ST502: Final Project - Part 1

Apostolos Stamenos & Tyler Pollard

4/19/2022

Part 1

Need an introduction

Let $Y_{i1}, \ldots, Y_{in}$ be systolic blood pressure measurements from a simple random sample, where $i = 1$ denotes that the individual was selected from the population of nonsmokers and $i = 2$ denotes that the individual was selected from the population of smokers. For the samples from each population, we assume the parametric model $Y_{i1}, \ldots, Y_{in} \overset{\text{iid}}{\sim} N\left(\mu_i, \sigma_i^2\right)$, where μ_i is the mean systolic blood pressure and σ_i^2 is the variance for population $i \in \{1, 2\}$. We tested the following hypotheses to determine whether or not the true population mean systolic blood pressure differs for smokers and nonsmokers:

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_A : \mu_1 \neq \mu_2$$

Table 1: **Summary of tests**: The confidence interval method and p-value method reject the null hypothesis for both the pooled and Satterthwaite versions of the t-test.

Test	Point Estimate	SE	df	Test Statistic	p-value	Confidence Interval
Pooled Variance	9.16	3.01	298	3.04	0.0026	(3.23, 15.08)
Satterthwaite						

Discuss results of each test (Tyler)

Both the histograms and the normal QQ plots indicate that the data are skewed to the right. The boxplots are fairly symmetrical if the outliers are excluded. We decided to keep the outliers in the analysis. By the Central Limit Theorem, even if the two datasets are not completely normal, their sample means are asymptotically normally distributed. Since the number of smokers and the number of nonsmokers are sufficiently large, the use of t-tests and confidence intervals is justified by the Central Limit Theorem.

The boxplots indicate that the distribution of systolic blood pressure for nonsmokers is more spread out than the distribution of systolic blood pressure for smokers. Based on the boxplots, there is no indication that the true population variances are equal. In addition to visually inspecting the distributions of systolic blood pressure for the two groups, we conducted a formal hypothesis test for equality of variances. We used the median-based extension of the Levene test, as specified in Brown & Forsythe (1974):

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad vs \quad H_A : \sigma_1^2 \neq \sigma_2^2$$

Unlike other tests for equality of variances, this version is more robust to deviations from Normality (Brown & Forsythe, 1974). With a p-value of 0.045, we reject the null hypothesis of equal variances at the 5% significance level. Thus, we conclude that the t-test with the Satterthwaite approximation is preferred.

Need a conclusion

Appendix A: Data Visualizations

Figure 1: **Histograms of systolic blood pressure for smokers and nonsmokers**

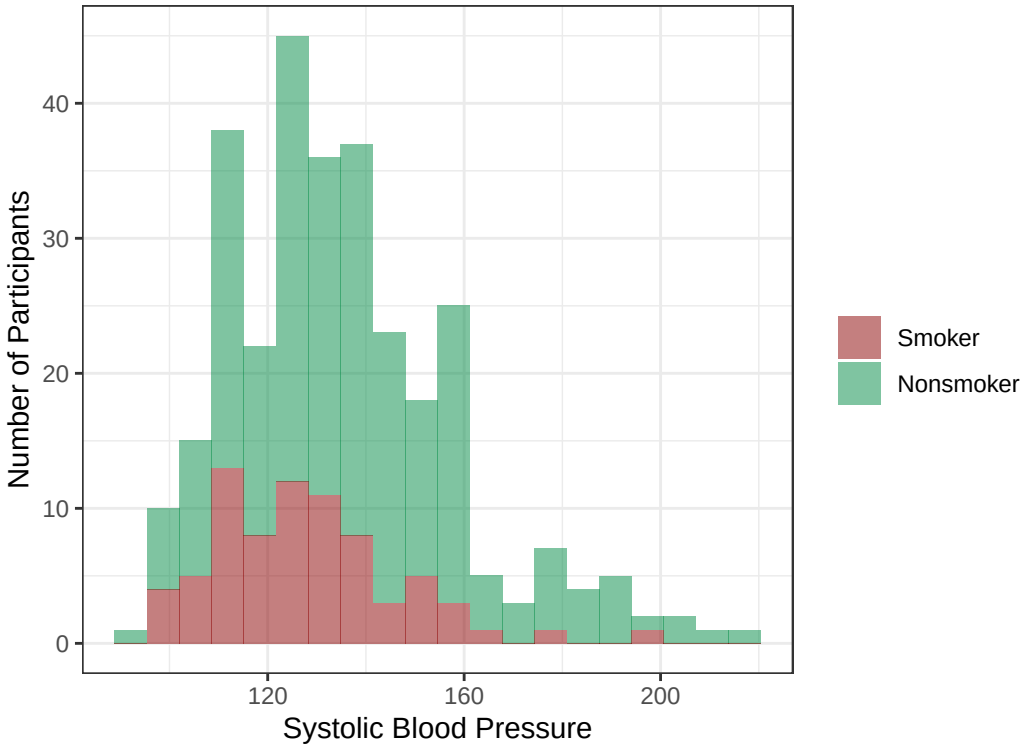


Figure 2: **Normal QQ plots of systolic blood pressure for smokers and nonsmokers**

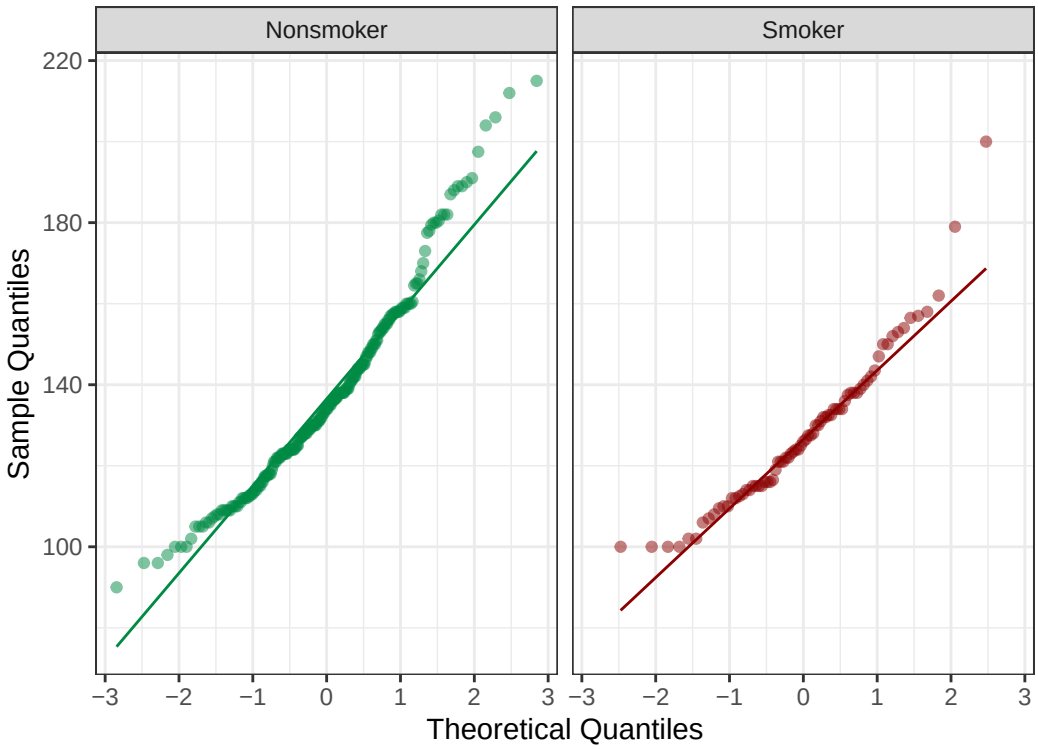
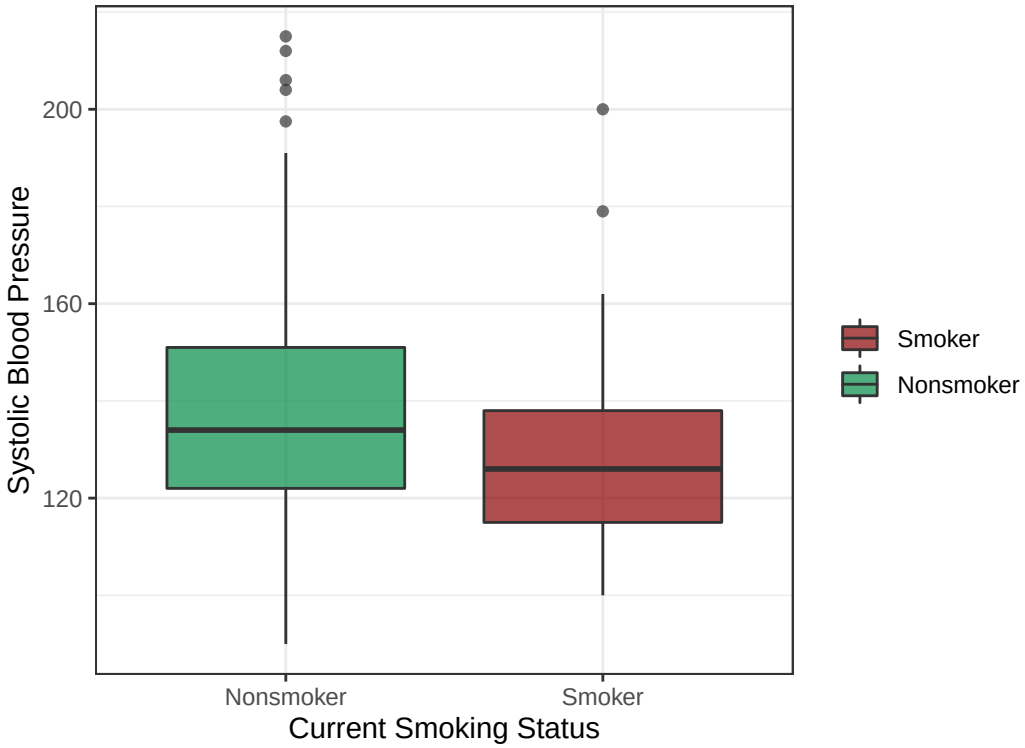


Figure 3: **Boxplots of systolic blood pressure for smokers and nonsmokers**


Systolic Blood Pressure
200
160
120
Nonsmoker
Smoker
Current Smoking Status
Smoker
Nonsmoker

Bibliography

Brown, M. B. and Forsythe, A. B. (1974), *Journal of the American Statistical Association*, 69, pp. 364-367