**Hands-on-Training in Data Analysis by Kreativstorm. Homework-W1**

Q1. What is the mean birth weight for babies of non-smoking mothers?

3.5095 Kg

Q2. What is the mean birth weight for babies of smoking mothers?

3.1341 Kg

Q3. What is the mean head circumference for babies of non-smoking mothers?

35.05 cm

Q4. What is the mean gestational age at birth for babies of smoking mothers?

38.95 weeks

Q5. What is the maximum head circumference for babies of non-smoking mothers? 39 cm

Q6. What is the minimum gestational age at birth for babies of smoking mothers? 33 weeks

Q7. Based on the dataset you have, out of the two, which one would be a better bet:

* Pregnancy period in smoking mothers is shorter.
* Pregnancy period in non-smoking mothers is shorter.

Based on the dataset, Pregnancy period in smoking mothers is shorter.

Q8. Justify the above choice in a few words.

The mean gestation period in smoking mothers and non-smoking mothers are 38.95 and 39.45 weeks respectively, while the minimum gestation period in smoking mothers and non-smoking mothers are both the same (33 weeks). Based on the analysis, Pregnancy period in smoking mothers is shorter

Q9. What is the baby birth weight range for babies of smoking mothers?

2.65 Kg

Q10. In your own words describe what the value of the above range for baby's birthweight tells us about smoking versus non-smoking mothers?

While the baby birth weight range for babies of smoking mothers is

2.65 Kg, and the baby birth weight range of babies of non-smoking mothers is 1.90 Kg, the difference indicates that on average, babies born to smoking mothers are heavier at birth compared to babies born to non-smoking mothers. However, this does not necessarily mean that smoking during pregnancy is beneficial for the baby’s health.

Q11. Are head circumference data for babies of smoking mothers normally distributed? No, they are not normally distributed. Besides, there is an outlier in the distribution.

Q12. What is the significance value for the above on the Shapiro-Wilk test?

The significance value is 0.372.

Q13. What is the standard score (Z-score) for head circumference of 35.05 (X=35.05) in non-smoking mothers? The standard score is 0.51994.

Q14. How are birth weight data of non-smoking mothers skewed?

They are positive / right skewed.

Q15. Are birth weight data for babies of smoking mothers normally distributed?

No, they are negative / left skewed.

Q16. What is the significance value for the above on the Shapiro-Wilk test?

The significance value of the Shapiro-Wilk test in the analysis is 0.949.

Q17. Based on the dataset you have; how confident can you be in saying that a baby's birth weight will be +/- 1 standard deviation from the mean?

95% confident.

Q18. Based on the dataset you have, what is the probability that the birth weight for a baby of a smoking mother will be less than 4.2 kg?

4.2/9 =0.44

Q19. Are data for length of baby of non-smoking mothers normally distributed?

No, they are not normally distributed.

Q20. What is the significance value for the above on the Shapiro-Wilk test?

The significance value is 0.70.

Q21. What is the standard score for the length of a baby of 48.5cm for non-smoking mothers? 0.14686.

Q22. Based on the dataset you have, what is the probability that the length of baby for non-smoking mothers will be more than 55 cm? 1/60 = 0.26.

ANSWERS

Q1. What is the mean birth weight for babies of non-smoking mothers?

*M* = 3.51

Q2. What is the mean birth weight for babies of smoking mothers?

*M* = 3.13

Q3. What is the mean head circumference for babies of non-smoking mothers?

*M* = 35.05

Q4. What is the mean gestational age at birth for babies of smoking mothers?

*M* = 38.95

Q5. What is the maximum head circumference for babies of non-smoking mothers?

Max = 39

Q6. What is the minimum gestational age at birth for babies of smoking mothers?

Min = 33

Q7. Based on the dataset you have, out of the two, which one would be a better bet:

* **The pregnancy period in smoking mothers is shorter.**
* Pregnancy period in non-smoking mothers is shorter.

Q8. Justify the above choice in a few words.

Due to data being normally distributed, the mean is a good measure of central tendency so we can use it to represent all values in the dataset. Since in normal data the mean value is also the most frequent value, and the most frequent value means the highest occurrence in the past which then means the most likely value to happen in the future, we can conclude that the mean value is also a value most likely to happen in each group. If we compare the most likely values to happen in each group (two means), the value of smoking group is lower which means shorter pregnancy period.

Q9. What is the baby birth weight range for babies of smoking mothers?

Range = 2.65

Q10. In your own words describe what the value of the above range for baby's birthweight tells us about smoking versus non-smoking mothers?

The range for the smoking group is 2.65 while for non-smoking is 1.90 which indicates that there is less variation in the non-smoking group.

Q11. Are head circumference data for babies of smoking mothers normally distributed?

Yes

Q12. What is the significance value for the above on the Shapiro-Wilk test?

*p* = .372

Q13. What is the standard score (Z-score) for head circumference of 35.05 (X=35.05) in non-smoking mothers?

Z = 0

Q14. How are birth weight data of non-smoking mothers skewed?

Positively skewed (skewness is .361).

Q15. Are birth weight data for babies of smoking mothers normally distributed?

Yes.

Q16. What is the significance value for the above on the Shapiro-Wilk test?

*p* = .949

Q17. Based on the dataset you have, how confident can you be in saying that a baby's birth weight will be +/- 1 standard deviation from the mean?

p = 0.6827 based on theory or around .69 is you calculate it

Q18. Based on the dataset you have, what is the probability that the birth weight for a baby of a smoking mother will be less than 4.2 kg?

*p* (X < 4.2) = *p* (Z = 1.69)

*p* (X < 4.2) = 0.95

Q19. Are data for length of baby of non-smoking mothers normally distributed?

Yes.

Q20. What is the significance value for the above on the Shapiro-Wilk test?

*p* = .070

Q21. What is the standard score for the length of a baby of 48.5cm for non-smoking mothers?

Z = -1.01

Q22. Based on the dataset you have, what is the probability that the length of baby for non-smoking mothers will be more than 55 cm?

*p* (X > 55) = 1 - *p* (X <= 55)

*p* (X > 55) = 1 - *p* (Z = 0.98)

*p* (X > 55) = 1 - .83646

*p* (X > 55) = 0.16