<u>R Code</u>: Q1-Q3

```
# Create the dataset using the Total Fleet statistics
data <- data.frame(
  Year = 1995:2015,
  Salaries = c(510, 553, 623, 616, 606, 650, 720, 758, 487, 430, 386, 386, 397, 435, 465, 473, 493, 521, 633,
738, 767),
  PilotTraining = c(29, 23, 26, 23, 26, 26, 32, 31, 19, 13, 12, 9, 11, 11, 15, 10, 11, 7, 11, 22, 35),
  Benefits = c(91, 103, 95, 128, 124, 157, 273, 342, 288, 220, 162, 155, 153, 181, 218, 203, 141, 194, 220,
237, 242),
  PerDiem = c(42, 46, 48, 50, 54, 56, 61, 55, 45, 43, 43, 46, 51, 55, 55, 55, 52, 52, 64, 102, 57),
  Maintenance = c(730, 762, 860, 857, 875, 890, 961, 936, 837, 799, 868, 936, 1050, 1047, 1088, 1161, 1369,
1034, 1018, 1014, 972),
  AircraftOwnership = c(760, 732, 684, 663, 645, 698, 750, 772, 678, 602, 533, 495, 512, 538, 587, 536, 305,
329, 314, 298, 259),
  BlockHours = c(10.04, 10.13, 10.32, 10.55, 10.68, 10.53, 9.95, 9.58, 9.18, 10.31, 10.86, 11.15, 11.17, 10.72,
10.79, 11.01, 10.71, 10.65, 10.46, 10.42, 10.39)
)

# Perform multiple linear regression
model <- lm(Salaries ~ PilotTraining + Benefits + PerDiem + Maintenance +
            AircraftOwnership + BlockHours, data = data)

# Display the regression equation
cat("The regression equation is:")
print(summary(model)$coefficients)

# Display the R-squared value
r_squared <- summary(model)$r.squared
cat("The R-squared value is:", r_squared, "\n")

# Display the standard error of the estimate
standard_error <- summary(model)$sigma
cat("The standard error of the estimate is:", standard_error, "\n")
```

Output:

```
The regression equation is:> print(summary(model)$coefficients)
                     Estimate  Std. Error    t value     Pr(>|t|)
(Intercept)       457.2682467 484.7225666  0.9433608 3.614924e-01
PilotTraining      11.2020948   2.0371500  5.4989052 7.838327e-05
Benefits            0.1258638   0.2643271  0.4761670 6.413009e-01
PerDiem             3.1470809   1.3157475  2.3918578 3.135390e-02
Maintenance         0.0748456   0.1245415  0.6009693 5.574658e-01
AircraftOwnership  -0.1428217   0.1113780 -1.2823146 2.205591e-01
BlockHours        -28.9059887  42.7235116 -0.6765827 5.096954e-01
>
```

The R-squared value is: 0.8550396 ; meaning it can change 85% accounted for the change in the independent variables

The standard error of the estimate is: 56.79296

The standard error of the forecast (often abbreviated as SEE) is a measure of the accuracy of predictions made by the regression model. Specifically, it represents the average distance that the observed values fall from the regression line (predicted values).

<u>R Code</u>: Q4

```
# Create the dataset using the Total Fleet statistics
data <- data.frame(
  Year = 1995:2015,
  Salaries = c(510, 553, 623, 616, 606, 650, 720, 758, 487, 430, 386, 386, 397, 435, 465, 473, 493, 521, 633,
738, 767),
  PilotTraining = c(29, 23, 26, 23, 26, 26, 32, 31, 19, 13, 12, 9, 11, 11, 15, 10, 11, 7, 11, 22, 35),
  Benefits = c(91, 103, 95, 128, 124, 157, 273, 342, 288, 220, 162, 155, 153, 181, 218, 203, 141, 194, 220,
237, 242),
  PerDiem = c(42, 46, 48, 50, 54, 56, 61, 55, 45, 43, 43, 46, 51, 55, 55, 55, 52, 52, 64, 102, 57),
  Maintenance = c(730, 762, 860, 857, 875, 890, 961, 936, 837, 799, 868, 936, 1050, 1047, 1088, 1161, 1369,
1034, 1018, 1014, 972),
  AircraftOwnership = c(760, 732, 684, 663, 645, 698, 750, 772, 678, 602, 533, 495, 512, 538, 587, 536, 305,
329, 314, 298, 259),
  BlockHours = c(10.04, 10.13, 10.32, 10.55, 10.68, 10.53, 9.95, 9.58, 9.18, 10.31, 10.86, 11.15, 11.17, 10.72,
10.79, 11.01, 10.71, 10.65, 10.46, 10.42, 10.39)
)
# Calculate the correlation matrix
correlation_matrix <- cor(data)

# Display the correlation matrix
print(correlation_matrix)

# Extract correlations with the dependent variable (Salaries)
salaries_correlations <- correlation_matrix["Salaries", ]
cat("\nCorrelations with Salaries:\n")
print(salaries_correlations)

# Determine strong and weak correlations
strong_correlations <- names(salaries_correlations[abs(salaries_correlations) > 0.5])
weak_correlations <- names(salaries_correlations[abs(salaries_correlations) <= 0.5])

cat("\nStrong correlations with Salaries (|r| > 0.5):", strong_correlations, "\n")
cat("Weak correlations with Salaries (|r| <= 0.5):", weak_correlations, "\n")
```

Output:

```
> print(correlation_matrix)
                          Year     Salaries PilotTraining    Benefits     PerDiem Maintenance AircraftOwnership  BlockHours
Year               1.000000000 -0.008135632    -0.4584546  0.42027560  0.49568875  0.71244857      -0.903647723  0.36422086
Salaries          -0.008135632  1.000000000     0.7779946  0.36622497  0.55124496 -0.09204659       0.004596691 -0.46694274
PilotTraining     -0.458454618  0.777994609     1.0000000  0.13273110  0.12992266 -0.45691507       0.440238973 -0.57850467
Benefits           0.420275602  0.366224966     0.1327311  1.00000000  0.34104964  0.17340086      -0.092957741 -0.47537489
PerDiem            0.495688752  0.551244965     0.1299227  0.34104964  1.00000000  0.31885632      -0.443478108  0.02526653
Maintenance        0.712448566 -0.092046587    -0.4569151  0.17340086  0.31885632  1.00000000      -0.617664711  0.45083519
AircraftOwnership -0.903647723  0.004596691     0.4402390 -0.09295774 -0.44347811 -0.61766471       1.000000000 -0.44817679
BlockHours         0.364220856 -0.466942742    -0.5785047 -0.47537489  0.02526653  0.45083519      -0.448176794  1.00000000
>

Correlations with Salaries:
> print(salaries_correlations)
             Year          Salaries     PilotTraining          Benefits           PerDiem       Maintenance AircraftOwnership        BlockHours
     -0.008135632       1.000000000       0.777994609       0.366224966       0.551244965      -0.092046587       0.004596691      -0.466942742
>
```

Strong correlations with Salaries (|r| > 0.5): Salaries PilotTraining PerDiem
Weak correlations with Salaries (|r| <= 0.5): Year Benefits Maintenance AircraftOwnership BlockHours

R Code: Q5

```
# Create the dataset using the Total Fleet statistics
data <- data.frame(
  Year = 1995:2015,
  Salaries = c(510, 553, 623, 616, 606, 650, 720, 758, 487, 430, 386, 386, 397, 435, 465, 473, 493, 521, 633,
738, 767),
  PilotTraining = c(29, 23, 26, 23, 26, 26, 32, 31, 19, 13, 12, 9, 11, 11, 15, 10, 11, 7, 11, 22, 35),
  Benefits = c(91, 103, 95, 128, 124, 157, 273, 342, 288, 220, 162, 155, 153, 181, 218, 203, 141, 194, 220,
237, 242),
  PerDiem = c(42, 46, 48, 50, 54, 56, 61, 55, 45, 43, 43, 46, 51, 55, 55, 55, 52, 52, 64, 102, 57),
  Maintenance = c(730, 762, 860, 857, 875, 890, 961, 936, 837, 799, 868, 936, 1050, 1047, 1088, 1161, 1369,
1034, 1018, 1014, 972),
  AircraftOwnership = c(760, 732, 684, 663, 645, 698, 750, 772, 678, 602, 533, 495, 512, 538, 587, 536, 305,
329, 314, 298, 259),
  BlockHours = c(10.04, 10.13, 10.32, 10.55, 10.68, 10.53, 9.95, 9.58, 9.18, 10.31, 10.86, 11.15, 11.17, 10.72,
10.79, 11.01, 10.71, 10.65, 10.46, 10.42, 10.39)
)

# Perform multiple linear regression
model <- lm(Salaries ~ PilotTraining + Benefits + PerDiem + Maintenance +
        AircraftOwnership + BlockHours, data = data)

# Conduct global F-test
summary_model <- summary(model)

# ANOVA table for global test
anova_result <- anova(model)
print("\nANOVA Table:")
print(anova_result)
```

Output:

```
Analysis of Variance Table

Response: Salaries
                  Df Sum Sq Mean Sq F value    Pr(>F)    
PilotTraining      1 188548  188548 58.4564 2.311e-06 ***
Benefits           1  21927   21927  6.7980  0.020682 *  
PerDiem            1  46963   46963 14.5602  0.001891 ** 
Maintenance        1   2840    2840  0.8805  0.363981    
AircraftOwnership  1   4597    4597  1.4253  0.252372    
BlockHours         1   1476    1476  0.4578  0.509695    
Residuals         14  45156    3225                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The F-statistic tests the null hypothesis that all regression coefficients for the independent variables are equal to zero (i.e., they do not affect *Salaries*). A high F-statistic value indicates that at least one of the coefficients is significantly.

If the p-value associated with the F-statistic is low (typically below 0.05), you can reject the null hypothesis, meaning that there is a statistically significant relationship between the dependent variable (*Salaries*) and at least one of the independent variables.

<u>R Code</u>: Q6

```
# Create the dataset using the Total Fleet statistics
data <- data.frame(
  Year = 1995:2015,
  Salaries = c(510, 553, 623, 616, 606, 650, 720, 758, 487, 430, 386, 386, 397, 435, 465, 473, 493, 521, 633, 738, 767),
  PilotTraining = c(29, 23, 26, 23, 26, 26, 32, 31, 19, 13, 12, 9, 11, 11, 15, 10, 11, 7, 11, 22, 35),
  Benefits = c(91, 103, 95, 128, 124, 157, 273, 342, 288, 220, 162, 155, 153, 181, 218, 203, 141, 194, 220, 237, 242),
  PerDiem = c(42, 46, 48, 50, 54, 56, 61, 55, 45, 43, 43, 46, 51, 55, 55, 55, 52, 52, 64, 102, 57),
  Maintenance = c(730, 762, 860, 857, 875, 890, 961, 936, 837, 799, 868, 936, 1050, 1047, 1088, 1161, 1369, 1034, 1018, 1014, 972),
  AircraftOwnership = c(760, 732, 684, 663, 645, 698, 750, 772, 678, 602, 533, 495, 512, 538, 587, 536, 305, 329, 314, 298, 259),
  BlockHours = c(10.04, 10.13, 10.32, 10.55, 10.68, 10.53, 9.95, 9.58, 9.18, 10.31, 10.86, 11.15, 11.17, 10.72, 10.79, 11.01, 10.71, 10.65, 10.46, 10.42, 10.39)
)

# Perform multiple linear regression
model <- lm(Salaries ~ PilotTraining + Benefits + PerDiem + Maintenance +
            AircraftOwnership + BlockHours, data = data)

# Summary of the model to display coefficients, t-values, and p-values
summary_results <- summary(model)
print(summary_results)

# Extract the p-values of each independent variable
p_values <- summary_results$coefficients[, "Pr(>|t|)"]

# Identify variables with p-values greater than 0.05 for potential deletion
cat("Independent variables with p-values > 0.05 (consider for deletion):\n")
for (i in 2:length(p_values)) {  # Skipping the intercept (index 1)
  if (p_values[i] > 0.05) {
    cat(names(p_values)[i], "- p-value:", p_values[i], "\n")
  }
}
```

Output:

```
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)       457.26825  484.72257   0.943   0.3615    
PilotTraining      11.20209    2.03715   5.499 7.84e-05 ***
Benefits            0.12586    0.26433   0.476   0.6413    
PerDiem             3.14708    1.31575   2.392   0.0314 *  
Maintenance         0.07485    0.12454   0.601   0.5575    
AircraftOwnership  -0.14282    0.11138  -1.282   0.2206    
BlockHours        -28.90599   42.72351  -0.677   0.5097    
```

```
Residual standard error: 56.79 on 14 degrees of freedom
Multiple R-squared:  0.855,     Adjusted R-squared:  0.7929
F-statistic: 13.76 on 6 and 14 DF,  p-value: 3.693e-05
```

Independent variables with p-values > 0.05 (consider for deletion):
Benefits - p-value: 0.6413009
Maintenance - p-value: 0.5574658
AircraftOwnership - p-value: 0.2205591
BlockHours - p-value: 0.5096954
Based on the model summary, which specific variables have p-values above 0.05 and are thus candidates for removal.

R Code: Q7

```
#only Benefits and Maintenance were significant, so fit the model again using only those variables.
# Create the dataset using the Total Fleet statistics
data <- data.frame(
  Year = 1995:2015,
  Salaries = c(510, 553, 623, 616, 606, 650, 720, 758, 487, 430, 386, 386, 397, 435, 465, 473, 493, 521, 633, 738, 767),
  PilotTraining = c(29, 23, 26, 23, 26, 26, 32, 31, 19, 13, 12, 9, 11, 11, 15, 10, 11, 7, 11, 22, 35),
  Benefits = c(91, 103, 95, 128, 124, 157, 273, 342, 288, 220, 162, 155, 153, 181, 218, 203, 141, 194, 220, 237, 242),
  PerDiem = c(42, 46, 48, 50, 54, 56, 61, 55, 45, 43, 43, 46, 51, 55, 55, 55, 52, 52, 64, 102, 57),
  Maintenance = c(730, 762, 860, 857, 875, 890, 961, 936, 837, 799, 868, 936, 1050, 1047, 1088, 1161, 1369, 1034, 1018, 1014, 972),
  AircraftOwnership = c(760, 732, 684, 663, 645, 698, 750, 772, 678, 602, 533, 495, 512, 538, 587, 536, 305, 329, 314, 298, 259),
  BlockHours = c(10.04, 10.13, 10.32, 10.55, 10.68, 10.53, 9.95, 9.58, 9.18, 10.31, 10.86, 11.15, 11.17, 10.72, 10.79, 11.01, 10.71, 10.65, 10.46, 10.42, 10.39)
)

# Re-run the regression with only significant variables
new_model <- lm(Salaries ~ Benefits + Maintenance, data = data)

# Display the summary of the new model, including coefficients
summary(new_model)
```

Output:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 547.4006   184.2035   2.972  0.00817 **
Benefits      0.7387     0.4114   1.795  0.08940 . 
Maintenance  -0.1370     0.1875  -0.731  0.47435   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 120.6 on 18 degrees of freedom
Multiple R-squared:  0.1591,    Adjusted R-squared:  0.06563 
F-statistic: 1.702 on 2 and 18 DF,  p-value: 0.2103
```