IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Hendi Kushta
Date: 02/12/2023

****Important** Copying and/or pasting anything from the textbook will not be acceptable for your chapter notes submissions. You must write your notes in your own words and generate your own code, results, and graphs in R. This is what forces your brain to process the material that you read.**

This chapter is discussing the use of t-tests in analyzing data, using the example of comparing fuel economy in cars with manual and automatic transmissions. It mentions that William Sealy Gosset, who developed the t-test, assumed that the difference in means between two groups would be zero, and that the sample data would provide evidence to the contrary. The article mentions that in the example of the mtcars data set, one group could be considered the control (manual transmission) and the other the treatment (automatic transmission), but this choice is arbitrary.
There is a discussion between Bayesian thinking and its comparison to traditional statistical methods such as t-tests. Bayesian thinking allows for the assertion of a prior belief about the difference between two groups, whereas t-tests begin with the assumption that the difference is zero. Bayesian thinking considers each new piece of evidence to update and modify previous beliefs about the difference. The method provides probabilities for various amounts of difference between the groups, whereas t-tests only provide upper and lower bounds to the mean difference. The chapter mentions Bayes' theorem, named after Thomas Bayes, which allows the estimation of the probability of a scenario given new evidence and prior probability information.
Lets get an example of a group of boys and girls that like or not chocolates with 51 boys and 49 girls. Bayes theorem will help to find the probability that a selected person from the group will be boy that likes chocolate, p(like chocolates | boy). This is translated as the probability of observing Like chocolates given that the person is a boy. p(boy) is the column total and p(like chocolates) is the row total.

Contingency table			
	Boy	Girl	Total
Like chocolates	43	30	73
Doesn't like chocolates	8	19	27
Total	51	49	100

The table below is representation in probabilities of the table above.

IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing

Originality Assertion: By submitting this file you affirm that this writing is your own.

Contingency Table with Frequencies Converted to Probabilities			
	Boy	Girl	Total
Like chocolates	0.43	0.3	0.73
Doesn't like chocolates	0.08	0.19	0.27
Total	0.51	0.49	1

p(Boy | Like chcolates) = 0.43 / 0.73 = 0.589

p(Like chcolates | Boy) = (0.589 * 0.73)/0.51 = 0.843 or 84.3%

IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing

Originality Assertion: By submitting this file you affirm that this writing is your own.

BOX ON P.69: NOTATION, FORMULA, AND NOTES ON BAYES' THEOREM

Formulas for the Confidence Interval

A confidence Interval is a range of values that is likely to contain the true value of a population parameter with a certain level of confidence.

Example: A 95% confidence (~~parameter~~) interval means that if the same sample was drawn multiple times and confidence level was interpreted every time, approximately 95% of the intervals would contain the true population parameter.

The confidence interval falls within Lower bound (left side) and Upper bound (right side)

$$\text{Lower bound} = \underbrace{(\bar{x}_1 - \bar{x}_2)}_{} - \underbrace{t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}_{}$$

→ is the difference between 2 sample means, so in the example from the horse power for automatic and manual transmissions,

$\bar{x}_1$ → sample mean for automatic

$\bar{x}_2$ → sample mean for manual

$\bar{x}_1 - \bar{x}_2 = 160,2632 - 126,8462$
$= 33,417$

→ calculates the width ~~[illegible]~~ of the interval of confidence by substructing it from the means difference so we can find the lower bound.

t^* => is a critical value which is found in t-distribution table.

$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ => Standard error → represents the average difference between the sample estimate and the population

s^2 are ~~variables~~ variances for each samples

n are the number of observations for each sample.

* => To find the Upper bound its the same thing, but now we add the second part.

$$\text{Upper bound} = (\bar{x}_1 - \bar{x}_2) + t^* \sqrt{\frac{s_1}{n_1} + \frac{s_2}{n_2}}$$

IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing

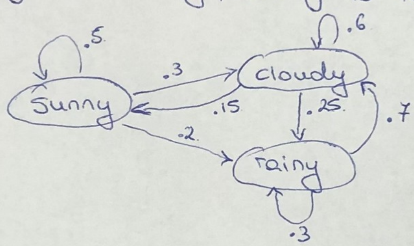
Originality Assertion: By submitting this file you affirm that this writing is your own.

BOX ON P.71: MARKOV-CHAIN MONTE CARLO OVERVIEW

Markov-Chain Monte Carlo Overview

Markov chain is named after Andrey Markov, a Russian mathematician. It is a mathematical model that describes a sequence of events in which the probability of each event depends only on the state attained in previous event.

Example | Consider a weather model that describes the weather conditions on a particular day based on the previous day's weather. If the weather today is sunny, there is a 50% chance that tomorrow will be sunny as well, 30% cloudy and 20% rainy. If it is cloudy, 60% cloudy, 25% rainy, 15% sunny, and if its rainy, 70% will be cloudy and 30% rainy.



Monte Carlo method involves generating random sampling. It averages results over many random trials.

Example | In finance it can be used to calculate the price movement of a stock overtime by modeling the underlying factors that influence the stock price. The basic idea is to generate a large number of random scenarios for the future price of the stock, taking into account factors such as interest rates, dividends, volatility and market trends.

When Bayesian analysis becomes too complexive, we use both Markov Chain and Monte Carlo together to sample, evaluate and improve the solution. The process starts with an initial estimate of the parameter being modeled. The posterior probability is evaluated based on the data and then the parameter is adjusted and the probability re-evaluated. This is repeated many times, building up a distribution of likely parameter values. that can be analyzed using histogram and descriptive statistics.

IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing

Originality Assertion: By submitting this file you affirm that this writing is your own.

BOX ON P.74: DETAILED OUTPUT FROM BESTMCMC()

The BEST package in R runs a simulation called Markov chain Monte Carlo to study the distances between the means of two groups. The install.packages and library functions are used to download and use the package, while the BESTmcmc function computes the probability distribution for the mean differences between the two groups. The plot function shows the specialized output from BESTmcmc and the resulting histogram indicates that manual transmissions in 1974 model year cars were more efficient in fuel economy compared to automatics. The 99.8% of the mean differences were negative, confirming that the chances of automatics being equal to or better than manual transmissions were close to zero. The highest density interval (HDI) shows that there is a 95% probability that the population mean difference between the two groups lies between -11.6 and -2.82 mpg, with the greatest likelihood being around -7.21. The HDI directly models the population parameters of interest, whereas the confidence interval uses sample data to provide information about the population mean difference.

I can not install BEST package in R Studio.
The output table shows the results of a Markov chain Monte Carlo (MCMC) simulation with 100,002 steps. It contains statistics (mean, standard deviation, median, lower and upper bound of the highest density interval, Rhat, and n.eff) for each population parameter (fuel efficiency of cars with automatic and manual transmissions) represented in the table. The posterior distributions of the population means and standard deviations can be seen in histograms produced by the command "plotAll(carsBest)". The "nu" value indicates the shape of each posterior distribution of means and Rhat and n.eff provide diagnostic information about the MCMC process. If Rhat is larger than 1.1 or n.eff is less than 10,000, it may indicate a problem with the MCMC process that could be addressed by running additional steps.

IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing
Originality Assertion: By submitting this file you affirm that this writing is your own.

THE NULL HYPOTHESIS SIGNIFICANCE TEST

This part discusses the generality of the null hypothesis significance test (NHST) as a statistical method in scientific research. NHST was developed by Ronald Fisher and involves the following steps:
1) making an assumption of no difference between two groups,
2) choosing an alpha level to determine the significance of the results,
3) collecting data and conducting a statistical test,
4) if p-value (calculated significance value) is less than the chosen alpha level, rejecting the null hypothesis,
5) if p-value is greater than alpha level, failing to reject the null hypothesis.
However, the NHST does not provide information about the alternative hypothesis or the probability of the null hypothesis being true.
The article discusses the continued use of the null hypothesis significance test (NHST) despite concerns about its reliability. A survey showed that 88% of scientists could not correctly identify the logic behind the NHST. The American Statistical Association has published guidelines in response to criticisms of the method. Despite these criticisms, NHST is still used because it is taught to many researchers as the main method of statistical inference. The author believes that it is important to understand the logic behind NHST and its criticisms. Three points of criticism against NHST are discussed without going into the mathematical details.

Statisticians have addressed the issue of interpreting statistical findings by developing measures of effect size. The effect size refers to the strength or magnitude of the statistical finding. There are both standardized measures, such as the "d" statistic developed by Jacob Cohen, and "rules of thumb" classifications for results. The "d" statistic provides a standardized measure of the difference between two sample means by dividing the mean difference by the pooled standard deviation of the two samples. The R programming language offers multiple methods to obtain effect size statistics. We can install the effsize() package.

```
> library(effsize)
> cohen.d(mtcars$hp[mtcars$am==0] ,mtcars$hp[mtcars$am==1])

Cohen's d

d estimate: 0.4943081 (small)
95 percent confidence interval:
     lower      upper
-0.2515338  1.2401500
```

The effect size of a statistical finding can be measured by the "d" statistic, developed by statistician Jacob Cohen. The "d" statistic provides a standardized measure of the size of the difference between two sample means by dividing the mean difference by the pooled standard deviation of the two samples. A finding of d = 0.494 is considered a small effect size, indicating that manual transmissions are nearly one and a half standard deviations less horsepower than automatic transmissions.

IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing

Originality Assertion: By submitting this file you affirm that this writing is your own.

BOX ON P.80: THE CALCULATION OF T

The Calculation of t

The t-distribution is a family of distributions used to model the distribution of population mean differences estimated from sample data. The t-distribution looks like a normal curve with "thicker" tails when the sample size is small, and becomes almost identical to the normal curve when sample is large. Statisticians developed this theory to take into account the uncertainty of population variance when it is estimated from a sample. The t-value is calculated using the means and std from two samples.

$$t_{obs} = \frac{(\bar{x}_1 - \bar{x}_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Difference between the means of two groups.

Standard deviation of the difference in the sample means.

The results is then compared to the t-distribution which takes into account the number of degree of fredom (the number of observations minus the number of variables)

Example | Lets say the sample data for manual transmissions gave a mean of 20 miles/gallon with a std of 3 mpg, and the sample data for automatic transmission gave a mean of 18 mpg and a std of 2 mpg.

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{20 - 18}{\sqrt{\frac{9}{n_1} + \frac{4}{n_2}}} \Rightarrow$$ Lets say $t = 2{,}5$.

We compare this value to the critical t-value from the t-distribution with degrees of freedom $n1-1$ and $n2-1$.

If the calculated t value > critical t-value we reject the null hypothesis and conclude that there is a significant difference in the means of fuel efficiency between manual and automatic transmissions.

IST772 Summary Template: Chapter 5 – Bayesian and Traditional Hypothesis Testing
Originality Assertion: By submitting this file you affirm that this writing is your own.

REPLICATION AND THE NHST

The alpha threshold of .05 means that a false positive result may occur in one out of 20 statistical tests. Research articles often conduct several tests, leading to a high likelihood of at least one irreproducible result in a published article. Studies on replications and retractions show that these false results are common, with research suggesting that fewer than 40% of articles in psychology can be replicated with similar results using the same design.
The practice of seeking statistically significant results, often with a threshold of $p < .05$, can lead to the manipulation of data and methods in a phenomenon known as "p-hacking." The pressure to obtain significant results can cause researchers to prioritize statistical significance over understanding the practical implications and meaningfulness of their findings. The availability of large data sets can also result in trivial differences appearing statistically significant, making the null hypothesis significance test (NHST) less useful. The scientific community's emphasis on statistical significance, combined with the temptation to manipulate data and methods, highlights a fundamental problem in the way scientific studies are conducted and published. It is important to critically evaluate the results of studies and consider the broader implications and understanding that can be gained beyond just the statistical significance of the findings.