confidence intervals, with family confidence coefficient 90%, for $\beta_2$ and $\beta_3$; then indicate how the intervals can be used to test a certain null hypothesis, please state the null and alternative, also the decision rule and conclusion.

(d) (10 points) Use the general linear test approach to test $H_0 : \beta_2 = \beta_3 = 0$, at level 0.10. Is the test conclusion the same as you have in (c)? Why?

(e) (10 points) The commercial real estate company would like to predict the response for the cases with the following readings:

| | BCS | PI | ET | LT | Age |
|---|---|---|---|---|---|
| case 1 | 6.2 | 76 | 67 | 1.8 | 45 |
| case 2 | 7.5 | 58 | 77 | 2.3 | 38 |
| case 3 | 8.3 | 82 | 91 | 3.1 | 50 |
| case 4 | 9.0 | 90 | 88 | 3.6 | 48 |

Find the simultaneous Bonferroni prediction intervals for these four cases, with family confidence coefficient 90%.

```
      <<   R  Outputs  For  Problem  5.  >>

> ch09TA01<-matrix(scan("ch09TA01.txt"),ncol=10, byrow=T) ;
Read 540 items
> y <- ch09TA01[,10]; > x1 <- ch09TA01[,1]; > x2 <- ch09TA01[,2];
> x3 <- ch09TA01[,3]; > x4 <- ch09TA01[,4]; > x5 <- ch09TA01[,5];
> dum <- data.frame("LST"=y,"BCS"=x1,"PI"=x2,"ET"=x3,"LT"=x4, "Age"=x5)
> fm <- lm(LST~BCS+PI+ET+LT+Age, dum)
> summary(fm)

Call:
lm(formula = LST ~ BCS + PI + ET + LT + Age, data = dum)

Residuals:
     Min       1Q   Median       3Q      Max
-0.38972 -0.18936  0.00492  0.17813  0.51173

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.047377   0.296662  13.643  < 2e-16 ***
BCS          0.090811   0.028959   3.136  0.00292 **
PI           0.012969   0.002300   (1)    8.91e-07 ***
ET           0.016130   0.002107   7.655 7.33e-10 ***
LT           0.011042   0.053012   0.208  0.83587
Age         -0.004579   0.003196  -1.433  0.15842
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2482 on 48 degrees of freedom
Multiple R-squared: (2),     Adjusted R-squared: 0.745
F-statistic: 31.97 on (3) and (4) DF,  p-value: 3.462e-14

> anova(fm)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq  F value    Pr(>F)
BCS        1 0.7763  0.7763  12.5986 0.0008756 ***
PI         1 2.5888  2.5888  42.0162 4.583e-08 ***
ET         1 6.3341  6.3341 102.8029 1.608e-13 ***
LT         1 0.0246  0.0246   0.3989 0.5306560
Age        1 0.1265  0.1265   2.0527 0.1584200
Residuals 48 2.9575  (5)
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> confint(fm,level=.9875)
                0.625 %    99.375 %
(Intercept)  3.277442921 4.817310996
BCS          0.015653577 0.165968561
PI           0.006998752 0.018938302
ET           0.010661441 0.021598034
LT          -0.126539790 0.148624750
Age         -0.012872825 0.003715463
> confint(fm,level=.975)
                 1.25 %     98.75 %
(Intercept)  3.360930651 4.733823266
BCS          0.023803273 0.157818865
PI           0.007646084 0.018290970
ET           0.011254395 0.021005079
LT          -0.111621068 0.133706028
Age         -0.011973450 0.002816088
> confint(fm,level=.95)
                  2.5 %      97.5 %
(Intercept)  3.450897714 4.643856202
BCS          0.032585453 0.149036686
PI           0.008343654 0.017593400
ET           0.011893367 0.020366107
LT          -0.095544532 0.117629491
Age         -0.011004276 0.001846915
> confint(fm,level=.9)
                    5 %          95 %
(Intercept)  3.549808011 4.5449459061
BCS          0.042240630 0.1393815083
PI           0.009110567 0.0168264868
ET           0.012595857 0.0196636173
LT          -0.077869897 0.0999548565
Age         -0.009938761 0.0007813996
> confint(fm,level=.875)
                 6.25 %       93.75 %
(Intercept)  3.584169801 4.5105841156
BCS          0.045594874 0.1360272651
PI           0.009376995 0.0165600586
ET           0.012839905 0.0194195698
LT          -0.071729665 0.0938146251
Age         -0.009568598 0.0004112359
> confint(fm,level=.75)
                 12.5 %        87.5 %
(Intercept)  3.701931148  4.3928227690
BCS          0.057090206  0.1245319330
PI           0.010290072  0.0156469820
ET           0.013676280  0.0185831943
LT          -0.050686469  0.0727714285
```

```
Age         -0.008300009 -0.0008573527

> # Estimation of mean response, its confidence intervals, prediction intervals

> new<- data.frame(BCS=c(6.2, 7.5, 8.3, 9),PI=c(76, 58, 82, 90),ET=c(67, 77, 91,
+ 88),LT=c(1.8, 2.3, 3.1, 3.6), Age=c(45, 38, 50, 48))

> new
  BCS PI ET  LT Age
1 6.2 76 67 1.8  45
2 7.5 58 77 2.3  38
3 8.3 82 91 3.1  50
4 9.0 90 88 3.6  48

> predict(lm(LST~BCS+PI+ET+LT+Age), new, se.fit = TRUE)

> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="prediction",level=.9875)
       fit      lwr      upr
1 6.490542 5.812892 7.168192
2 6.574032 5.887388 7.260676
3 7.137632 6.438317 7.836947
4 7.271237 6.562157 7.980318
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="confidence",level=.9875)
       fit      lwr      upr
1 6.490542 6.280318 6.700765
2 6.574032 6.336409 6.811655
3 7.137632 6.865554 7.409709
4 7.271237 6.974961 7.567513
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="prediction",level=.975)
       fit      lwr      upr
1 6.490542 5.886373 7.094711
2 6.574032 5.961844 7.186220
3 7.137632 6.514147 7.761117
4 7.271237 6.639046 7.903428
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="confidence",level=.975)
       fit      lwr      upr
1 6.490542 6.303114 6.677970
2 6.574032 6.362176 6.785889
3 7.137632 6.895057 7.380207
4 7.271237 7.007088 7.535386
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="prediction",level=.95)
       fit      lwr      upr
1 6.490542 5.965556 7.015527
2 6.574032 6.042079 7.105986
3 7.137632 6.595862 7.679402
4 7.271237 6.721902 7.820572
```

```
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="confidence",level=.95)
       fit      lwr      upr
1 6.490542 6.327679 6.653405
2 6.574032 6.389942 6.758122
3 7.137632 6.926849 7.348414
4 7.271237 7.041708 7.500766
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="prediction",level=.9)
       fit      lwr      upr
1 6.490542 6.052611 6.928472
2 6.574032 6.130289 7.017775
3 7.137632 6.685700 7.589563
4 7.271237 6.812995 7.729479
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="confidence",level=.9)
       fit      lwr      upr
1 6.490542 6.354685 6.626398
2 6.574032 6.420468 6.727596
3 7.137632 6.961802 7.313462
4 7.271237 7.079769 7.462705
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="prediction",level=.875)
       fit      lwr      upr
1 6.490542 6.082855 6.898229
2 6.574032 6.160934 6.987131
3 7.137632 6.716910 7.558353
4 7.271237 6.844641 7.697834
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="confidence",level=.875)
       fit      lwr      upr
1 6.490542 6.364067 6.617016
2 6.574032 6.431073 6.716991
3 7.137632 6.973945 7.301319
4 7.271237 7.092992 7.449482
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="prediction",level=.75)
       fit      lwr      upr
1 6.490542 6.186501 6.794583
2 6.574032 6.265956 6.882108
3 7.137632 6.823870 7.451393
4 7.271237 6.953094 7.589380
> predict(lm(LST~BCS+PI+ET+LT+Age), new, interval="confidence",level=.75)
       fit      lwr      upr
1 6.490542 6.396221 6.584863
2 6.574032 6.467418 6.680646
3 7.137632 7.015559 7.259705
4 7.271237 7.138307 7.404167


> # General Linear Test Approach

> fm12 <- lm(LST~BCS+PI, dum)
```

```
> fm123 <- lm(LST~BCS+PI+ET, dum)
> fm23 <- lm(LST~PI+ET, dum)
> fm234 <- lm(LST~PI+ET+LT, dum)
> fm13 <- lm(LST~BCS+ET, dum)
> fm134 <- lm(LST~BCS+ET+LT, dum)
> fm24 <- lm(LST~PI+LT, dum)
> fm245 <- lm(LST~PI+LT+Age, dum)
> fm14 <- lm(LST~BCS+LT, dum)
> fm145 <- lm(LST~BCS+LT+Age, dum)
> fm45 <- lm(LST~LT+Age, dum)
> fm345 <- lm(LST~ET+LT+Age, dum)


> anova(fm12)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
BCS        1 0.7763 0.77626  4.1926 0.0457685 *
PI         1 2.5888 2.58880 13.9821 0.0004681 ***
Residuals 51 9.4427 0.18515
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(fm123)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
BCS        1 0.7763  0.7763  12.486 0.0008931 ***
PI         1 2.5888  2.5888  41.640 4.307e-08 ***
ET         1 6.3341  6.3341 101.883 1.174e-13 ***
Residuals 50 3.1085  0.0622
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(fm23)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
PI         1 2.8285  2.8285  33.451 4.467e-07 ***
ET         1 5.6667  5.6667  67.015 7.439e-11 ***
Residuals 51 4.3125  0.0846
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> anova(fm234)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
PI         1 2.8285  2.8285 39.1317 8.757e-08 ***
ET         1 5.6667  5.6667 78.3963 8.178e-12 ***
LT         1 0.6984  0.6984  9.6615  0.003102 **
Residuals 50 3.6141  0.0723
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(fm13)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
BCS        1 0.7763  0.7763  6.8482   0.01165 *
ET         1 6.2505  6.2505 55.1423 1.158e-09 ***
Residuals 51 5.7810  0.1134
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(fm134)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
BCS        1 0.7763  0.7763  7.8128  0.007339 **
ET         1 6.2505  6.2505 62.9099 2.122e-10 ***
LT         1 0.8131  0.8131  8.1841  0.006153 **
Residuals 50 4.9678  0.0994
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(fm24)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
PI         1 2.8285  2.8285  21.784 2.247e-05 ***
LT         1 3.3572  3.3572  25.855 5.321e-06 ***
Residuals 51 6.6220  0.1298
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(fm245)
```

```
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
PI         1 2.8285  2.8285 21.3712 2.695e-05 ***
LT         1 3.3572  3.3572 25.3652 6.555e-06 ***
Age        1 0.0043  0.0043  0.0329    0.8569
Residuals 50 6.6177  0.1324
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(fm14)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
BCS        1 0.7763  0.7763  5.4239   0.02386 *
LT         1 4.7324  4.7324 33.0664 5.034e-07 ***
Residuals 51 7.2991  0.1431
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(fm145)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
BCS        1 0.7763  0.7763  5.3175   0.02530 *
LT         1 4.7324  4.7324 32.4181 6.524e-07 ***
Age        1 0.0000  0.0000  0.0002   0.98943
Residuals 50 7.2990  0.1460
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(fm45)
Analysis of Variance Table

Response: LST
          Df Sum Sq Mean Sq F value    Pr(>F)
LT         1 5.3990  5.3990 37.1726 1.445e-07 ***
Age        1 0.0014  0.0014  0.0098    0.9215
Residuals 51 7.4073  0.1452
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> anova(fm345)
Analysis of Variance Table

Response: LST
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)    
ET         1 5.4762  5.4762 53.6548 1.863e-09 ***
LT         1 2.2019  2.2019 21.5737 2.503e-05 ***
Age        1 0.0266  0.0266  0.2603    0.6122    
Residuals 50 5.1031  0.1021                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

```
> c(qf(.95,1,50),qf(.95,2,50),qf(.95,3,50),qf(.95,4,50))
[1] 4.034310 3.182610 2.790008 2.557179
> c(qf(.95,1,48),qf(.95,2,48),qf(.95,3,48),qf(.95,4,48))
[1] 4.042652 3.190727 2.798061 2.565241
> c(qf(.95,1,46),qf(.95,2,46),qf(.95,3,46),qf(.95,4,46))
[1] 4.051749 3.199582 2.806845 2.574035
> c(qf(.9,1,50),qf(.9,2,50),qf(.9,3,50),qf(.9,4,50))
[1] 2.808658 2.411955 2.196730 2.060816
> c(qf(.9,1,48),qf(.9,2,48),qf(.9,3,48),qf(.9,4,48))
[1] 2.813081 2.416660 2.201591 2.065805
> c(qf(.9,1,46),qf(.9,2,46),qf(.9,3,46),qf(.9,4,46))
[1] 2.817901 2.421788 2.206890 2.071244
> c(qf(.8,1,50),qf(.8,2,50),qf(.8,3,50),qf(.8,4,50))
[1] 1.686657 1.662374 1.604763 1.558084
> c(qf(.8,1,48),qf(.8,2,48),qf(.8,3,48),qf(.8,4,48))
[1] 1.688541 1.664629 1.607214 1.560677
> c(qf(.75,1,46),qf(.8,2,46),qf(.8,3,46),qf(.8,4,46))
[1] 1.357370 1.667085 1.609884 1.563500
> c(qf(.75,1,50),qf(.75,2,50),qf(.75,3,50),qf(.75,4,50))
[1] 1.354597 1.425451 1.412788 1.392661
> c(qf(.75,1,48),qf(.75,2,48),qf(.75,3,48),qf(.75,4,48))
[1] 1.355925 1.427114 1.414628 1.394626
> c(qf(.75,1,46),qf(.75,2,46),qf(.75,3,46),qf(.75,4,46))
[1] 1.357370 1.428925 1.416631 1.396764
>
```