SIT191 Problem Solving Task 3

Michael Alford 221147552

Due: by 8pm Monday October 4th
Total marks: 80, Weighting: 20%

Question 3.1

The International Tennis Federation (ITF) requires that tennis balls have an average diameter of 6.70 centimetres. Tennis balls being produced by one manufacturer are tested to see if they meet the ITF standard. A random sample of 15 tennis balls is taken and the diameters (in cm) are shown below:

```
## [1] 6.74 6.64 6.71 6.57 6.91 6.65 6.80 6.72 6.50 6.62 6.54 6.85 6.73 6.79 6.67
```

a. Create a 98% confidence interval for the mean diameter of tennis balls produced by the manufacturer and interpret your interval in context.

$$\bar{x} \pm t^*_{n-1} \times \frac{s}{\sqrt{n}}$$
$$6.696 \pm 0.135 \times \frac{0.114}{\sqrt{15}}$$
$$= 6.62cm, 6.77cm$$

Based on the tested sample, we can be 98% confident that all manufactured tennis are between 6.62cm to 6.77cm in diameter.

b. Perform a hypothesis test to see if there is evidence that the manufacturer's tennis balls are smaller than the ITF standard. Use $\alpha = 0.01$.

$$H_0 : \mu = \mu_0$$
$$H_a : \mu < \mu_0$$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$
$$= \frac{6.696 - 6.7}{\frac{0.114}{\sqrt{15}}}$$
$$= -0.136$$

0.25 < p-value < 0.5

In this case, the p-value is greater than 0.01, confirming that the result is not significant.

c. Explain how your confidence interval and hypothesis test conclusion support each other.

The confidence interval of the range of sizes includes the mean and are therefore, not all smaller than the 6.7cm requirement, while the hypothesis test confirms that only 25% to 50% of manufactured tennis balls are smaller than the 6.7cm requirement.

d. Are the assumptions required for the test you performed met? Discuss briefly.

Yes, the data gathered satisfies the requirements of independence and normality.

Question 3.2

Exercise vigorous enough to increase heart rate to between 50% and 85% of maximum heart rate is recommended for good health. Does boxing high intensity interval training (HIIT) or cardio tennis increase heart rate the most?

Ten subjects participated in two different 45 minute exercise classes – boxing HIIT and cardio tennis. At the end of each class their heart rate (beats per minute) was recorded.

Boxing.HIIT	Cardio.Tennis
145	152
168	159
157	143
129	122
189	162
135	130
165	154
172	175
127	132
148	146

a. Should the results best be analysed as paired data or two independent means? Explain.

The data in this example should be analysed as two independent means as we are analysing the differences in the effect side of the relationship. Paired data on the other hand, would specifically analyse the cause and effect relationship before and after a single scenario exclusively.

b. Briefly describe how this experiment could be designed to help avoid biased results.

The best design method for this experiment would be to include the same participants for each exercise, with a suitable rest period in between.

c. Is there a significant difference in heart rate between boxing HIIT and cardio tennis? Carry out a hypothesis test using $\alpha = 0.05$, showing all steps.

$$
\begin{aligned}
t &= \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \\
&= \frac{6.0}{\frac{10.154}{\sqrt{10}}} \\
&= 1.87
\end{aligned}
$$

$$
\begin{aligned}
H_0 &: \mu_d = 0 \\
H_a &: \mu \neq 0
\end{aligned}
$$

$0.05 < p - value < 0.10$

Reject null hypothesis as p-value is between 0.05 and 0.10 confirming the differences are not 0.

d. Calculate a 95% confidence interval for the difference in heart rate between the two classes and interpret your interval in context.

$$\bar{d} \pm t^*_{n-1} \times \frac{s_d}{\sqrt{n}}$$
$$6.0 \pm 1.87 \times \frac{10.154}{\sqrt{10}}$$
$$-1.26,\ 13.26$$

We can be 95% confident that the differences in heart rate between the two exercise regimes are from -1.26 to 13.26bpm.

Based on your hypothesis test result, did you expect to find 0 in the interval? Explain.

Yes, as the zero value indicates that there is no difference to heart rate between the two exercise regimes for some people

Question 3.3

Resting heart rate was compared for a random sample of healthy adult men and women of the same age. The 15 women sampled had a resting heart rate of 78 beats per minute with standard deviation 5 bpm. The resting heart rate for the sampled 18 men was 71 bpm with standard deviation 3 bpm.

a. Should the results best be analysed as paired data or two independent means? Explain.

These results should be analysed as independent as the test subjects are from different groups (i.e. men and women).

b. Do women of this age have a significantly higher resting heart rate than men? Carry out a hypothesis test with a significance level of 5%, including all steps.

$$\begin{aligned} t &= \frac{(78-71)}{\sqrt{\frac{5^2}{15}+\frac{3^2}{18}}} \\ &= \frac{7}{1.414} \\ &= 4.95 \end{aligned}$$

---

c. Create a 90% confidence interval for the mean difference in heart rate, and interpret the interval in context.

$$(\bar{x}_1 - \bar{x}_2) \pm t^*_{n_1+n_2-2} \times \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$
$$(78-71) \pm 1.697 \times \sqrt{\frac{5^2}{15}+\frac{3^2}{18}}$$
$$= 4.502, 9.498$$

d. Is the confidence interval consistent with your hypothesis test conclusion? Explain.

e. Is it necessary for the test you used that resting heart rate is normally distributed? Explain.
Yes, the results of both variables need to both be normally distributed to ensure the data are comparable.

Question 3.4

A hospital is comparing three different drugs to determine if there is a difference in the number of hours of pain relief each provides on average for patients with pain after surgery. The mean pain relief (in hours) for each drug and SPSS ANOVA output are given below.

Drug	Mean pain relief
1	8.28
2	7.52
3	8.36

ANOVA

time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.385	2	1.693	2.662	.093
Within Groups	13.352	21	.636		
Total	16.737	23			

Multiple Comparisons

Dependent Variable: time

Bonferroni

					95% Confidence Interval	
(I) Drug	(J) Drug	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
1.00	2.00	.75903	.38746	.191	-.2489	1.7669
	3.00	-.07937	.40185	1.000	-1.1247	.9660
2.00	1.00	-.75903	.38746	.191	-1.7669	.2489
	3.00	-.83839	.41269	.165	-1.9119	.2352
3.00	1.00	.07937	.40185	1.000	-.9660	1.1247
	2.00	.83839	.41269	.165	-.2352	1.9119

a. Does the average length of pain relief differ between any of the drugs? State the appropriate hypotheses for this test.

b. Give the values of the test statistic and P-value, and your conclusion for the test. Use $\alpha = 0.05$.

c. Would it be appropriate to perform a Bonferroni test to see if any of the drugs differ in the mean hours of pain relief? Explain.

d. Describe a well designed experiment to obtain this data.

Question 3.5

At an athletics meet, organisers were interested in whether the standard of the competitors running in the 400m event varied significantly between heats. There were four heats for the 400m event, with 6 competitors in heats 1 and 3, 8 runners in heat 2 and 7 in heat 4. The time (in seconds) each runner took to complete the event was recorded, with the data given in the following table.

X1	X2	X3	X4
51.92	53.25	53.10	51.22
51.29	52.17	53.20	50.10
52.90	53.10	53.12	51.19
52.97	51.53	52.96	50.34
52.10	54.22	53.91	51.78
51.89	52.85	52.55	52.25
NA	54.17	NA	51.22
NA	54.58	NA	NA

a. State the null and alternative hypotheses required to test if the average 400m running times were significantly different between any of the heats.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_a : \textit{Not all means are equal}$$

b. Carry out the ANOVA (with technology) and report the test statistic and P-value.

The test statistic and P-value are 10.67 and 0.0001 respectively

c. Write a brief conclusion of the ANOVA analysis, in context. Use $\alpha = 0.05$.

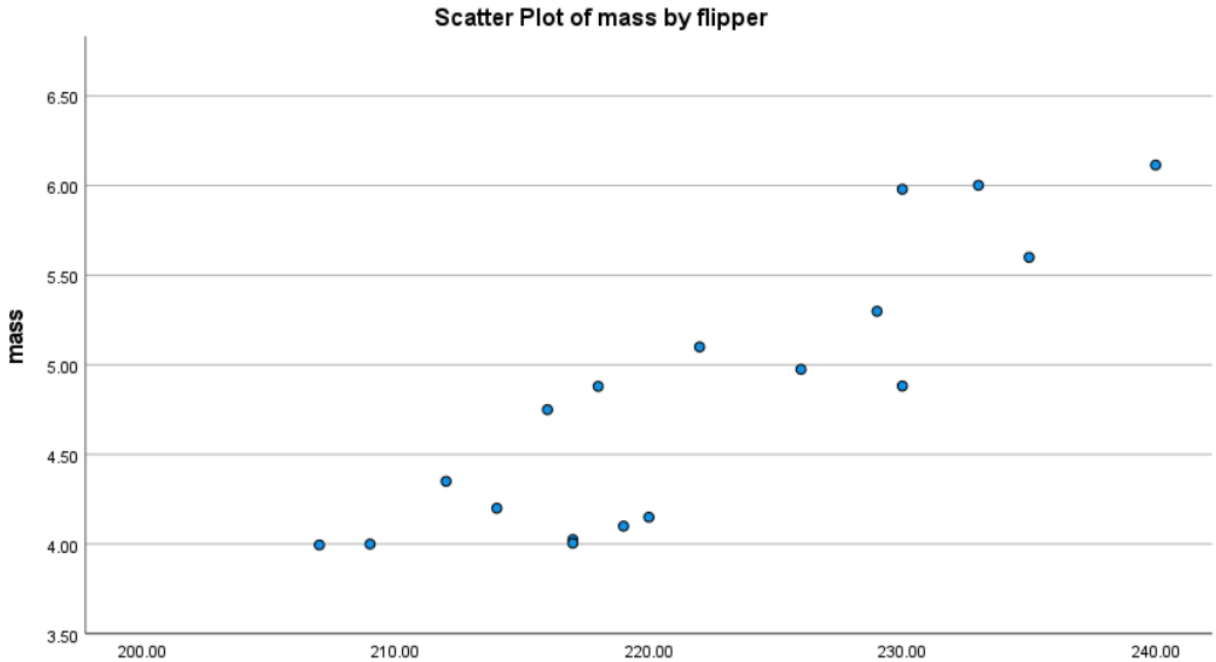
We can be 95% certain that the means of the 4 groups studied are not the same.

d. Carry out the Bonferroni's multiple comparisons (if appropriate) and summarise the results.

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## Q3_5.Test$ind  3  19.91   6.637   10.67 0.000137 ***
## Residuals     23  14.31   0.622
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 5 observations deleted due to missingness
```

Question 3.6

At a station in Antarctica, researchers measured flipper length and body mass for different penguin species. Below are a scatter plot and SPSS regression output for measurements made on 18 Gentoo penguins. Flipper length was measured in millimetres, and body mass in kilograms.



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.878[a]	.771	.757	.37042

a. Predictors: (Constant), flipper

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-10.882	2.138		-5.089	<.001
	flipper	.071	.010	.878	7.340	<.001

a. Dependent Variable: mass

a. Describe the apparent relationship between the Gentoo penguins' flipper length and body mass.

b. State the value of the correlation between the variables and explain how it supports the relationship you described in a.

c. Give the regression equation.

d. State the R^2 value and explain what it represents in this context.

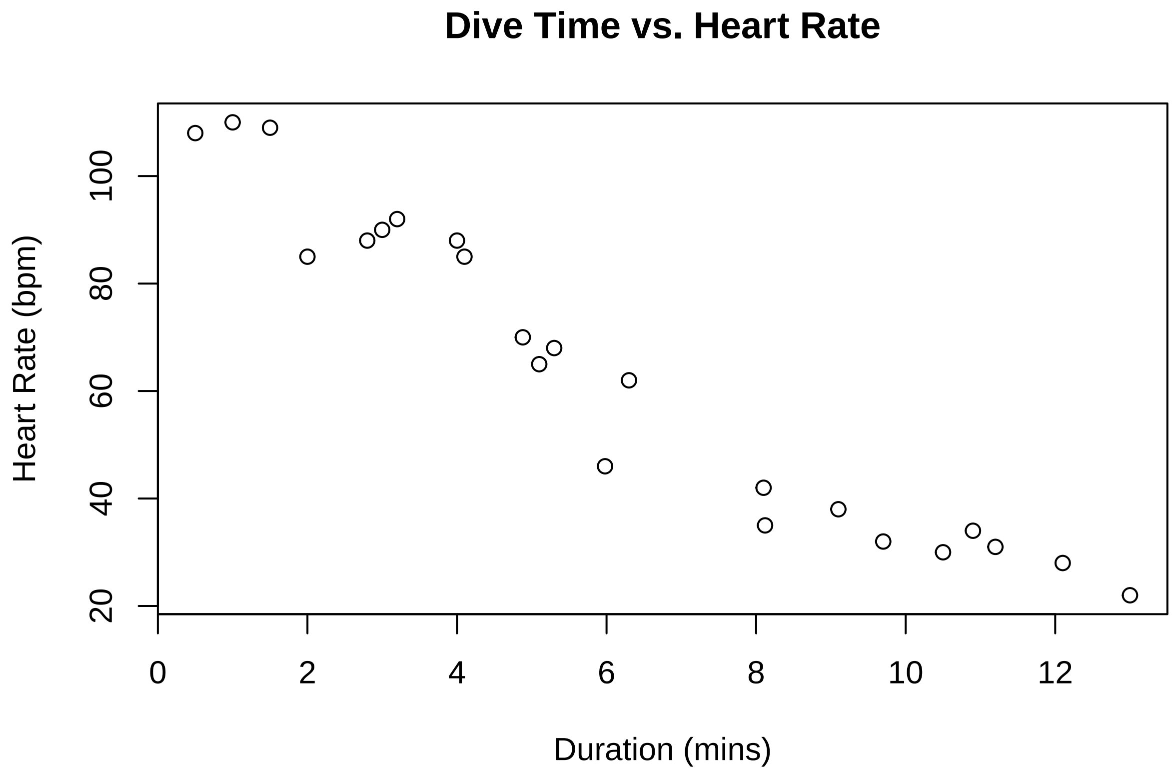
Question 3.7

To conserve oxygen, the heart rate of penguins slows when they dive. For 23 Emperor penguins, researchers have measured the dive heart rate (bpm) and the duration of the dive (minutes) for each penguin. The researchers want to further explore a possible relationship between dive duration and heart rate to predict heart rate from the length of a penguin's dive. Below are the data for each penguin:

duration	heart.rate
2.80	88
1.00	110
1.50	109
2.00	85
0.50	108
3.00	90
3.20	92
4.00	88
4.10	85
4.88	70
5.10	65
5.30	68
6.30	62
5.98	46
8.10	42
8.12	35
12.10	28
9.70	32
10.50	30
11.20	31
10.90	34
13.00	22
9.10	38

Use SPSS or other technology to produce the output necessary to answer the following questions.

a. Produce a scatterplot and use it to describe the apparent relationship between the dive duration and dive heart rate.



b. State the regression equation.

$$\hat{y} = a + bx$$

c. Is it meaningful to interpret the y-intercept for this study? Explain.

_The y-intercept for this study is > 0, therefore the findings of the study are more likely to be repl

d. Interpret the slope of the regression equation in the context of these variables.

_The slope of this regression suggests that the heart rate of a penguin decreases at a predictable ra

e. Predict the heart rate of a penguin that dives for 16 minutes. Is this prediction reliable? Explain.

$$\begin{aligned}\hat{y} &= a + bx \\ &= 109.7535 + (-7.49 \times 16) \\ &= -10.08\end{aligned}$$

```
##
## Call:
## lm(formula = Q3_7$heart.rate ~ Q3_7$duration)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.9673  -4.3538  -0.5707   6.0825  10.4805
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)   109.7535     3.2513   33.76  < 2e-16 ***
## Q3_7$duration  -7.4893     0.4503  -16.63 1.44e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.026 on 21 degrees of freedom
## Multiple R-squared:  0.9294, Adjusted R-squared:  0.9261
## F-statistic: 276.6 on 1 and 21 DF,  p-value: 1.445e-13
```

f. Is the relationship between the variables significant?

Perform the relevant hypothesis test (include all steps) and write your conclusion in context. Use $\alpha = 0.05$.