

```

1      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
70
71      *****
72      Exam 2
73      Name: Gavin Frias
74      Version: 1
75      *****;
76
77
78      *****
79      ***** Task 1: DATA *****
80      *****;
81
82      /* Question 1: Import Data */
83      TITLE 'Task1 Q1: Import Data';
84
85      %web_drop_table(WORK.IMPORT);
86
87
88      FILENAME REFFILE '/home/u61397358/sasuser.v94/coaster1.csv';
89
90      PROC IMPORT DATAFILE=REFFILE
91      DBMS=CSV
92      OUT=coaster1;
93      GETNAMES=YES;
94      RUN;

```

NOTE: Import cancelled. Output dataset WORK.COASTER1 already exists. Specify REPLACE option to overwrite it.

NOTE: The SAS System stopped processing this step because of errors.

NOTE: PROCEDURE IMPORT used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	306.34k
OS Memory	41908.00k
Timestamp	05/03/2022 11:35:29 PM
Step Count	250 Switch Count 0
Page Faults	0
Page Reclaims	14
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	0

95

```

96      PROC CONTENTS DATA=coaster1; RUN;

```

NOTE: PROCEDURE CONTENTS used (Total process time):

real time	0.07 seconds
user cpu time	0.07 seconds
system cpu time	0.01 seconds
memory	3222.43k
OS Memory	42428.00k
Timestamp	05/03/2022 11:35:29 PM
Step Count	251 Switch Count 0
Page Faults	0
Page Reclaims	102
Page Swaps	0
Voluntary Context Switches	2
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	24

```

97
98
99      %web_open_table(WORK.IMPORT);
100
101      /* Question 2: Remove the rows that contain missing data (see PDF for column) */
102      TITLE 'Task1 Q2: Remove Missing Data';
103
104      DATA Coaster1_Task1;
105      SET Coaster1;
106      IF Drop = . THEN DELETE;
107      RUN;

```

NOTE: There were 200 observations read from the data set WORK.COASTER1.

NOTE: The data set WORK.COASTER1_TASK1 has 110 observations and 11 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	859.03k
OS Memory	42428.00k
Timestamp	05/03/2022 11:35:29 PM
Step Count	252 Switch Count 2
Page Faults	0
Page Reclaims	121
Page Swaps	0
Voluntary Context Switches	14
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	272

```

108
109      /* Question 3: Create a new character variable */
110      TITLE 'Task1 Q3: Create Character Variable';
111
112      DATA Coaster1_Task1;
113      SET Coaster1;
114      LENGTH LengthGroup $6.;
115      IF Length<2500 THEN LengthGroup="Short";
116      IF Length>=2500 AND Length<4000 THEN LengthGroup="Medium";
117      IF Length>=4000 THEN LengthGroup="Long";
118      RUN;

```

NOTE: There were 200 observations read from the data set WORK.COASTER1.

NOTE: The data set WORK.COASTER1_TASK1 has 200 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	1088.87k
OS Memory	42428.00k
Timestamp	05/03/2022 11:35:29 PM
Step Count	253 Switch Count 2
Page Faults	0
Page Reclaims	117
Page Swaps	0
Voluntary Context Switches	16
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

119
120      /* Question 4: Create a new variable called Ratio */
121      TITLE 'Task1 Q4: Create Ratio';
122
123      DATA Coaster1_Task1;
124      SET Coaster1;
125      Ratio=Height/Drop;
126      RUN;

```

NOTE: Missing values were generated as a result of performing an operation on missing values.

Each place is given by: (Number of times) at (Line):(Column).

90 at 125:14

NOTE: There were 200 observations read from the data set WORK.COASTER1.

NOTE: The data set WORK.COASTER1_TASK1 has 200 observations and 12 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	926.78k
OS Memory	42428.00k
Timestamp	05/03/2022 11:35:29 PM
Step Count	254 Switch Count 2
Page Faults	0
Page Reclaims	113
Page Swaps	0
Voluntary Context Switches	13
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

127
128

```

129      /* Question 5: Create a New Dataset called High_Ratio and Print it */
130      TITLE 'Task1 Q5: Create Dataset High_Ratio';
131
132      DATA High_Ratio;
133      SET Coaster1_Task1;
134      WHERE Ratio>1.15;
135      KEEP Track Height Drop Length;
136      RUN;

```

NOTE: There were 18 observations read from the data set WORK.COASTER1_TASK1.
WHERE Ratio>1.15;

NOTE: The data set WORK.HIGH_RATIO has 18 observations and 4 variables.

NOTE: DATA statement used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	981.31k
OS Memory	42428.00k
Timestamp	05/03/2022 11:35:29 PM
Step Count	255 Switch Count 2
Page Faults	0
Page Reclaims	125
Page Swaps	0
Voluntary Context Switches	11
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	264

```

137
138
139      *****
140      ***** Task 2: INTRODUCTORY ANALYSIS *****
141      *****;
142
143      /* Question 6: Compute values of sample mean / median / std dev / IQR
144      / # Observations / # Missing */
145      TITLE 'Task2 Q6: Summary Statistics';
146
147      PROC SORT DATA=Coaster1; by Duration; RUN;

```

NOTE: Input data set is already sorted, no sorting done.

NOTE: PROCEDURE SORT used (Total process time):

real time	0.00 seconds
user cpu time	0.00 seconds
system cpu time	0.00 seconds
memory	557.06k
OS Memory	42168.00k
Timestamp	05/03/2022 11:35:29 PM
Step Count	256 Switch Count 0
Page Faults	0
Page Reclaims	50
Page Swaps	0
Voluntary Context Switches	0
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	0

```

148      PROC MEANS DATA=Coaster1 MEAN MEDIAN STDDEV QRANGE N NMISS;
149      by Duration;
150      RUN;

```

NOTE: There were 200 observations read from the data set WORK.COASTER1.

NOTE: PROCEDURE MEANS used (Total process time):

real time	1.65 seconds
user cpu time	1.66 seconds
system cpu time	0.00 seconds
memory	2619.93k
OS Memory	43196.00k
Timestamp	05/03/2022 11:35:31 PM
Step Count	257 Switch Count 59
Page Faults	0
Page Reclaims	182
Page Swaps	0
Voluntary Context Switches	150
Involuntary Context Switches	5
Block Input Operations	0
Block Output Operations	328

```

152
153      /* Question 7: Histogram with density kernel */
154      TITLE 'Task2 Q7: Histogram with Density Kernel';
155
156      PROC SGPLOT DATA=Coaster1;
157      HISTOGRAM Height;
158      DENSITY Height / type=kernel;
159      RUN;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

real time	0.15 seconds
user cpu time	0.08 seconds
system cpu time	0.01 seconds
memory	13492.78k
OS Memory	47804.00k
Timestamp	05/03/2022 11:35:31 PM
Step Count	258 Switch Count 1
Page Faults	0
Page Reclaims	2022
Page Swaps	0
Voluntary Context Switches	231
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	912

NOTE: There were 200 observations read from the data set WORK.COASTER1.

```

160
161
162      /* Question 8: Bar Chart */
163      TITLE 'Task2 Q8: Bar Chart';
164
165      PROC SGPLOT DATA=Coaster1;
166      VBAR SpeedGroup;
167      RUN;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

real time	0.10 seconds
user cpu time	0.05 seconds
system cpu time	0.01 seconds
memory	2174.28k
OS Memory	49348.00k
Timestamp	05/03/2022 11:35:31 PM
Step Count	259 Switch Count 2
Page Faults	0
Page Reclaims	626
Page Swaps	0
Voluntary Context Switches	201
Involuntary Context Switches	1
Block Input Operations	0
Block Output Operations	416

NOTE: There were 200 observations read from the data set WORK.COASTER1.

```

168
169
170      /* Question 9: Boxplot */
171      TITLE 'Task2 Q9: Boxplot';
172      /* CODE */
173
174      PROC SGPLOT DATA=Coaster1;
175      HBOX Drop;
176      RUN;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

real time	0.10 seconds
user cpu time	0.05 seconds
system cpu time	0.00 seconds
memory	2259.50k
OS Memory	49220.00k
Timestamp	05/03/2022 11:35:31 PM
Step Count	260 Switch Count 1
Page Faults	0
Page Reclaims	297
Page Swaps	0
Voluntary Context Switches	221
Involuntary Context Switches	0
Block Input Operations	0
Block Output Operations	440

NOTE: There were 200 observations read from the data set WORK.COASTER1.

```

177
178
179      /*
180      Are there outliers?
181      Yes, there is at least one outlier to the far right based on the box plot.
182      */
183
184
185      *****
186      ***** Task 3: INFERENCE *****
187      *****;
188
189      TITLE 'Task3 Q10, Q11: Inference';
190      /* CODE */
191      proc ttest data=Coaster1 h0=0 sides=2 ALPHA=0.017 plots;
192      var Length;
193      run;

```

NOTE: PROCEDURE TTEST used (Total process time):

```

real time          0.35 seconds
user cpu time      0.17 seconds
system cpu time    0.04 seconds
memory            10717.21k
OS Memory          56804.00k
Timestamp          05/03/2022 11:35:32 PM
Step Count                261  Switch Count  24
Page Faults              0
Page Reclaims           13643
Page Swaps              0
Voluntary Context Switches 729
Involuntary Context Switches 1
Block Input Operations   0
Block Output Operations  1208

```

```

194
195      proc ttest data=Coaster1 ho=-800 sides=u ALPHA=0.017 plots;
196      var Length;
197      RUN;

```

NOTE: PROCEDURE TTEST used (Total process time):

```

real time          0.35 seconds
user cpu time      0.16 seconds
system cpu time    0.05 seconds
memory            9444.56k
OS Memory          57320.00k
Timestamp          05/03/2022 11:35:32 PM
Step Count                262  Switch Count  24
Page Faults              0
Page Reclaims           13216
Page Swaps              0
Voluntary Context Switches 717
Involuntary Context Switches 0
Block Input Operations   0
Block Output Operations  1256

```

```

198
199      /* Question 10: Equal Variance Test */
200      /* Hypotheses
201      H0: Steel Tracks - Wood Tracks = 0
202      H1: Steel Tracks - Wood Tracks != 0
203      Test Statistic: 23.31
204      P-Value: <0.0001
205      Decision: Reject H0
206      Conclusion: There is enough evidence to suggest a difference in length between wood and steel track roller coasters.
207      */
208
209
210      /* Question 11: Mean Testing */
211      /*Hypotheses
212      H0: Steel Tracks - Wood Tracks = -800
213      H1: Steel Tracks - Wood Tracks < -800
214      Test Statistic: 30.23
215      P-Value: <0.0001
216      Decision: Reject H0
217      Conclusion: There is enough evidence to suggest that the mean length of Steel Tracks - Wood Tracks is less than -800.
218      */
219
220
221      *****
222      ***** Task 4: REGRESSION *****

```

```

223 *****;
224 TITLE 'Task4 Q12: Multiple Linear Regression';
225 /* CODE */
226
227 PROC REG DATA=Coaster1 ALPHA=0.04 ;
228 MODEL Duration = Length Type / corrb;
229 RUN;
230
231 /*
232
233 Part a - Check model assumptions
234 Linearity
235 Graph / results looked at: Plot of residuals vs Length and Type.
236 Is the linearity condition met or not? Yes.
237
238 Normality
239 Graph / results looked at: Plots of residual vs quantile and percent vs residual.
240 Is the normality of residuals condition met or not? Yes
241
242 Equal Variance
243 Graph / results looked at: Plot of residual vs predicted value
244 Is the equal variance of residuals condition met or not? Yes.
245
246
247 Part b - Give the equation of the Multiple Linear Regression line
248
249 Duration = B0+B1Length+B2Type
250 Y = 45.15060 + 0.02386Length + 12.28970Type
251
252
253 Part c - Does the model in total explain variability in Duration?
254 Hypotheses
255 H0: beta_length = beta_type = 0
256 H1: beta_length = beta_type != 0
257 Test Statistic: 146.85
258 P-Value: <0.0001
259 Decision: Reject H0
260 Conclusion: There is enough evidence to suggest that at least one variable explains the variability in Duration.
261
262
263 Part d (If needed. If not needed, state why.)
264
265 Testing Individual Variables (Variable 1)
266 Hypotheses
267 H0: beta_length = 0
268 H1: beta_length != 0
269 Test Statistic: 16.89
270 P-Value: <0.001
271 Decision: Reject H0
272 Conclusion: There is enough evidence to suggest that Length explains some variability in Duration.
273
274
275 Testing Individual Variables (Variable 2)
276 Hypotheses
277 H0: beta_type = 0
278 H1: beta_type != 0
279 Test Statistic: 1.95
280 P-Value: 0.0532
281 Decision: Do Not Reject H0
282 Conclusion: There is not enough evidence to suggest that Type explains some variability in Duration.
283
284
285 Part e - Value of R^2 and interpretation
286 R^2: 0.6835
287 Interpretation: We can interpret this as 68.35% of the variability observed in Duration is explained by the model.
288 */
289
290
291 *****
292 ***** Task 5: 1-way ANOVA *****
293 *****
294 TITLE 'Task5 Q13: 1-Way ANOVA';
295
296 TITLE2 'Part a: Mean Duration for each Group';
297 /* CODE */

```

NOTE: PROCEDURE REG used (Total process time):

real time	0.50 seconds
user cpu time	0.25 seconds
system cpu time	0.05 seconds
memory	11436.50k

```

OS Memory          59392.00k
Timestamp          05/03/2022 11:35:32 PM
Step Count         263  Switch Count  24
Page Faults        0
Page Reclaims      12326
Page Swaps         0
Voluntary Context Switches 821
Involuntary Context Switches 1
Block Input Operations 0
Block Output Operations 1184

```

```

298      PROC MEANS; CLASS SpeedGroup;
299      RUN;

```

NOTE: There were 200 observations read from the data set WORK.COASTER1.

NOTE: PROCEDURE MEANS used (Total process time):

```

real time          0.07 seconds
user cpu time      0.07 seconds
system cpu time    0.01 seconds
memory            7268.06k
OS Memory          58072.00k
Timestamp          05/03/2022 11:35:32 PM
Step Count         264  Switch Count  1
Page Faults        0
Page Reclaims      1373
Page Swaps         0
Voluntary Context Switches 15
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 32

```

```

300
301      /* Detail any difference by group.
302      Some differences to note are that the Fast rollercoasters have a higher mean duration
303      and also have the highest Duration of any rollercoaster.
304      */
305
306
307      TITLE2 'Part b: Side by Side Boxplots';
308      /* CODE */
309      PROC SGPLOT DATA=Coaster1;
310      HBOX Duration / Category=SpeedGroup;
311      RUN;

```

NOTE: PROCEDURE SGPLOT used (Total process time):

```

real time          0.12 seconds
user cpu time      0.05 seconds
system cpu time    0.01 seconds
memory            2193.21k
OS Memory          53836.00k
Timestamp          05/03/2022 11:35:33 PM
Step Count         265  Switch Count  1
Page Faults        0
Page Reclaims      313
Page Swaps         0
Voluntary Context Switches 398
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 496

```

NOTE: There were 200 observations read from the data set WORK.COASTER1.

```

312
313
314      /* Detail any difference by group.
315      The boxplot for the Fast variable has several outliers, while the Middle variable had the widest interval.
316      Another thing to note is that the Small variable boxplot seemed to be the most normal.
317      */
318
319
320      TITLE2 'Part c: Run a 1-way ANOVA model';
321
322      PROC GLM DATA=Coaster1 ALPHA=0.015;
323      CLASS SpeedGroup;
324      MODEL Duration = SpeedGroup;
325      MEANS SpeedGroup / BON CLDIFF HOVTEST=LEVENE;
326      OUTPUT OUT = ANOVA13 r = residual;
327      RUN;

```

328

```

329     TITLE2 'Part d: Normality Test';
330     /* Will you test the normality assumption using the overall dataset, or for each group individually?
331     The overall dataset. */
332
333
334     /* CODE, if needed */

```

NOTE: The data set WORK.ANOVA13 has 200 observations and 12 variables.

NOTE: PROCEDURE GLM used (Total process time):

```

real time      0.31 seconds
user cpu time   0.20 seconds
system cpu time 0.01 seconds
memory         4293.06k
OS Memory      54876.00k
Timestamp      05/03/2022 11:35:33 PM
Step Count     266  Switch Count  5
Page Faults    0
Page Reclaims  748
Page Swaps     0
Voluntary Context Switches 810
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 1136

```

```

335     PROC UNIVARIATE NORMAL PLOT DATA=Coaster1 ALPHA=0.015;
336     VAR Duration;
337     RUN;

```

NOTE: PROCEDURE UNIVARIATE used (Total process time):

```

real time      0.21 seconds
user cpu time   0.14 seconds
system cpu time 0.00 seconds
memory         3382.65k
OS Memory      54136.00k
Timestamp      05/03/2022 11:35:33 PM
Step Count     267  Switch Count  0
Page Faults    0
Page Reclaims  365
Page Swaps     0
Voluntary Context Switches 263
Involuntary Context Switches 0
Block Input Operations 0
Block Output Operations 504

```

```

338
339
340     /* Conclusion(s): The data passes the normality check. Shapiro-Wilk = 0.0955 which is greater than 0.05. */
341
342

```

```

343     TITLE2 'Part e: Equal Variance Assumption Check';
344     /* Conclusion: The data passes the equal variance check.*/
345
346

```

```

347     TITLE2 'Part f: Is there a significant evidence of an effect?';
348     /*Hypotheses
349     H0: = 0
350     H1: != 0
351     Test Statistic: 23.22
352     P-Value: <0.0001
353     Decision: Reject H0
354     Conclusion: There is enough evidence to suggest that Speed Group explains some variability in Duration.
355     */
356
357

```

```

358     TITLE2 'Part g: Bonerroni or Tukey';
359     /* Are you providing Bonferroni or Tukey Intervals?
360     Bonferroni Intervals */
361
362

```

```

363     /* Provide confidence intervals for each difference
364     (make sure to indicate the difference you are writing a confidence interval for):
365     Fast - Middle (5.348,55.575)
366     Fast - Slow(33.885,83.587)
367     Middle - Fast (-55.575,-5.348)
368     Middle - Slow (6.655,49.894)
369     Slow - Fast(-83.587,-33.885)
370     Slow - Middle (-49.894,-6.655)
371     */
372
373

```



```
374      /* For each pair, state whether the difference is significant or not
375      According to my output the difference of each of these confidence intervals are significant.
376      */
377
378
379      TITLE;
380      TITLE2;
381
382      OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
393
```