Week 2 Solutions
Submitted by: Dhanashree N P

Q1. What is the mean father's age?
Ans: 28.90

Q2. What is the mean father's age for low birthweight babies?
Ans: 24.83

Q3. Is the father's age normally distributed? Justify your answer.
Ans: No.
Shapiro-wilk test on fage gives the following results.
W = 0.94369, p-value = 0.0385
Here p-value < 0.05. For data to be normal, the value should be > 0.05.

Q4. If you apply the log transformation to the father's age, what is the mean score of the transformed variable?
Ans: 1.4493

Q5. Is the above mean score a good representation of the real value? Justify your answer.
Ans: No.
The mean score of 1.4493 from the log-transformed father's age is a valid summary for the transformed data, but it doesn't represent the actual original mean of the father's age unless it's back-transformed. For real-world interpretation, it would be more insightful to transform the value back to the original scale to understand the true central tendency.

Q6. Is the new variable (log transform of father's age) normally distributed? Justify your answer.
Ans: Yes
Shapiro-wilk test on new variable gives the following results.
W = 0.95833, p-value = 0.1287
Here p-value > 0.05. For data to be normal, the value should be > 0.05.

Q7. Is the variable "years father was in education" normally distributed?
Ans: No

Q8. Mentioning the null and alternative hypotheses, explain the above answer.
Ans: Applying Shapiro-Wilk test on the variable.

Null Hypothesis (H_0):

- The sample data follows a normal distribution.

H_0: The data is normally distributed.
Alternative Hypothesis (H_1):

- The sample data does not follow a normal distribution.

H_1: The data is not normally distributed.

If the p-value is small (usually less than 0.05): You reject the null hypothesis (H_0), suggesting that the data is significantly different from a normal distribution.
If the p-value is large (usually greater than 0.05): You fail to reject the null hypothesis, suggesting that there is not enough evidence to claim that the data is not normally distributed.

Here: W = 0.84415, p-value = 4.484e-05
p-value << 0.05. We reject the null hypothesis.

Q9. What is the mean score for the variable "years father was in education" after you apply the Box-Cox transformation?
Ans: 95.17 (using R) and 13.71 (SPSS)

Q10. Is this new variable normally distributed? Explain.
Ans: No
Shapiro-Wilk test results for this variable:
W = 0.84392, p-value = 4.427e-05.
The significance value should be > 0.05 for the data to be normal

Q11. What is the mean score for this new variable (B-C transformed fathers' years in education) for mothers aged under 35?
Ans: 92.55 (using R) and 13.55 (SPSS)

Q12. Which test would you use to investigate the relationship between birth weight and father's age?

- Pearson product-moment correlation
- Spearman's Rank order correlation
- Point-Biserial correlation
- Phi-Coefficient

Ans: Spearman's Rank order correlation

Q13. Justify the above choice in terms of the distribution of data and the nature of the test.
Ans: Birth weight – follows normal distribution
Father's age – does not follow normal distribution
Spearman's correlation is a non-parametric test. It does not assume that the data follows normal distribution. It works with the rank of the data instead of the actual data.
Spearman's correlation measures the monotonic relationship between the data, that is it doesn't assume a linear relationship – but the variable may increase/decrease as the other variable increases/decreases although not at a constant rate.
Spearman's Rank Order Correlation is the most appropriate test for investigating the relationship between birth weight and father's age because it does not assume normality, handles monotonic relationships, and works with continuous data.

Q14. What is the direction of that relationship?
Ans: Positive

Q15. What is the form of that relationship?

Ans: Monotonic

Q16. What is the degree of that relationship?
Ans: Weak

Q17. What test would you use to investigate the relationship between smoking and birth weight?

- Pearson product-moment correlation
- Spearman's Rank order correlation
- Point-Biserial correlation
- Phi-Coefficient

Ans: Point-Biserial correlation.
Smoking is a dichotomous variable, birthweight is continuous

Q18. Report on the above results including information about direction/form/degree of the relationship.
Ans:
t = -2.0934, df = 40, p-value = 0.0427
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.56427401 -0.01139044
sample estimates:
cor
-0.3142339

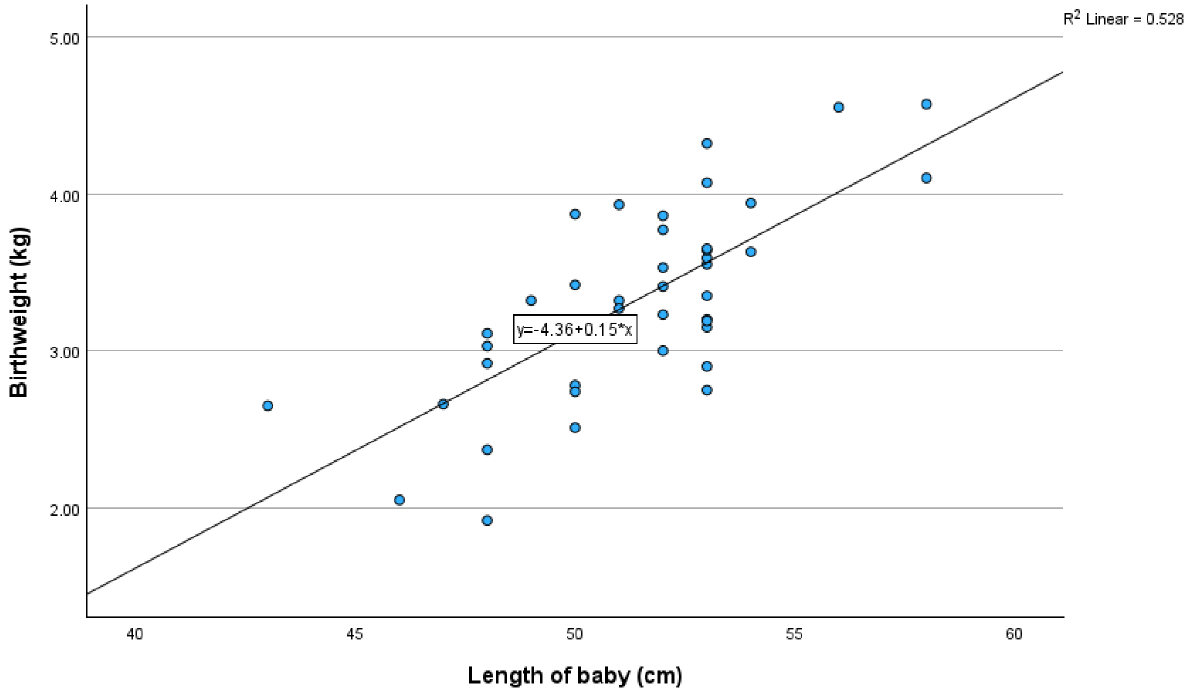
Direction: The relationship is negative, meaning that as one variable increases, the other tends to decrease.
Form: The relationship is monotonic, though not necessarily linear.
Degree: The strength of the relationship is weak, with a correlation of **-0.314**.
Significance: The correlation is statistically significant (p-value = 0.0427), indicating that the observed relationship is unlikely to have occurred by chance.
Confidence: The 95% confidence interval for the correlation excludes zero, suggesting a genuine negative relationship between the variables.

Q19. If you wanted to see the effect of the length of a baby on birthweight, what would your independent variable be?

- Length of baby
- Birthweight

Ans: Length of baby

Q20. In statistics, when creating a scatterplot, it is a common practice to put the independent variable on the x-axis and the dependent variable on the y-axis. With this in mind, create a scatterplot for the above case and provide the regression line. For homework submitted using MS Word, insert a picture of the scatterplot.
Ans: Using SPSSS



Q21. Is the relationship between the length of baby and birthweight linear?

- Yes
- No

Ans: Yes

Q22. Justify the above choice.

Ans: Pearson correlation test statistics between length and birthweight is as follows:

data: df$Length and df$Birthweight
t = 6.6931, df = 40, p-value = 5.029e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5428137 0.8442612
sample estimates:
      cor
0.7268335

The correlation score is 0.7268335 which shows a Strong, Positive and from the scatter plot above – Linear relationship between the two variables

Q23. Is there any evidence to suggest that the birth weight, length of baby, and head circumference are related?

- Yes

- No

Ans: Yes

Q24. Justify the above choice.

Ans: Results of partial correlation test between Birthweight and Head Circumference of baby keeping Length as constant:

estimate	p.value	statistic	n	gp	Method
0.4850457	0.001308298	3.463863	42	1	pearson

Based on the results of the partial correlation tests, there is evidence suggesting that **birth weight**, **length of the baby**, and **head circumference** are related. We observed a significant partial correlation of **0.4850** between **head circumference** and **birth weight**, after controlling for **length**. The p-value of **0.0013** for this test indicates that the relationship is statistically significant, meaning that after adjusting for length, head circumference and birth weight are moderately positively correlated.

it's reasonable to expect that these three variables may be interrelated, as both birth weight and head circumference are typical measures of infant growth and development, and length can also reflect overall growth.

Q25. Describe the above relationship in your own words and provide evidence for your claims.

Ans: The relationship between **birth weight**, **length**, and **head circumference** suggests that they are all interconnected measures of infant growth. Typically, larger babies tend to have a larger head circumference and longer length, so it is reasonable to assume that these variables are related in some way.

The specific evidence for this is found in the **partial correlation analysis** results:

- A **moderate positive relationship** was found between **head circumference** and **birth weight** after accounting for the effect of **length**. This suggests that as birth weight increases, head circumference tends to increase as well, even when length is held constant.
- The **p-value of 0.0013** from the test indicates that the correlation between head circumference and birth weight is **statistically significant**, meaning the observed relationship is unlikely to be due to chance.

Additionally, it is logical to consider that babies with higher birth weights generally also have larger head circumferences and greater lengths, reflecting overall development and health.

Evidence:

- From the partial correlation result: **cor = 0.4850**, **p-value = 0.0013**.
- The **moderate positive correlation** between head circumference and birth weight after controlling for length suggests a meaningful relationship between these variables.

Thus, the relationship between birth weight, length, and head circumference is supported by statistical evidence showing that head circumference and birth weight are moderately positively related after accounting for length, providing strong evidence that these measures of growth are interrelated.