DSB113

Exercise, Gender and Mortality Indicators

Andrew Reece

Contents

Introduction

Previous research has shown that simple everyday tasks such as getting out of a chair and standing on one leg can be used to predict which middle aged people are at risk of an early death. It is also apparent that increasing activity just a little when you're younger could pay dividends when you're older. (www.telegraph.co.uk (April 2014) from Dr Rachel Cooper, BMJ April 2014).

This study builds upon this research in order to investigate whether performance in these two tasks varies depending upon the amount of exercise (high or low) undertaken by young people (18-25), in this case students at Loughborough University.

Method

Apparatus

The apparatus required was a chair that did not swivel and did not have armrests, a blindfold, and a stopwatch.

Procedure

The participant was first asked for all of the background details required, including age, gender, and number of hours spent exercising per week. 'Exercise' was operationalised as activities that noticeably increased their breathing rate.

A coin was then flipped for each participant to determine, with an equal probability for each, which condition would be done first. This was done as a counterbalancing measure to minimise the impact of any order effects on the results.

Condition 1: Single-leg balance

Participants were asked to choose their dominant leg, which was determined by asking them with which they would kick a ball, and if they did not have an answer to that, the leg was determined by which they preferred to jump off. They were then asked to put on the blindfold and close their eyes, then stand on just their dominant leg with their hands on their hips for as long as they could. They were told that the timer would be stopped when their non-dominant leg touched the floor, their hands left their hips or they started hopping around. The stopwatch was started as they took their non-dominant foot off the ground and stopped when one of the above conditions occurred.

Condition 2: Stand-up/Sit-down

The non-swivel chair without armrests was positioned with its back to a wall so that it would not be pushed backwards. The participants were shown the chair and told that they would have to stand up and sit down on the chair as many times as they could in one minute. They were also told that their hips and knees had to be straight and vertical to be counted as standing, they had to put their full weight on the chair when sitting, and they could not use their arms to push on their legs or the chair to help standing up. Starting with the participant standing up, the stopwatch was started, counting down from one minute, and the participant was simultaneously told to begin. A count was kept that incremented by 1 for each time the participant fully stood up.

Participants

The n=24 participants were quota sampled to ensure an even mix of genders (12 male, 12 female). The subsets of participants were convenience sampled from Loughborough students living near the researchers and the Loughborough University library. All participants were aged 18-23.

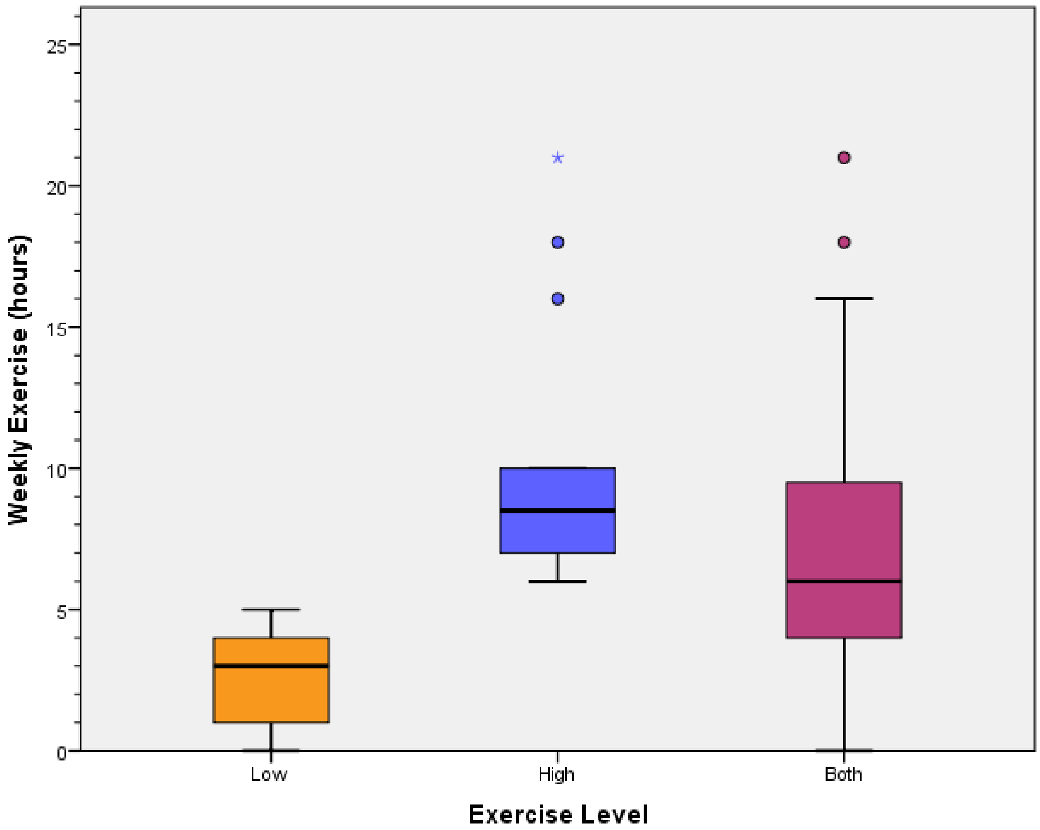
Results

For all statistical tests performed, the alpha value was set as $\alpha = 0.05$.
The NHS (2011), in describing a study with regard to mortality, classified 6 hours of moderate exercise as a “high level”. Based on this, the participants were divided into groups with those who reported doing 6 or more hours of exercise per week classified as a “high” level of exercise, and the rest were classified as “low”. This resulted in a sufficiently even split between groups with 10 in “high” and 14 in “low”.

Descriptive statistics

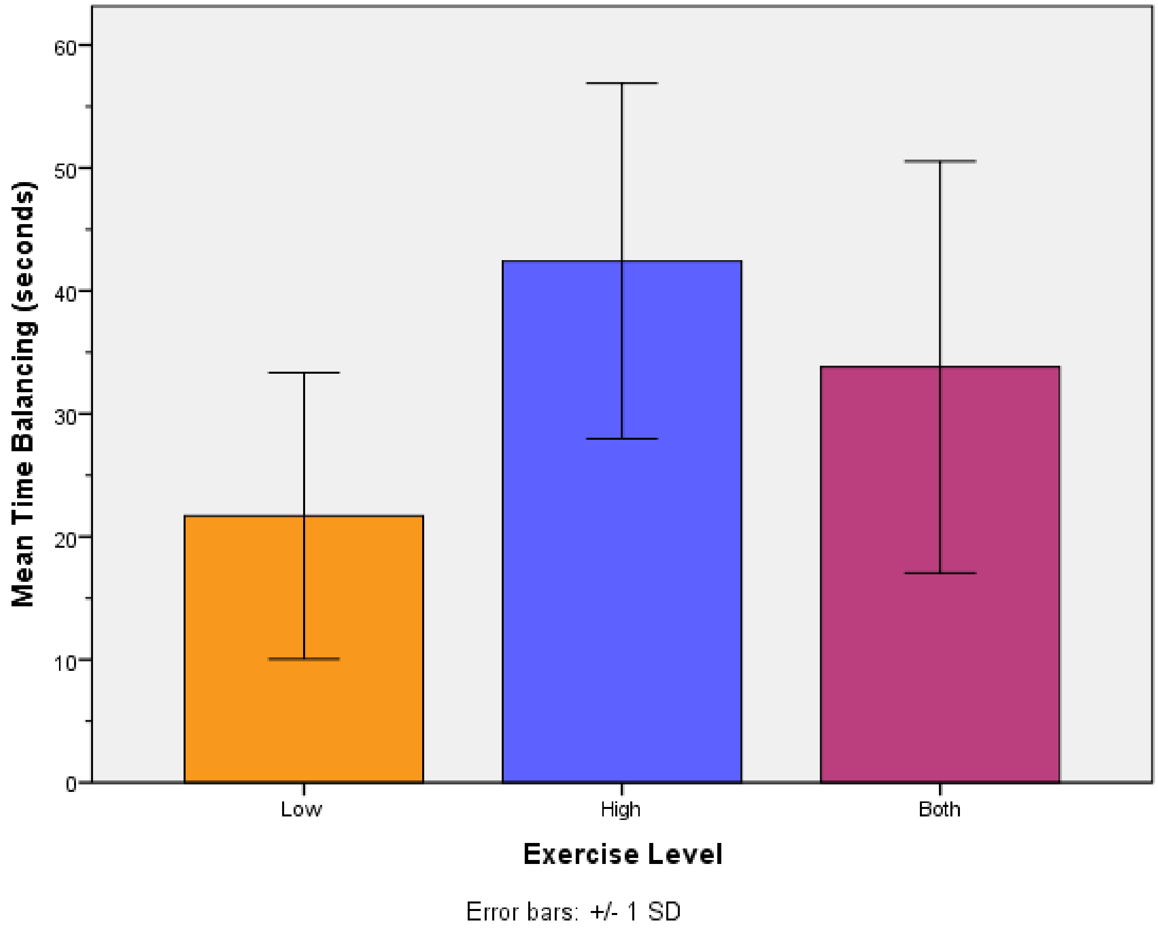
The mean age of the participants was 20.04 years old with a standard deviation of 1.16 years. This is a very narrow age band, and so the conclusions of this paper should not be uncritically generalized to other age ranges.

Figure 1: boxplot of the hours of exercise per week for the different exercise levels



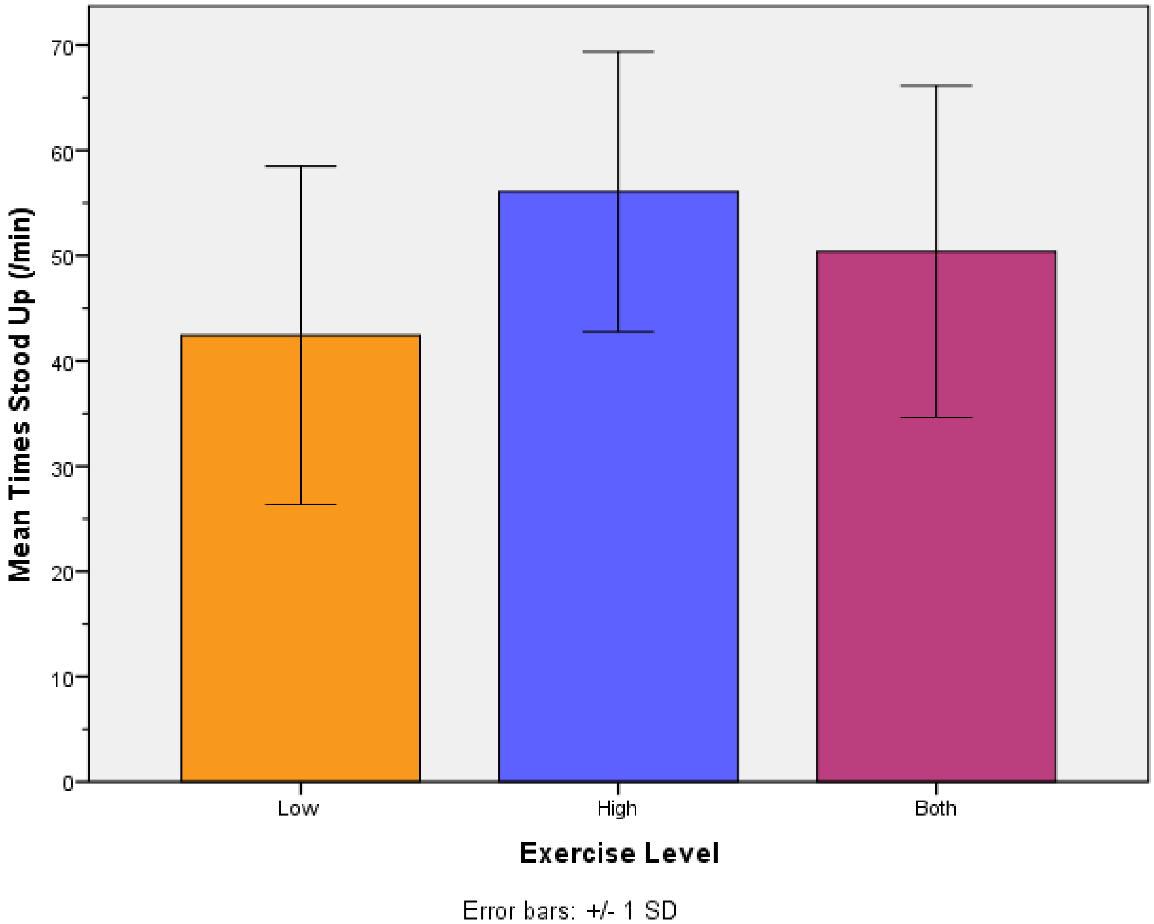
As shown in Figure 1 there are 2 suspected outliers and one true outlier for the weekly levels of exercise by those categorized in the high exercise level as well as 2 suspected outliers in the overall data. As a result, the median was used for the corresponding measure of central tendency. The median weekly exercise for the low exercise level group was 3 hours, compared to 8.5 hours in the high exercise level group. The median overall was 6 hours of exercise per week. The range in the low group is 0-5 hours, and 6-21 in the high group (6-10 if discounting possible outliers).

Figure 2: bar chart of mean time balancing for both exercise levels and overall



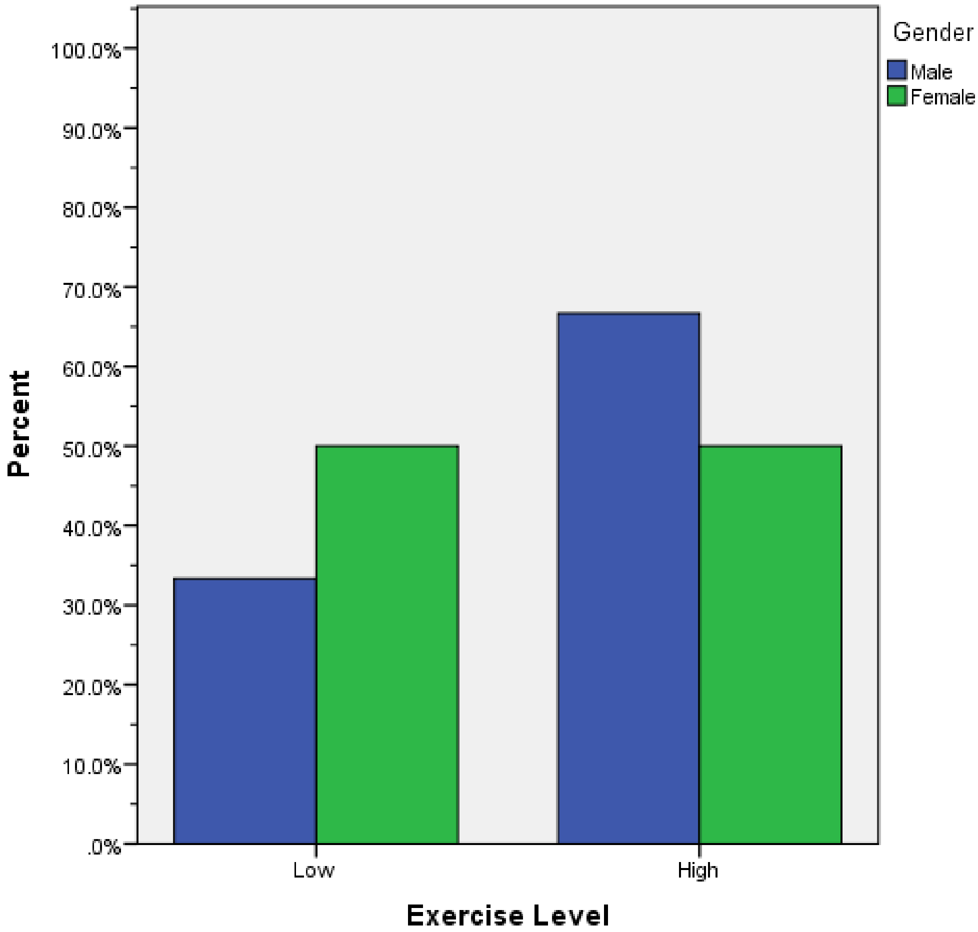
As depicted in Figure 2, for the time balancing on one leg, the mean [standard deviation] overall was 33.79 [16.75], for the low exercise level was 21.70 [11.662] and for the high exercise level was 42.43 [14.469]. Participants in the high exercise level balanced on one leg for a longer mean time than those in the low exercise level.

Figure 3: bar chart of mean times stood up for both exercise levels and overall



As shown in Figure 3, for the number of times stood up from a chair, the mean [standard deviation] overall was 50.38 [15.76], for the low exercise level was 42.40 [16.071] and for the high exercise level was 56.07 [13.304]. Participants in the high exercise level sat down on- and stood up from a chair more times on average than those in the low exercise level.

Figure 4: clustered bar chart showing exercise level by gender



As shown in Figure 4, females are evenly divided, with half exercising at each level. More males exercise at a high level (⅔) than at low level (⅓). Based on this, it would appear that there may be an association between exercise level and gender such that males are more likely to exercise 6 or more hours per week.

Inferential statistics

Association between gender and exercise group

The test for determining association depends on the variable types being tested. Gender is a nominal variable, and exercise group, although strictly ordinal, can be used as a nominal variable as well. The variables are also meaningfully categorised - each participant is either male or female and categorized as either high or low exercise level (no participant is in both groups for one variable). Because of these conditions, the Chi-Square test of association must be used to determine whether there is a significant association between gender and exercise level. For this, the null hypothesis is:

H_0: There is no association between gender and exercise level.

Leaving the two-tailed alternative hypothesis:

H_1: There is an association between gender and exercise level.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.686[a]	1	.408		
Continuity Correction[b]	.171	1	.679		
Likelihood Ratio	.689	1	.406		
Fisher's Exact Test				.680	.340
Linear-by-Linear Association	.657	1	.418		
N of Valid Cases	24				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.00.

b. Computed only for a 2x2 table

SPSS was used to determine the significance and effect size. No significant association was found between gender and exercise level (χ^2 (1, N = 24) = 0.686, p = 0.679)) with the continuity correction applied. $p > \alpha$. This means that the null hypothesis cannot be rejected, and no association can be claimed between gender and exercise level as defined in this paper.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.169	.408
	Cramer's V	.169	.408
N of Valid Cases		24	

As the data can form a 2x2 contingency table, the effect size can be determined using Cramer's V. This results in a value of 0.169, suggesting that even if there were a significant association between gender and exercise level, it would be minor.

The Chi-Square test assumes simple random sampling, so as the participants were quota sampled with regard to gender and convenience sampled from Loughborough University students otherwise, the validity of these results may be lower than preferred.

Balance difference between High and Low exercise groups

The participants in each exercise group have no crossover, and so an independent samples test of difference must be used to determine the significance of any difference in balancing time between them. The other factor in choice of test is whether the data for each exercise level is normally distributed. For normality testing, the null hypothesis is:

H_0: Data do not differ from normal.

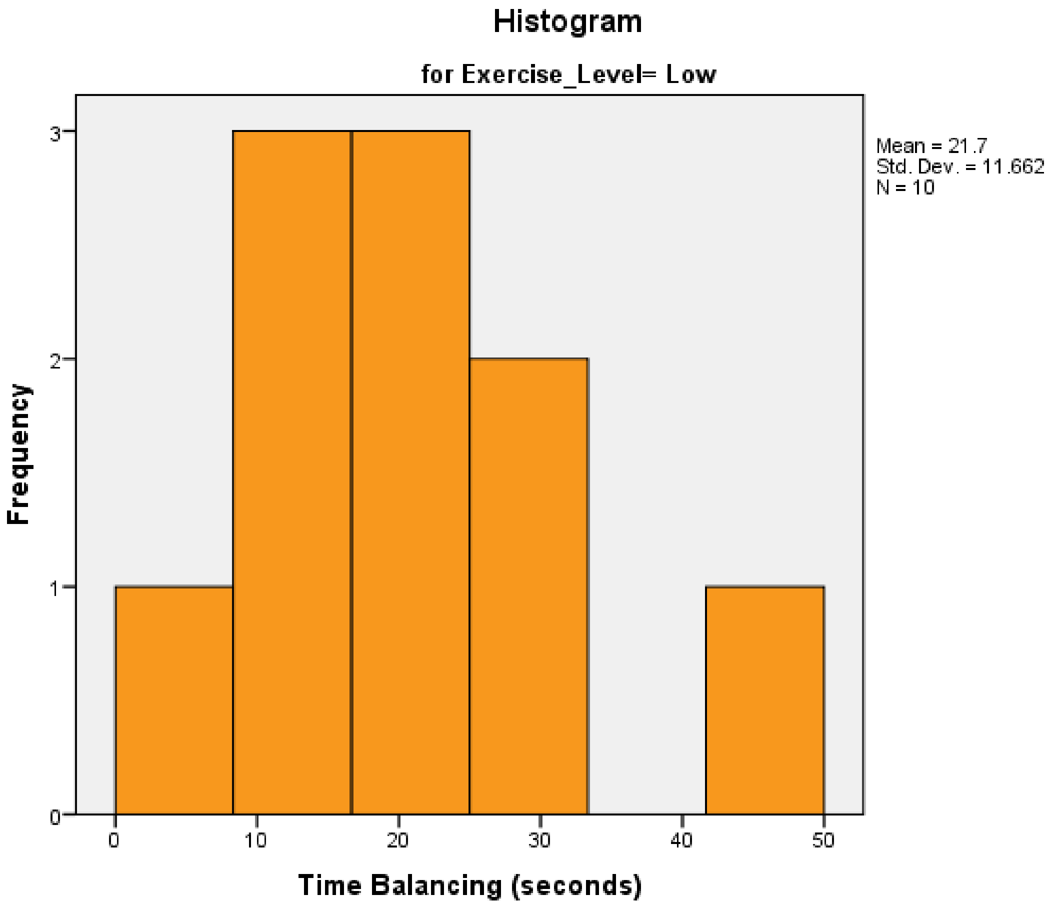
Leaving the two-tailed alternative hypothesis as:

H_1: Data do differ from normal.

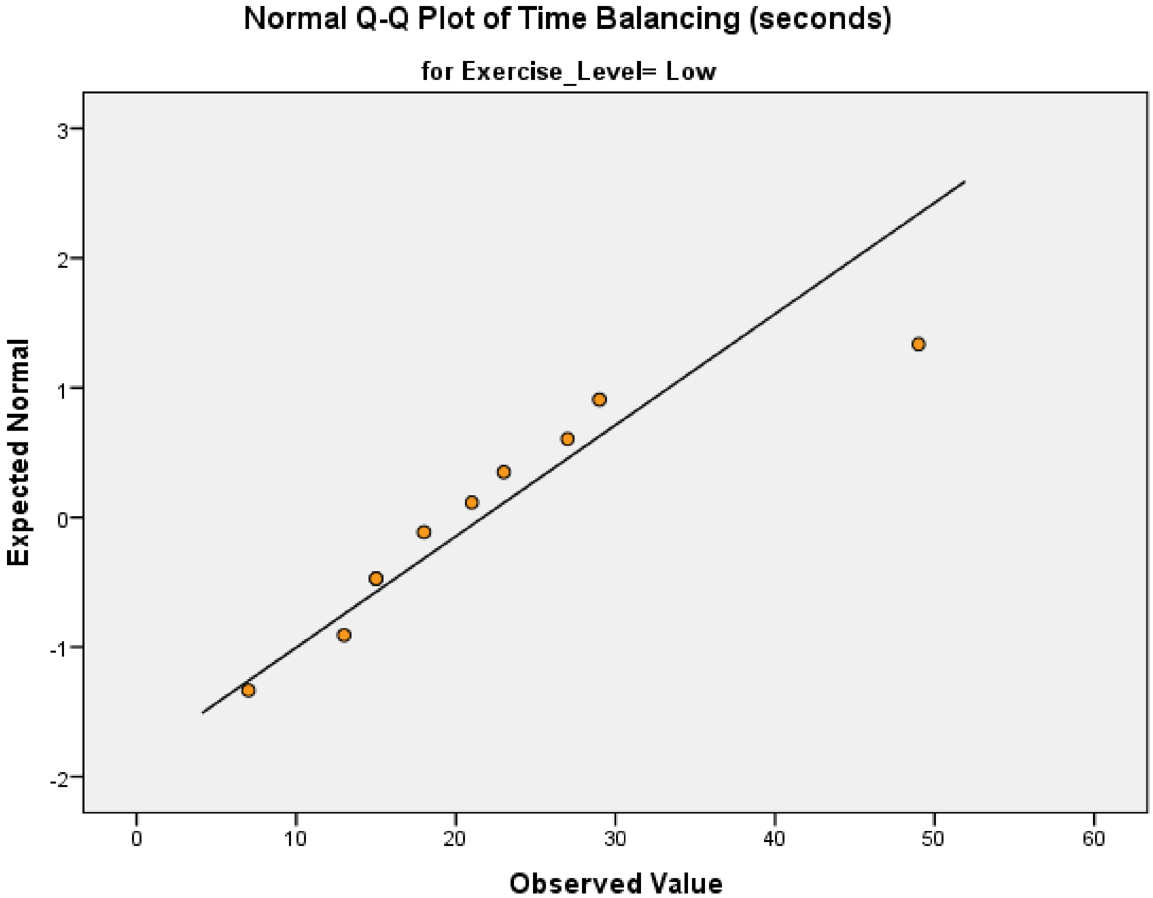
Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Exercise Level	Statistic	df	Sig.	Statistic	df	Sig.
Time Balancing (seconds)	Low	.166	10	.200*	.891	10	.176
	High	.175	14	.200*	.960	14	.717

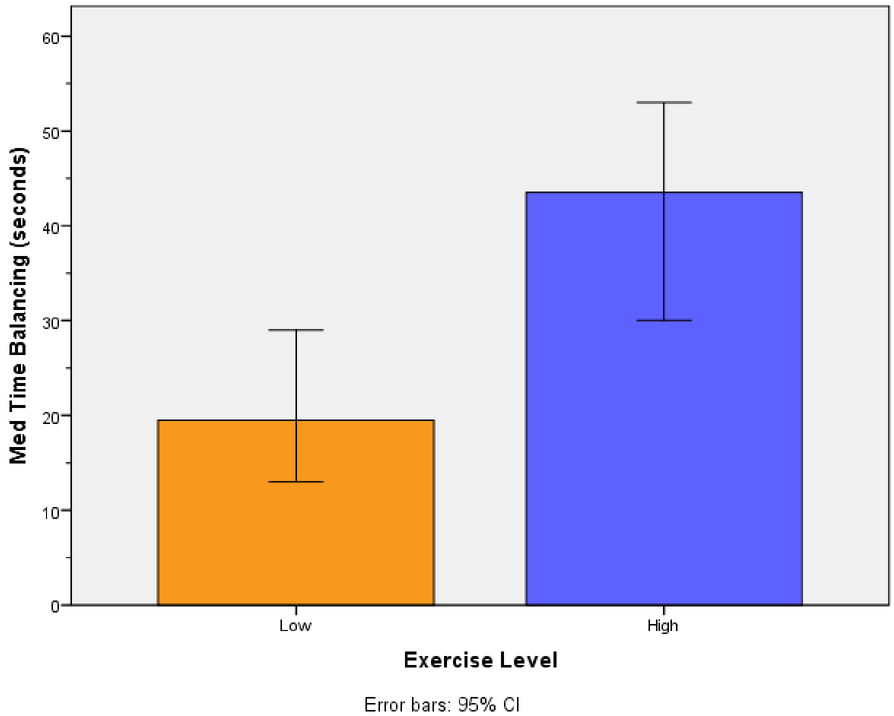
From the SPSS output above, for the high exercise level the test statistic = 0.960 ($p = 0.717$); $p > \alpha$ and so the null hypothesis should not be rejected. For the low exercise level the test statistic = 0.891 ($p = 0.176$); $p > \alpha$ (although not by a large amount) and so the null hypothesis should not be rejected based on this test. Both exercise levels appear at this point to be normally distributed. This was to be checked by analysing the histogram and Q-Q plot for each condition.



The above histogram for time balanced for the low exercise group does not deviate too far from normality, however the longest balance points do seem to skew the data slightly.



The point from the longest balance time appears to deviate considerably from the expected line. Considering this and the low p score for normality, this data set was considered not normally distributed. This means that the non-parametric Mann-Whitney U test would have been used regardless of the normality of the other dataset.



For this test, the hypotheses are:

H_0: For balancing on one leg, the median time balanced by participants who exercise for 6 or more hours per week will not differ from the median time balanced by participants who exercise for less than 6 hours per week.

H_1: For balancing on one leg, the median time balanced by participants who exercise for 6 or more hours per week will be longer than the median time balanced by participants who do less than 6 hours per week of exercise.

Test Statistics[a]

	Time Balancing (seconds)
Mann-Whitney U	21.500
Wilcoxon W	76.500
Z	-2.841
Asymp. Sig. (2-tailed)	.004
Exact Sig. [2*(1-tailed Sig.)]	.003[b]

a. Grouping Variable: Exercise Level

b. Not corrected for ties.

Based on the above output from SPSS, the (one-tailed) results are: U = 21.5 ($z = -2.841$, $p = 0.0015$). $p < \alpha$, and so the null hypothesis can be rejected in favour of the alternative hypothesis. This is very highly significant, meaning that there is a very high probability that exercise level affects one-legged balance time.

The effect size for the Mann-Whitney U test can be determined using the formula:

$$r = \frac{z}{\sqrt{n}}$$

With n as the total number of observations. Putting the calculated values into this formula:

$$r = \frac{-2.841}{\sqrt{24}}$$

$$r = -0.5799$$

This suggests a moderately large effect size with regard to the effect of exercise level on time balanced on one leg with a blindfold.

Difference in number of times standing and sitting in one minute between High and Low exercise groups

As before, the participants in each exercise group have no crossover, and so an independent samples test of difference must be used to determine the significance of any difference in times stood-up between them. A normal distribution must again be checked for, in order to determine if a parametric test can be used. The histogram, boxplot and Q-Q plot were again used in conjunction with the Shapiro-Wilks normality test. For this, the hypotheses are:

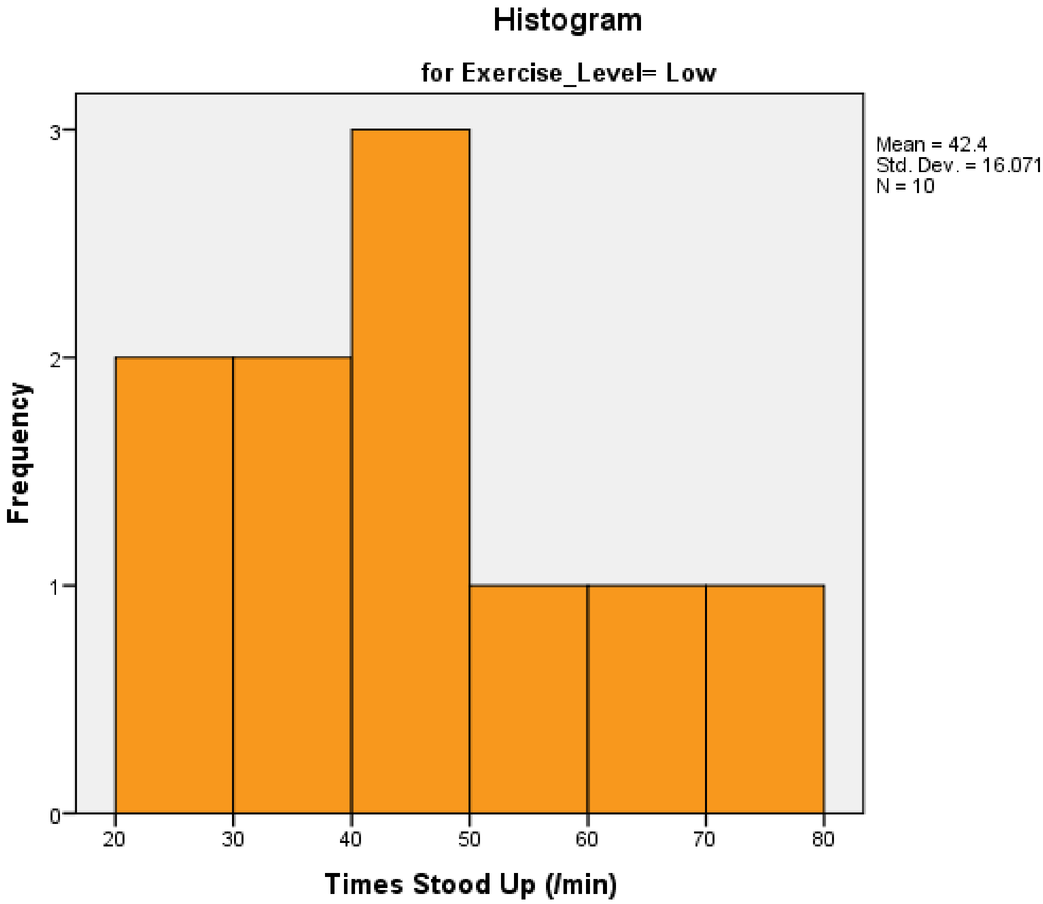
H_0: Data do not differ from normal.

H_1: Data do differ from normal.

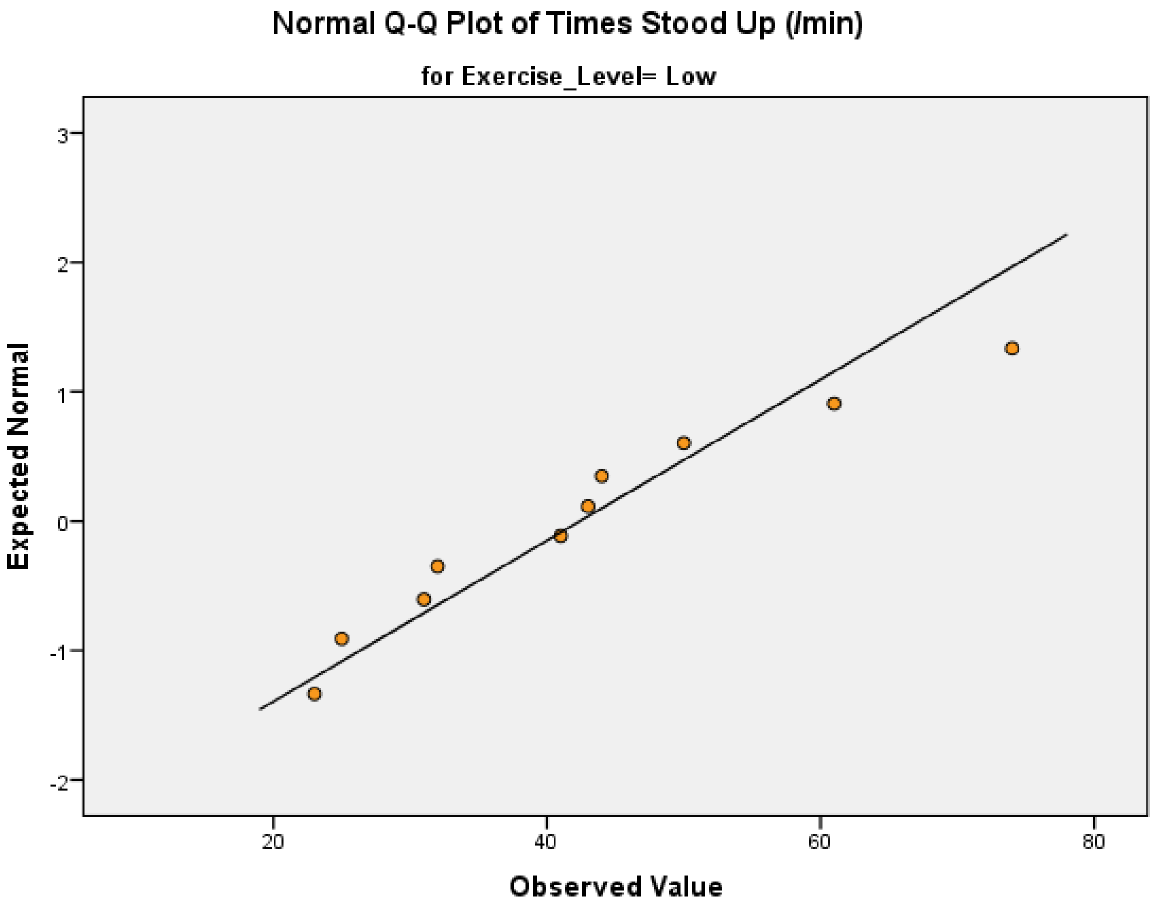
Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Exercise Level	Statistic	df	Sig.	Statistic	df	Sig.
Times Stood Up (/min)	Low	.160	10	.200*	.939	10	.546
	High	.129	14	.200*	.956	14	.660

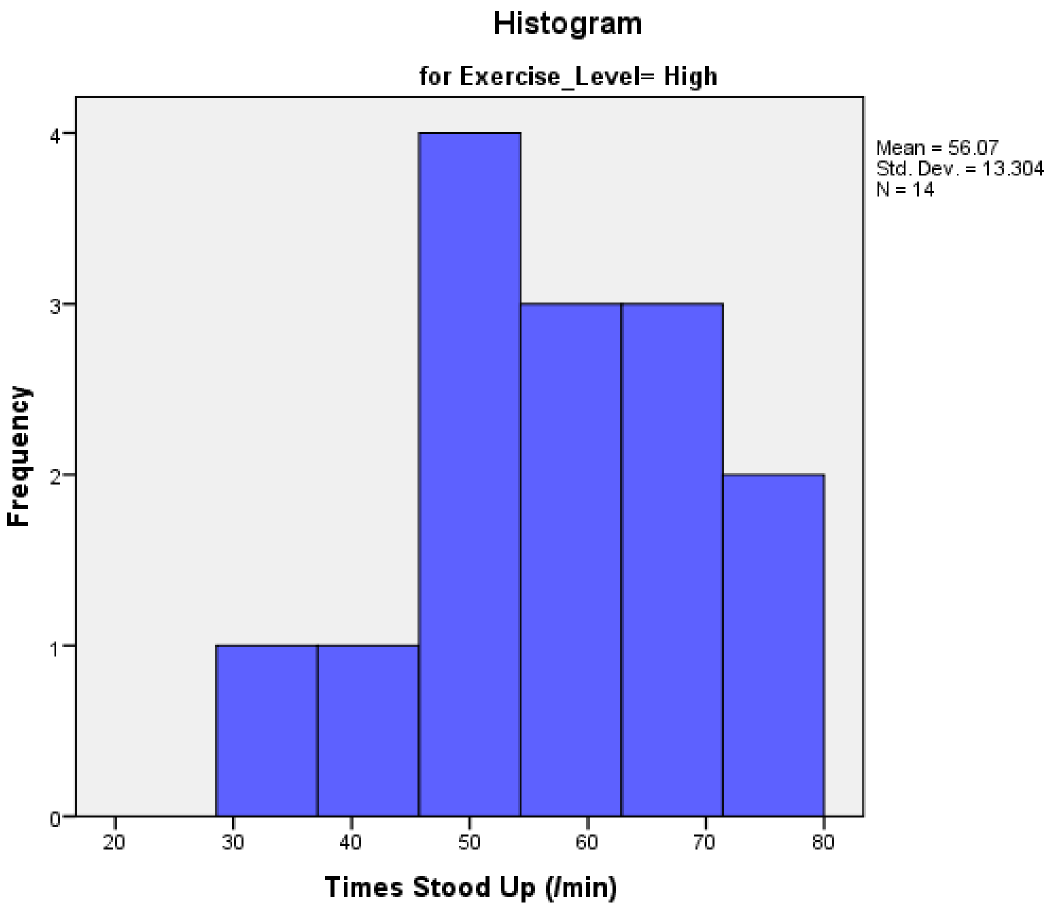
From the SPSS output above, for the low exercise level the test statistic = 0.939 ($p = 0.546$); $p > \alpha$ and so the null hypothesis should not be rejected. For the high exercise level the test statistic = 0.956 ($p = 0.660$); $p > \alpha$ and so the null hypothesis should not be rejected. Both exercise levels appear to be normally distributed. This was checked by analysing the histogram and Q-Q plot for each condition.



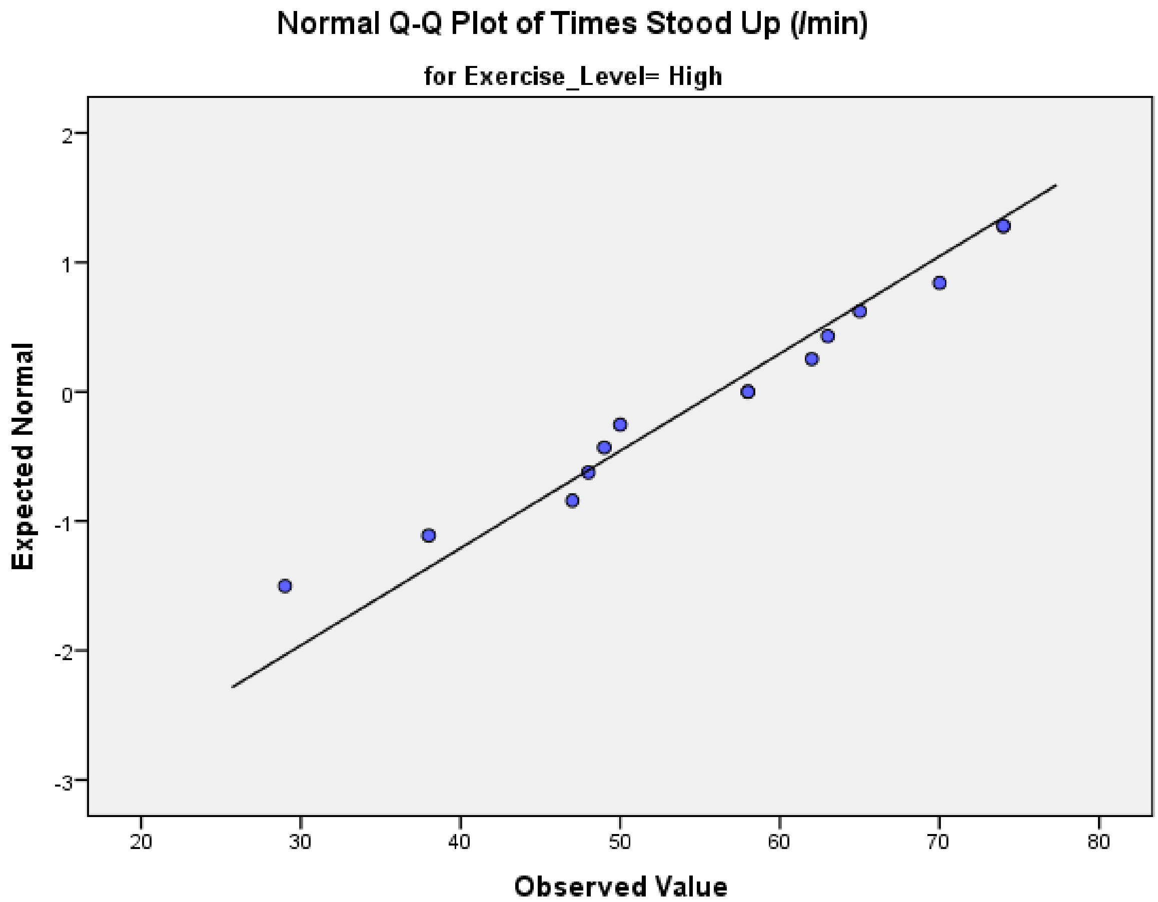
The above histogram for stand/sit times for the low exercise level shows a slight positive skew but looks approximately normally distributed.



The points on the above Q-Q plot for stand/sit times for the low exercise level do not appear to deviate much from the line of best fit.

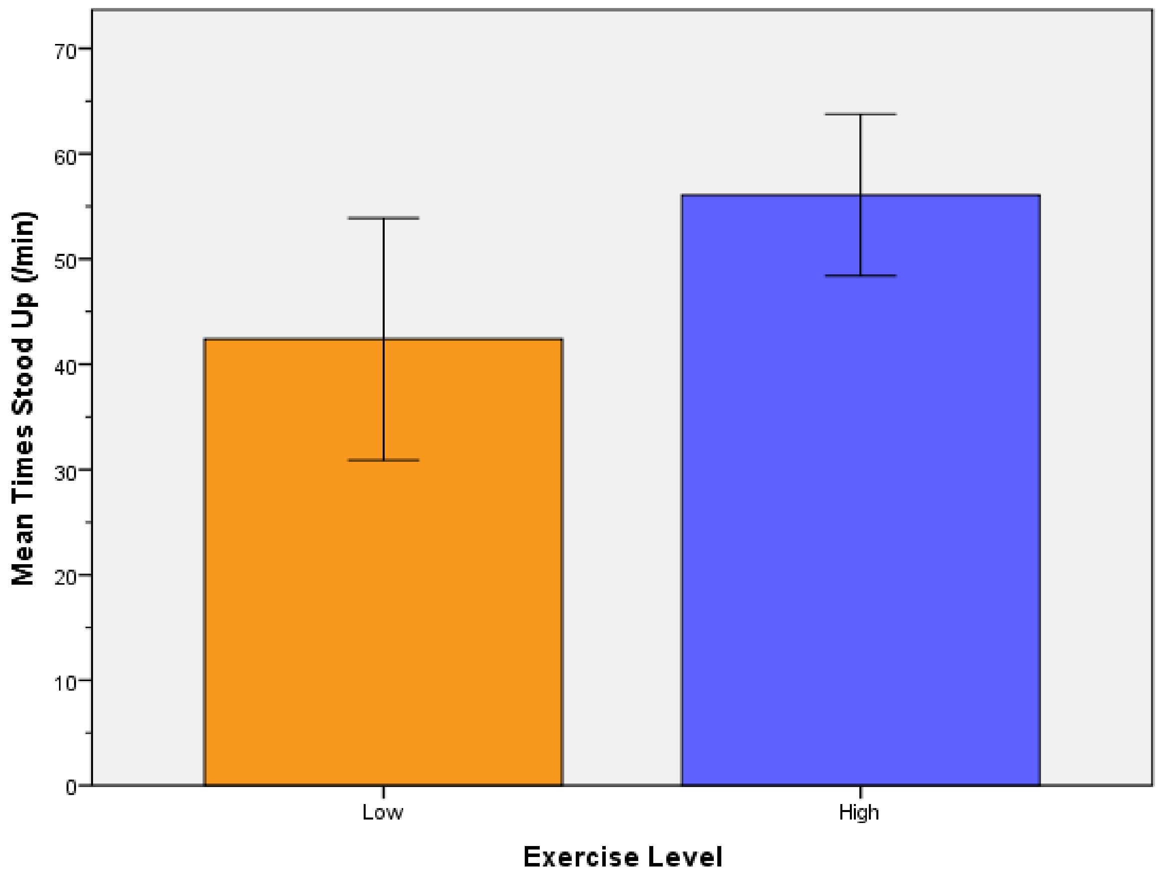


The above histogram for stand/sit times for the high exercise level shows a slight negative skew but looks approximately normally distributed.



The points on the above Q-Q plot for stand/sit times for the high exercise level do not appear to deviate much from the line of best fit.

As the Shapiro-Wilk tests suggest normality and the graphs do not seem to strongly disagree, the dataset was be treated as normally distributed. As such the independent samples t-test was used to determine the significance of difference between number of times stood up in a minute at the different exercise levels.



Error bars: 95% CI

For this test, the hypotheses are:

H_0: The mean number of times participants who exercise for 6 or more hours per week can sit and stand in one minute will not differ from the mean number of times participants who exercise for less than 6 hours per week can sit and stand in one minute.

H_1 (one-tailed): The mean number of times participants who exercise for 6 or more hours per week can sit and stand in one minute will be greater than the mean number of times participants who exercise for less than 6 hours per week can sit and stand in one minute.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the ...	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Times Stood Up (/min)	Equal variances assumed	.123	.729	2.277	22	.033	13.671	6.003	1.221	26.12
	Equal variances not assumed			2.204	17.13	.041	13.671	6.202	.593	26.75

Levene's (one-tailed) test for equality of variances (H_0: variances do not differ; H_1: variances do differ) was F = 0.123 (p = 0.729), which, as p > α, is not significant. As a result, the null hypothesis cannot be rejected and equal variances can be assumed.

SPSS calculates the result of the one-tailed t-test as t = 2.277 (p = 0.017). With α = 0.05, p < α, and so the result is significant. The 95% confidence interval for the population mean is between 1.221 and 26.12. The null hypothesis can be rejected in favour of the alternative hypothesis. On average, people who exercise at a greater level appear to be able to stand up and sit down more times in a minute.

The effect size for normally distributed datasets can be calculated using Cohen's d, which is found using the formula:

$$d = \frac{\mu_1 - \mu_2}{(\sigma_1 + \sigma_2) \div 2}$$

Thus:

$$d = \frac{\mu_H - \mu_L}{(\sigma_H + \sigma_L) \div 2}$$

$$d = \frac{56.07 - 42.4}{(13.304 + 16.071) \div 2}$$

$$d = \frac{13.67}{14.6875}$$

$$d = 0.9307234$$

This is a large effect size, meaning that level of exercise has a large effect on the number of times one can stand up and sit down on a chair.

A post hoc power analysis was performed using G*Power, with the output:

t tests – Means: Difference between two independent means (two groups)			
Analysis:	Post hoc: Compute achieved power		
Input:	Tail(s)	=	One
	Effect size d	=	0.9307234
	α err prob	=	0.05
	Sample size group 1	=	14
	Sample size group 2	=	10
Output:	Noncentrality parameter δ	=	2.2479106
	Critical t	=	1.7171444
	Df	=	22
	Power (1-β err prob)	=	0.7030131

The result for this analysis is power = 0.703 leaving:

$$\beta = 1 - power = 1 - 0.703 = 0.297$$

The power is smaller than Cohen's (1992) recommendation of having power close to 0.80, although it is not very far off.

Association between balance time and sit/stand frequency

Both balance time and count of stands in a minute are scalar variables, so in order to check for association, a correlation test will be used. In order to determine which, the first step was again to determine the normality of each dataset. A Shapiro-Wilks test performed with the following hypotheses:

H_0: Data do not differ from normal.

H_1: Data do differ from normal.

Tests of Normality

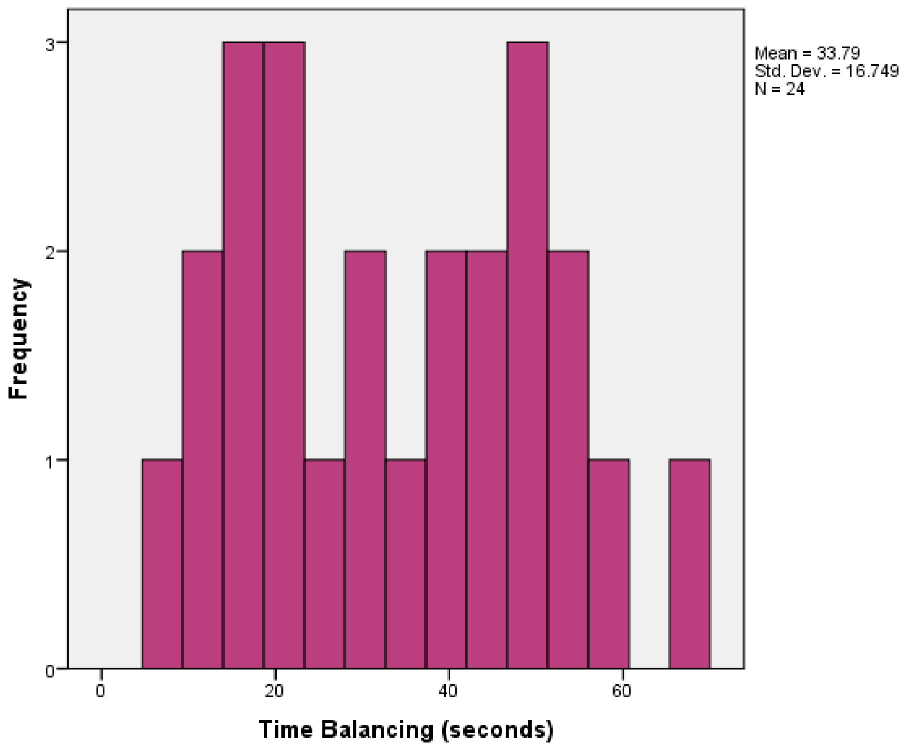
	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Time Balancing (seconds)	.125	24	.200*	.957	24	.373
Times Stood Up (/min)	.102	24	.200*	.955	24	.346

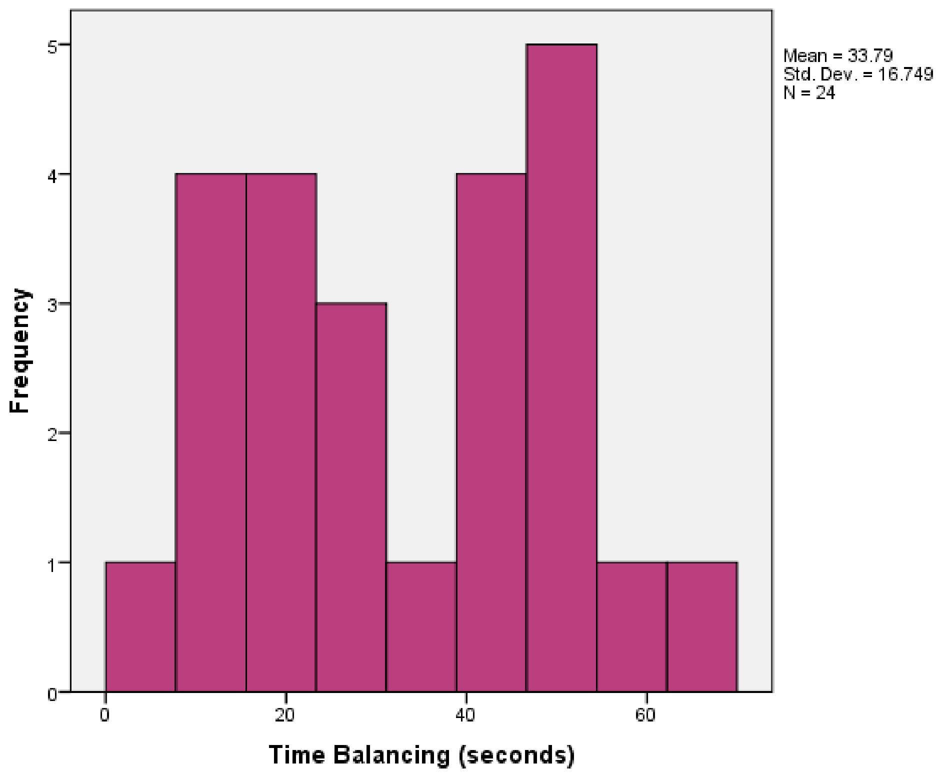
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

For the balancing variable. the result was test statistic = 0.957 ($p = 0.373$). $p > \alpha$, and so the null hypothesis should not be rejected. For the sit/stand variable, SPSS calculated the result as test statistic = 0.955 ($p = 0.346$). $p > \alpha$, and so the null hypothesis should not be rejected. The data for both variables do not appear to differ significantly from normal. The next step would be to look at the histogram and Q-Q plot for each variable.

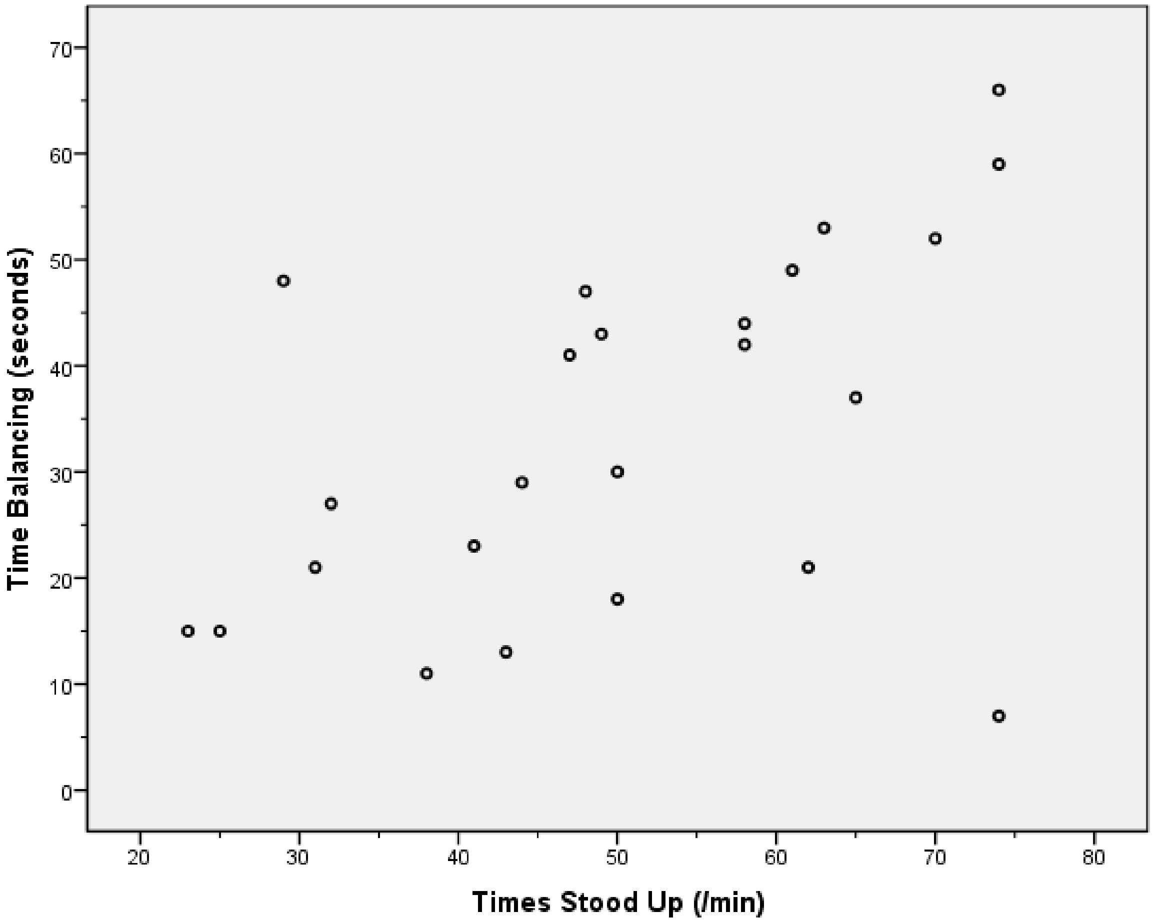
Figure 5: Histograms of time spent balancing, with interval size at 15 above and 9 below





The first graph observed was the histogram for balancing time (Figure 5), which appears to show a bimodal dataset (at multiple interval sizes). To be conservative, this was considered to exclude the dataset from being normal. As this one variable was considered not normally distributed from its histogram, there was no need to further test the normality of the sit/stand variable and a non-parametric test had to be used. Spearman's ρ (rho) was the most likely test, but requires the relationship between the 2 variables appears to be monotonic. A simple scatter plot should demonstrate monotonicity:

Figure 6: Scatter plot of sit/stand frequency and balance time.



As Figure 6 shows, although there are data points that do not appear to fit the general trend (top-left and bottom-right), there does appear to be a positive monotonic trend, possibly even linear. As a result, Spearman's ρ can be used. For this, the hypotheses are:

H_0: There is no correlation between times stood up and time balanced

H_1: There is a correlation between times stood up and time balanced

Correlations

			Time Balancing (seconds)	Times Stood Up (/min)
Spearman's rho	Time Balancing (seconds)	Correlation Coefficient	1.000	.493*
		Sig. (2-tailed)	.	.014
		N	24	24
	Times Stood Up (/min)	Correlation Coefficient	.493*	1.000
		Sig. (2-tailed)	.014	.
		N	24	24

*. Correlation is significant at the 0.05 level (2-tailed).

The result from SPSS is that $\rho = 0.493$ ($p = 0.014$). $p < \alpha$, and so the null hypothesis must be rejected in favour of the alternative hypothesis. There is a moderate positive correlation between time balanced on one leg and number of times sitting/standing from a chair.

ρ^2, which shows the common variance between the 2 variables is equal to 0.243. This means that approximately a quarter of the variance in number of times stood up can be accounted for by the variance in time balancing (and vice versa) and that approximately three quarters of their variances are unaccounted for.

Conclusions

For students at Loughborough University:

- Gender appears to have no significant association with the level of exercise performed, with a small effect if there is any association.
- Exercise level seems to have a highly significant but moderate effect on the time for which they can balance on one leg.
- Exercise level seems to have a highly significant and strong effect on the number of times they can sit and stand from a chair in one minute.
- There is a moderately strong positive correlation between the length of time for which they can balance on one leg and the number of times they can sit and stand from a chair in one minute.

To conclude, from the results gathered here, it does appear that performance in both the balancing on one leg and the sitting down and standing up as many times as possible in a minute are indeed associated with the level of exercise per week participants reported. No causal statements can be made, however, as independent variables were not manipulated by an experimenter, and many extraneous variables were not controlled for.

Evaluation

There were some limitations with the study that could be improved in future.

The exercise level was a simple cutoff based on number of hours of exercise per week reported. Calculations may have been more informative if they used the full spread of data, so that the shape of the relationship between the variables could have been better observed.

Exercise intensity was not properly accounted for. People who do light jogging may be affecting their bodies differently from those doing hill sprints, for example. Including the MET intensities as described by Ainsworth et al. (2000) into the calculations of exercise level would improve the precision of results.

The type of exercise may have affected the results in another way - there were a number of gymnasts in the high level group, but not in the low level group, and one might expect them to have better balance than many other people, likewise with weightlifters who squat frequently and sitting/standing as fast as possible. Intensity of exercise, then, may not be only major variable affecting results. Limiting the scope of a study to comparing only practitioners of only one activity should lead to more internally valid results, particularly if attempting to find causal links.

With the relatively small number of participants, Shapiro-Wilks tends to ascribe normality liberally. This could be improved by including more participants. The Chi-Square test assumes randomly sampled data, so the validity of its results would have been improved by sampling the participants more randomly.

References

Ainsworth, B.E., Haskell, W.L., Whitt, M.C., Irwin, M.L., Swartz, A.M., Strath, S.J., O Brien, W.L., Bassett, D.R., Schmitz, K.H., Emplaincourt, P.O. and Jacobs, D.R., 2000. Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and science in sports and exercise*, *32*(9; SUPP/1), pp.S498-S504.

Cohen, J., 1992. A power primer. *Psychological bulletin*, *112*(1), 155.

NHS, 2011. *Exercise more and 'live longer' - Health News - NHS Choices*. [online] Available at: http://www.nhs.uk/news/2011/08August/Pages/physical-activity-and-lifespan.aspx [Accessed 16 Dec. 2015].

Appendix A: SPSS Screenshots

Variable view:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	P#	Numeric	8	0	Participant Number	None	None	2	Right	Nominal	None
2	Exercise_Hours	Numeric	8	0	Weekly Exercise (hours)	None	None	11	Right	Scale	Input
3	Exercise_Level	Numeric	8	0	Exercise Level	{0, Low}...	None	10	Right	Ordinal	Input
4	Gender	Numeric	8	0	Gender	{1, Male}...	None	8	Right	Nominal	Input
5	Age	Numeric	8	0	Age	None	None	8	Right	Scale	Input
6	Balance_Time	Numeric	8	0	Time Balancing (seconds)	None	None	8	Right	Scale	Input
7	Sit_Stand	Numeric	8	0	Times Stood Up (/min)	None	None	8	Right	Scale	Input

Data View:

	P#	Exercise_Hours	Exercise_Level	Gender	Age	Balance_Time	Sit_Stand
1	1	7	High	Male	20	52	70
2	2	10	High	Male	19	30	50
3	3	5	Low	Male	19	15	25
4	4	21	High	Male	23	44	58
5	5	4	Low	Female	19	49	61
6	6	0	Low	Female	20	23	41
7	7	6	High	Female	20	41	47
8	8	2	Low	Female	19	29	44
9	9	8	High	Female	22	59	74
10	10	4	Low	Male	20	18	50
11	11	8	High	Male	21	66	74
12	12	10	High	Female	18	43	49
13	13	6	High	Male	20	21	62
14	14	16	High	Female	20	47	48
15	15	5	Low	Female	20	7	74
16	16	1	Low	Male	21	13	43
17	17	2	Low	Male	19	21	31
18	18	6	High	Male	21	53	63
19	19	18	High	Male	19	37	65
20	20	10	High	Male	20	42	58
21	21	8	High	Female	20	48	29
22	22	9	High	Female	22	11	38
23	23	0	Low	Female	19	15	23
24	24	4	Low	Female	20	27	32

The data for Exercise_Hours, Balance_Time & Sit_Stand was copied again below for 'Both' exercise levels for some of the graphs, but all of the calculations were done with the data as shown here.