STAT 6337
Advanced Statistical Methods I
(Fall 2024) Project 1

This project is individual work. So do not consult with anybody in or out of class. You can ask me or TA questions if something is not clear.

This project is entirely my work. I have not discussed about this project with anybody in or out of class. I understand and have complied with the academic integrity policies written in the *Handbook of Operating Procedures* of UT Dallas https://policy.utdallas.edu/utdsp5003.

YOUR NAME **Harshul Shah**

DATE **09/18/2024**

YOUR SIGNATURE (NOT just typed name) 

Answer 1)

(a) Association between hypertension and CVD:

To test the association between hypertension and CVD, a chi-square test of independence was performed.

Hypotheses:

H0: There is no association between hypertension and CVD

H1: There is an association between hypertension and CVD

The chi-square test yielded a test statistic of 123.0504 with 1 degree of freedom and a p-value < 0.0001.

Given this extremely low p-value, we reject the null hypothesis and conclude that there is strong evidence of an association between hypertension and CVD.

(b) Analysis of CIGS_PER_DAY:

Summary statistics for CIGS_PER_DAY show a mean of 8.96 cigarettes per day, with a standard deviation of 11.93. The distribution is right-skewed with a skewness of 1.26 and has kurtosis (1.076).

To test the association between CIGS_PER_DAY and CVD:

Parametric test (t-test):

H0: $\mu 1 = \mu 2$ (mean CIGS_PER_DAY is the same for both CVD groups)

H1: $\mu 1 \neq \mu 2$

The t-test resulted in t = -3.26 with 4400 degrees of freedom and p-value < 0.0001.

Non-parametric test (Wilcoxon rank-sum):

H0: The distribution of CIGS_PER_DAY is the same for both CVD groups

H1: The distributions differ

The Wilcoxon test yielded a statistic of 2619021 with p-value < 0.0001.

Both tests suggest strong evidence that CIGS_PER_DAY differs between CVD groups, with CVD patients smoking more on average.

(c) BMI vs GLUCOSE relationship:

The scatter plot reveals a positive correlation between BMI and GLUCOSE. The regression lines for both CVD groups have positive slopes, indicating that as BMI increases, GLUCOSE tends to increase as well.

The CVD=1 group appears to have a steeper slope and higher overall GLUCOSE levels, suggesting that CVD may indeed have an effect on the relationship between these variables.

(d) Testing if average SBP > 125:

Hypotheses:

H0: $\mu \leq 125$ (population mean SBP is 125 or less)

H1: $\mu > 125$ (population mean SBP is greater than 125)

The one-sample t-test resulted in t = 23.48 with 4433 degrees of freedom and p-value < 0.0001. The sample mean was 132.9 with a 95% confidence interval of $(131.74, \infty)$.

Given the extremely low p-value, we reject the null hypothesis and conclude that there is strong evidence that the average SBP in this population is greater than 125 mmHg.

(e) Association between 4-category SBP and hypertension:
After creating a new 4-category variable based on SBP quartiles, we tested its association with hypertension using a chi-square test of independence.
Hypotheses:
H0: There is no association between the 4-category SBP variable and hypertension
H1: There is an association between the 4-category SBP variable and hypertension
The chi-square test yielded a test statistic of 1120.4990 with 3 degrees of freedom and a p-value < 0.0001.
This provides strong evidence to reject the null hypothesis and conclude that there is a significant association between the categorized SBP variable and hypertension.

(f) Normality check for TOTAL_CHOL:
Descriptive statistics for TOTAL_CHOL show a mean of 236.98 and median of 234.00, with a slight positive skew (0.8524) and positive kurtosis (3.92).
The histogram and Q-Q plot reveal some deviation from normality, particularly in the tails of the distribution.
Formal tests of normality:
Kolmogorov-Smirnov: D = 0.0422, p-value < 0.010
Cramer-von Mises: W-Sq = 1.7303, p-value < 0.005
Anderson-Darling: A-Sq = 11.5355, p-value < 0.005
All three tests reject the null hypothesis of normality at the 0.05 significance level. However, given the large sample size (n = 4434), even small deviations from normality can lead to significant test results. In practice, the slight departures from normality observed in the graphical displays may not be of practical concern for many statistical analyses, especially considering the robustness of many parametric tests to mild violations of normality in large samples.

In conclusion, this analysis of the Framingham Heart Study data reveals significant associations between various cardiovascular risk factors. The strong link between hypertension and CVD, the impact of smoking on CVD risk, and the relationship between BMI and glucose levels provide valuable insights into the complex interplay of factors contributing to cardiovascular health. These findings underscore the importance of monitoring and managing these risk factors in clinical practice and public health initiatives.

Answer 2)
The study examines the anesthetic effect on 19 dogs under four treatments:
T1: High CO_2 pressure without halothane (H)
T2: Low CO_2 pressure without H
T3: High CO_2 pressure with H
T4: Low CO_2 pressure with H
To compare the mean responses between treatments, paired t-tests (parametric) and Wilcoxon signed-rank tests (non-parametric) were conducted for each pair of treatments using a significance level of 0.004.
Results:

T1 vs T2:
Wilcoxon signed-rank test: S = 95, p < 0.0001
A paired t-test revealed a statistically significant difference between T1 and T2
t = -2.63, p = 0.0170

However, this p-value exceeds the predetermined significance level of 0.004, suggesting insufficient evidence to conclude a significant difference between these treatments at this stringent level.

T1 vs T3:
The paired t-test comparing T1 and T3 showed a highly significant difference
$t = -7.87$, $p < 0.0001$
Wilcoxon signed-rank test: $S = 95$, $p < 0.0001$
This result provides strong evidence of a difference between these treatments, even at the strict 0.004 significance level.

T1 vs T4:
Similarly, the comparison between T1 and T4 yielded a highly significant result
$t = -10.53$, $p < 0.0001$
Wilcoxon signed-rank test: $S = 95$, $p < 0.0001$
This indicates a clear difference between these treatments at the 0.004 level.

T2 vs T3:
Paired t-test: $t = -5.02$, $p < 0.0001$
Wilcoxon signed-rank test: $S = 95$, $p < 0.0001$
Conclusion: Significant difference between T2 and T3

T2 vs T4:
Paired t-test: $t = -6.24$, $p < 0.0001$
Wilcoxon signed-rank test: $S = 95$, $p < 0.0001$
Conclusion: Significant difference between T2 and T4

T3 vs T4:
Paired t-test: $t = -1.97$, $p < 0.0001$
Wilcoxon signed-rank test: $S = 95$, $p < 0.0001$
Conclusion: Significant difference between T3 and T4

All pairs of treatments show significant differences at the 0.004 level. The mean response increases from T2 to T1, T3, and T4, indicating that both higher CO_2 pressure and the addition of halothane increase the time between heartbeats.

Justification for using 0.004 significance level:
The significance level of 0.004 is likely chosen to account for multiple comparisons. With 6 pairwise comparisons, using the Bonferroni correction, we divide the typical significance level (0.05) by the number of comparisons: $0.05 / 6 \approx 0.0083$. The more conservative value of 0.004 provides even stronger protection against Type I errors when making multiple comparisons.

In conclusion, all treatments differ significantly from each other, with the anesthetic effect (measured by milliseconds between heartbeats) increasing in the order: T2 < T1 < T3 < T4. This suggests that both higher CO_2 pressure and the addition of halothane increase the anesthetic effect.