How to Report Statistical Results

ONE WAY ANOVA

There was a statistically significant difference between groups as determined by one-way ANOVA ($F(2,27) = 4.467$, $p = .021$).
A Modified Stutent's test post-hoc test revealed that the time to complete the problem was statistically significantly lower after taking the intermediate...
There were no statistically significant differences between the intermediate and advanced groups ($p = .989$).

The ability to cope with workplace-related stress (CWWS score) increased from the sedentary ($n = 7$, $M = 4.2$, $SD = 0.8$), to low ($n = 9$, $M = 5.9$, $SD = 1.7$), to moderate ($n = 8$, $M = 7.1$, $SD = 1.6$) to high ($n = 7$, $M = 7.5$, $SD = 1.2$) physical activity groups, in that order.

The ability to cope with workplace-related stress (CWWS score) was statistically significantly different for different levels of physical activity group, $F(3, 27) = 8.316$, $p < .0005$.

There were no statistically significant differences in CWWS score between the different physical activity groups, $F(3, 27) = 1.116$, $p = .523$.

A one-way ANOVA was conducted to determine if the ability to cope with workplace-related stress (CWWS score) was different for groups with different physical activity levels. Participants were classified into four groups: sedentary ($n = 7$), low ($n = 9$), moderate ($n = 8$) and high levels of physical activity ($n = 7$). There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .120$). CWWS score increased from the sedentary ($M = 4.2$, $SD = 0.8$), to low ($M = 5.9$, $SD = 1.7$), to moderate ($M = 7.1$, $SD = 1.6$) to high ($M = 7.5$, $SD = 1.2$) physical activity groups, in that order, but the differences between these physical activity groups was not statistically significant, $F(3, 27) = 1.116$, $p = .523$.

The group means were not statistically significant different ($p > .05$) and, therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis.

REPEATED MEASURES ANOVA

There was a statistically significant effect of time on exercise-induced fitness, $F(2, 10) = 12.53$, $p = .002$.

The six-month exercise-training programme had a statistically significant effect on fitness levels, $F(2, 10) = 12.53$, $p = .002$.

Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, $\chi2(2) = 22.115$, $p < .0005$, and therefore, a Greenhouse-Geisser correction was used. There was a significant effect of time on cholesterol concentration, $F(1.171, 38) = 21.032$, $p < .0005$.

SPHERICITY

Mauchly's Test of Sphericity indicated that the assumption of sphericity had not been violated, $\chi2(2) = 3.343$, $p = .188$.

This study found that overweight, physically inactive male participants had statistically significantly lower cholesterol concentrations (5.80 ± 0.38 mmol/L) at the end of an exercise-training programme compared to after a calorie-controlled diet (6.15 ± 0.52 mmol/L), t(38) = 2.428, p = 0.020.

TWO-WAY ANOVA

A two-way ANOVA was conducted that examined the effect of gender and education level on interest in politics. There was a statistically significant interaction between the effects of gender and education level on interest in politics, F (2, 54) = 4.643, p = .014.

Simple main effects analysis showed that males were significantly more interested in politics than females when educated to university level (p = .002), but there were no differences between gender when educated to school (p = .465) or college level (p = .793).

THREE-WAY ANOVA

There was a statistically significant three-way interaction between gender, risk and drug, F(2, 60) = 7.406, p = .001.

INDEPENDET TWO SAMPLE T TEST

There was a statistically significant difference in mean engagement score between males and females, t(38) = 2.365, p = .023.

There was a statistically significant difference in mean engagement score between males and females, with males scoring higher than females, 0.26 ± 0.11 [mean ± standard error], t(38) = 2.365, p = .023.

There was a statistically significant difference between means (p < .05), and therefore, we can reject the null hypothesis and accept the alternative hypothesis.

An independent-samples t-test was run to determine if there were differences in engagement to an advertisement between males and females. The advertisement was more engaging to male viewers (M = 5.56, SD = 0.35) than female viewers (M = 5.30, SD = 0.35), a statistically significant difference, M = 0.26, 95% CI [0.04, 0.48], t(38) = 2.365, p = .023.

There were 20 male and 20 female participants. An independent-samples t-test was run to determine if there were differences in engagement to an advertisement between males and females. There were no outliers in the data, as assessed by inspection of a boxplot. Engagement scores for each level of gender were normally distributed, as assessed by Shapiro-Wilk's test (p > .05), and there was homogeneity of variances, as assessed by Levene's test for equality of variances (p = .174). The advertisement was more engaging to male viewers (M = 5.56, SD = 0.35) than female viewers (M = 5.30, SD = 0.35), a statistically significant difference, M = 0.26, 95% CI [0.04, 0.48], t(38) = 2.365, p = .023.

SAMPLE

Data are mean ± standard deviation, unless otherwise stated. There were 20 male and 20 female participants. Mean male engagement score (5.56 ± 0.29) was higher than mean female engagement score (5.30 ± 0.39).

PLOTS

There were no outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box.

Engagement scores were approximately normally distributed for both males and females, as assessed by visual inspection of their histograms.

Engagement scores were normally distributed for both males and females, as assessed by visual inspection of Normal Q-Q Plots.

Engagement scores were normally distributed for males with a school-level education with a skewness of 0.552 (standard error = 0.687) and kurtosis of 1.986 (standard error = 1.334).

The assumption of normality for 'interest in politics' scores was satisfied for all group combinations of gender and education level, as assessed by visual inspection of Normal Q-Q Plots.

ASSUMPTIONS

There was homogeneity of variances for engagement scores for males and females, as assessed by Levene's test for equality of variances ($p = .174$).

The assumption of homogeneity of variances was violated, as assessed by Levene's test for equality of variances ($p = .024$).

Engagement scores were normally distributed for males with a skewness of -0.173 (standard error = 0.512) and kurtosis of -0.311 (standard error = 0.992) and for females with a skewness of -0.318 (standard error = 0.512) and kurtosis of -0.859 (standard error = 0.992).

The assumption of normality for 'interest in politics' scores was satisfied for all group combinations of gender and education level, as assessed by Shapiro-Wilk's test ($p > .05$).

CONFIDENCE INTERVALS

Male mean engagement score was 0.26 (95% CI, 0.04 to 0.48) higher than female mean engagement score.

Male mean engagement score was 0.26 ± 0.11 [mean ± standard error] higher than female mean engagement score.

POWER ANALYSIS

A statistical power analysis was performed for sample size estimation, based on data from pilot study/published study X (N=...), comparing A to B The effect size (ES) in this study was X, considered to be extremely large/large/medium/small using Cohen's (1988) criteria. With an alpha = .05 and power = 0.80, the projected sample size needed with this effect size is approximately N = X for this simplest between/within group comparison. Thus, our proposed sample size of N+ will be more than adequate for the main objective of this study and should also allow for expected attrition and our additional objectives of controlling for possible mediating/moderating factors/ subgroup analysis, etc.

A post hoc power analysis was conducted. The sample size of N=322 was used for the statistical power analyses and a 7 predictor variable equation was used as a baseline. The recommended effect sizes used for this assessment were as follows: small ($f2 = .02$), medium ($f2 = .15$), and large ($f2 = .35$) (see Cohen 1977). The alpha level used for this analysis was $p < .05$. The post hoc analyses revealed the statistical power for this study was .40 for detecting a small effect, whereas the power exceeded .99 for the detection of a moderate to large effect size. Thus, there was more than adequate power (i.e., power * .80) at the moderate to large effect size level, but less than adequate statistical power at the small effect size level.

There was no significant difference between the mean biases of experts and novices, $t(78) = -.149$, $p = .882$, Cohen's $d = 0.03$. Power analysis revealed that in order for an effect of this size to be detected (80% chance) as significant at the 5% level, a sample of 34886 participants would be required. [With such a large N we would start having significant differences between expert and novices?]

The minimum number of participants required was determined by an a priori power analysis. A total of 152 participants (all Caucasian) were included in the data analysis.

REGRESSION

A linear regression established that daily time spent watching TV could statistically significantly predict cholesterol concentration, $F(1, 98) = 17.47$, $p = .0001$ and time spent watching TV accounted for 14.3% of the explained variability (R^2) in cholesterol concentration. The regression equation was: predicted cholesterol concentration = -2.135 + 0.044 x (time spent watching tv).

CORRELATION

A Pearson product-moment correlation coefficient was computed to assess the relationship between the amount of water that one consumed and rating of skin elasticity. There was a positive correlation between the two variables, $r = 0.985$, $n = 5$, $p = 0.002$. A scatterplot summarizes the results (Figure 1) Overall, there was a strong, positive correlation between water consumption and skin elasticity. Increases in water consumption were correlated with increases in rating of skin elasticity

The two variables were strongly correlated, $r(df=n-2=55) = .49$, $p < .01$.

There was a positive correlation between the two variables, $r = .35$, $p = < .001$.

There was a positive correlation between height ($M = 55.39$ $SD = 16.33$) and weight ($M = 145.22$ $SD = 15.54$) , $r = .35$, $p = < .001$, $n = 100$.

A Spearman's rank-order correlation was run to determine the relationship between 10 students' English and maths exam marks. There was a strong, positive correlation between English and maths marks, which was statistically significant ($r_s(df=n-2=8) = .669$, $p = .035$)

FRIEDMAN'S TEST & WILCOXON

There was a statistically significant difference in perceived effort depending on which type of music was listened to whilst running, $\chi2(2) = 7.600$, $p = 0.022$.

There was a statistically significant difference in perceived effort depending on which type of music was listened to whilst running, $\chi2(2) = 7.600$, $p = 0.022$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p < 0.017$. Median (IQR) perceived effort levels for the no music, classical and dance music running trial were 7.5 (7 to 8), 7.5 (6.25 to 8) and 6.5 (6 to 7), respectively. There were no significant differences between the no music and classical music running trials ($Z = -0.061$, $p = 0.952$) or between the classical and dance music running trials ($Z = -1.811$, $p = 0.070$), despite an overall reduction in perceived effort in the dance vs classical running trials. However, there was a statistically significant reduction in perceived effort in the dance music vs no music trial ($Z = -2.636$, $p = 0.008$).