

Project: Comparison of Multiple Distributions

Author: Opeyemi Ayanwale

June 9, 2023

Contents

1	Introduction	1
2	Problem Statement	2
2.1	Dataset and Data Quality.....	2
2.2	Project Objectives	2
3	Statistical Methods	3
3.1	Hypothesis Testing.....	3
3.1.1	Methods for testing hypothesis	4
3.2	Analysis of Variance (ANOVA).....	4
3.2.1	Assumptions of ANOVA for comparing two or more means.....	6
3.3	T-Test	7
3.3.1	Independent t-tests and assumptions	7
3.4	Multiple Pairwise-Comparison Between the Means of Groups	8
3.4.1	Pairwise t-test	9
3.4.2	The Scheffé Test and Tukey's Honest Significant Difference (HSD)	9
3.4.3	Bonferroni correction	9
4	Statistical Analysis	10
4.1	Descriptive Statistics of Babies Weight and Smoking Status	10
4.2	Assumptions of ANOVA: normality assumption	11
4.3	Assumptions of ANOVA: homogeneity of variance assumption	12
4.4	Global Test: ANOVA	13
4.5	Pairwise Differences: T-test	13
4.6	Bonferroni Adjustment.....	14
4.7	Scheffé Test	14
5	Summary	15
	Bibliography	16
	Appendix	18
A	Additional figures	18
B	Additional tables.....	18

1 Introduction

Smoking before and during pregnancy has been linked to an increased risk of health complications for women. Previous research has shown that maternal smoking has several negative health complications for babies. One of the most frequently reported effects of smoking during pregnancy is low birth weight (CDC, 2020). The World Health Organization defines low birthweight as a birthweight of less than 2,500 grams (5.5 pounds) (WHO, 2004).

The primary goal of this report is to examine the relationship between maternal smoking and babies' weight, as well as whether different smoking conditions lead to changes in the weight of different groups of neonates, using the Stat Lab's mothers and their babies dataset. The independent variables include infant survival, birth weight, date of birth, sex, mother's ethnicity, age, education level, height, weight, and smoking status (Stat Labs, 2002). For the purpose of this report, only the variable weight (wt) of the babies and mothers' smoking status/history will be considered for the analysis.

The distribution of the babies' weight and the different smoking categories were described using descriptive statistics. Analysis of variance (ANOVA) is used for multiple comparisons to determine whether the babies' birth weight (in ounces) between the categories is the same. All the assumptions of ANOVA were checked before the test was conducted, the QQ-plot, Shapiro test, and Levene test are used to validate the ANOVA assumptions.

The t-test was used to test for pairwise differences between the resulting birth weight, taking into account all possible pairs of categories. The Bonferroni correction method, commonly known as the Bonferroni type adjustment procedure and Tukey's Honest significant difference test/Scheffé test was used to correct multiple comparison problems.

The second section provides a more detailed overview of the dataset, description of variables, and information on data quality. The statistical analysis methods are presented and explained in the third section. And in the fourth section, the presentation, analysis, and interpretation of the results are presented. Finally, in the fifth section, the main findings are summarized.

2 Problem Statement

2.1 Dataset and Data Quality

The dataset used in this report is an extract from the Stat Labs data of mothers and their babies. It includes 1236 samples and 23 independent variables such as infant survival, birth weight, date of birth, sex, mother's ethnicity, age, education level, height, weight, and smoking status. For the purposes of this report, only the variable babies' weight (wt) and mothers' smoking status/history will be considered. The variable wt contains babies' birth weight in ounces (999 = unknown) and smoke contains mothers' smoke history (0 = never, 1 = smokes now, 2 = until current pregnancy, 3 = once did, not now, 9 = unknown). The original data are available for educational/research stat labs data website (Stat Labs, 2002).

The two variables of interest (babies weight and smoking status) are numerical but smoking status was changed to factor since its a categorical variable. Weight had 10 NA, smoke have 10 unknown values, and all the missing and unknown rows were removed before proceeding to the analysis.

2.2 Project Objectives

The primary goal of this report is to examine the relationship between maternal smoking and baby weight, as well as whether different smoking conditions cause weight changes in different groups of neonates and pairwise differences between the resulting birth weights. The distribution of the babies' weight and the different smoking categories will be described using descriptive statistics. One-way analysis of variance (ANOVA) will be used to examine if the babies' birth weight differs between the smoking categories. T-tests for independent samples with equal variances will also be used to examine if there are pairwise differences between the resulting babies' birth weights. All the assumptions will be checked before conducting the test and to correct the multiple comparison problems associated with utilizing multiple testing, the Bonferroni method for p-adjustment and Tukey's Honest significant difference/Scheffé test based on equal or unequal sample size will be applied. The statistical method session will go over mean hypothesis testing and other methods used to meet the objectives of this report in detail.

3 Statistical Methods

This section presents several statistical methods that will be used to analyze the dataset based on the objectives of this report. All analyses and visualizations were performed using the statistical software R, version 4.2.3 (R Development Core Team, 2022), utilizing the basic r packages with desctools (Signorell, 2023) and car packages respectively (Fox and Weisberg, 2019).

3.1 Hypothesis Testing

A decision-making procedure for examining assumptions about a population parameter is known as ***hypothesis testing***. It is possible to investigate hypotheses involving parameters such as means and proportions. There are two types of statistical hypotheses: the null hypothesis and the alternative hypothesis (Allan G. Bluman, 2009, p. 401).

The null hypothesis (H_0) is a statistical hypothesis that claims there is no difference between a parameter and a certain value, or between two parameters (Allan G. Bluman, 2009, p. 401).

$$H_0 : \mu = k$$

The alternative hypothesis (H_1) is a statistical hypothesis that states a difference between a parameter and a certain value, or between two parameters (Allan G. Bluman, 2009, p. 401).

$$H_1 : \mu \neq k$$

A two-tailed test state that the mean is equal $H_0 : \mu = k$ vs $H_1 : \mu \neq k$, right-tailed test state that the mean is greater $H_0 : \mu = k$ vs $H_1 : \mu > k$ and left-tailed test state that the mean is less than $H_0 : \mu = k$ vs $H_1 : \mu < k$, where k represents a specified number (Allan G. Bluman, 2009, p. 402).

The value obtained from a statistical test is called the ***test value***. The decision to reject or not to reject the null hypothesis is based on the test value. The null hypothesis is rejected if there is a significant difference otherwise, the null hypothesis is accepted (Allan G. Bluman, 2009, p. 404).

3.1.1 Methods for testing hypothesis

Confidence interval (CI): A confidence interval is a specific interval within which the true population parameter is expected to fall with a given level of certainty (confidence level). A confidence level is a measure of certainty associated with a confidence interval which is typically chosen prior to conducting the hypothesis test. The most commonly used confidence intervals are 90%, 95%, and 99% CI. The desired level of certainty or precision for the analysis determines the confidence level (Allan G. Bluman, 2009, p. 358).

Critical value (CV): Critical values are basically cut-off values that identify region where the test statistic is unlikely to lie. The critical value can inform us how likely it is that two sample means belong to the same distribution. The null hypothesis is rejected if the test statistics is found to be greater than the critical value. If it is lower, then the null hypothesis is not rejected (Allan G. Bluman, 2009, p. 406). Furthermore, if a null hypothesis is true and it is rejected, then a type I error has been made. Similarly, if the null hypothesis is false and it is not rejected, then a type II error has been made (Allan G. Bluman, 2009, p. 405).

Level of Significance: The level of significance determines how large of a difference is required to reject the null hypothesis and is the maximum probability of committing a type I error. The probability of type I error is denoted by α and β for the probability of type II error. The most frequent significant levels used by statisticians are 0.01, 0.05, and 0.10, which means that if the null hypothesis is rejected, the probability of a type I error is 10%, 5%, or 1% respectively (Allan G. Bluman, 2009, p. 406).

P-value: The P-value represents the probability of a specific sample statistic or a more extreme sample statistic occurring if the null hypothesis is true. In hypothesis testing, a p-value is also used to decide whether to accept or reject the null hypothesis. The lower the p-value, the more evidence you have that the null hypothesis should be rejected. A small p (≤ 0.05), rejects the null hypothesis. This is significant evidence that the null hypothesis is incorrect. A large p (> 0.05) indicates that the alternative hypothesis is weak, therefore you do not reject the null hypothesis (Allan G. Bluman, 2009, p. 418).

3.2 Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) is a statistical test used to compare means of three or more groups and to test whether the differences between the means are statistically

significant. A one-way analysis of variance is used to compare three or more means because it contains only one variable, whereas a two-variable analysis of variance is known as a two-way analysis of variance. Although we are comparing three or more means, variances are used instead of means for the test, hence the name analysis of variance. Two estimates are produced from the test, between-group variance (finding the variance of means) and within-group variance (computing variance using all the data and less affected by mean differences). If the means do not differ, the between-group variance estimate will be approximately equal to the within-group variance estimate, and the F test value will be approximately equal to 1 (Allan G. Bluman, 2009, p. 631).

Mathematical formula:

$$F = \frac{MS_{between}}{MS_{within}}$$

where:

F = ANOVA coefficient

$MS_{between}$ = Mean sum of square between group variation

MS_{within} = Mean sum of square within group variation

$$MS_{between} = \frac{SS_{between}}{k - 1}$$

$$SS_{between} = n_1(\bar{x}_1 - \bar{x}_G)^2 + n_2(\bar{x}_2 - \bar{x}_G)^2 + n_3(\bar{x}_3 - \bar{x}_G)^2 + \dots + n_k(\bar{x}_k - \bar{x}_G)^2$$

where $SS_{between}$ is the sum of squares between groups, \bar{x}_G is the grand mean, $\bar{x}_1 \dots \bar{x}_k$ is the mean of each sample, $n_1 \dots n_k$ are the sample sizes, k is the number of groups and $k - 1$ is degree of freedom is the number of sample means(k) minus one (Allan G. Bluman, 2009, p. 633-636).

$$MS_{within} = \frac{\sum (x_{ij} - \bar{x}_j)^2}{N - k}$$

$$SS_{within} = \sum (x_{i1} - \bar{x}_1)^2 + \sum (x_{i2} - \bar{x}_2)^2 + \dots + \sum (x_{ik} - \bar{x}_k)^2 = \sum (x_{ij} - \bar{x}_j)^2 \text{ where}$$

SS_{within} is the sum of squares within groups, x_{i1} is the i th value of first sample, x_{i2} is the i th value of second sample, x_{ij} is the j th value from the j th sample, k is the

number of groups and $N = n_1 + n_2 + \dots + n_k$ which is the sum of sample sizes of the groups (Allan G. Bluman, 2009, p. 633-636).

Hypotheses of a One-Way ANOVA: there are two possible hypotheses.

H_0 : there is no difference between the groups and equality between means versus H_1 : at least one of the means is different from others.

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{ vs } H_1 : \mu_1 = \mu_2 \neq \mu_3$$

3.2.1 Assumptions of ANOVA for comparing two or more means

- The populations from which the samples were obtained must be normally distributed.
- The samples must be independent of one another.
- The variances of each sample are assumed to be equal.

Methods for testing the assumptions of ANOVA

QQ-plot: A "quantile-quantile" plot (QQ-plot) can be used to determine whether a sample violates the normality assumption. This allows visual inspection for any systemic infractions. In a QQ-plot, each observation is represented by a single dot. A 45-degree reference line is also plotted and if the data is normal, the dots should form a straight line (Mario F. Triola, 2018, p. 53).

Levene test: The assumption of equal variance can also be tested formally via the Levene test. An F-test is used in Levene test to test the null hypothesis that variance is equal across groups versus the alternate hypothesis that the variances are not equal for at least one pair. If the p-value is less than .05, the assumption of homogeneity has been violated and then there is a significant difference between the variances but if the p-value is greater than .05, then the variances are not significantly different from each other then the homogeneity assumption of the variance is met (Danielle Navarro, 2020a, p. 8276).

Levene test hypothesis: $H_0 : \sigma_1 = \sigma_2 = \sigma_3$ vs $H_1 : \sigma_1 \neq \sigma_2 \neq \sigma_3$

Levene test statistics:

$$W = \frac{(N - k)}{(k - 1)} \frac{\sum_{i=1}^k N_i (Z_i - Z_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} (Z_{ij} - Z_{i.})^2}$$

where k is the number of different groups to which the sampled cases belong, N_i is the number of cases in the i th group, N is the total number of cases in all groups, Z_{ij} can have three different interpretations depending on whether the mean, median, or trimmed mean of any subgroup is used (Danielle Navarro, 2020a, p. 8276).

Sharpiro-Wilk test: The Shapiro-Wilk test is a hypothesis test used to determine whether a dataset is normally distributed. Small W values indicate that the sample is not normally distributed (we can reject the null hypothesis that our population is normally distributed if the values are less than a certain threshold)(Danielle Navarro, 2020b, p. 8267).

Sharpiro-Wilk test statistics:

$$W = \frac{\sum_{i=1}^n a_i x(i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

where x_i are the ordered sample values and a_i are constants that are generated by the expression.

3.3 T-Test

The t-test is a statistical test used for finding the mean of a population where σ is unknown and the population is normally or essentially normally distributed. Because the population standard deviation σ is unknown, the sample standard deviation s is used instead. A t-test can be computed using the mean difference between the mean values of each dataset, the standard deviation of each group, and the number of data values in each group (Allan G. Bluman, 2009, p. 427).

Mathematical formula:

$$t = \frac{\bar{X} - \mu}{s \div \sqrt{n}}$$

Where t is the student's t-test, \bar{X} is the sample mean, μ is hypothesized population mean, s is the sample standard deviation, and n is the sample size.

3.3.1 Independent t-tests and assumptions

There are two types of t-tests, the independent t-tests (uncorrelated) and paired t-tests (correlated). The independent sample t-test with equal variances is used for this project.

The independent t-test, also known as the uncorrelated t-test, is a statistical test used to evaluate the difference between two means when the two samples are unrelated and are taken from two normally or approximately normally distributed populations (Allan G. Bluman, 2009, p. 484-485).

Assumptions for the t-tests for two independent means when σ_1 and σ_2 are unknown

- The samples are random samples.
- The sample data are independent of one another.
- If the sample size is less than 30, the population must be normally or approximately normally distributed.
- The variances of the samples are equal

Mathematical formula when variances are assumed to be equal:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

where $\bar{X}_1 - \bar{X}_2$ = observed difference between sample means, $\mu_1 - \mu_2$ = expected difference between population means, n_1 = sample size of sample 1, n_2 = sample size of sample 2, S_p^2 = pooled estimate of the variance. S_p is the pooled standard deviation of the two samples obtained by taking the square root of the weighted average of the variance using the variances of the two samples with their degrees of freedom as weights (Allan G. Bluman, 2009, p. 484-487).

3.4 Multiple Pairwise-Comparison Between the Means of Groups

A significant p-value in a one-way ANOVA test indicates that some group means differ, although the exact pairs of groups that differ may not be known. Multiple pairwise comparisons can be performed to evaluate if the mean difference between specified pairs of groups is statistically significant (Nordstokke and Stelnicki, 2014).

3.4.1 Pairwise t-test

When numerous t-tests are run, the issue of multiple testing (also known as multiplicity) occurs. While there are many techniques for resolving the challenges caused by multiple testing, they all seek to apply an adjusted p-value to each test or drop the p-value threshold from 5% to a more appropriate amount. The *pairwise.t.test()* function in R is used to compute pairwise comparisons between group levels with corrections for multiple testing (Danielle Navarro, 2020c, p. 4032).

3.4.2 The Scheffé Test and Tukey's Honest Significant Difference(HSD)

The Tukey and Scheffe test can also be used after the analysis of variance to make pairwise comparisons between means to determine if the relationship between groups is statistically significant (Allan G. Bluman, 2009, p. 642). For this report, Scheffé test will be used since the samples are of different sizes.

Mathematical formula:

$$F_s = \frac{(\bar{X}_i - \bar{X}_j)^2}{s_w^2 \frac{1}{n_i} + \frac{1}{n_j}}$$

where \bar{X}_i and \bar{X}_j are the means of the samples being compared, n_i and n_j are the sample sizes, and s_w^2 is the within-group variance. There is a significant difference between the two means being compared if F_s is greater than F' . To find the critical value F' for the Scheffé test, multiply the critical value (CV) for the F_s test by $k - 1$ (Allan G. Bluman, 2009, p. 643).

3.4.3 Bonferroni correction

The Bonferroni correction, also known as the Bonferroni type adjustment is one of the most basic methods used in statistical analysis for multiple comparison testing (Herzog et al., 2019). In hypothesis testing, the Bonferroni correction procedure is a straightforward method of controlling the Type I error rate. In the context of hypothesis testing, Type I error is the possibility of discovering a false-positive result and so rejecting a true null hypothesis. The Bonferroni test is intended to limit the occurrence of false positives. The Bonferroni test has been criticized for being too cautious and hence missing certain significant discoveries, yet it remains the most commonly used method to account for multiplicity due to its simplicity. Bonferroni's correction is computed by dividing the

initial alpha level (usually set to 0.05) by the number of tests being performed (Danielle Navarro, 2020c, p. 4032).

Mathematical formula:

$$BCP = \frac{\alpha}{n}$$

Where BCP = Bonferroni corrected p-value, α = the original p-value, n = number of tests performed.

4 Statistical Analysis

4.1 Descriptive Statistics of Babies Weight and Smoking Status

The weight of babies varies across the 4 smoking groups (0 = never, 1 = now, 2 = until current pregnancy, 3 = once did, not now). Mothers who never smoke are 540, those who smoke now are 481, and mothers who smoke until pregnancy are 95 compared to those who once did but no longer do, who are 100 in number. From Table 1, the highest mean weight is among those mothers who used to smoke but not at the moment followed by those who did until current pregnancy. This suggests that these two groups tend to have higher average baby weights compared to the other groups (see Table1).

Table 1:Descriptive Statistics of Babies Weight

smoke	n()	mean_weight	variance_weight	std_weight
0	540	122.8611	291.0512	17.06022
1	481	114.1081	323.0175	17.97269
2	95	123.0842	316.9716	17.80370
3	100	124.6300	344.8617	18.57045

Figure 1 depicts a histogram of distribution of babies weight. The graph shows that babies' weights are normally distributed from the shape of the histogram, we can assume that the sample data come from a population that is approximately a normal distribution (see Figure 1).

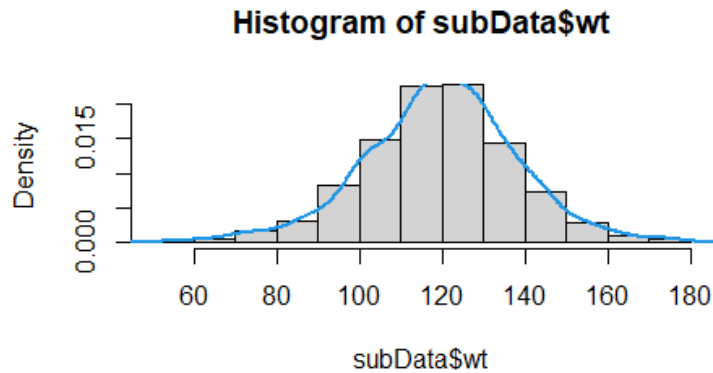


Figure 1:Histogram of Babies Weight.

4.2 Assumptions of ANOVA: normality assumption

The normality assumption is visually examined using the quantile-quantile plot (QQ- plot). Figure 1 shows the QQ-plot of mothers smoking history based on the four groups (Never smoke, Smoke now, Until pregnancy, Once did, now) while Figure 2 shows the QQ-plot for each city separately. The null and alternative hypotheses are stated below:

H_0 : The data are from a normal distribution vs

H_1 : The data are not from a normal distribution

Figures 2 and 3 in the Appendix on page 18 show the plot of the residual quantiles against the normal distribution quantiles with a 45-degree reference line. Almost all of the points fall on a straight line or are extremely close to the reference line, implying that the samples for each of the groups are normally distributed (see Figures 2 and 3 in Appendix on page 18).

The Shapiro-Wilk test was also used to test the assumption of normality. For the first group: the p-value is 2.959e-06 suggesting strong evidence against the null hypothesis of normality. Therefore, we would reject the assumption of normality for this group.

For the second group: the p-value is 0.5476 (> 0.05), this suggests that there is not enough evidence to reject the null hypothesis of normality for this group. Therefore, we can assume that the data for this group follows a normal distribution.

For the third group: the p-value is 0.2456 (> 0.05), indicating that there is not enough evidence to reject the null hypothesis of normality for this group too. Hence, we can assume that the data for this group follows a normal distribution.

For the fourth group: the p-value is 0.5079 (> 0.05). This suggests that there is not enough evidence to reject the null hypothesis of normality for this group as well. Therefore, we can assume that the data for this group follows a normal distribution. Since small W-values also indicate that the sample is not normally distributed and from the test, we have large W of 0.99 for the four groups so we assumed that the data is normally distributed (see Table 2).

Table 2: Shapiro-Wilk normality test by groups.

Groups	W-value	p-value
group0wt	0.98187	2.959e-06
group1wt	0.99707	0.5476
group2wt	0.98275	0.2456
group3wt	0.988	0.5079

4.3 Assumptions of ANOVA: homogeneity of variance assumption

Levene Test is used to test for homogeneity of variance assumption and the output is summarised in Table 3. The null and alternative hypotheses for the test are stated below:

H_0 : variances are equal vs H_1 : at least one variance is different

Table 3: Levene Test for Homogeneity of Variance (center = median).

	Df	F-value	Pr(>F)
Group	3	2.0519	0.1049
	1212		

The test was performed on three groups, and the calculated test statistic (F value) of 2.0519 with a p-value of 0.1049 shows that there is no strong evidence to reject the null hypothesis of equal variances because the p-value is greater than the significance level (0.05). This shows that the variances across the three groups are not significantly different from one another, indicating variance homogeneity (see Table 3).

4.4 Global Test: ANOVA

To test if the babies' birth weights differ between the groups/categories, analysis of variance (ANOVA) was performed and summarised in Table 4. The hypothesis is stated as follows

H_0 : The mean birth weights are equal across all categories of smoking status vs

H_1 : The mean birth weights differ between at least one pair of categories

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

where μ_1 = never smoke μ_2 = smoke now μ_3 = until current pregnancy μ_4 = once did, not now

The p-value from the analysis of variance is extremely small(< 0.05), indicating a significant difference in the mean birth weights between at least one pair of smoking status categories. As a result, we can conclude that there is sufficient evidence to reject the null hypothesis and support the alternative hypothesis that the birth weights of the babies differ across smoking status categories (see Table 4).

Table 4: Summary of analysis of variance test.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
smoke	3	239312	7977	25.72	3.91e-16
Residuals	1212	375862	310	NA	NA

4.5 Pairwise Differences: T-test

The results of the analysis of variance do not indicate which group(s) are different from the others. The t-test, which compares the means of two independent samples with equal variances, was used to check for pairwise variations to further analyze where this difference in mean occurs. All possible combination was considered. The hypothesis is the same and is presented below where i and k are the means of the pairs that are being tested:

$$H_0 : \mu_i = \mu_k \text{ vs } H_1 : \mu_i \neq \mu_k$$

Table 5: Pairwise comparisons using t-tests without adjusting for multiple testing.

	0	1	2
1	5.0e-	-	-
2	0.91	6.2e-06	-
3	0.36	6.6e-08	0.54

From Table 5, the pairs of never-smokers (0) and current smokers (1), current smokers (1) and those who smoked until the current pregnancy (2), current smokers (1) and once smoked but not now (3) have low p-values shows strong evidence to reject the null hypothesis and conclude that there is a significant difference in mean birth weights between these groups. And the pairwise comparison between the pairs of never-smokers (0) and those who smoked until the current pregnancy (2), never-smokers (0) and once smoked but not now (3), those who smoked until the current pregnancy (2) and once smoked but not now (3) p-values are greater than 0.05 shows that there no strong evidence to reject the null hypotheses and we can conclude that there is no significant difference in mean birth weights between these two groups.

4.6 Bonferroni Adjustment

Table 6 presents the adjusted p-values obtained by the Bonferroni method from pairwise comparisons of all possible pairs at $\alpha = 0.05$. Based on the result, there are significant differences in mean birth weights between several pairs of categories: 0 and 1, 1 and 2, 2 and 3. However, there is no significant difference between categories 0 and 2 or between categories 1 and 2.

Table 6: Pairwise comparisons using t-tests with pooled SD and adjusting for multiple testing.

	0	1	2
1	3.0e-	-	-
2	1	3.7e-05	-
3	1	3.9e-07	1

4.7 Scheffé Test

Based on the Scheffé test results in Table 7 in the Appendix on page 19, groups 1-0, 2-1, and 3-1 with the lowest p-values show strong evidence to reject the null hypoth-

esis and conclude that there is a significant difference in mean birth weights between these groups while groups of 2-0, 3-0 and 3-2 have no significant difference in mean birth weight between these groups (see Table 7). When we compare the results of the two correction methods (Bonferroni and Scheffe) in the non-adjusted test, there are several pairwise comparisons that have p-values that are less than the significance level in the non-adjusted test. This suggests that there are significant differences between the groups being compared. The p-values of the Scheffé correction, which provides a balance between controlling the family-wise error rate are smaller than the p-values of the Bonferroni correction, which is a conservative method that strictly controls the overall type I error rate (Danielle Navarro, 2020c, p. 4032).

5 Summary

The dataset used in this report is an excerpt from Stat Labs data on mothers and their babies. This report focuses on two variables: baby weight (wt) and smoking status/history of mothers. The objective of this report is to examine the relationship between maternal smoking and baby weight, as well as whether different smoking conditions cause weight changes in different groups of neonates and pairwise differences between the resulting birth weights.

Babies' weights differ across the four smoking categories/groups, with the highest mean weight among mothers who used to smoke but no longer do, followed by those who did until the current pregnancy. The quantile-quantile plot, Shapiro-Wilk test, and Levene test were used to further validate the normality assumption. Based on the test results, all groups except group0wt are normally distributed.

The Levene Test also shows that the variances between the three groups are not significantly different, indicating variance homogeneity. The analysis of variance p-value also shows a significant difference in mean birth weights between at least one pair of smoking status groups. According to the findings from the post-hoc analysis, there are significant differences in mean birth weights between several pairs of groups/categories: never smoked and now smokes, now smokes and until current pregnancy, now smokes and until current pregnancy, and once did but not now. There is, however, no statistically significant difference between the categories never and until current pregnancy, or between smokers now and until current pregnancy.

Bibliography

- Allan G. Bluman (2009). *Elementary Statistics: a step by step approach - 8th edition*. McGraw-Hill Higher Education.
- CDC (2020). Smoking and tobacco use. https://www.cdc.gov/tobacco/basic_information/health_effects/pregnancy/index.htm(visited on 1st June 2023).
- Danielle Navarro (2020a). *Learning Statistics with R: Checking the Homogeneity of Variance Assumption*. University of New South Wales. <https://stats.libretexts.org/@go/page/8276>.
- Danielle Navarro (2020b). *Learning Statistics with R: Checking the Normality of a sample*. University of New South Wales. <https://stats.libretexts.org/@go/page/8267>.
- Danielle Navarro (2020c). *Learning Statistics with R: Multiple Comparisons and Post Hoc Tests*. University of New South Wales. <https://stats.libretexts.org/@go/page/4032>.
- Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, third edition.
- Herzog, M. H., Francis, G., and Clarke, A. (2019). *The Multiple Testing Problem*, pages 63–66. Springer International Publishing, Cham.
- Mario F. Triola (2018). *Elementary Statistics*. Pearson Education, Inc Boston, 13th edition.
- Nordstokke, D. and Stelnicki, A. M. (2014). *Pairwise Comparisons*, pages 4575–4576. Springer Netherlands, Dordrecht. https://link.springer.com/referenceworkentry/10.1007/978-94-007-0753-5_2059 (visited on 6th June 2023).
- R Development Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/> (visited on 1st June 2023).
- Signorell, A. (2023). *DescTools: Tools for Descriptive Statistics*. R package version 0.99.48.

Stat Labs (2002). Stat labs data. <https://www.stat.berkeley.edu/users/statlabs/labs.html> (visited on 1st June 2023).

WHO (2004). Low birth-weight : country, regional and global estimates. <https://apps.who.int/iris/handle/10665/43184> (visited on 1st June 2023).

Appendix

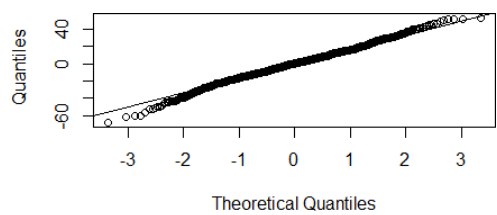
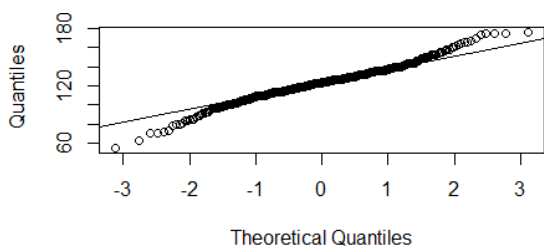
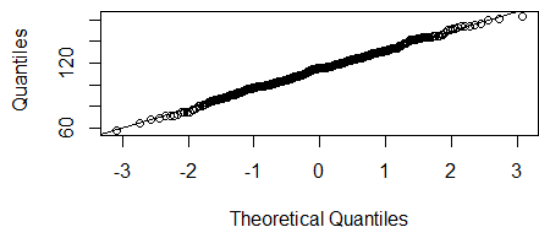


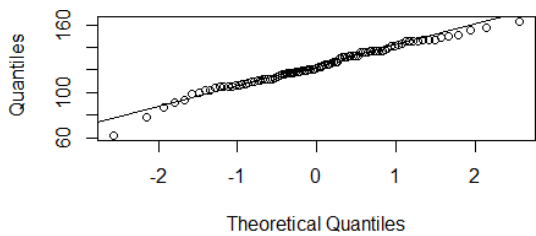
Figure 2:QQ-plot of babies' birth weight for all the 4 groups.



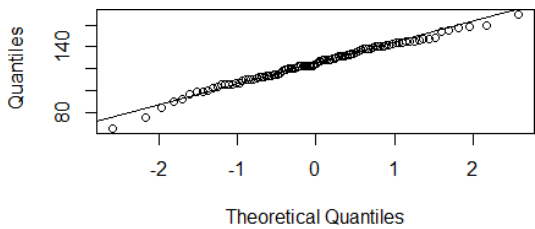
(a)Never smoke



(b)Smokes now



(c)Until Pregnancy



(d)Once did

Figure 3:QQ-plots of babies' birth weight for each of the 4 groups.

Table 7:Scheffe test of all pairs of maternal smoking status

SmokingGroup	Diff	lowerCI	upperCI	P-value
1-0	-8.7530030	-11.843823	-5.662183	3.1e-13 ***
2-0	0.2230994	-5.261671	5.707870	0.99961
3-0	1.7688889	-3.598009	7.135787	0.83715
2-1	8.9761024	3.441236	14.510969	0.00014 ***
3-1	10.5218919	5.103807	15.939977	2.0e-06 ***
3-2	1.5457895	-5.517155	8.608734	0.94526