p8130_hw4_xx2485

Xiaoni Xu

2024-11-16

Loading needed packages

```
library(readxl)
library(janitor)
```

Problem 1

Perform sign test.

```
data <- c(125, 123, 117, 123, 115, 112, 128, 118, 124, 111, 116, 109,
          125, 120, 113, 123, 112, 118, 121, 118, 122, 115, 105, 118, 131)

# Define the hypothesized median
median_hypothesis <- 120

# Count values above and below the hypothesized median
above <- sum(data > median_hypothesis)
below <- sum(data < median_hypothesis)

# Perform the sign test
binom_test <- binom.test(below, below + above, p = 0.5, alternative = "less")

# Report results
binom_test
```

```
##
##  Exact binomial test
##
## data:  below and below + above
## number of successes = 14, number of trials = 24, p-value = 0.8463
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.7536114
## sample estimates:
## probability of success
##              0.5833333
```

We conducted an exact binomial test to evaluate whether the median blood sugar level in the population is less than 120. The hypotheses were as follows:

$$H_0 : \text{The true median is } \geq 120 \quad (\text{probability of success } = 0.5).$$
$$H_a : \text{The true median is } < 120 \quad (\text{probability of success } > 0.5).$$

The test results are: - Number of successes (blood sugar readings below 120): 14 - Number of trials: 24 - p-value: 0.8463 - 95% confidence interval for the probability of success: $[0.000, 0.7536]$

Since the p-value (0.8463) is greater than the significance level ($\alpha = 0.05$), we fail to reject the null hypothesis.

Conclusion: There is no statistically significant evidence to suggest that the median blood sugar level in the population is less than 120.

The test statistic is 14.

Perform Wilcoxon signed-rank test.

```r
# Calculate differences from the hypothesized median
differences <- data - 120

# Remove zero differences
nonzero_differences <- differences[differences != 0]

# Rank absolute differences, handling ties with average ranks
abs_differences <- abs(nonzero_differences)
ranks <- rank(abs_differences)

# Sum ranks for negative differences
negative_ranks <- ranks[nonzero_differences < 0]
W_minus <- sum(negative_ranks)

# Calculate p-value using the normal approximation
n <- length(nonzero_differences)
mean_W <- n * (n + 1) / 4
sd_W <- sqrt(n * (n + 1) * (2 * n + 1) / 24)

z <- (W_minus - mean_W) / sd_W
p_value <- pnorm(z)

# Results
list(
  test_statistic = W_minus,
  p_value = p_value,
  z_score = z
)
```

```
## $test_statistic
## [1] 187.5
##
## $p_value
## [1] 0.8580116
##
## $z_score
## [1] 1.071429
```

Since the p-value (0.8580) is greater than the significance level ($\alpha = 0.05$), we fail to reject the null hypothesis.

There is no statistically significant evidence to suggest that the median blood sugar level in the population is less than 120.

Problem 2

(a)

```
# Load the data
brain <- read_excel("Brain.xlsx") %>%
  clean_names()


# Filter out the human species (Homo sapiens)
nonhuman_data <- subset(brain, species != "Homo sapiens")

# Fit a regression model using ln_brain_mass as the predictor for glia_neuron_ratio
model <- lm(glia_neuron_ratio ~ ln_brain_mass, data = nonhuman_data)

# Summary of the regression model
summary(model)
```

```
##
## Call:
## lm(formula = glia_neuron_ratio ~ ln_brain_mass, data = nonhuman_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.24150 -0.12030 -0.01787  0.15940  0.25563
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.16370    0.15987   1.024 0.322093
## ln_brain_mass  0.18113    0.03604   5.026 0.000151 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1699 on 15 degrees of freedom
## Multiple R-squared:  0.6274, Adjusted R-squared:  0.6025
## F-statistic: 25.26 on 1 and 15 DF,  p-value: 0.0001507
```

The regression model shows that ln(Brain Mass) is a significant predictor of the glia-neuron ratio in non-human species ($p < 0.001$). The positive slope (0.1811) indicates that species with larger brain masses (in terms of the natural logarithm) tend to have higher glia-neuron ratios. The model explains a substantial portion of the variability in the data ($R^2 = 0.6274$), making it a good fit for the observed relationship.

(b)

```
# Given human brain mass
human_brain_mass <- 1373.3
```

```r
# Calculate the natural logarithm of human brain mass
ln_human_brain_mass <- log(human_brain_mass)

# Use the regression coefficients from the model to predict glia-neuron ratio
intercept <- 0.1637  # From the regression model
slope <- 0.1811      # From the regression model

# Predicted glia-neuron ratio
predicted_glia_neuron_ratio <- intercept + slope * ln_human_brain_mass

# Output the result
predicted_glia_neuron_ratio
```

```
## [1] 1.472142
```

The predicted glia-neuron ratio for humans, given their brain mass is {r predicted_glia_neuron_ratio}.