Brewing Lower Acidity Coffee with a Single-Serve Capsule Coffee Maker

by

Danilo Petti And Luisa Ferrari

February 2020

Summary

Consumption of coffee can trigger heartburn due to both caffeine and its acidity, the latter being affected by a wide range of factors not necessarily well-understood by consumers. This research examines treatments of three common ingredients – the intensity rating of coffee grounds, addition of sugar and carbonation of water – on the pH level of coffee using a single-serve capsule home espresso machine. Capsule machines are ideal in an experimental setting, providing consistent conditions for brewing and removing much of the operator variation. The experiment is conducted as a fully randomised 2^3 full factorial design and analysed using ANOVA. Results suggest that lower acidity coffee is achievable with careful choice of water and may be a viable solution for those who suffer from heartburn. Although coffee intensity and sugar strongly influence perceived acidity, insufficient evidence is found to suggest they alter the actual acidity of coffee. Consumers would therefore benefit from standardised acidity information on packaging.

1 Introduction

Heartburn is a common symptom of Gastroesophageal Reflux Disease (GERD) and has high prevalence in Western populations [1, 2]. Coffee is an example food often cited in medical advice [3] that may trigger or worsen the effects. In [4], a correlation between pH (acidity) and heartburn score is found within beverages, with coffee having among the highest scores. It is therefore hypothesised that coffee acidity may be linked to an increased risk of heartburn.

For consumers who suffer heartburn, reliance on perceived acidity – which may not be a good indicator of pH [5] – can lead to failed attempts to brew lower acidity coffee. This study aims to identify, within an experimental setting, whether three accessible treatments can alter the pH level of brewed coffee and not just perceived acidity.

2 Materials and Methods

2.1 Choice of Design Factors

Three design factors are considered for the experiment: coffee intensity, type of water and addition of sugar.

A more intense coffee is anecdotally associated with bitter notes and lower perceived acidity. However, the intensity scales advertised on packaging are arbitrary (brand-specific) and how it relates to actual acidity may not be clear to consumers.

Entrapped CO_2 within coffee beans are hypothesised in [6] to affect the physical mechanisms of brewing when released. This study examines whether a higher volume of CO_2 released when brewing with carbonated water affects the amount of acid extracted. However, it was not known whether the lower pH of carbonated waters would mask this effect.

Finally, sugar is a common additive used to offset the bitter or sour notes in coffee, directly altering acidity perception. Regular white sugar is not expected to affect pH because it does not dissociate into ions when dissolved in water [7].

2.2 Materials

Two coffee capsules, each containing 5 g of ground coffee, produced by *Caffè Borbone Srl* are chosen for the experiment: Black, marketed as intense and Blue, marketed as balanced. Capsules are hermetically sealed to prolong the coffee integrity compared to brewing methods where beans may have more exposure to oxygen. Nevertheless, dates of manufacture for all samples were matched to the nearest month as a precaution.

Efforts were made to select waters with minimal differences in mineral composition, although complete elimination of differences was not possible. One carbonated water *Acqua Sant'Agata*, and one still water *Acqua Santo Stefano* are selected, with an advertised pH of 6 and 7.7 respectively. Water was tested prior to use and pH levels were found to be within ± 0.03 of the specified value.

3 g of white caster sugar is measured with digital weighing scales (± 0.1 g accuracy). Caster sugar is selected as the smaller grain size aids dissolution in the coffee.

2.3 Pilot trials

Numerous small-scale trials were conducted to refine the experimental procedure and reduce the effects of extraneous factors. A summary of identified factors and the control methods taken to mitigate their effects can be found in Table 2.1.

A larger trial run was used to conduct preliminary checks of modelling assumptions as well as estimate the residual variance for sample size calculation. No significant concerns were raised regarding modelling assumptions.

2.4 Coffee Sample Preparation

Coffee samples were prepared with a Nespresso® Inissia D40 machine and is a method well-suited to an experimental setting, reducing variation caused by human factors. The machine was de-scaled and underwent a warm-up procedure consisting of running $2\times$ with no capsule; this ensured the system was fully-primed to produce consistent conditions during extraction. For repeatable dilution, the machine was programmed to extract coffee for a fixed 5 s.

All samples were extracted into homogeneous vessels that had been cleaned thoroughly with mild soap and distilled water. Glass vessels were selected to have enough thermal mass to balance cooling the coffee to a measurable temperature range whilst allowing the sugar to dissolve completely.

The weighed sugar was immediately added following extraction and stirred with a plastic spoon for 10 s for complete dissolution. Mechanical stirring was applied to all samples, with or without sugar, to ensure process similarity.

2.5 Measurement of Acidity

Measurement of pH "power of Hydrogen" quantifies the acidity of a solution, and is therefore a suitable response variable. It is measured on a base-10 logarithmic scale between 0 (very acidic) and 14 (very alkaline), with coffee typically between 3 and 6. This narrow range eliminated pH test strips which can only provide coarse estimates to the nearest 0.5 or 1 pH.

Sufficient measurement capability is obtained through use of a Dr. Meter 838 pH meter, with a resolution of pH 0.01 across the full pH scale and measurement error of pH ± 0.03. The meter is equipped with Automatic Temperature Compensation (ATC) to account for the effect of temperature on pH measurement. This removed the need to measure all samples at 25 °C, reducing the overall time required for the experiment.

The pH meter was re-calibrated before running the experiment for accurate measurements. Measurement was taken by immersing the meter in the coffee sample immediately after mechanical stirring with a wait time of 40 s to allow the reading to stabilise. In between measurements, the meter electrode was cleaned thoroughly with distilled water to prevent excessive wear and contamination between samples.

Table 2.1: Control methods for extraneous factors.

Factor (units)	**Experimental Level /Allowable Range**	**Measurement precision** (how known)	**Method of Control**
Brewing Temperature (°C)	78 ± 5		Do runs only after 2× dummy run machine warm-up procedure. Remaining variation controlled by randomisation.
Brewing Pressure (bar)	max = 19	Specified by manufacturer. Assumed adequate.	As above.
Brewing volume (ml)	60 ± 5		As above. Additionally, programmed to extract for a fixed 5 s.
Machine wear	-	-	Randomisation
Measurement Temperature (°C)	40 - 60	±0.1. (pH meter equipped with temperature probe.)	Automatic Temperature Compensation adjusts pH reading based on temperature; fixed wait time to allow sufficient cooling; same vessel for consistent cooling rate.
pH meter wear	-	-	Re-calibration before experiment. Clean with distilled water immediately after measuring.
Contaminants	-	-	Descale machine. Flush out remaining water and coffee from prior trial before next trial. Clean measuring equipment, cups etc. with distilled water.
Water Mineral Content (mg/litre)	-	Specified by water brand. Assumed accurate.	Minimise the difference between selected waters.
Carbonation of water (based on pH)	+0.05pH	Estimated with pH meter	Use only unopened bottles of carbonated water; limit excitation of water when transferring to machine. Use new bottle if too much CO_2 is lost.
Capsule variation	-	-	Match manufacture dates of capsules. Randomisation handles capsule-to-capsule variation.
Sugar Dissolved (g)	3 ± 0.1	Weighed using digital scales	Caster sugar with small grain size; add sugar immediately after coffee extraction finished; stir for 10 s for complete dissolution.

2.6 Randomisation

Extraneous factors identified during pre-experimental planning are unlikely to be exhaustive. There also exist nuisance factors that cannot be practically controlled in the proposed experimental setup. For these factors, randomisation in the experimental run order is employed to reduce potential sources of bias.

Full randomisation of the run order is made possible through use of two water tanks, enabling efficient swapping between water types and avoiding a split-plot design. The run order was determined by assigning a randomly generated number to each combination/ replicate and then sorting according to this order.

2.7 Statistical Analysis

A 2^3 full factorial design was used with all factors treated as categorical. As all interaction terms are included, replication was necessary to have enough degrees of freedom to calculate an independent estimate of σ^2. A balanced design was used for replicates; aside from maximised power, this offered advantages such as increased robustness to mild violations of the homogeneous variance assumption [8].

Assumptions were checked using a combination of graphical methods and statistical tests. For checking normality, the Shapiro-Wilk test [9] was used due to its superior power compared to other normality tests [10]. For testing the null hypothesis for equality of variances, the modified Levene (Brown-Forsythe) test [11] was chosen for its robustness to minor departures in normality. For the assumption of independence, the total and partial autocorrelation functions were visually inspected.

After confirming assumptions were met, results were analysed using analysis of variance (ANOVA) with a fixed effects model and sum-to-zero constraints. A significance level of 5% was applied unless otherwise specified.

2.8 Sample Size Determination

A power analysis of the 2^3 design was performed in Minitab [12] to determine the appropriate number of replicates. For the analysis, the desired significance level $\alpha = 0.05$, and desired power for all terms $1 - \beta = 0.95$ were set. An effect size of 0.1 pH is considered to be practically significant and is purposely set higher than the repeatability of the meter. Using the residual variance $\hat{\sigma}^2 = 0.0057$ estimated from the pilot run, power curves in Figure 2.1 were generated for 3 up to 7 replicates. To meet the desired 95% power on all terms, 5 replicates were necessary, totalling 40 runs for the entire experiment.

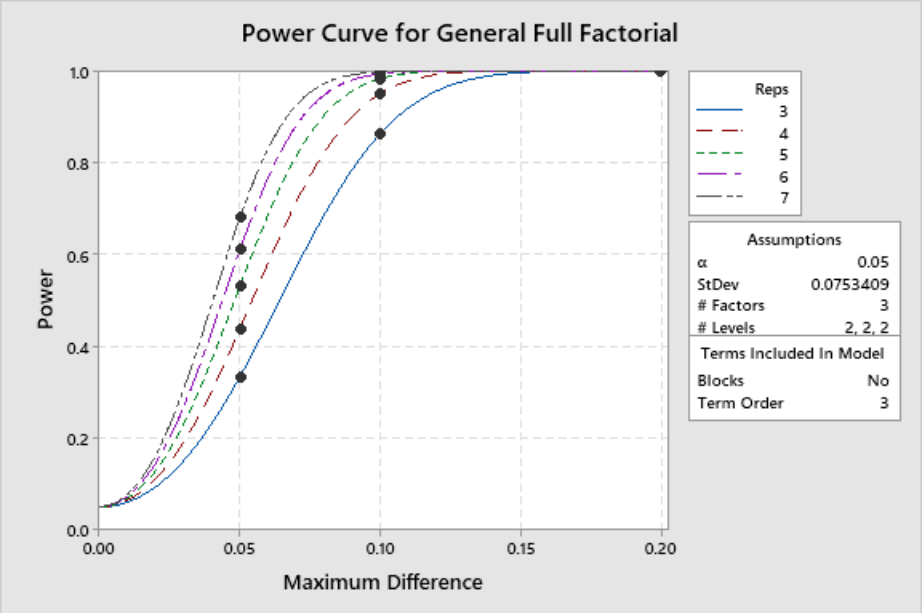


Figure 2.1: Power curves estimated using pilot run data [12]

3 Results and Analysis

The experiment was run in a single session with all 40 trials conducted according to the rigorous experimental procedure and no missing data.

Visualisation of the data in Figure 3.1 reveals a potentially dominant effect of water treatment on the pH of coffee. All eight treatment means are summarised in Figure 3.2.

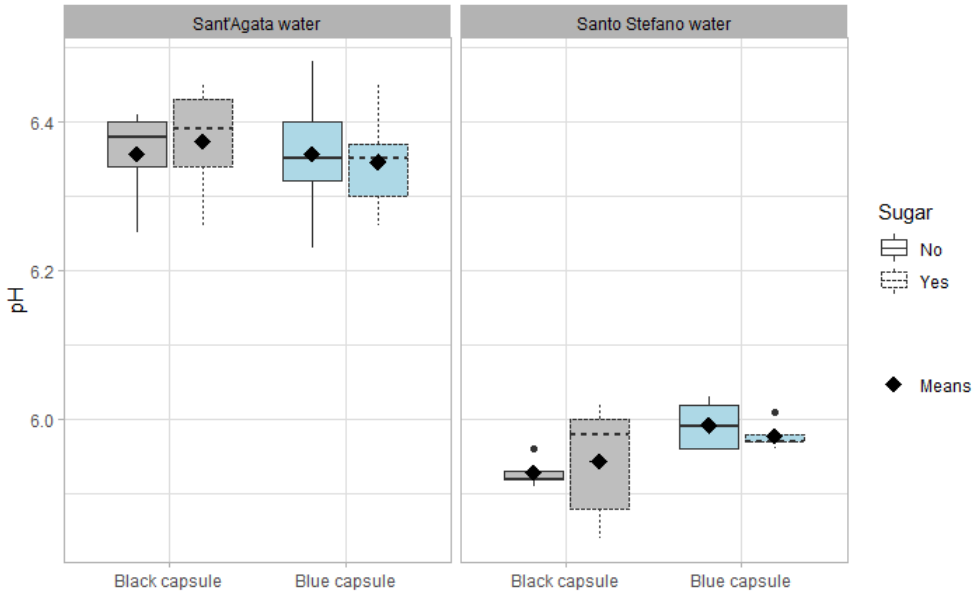


Figure 3.1: Boxplot of pH across treatment combinations

Model adequacy checks were carried out according to Section 2.7. On normality of residuals, the Shapiro-Wilk test returned a p-value of 0.453, indicating insufficient evidence to reject the assumption of normality. The construction of a normal probability plot of the residuals in Figure 3.3a supports this as the error distribution approximately follows a normal distribution except for minor deviation in the longer left tail. Three standardised residuals were found to be slightly < -2, but this is not far from typical for 40 data points ($\approx 5\%$ of points outside $\pm 2\sigma$).

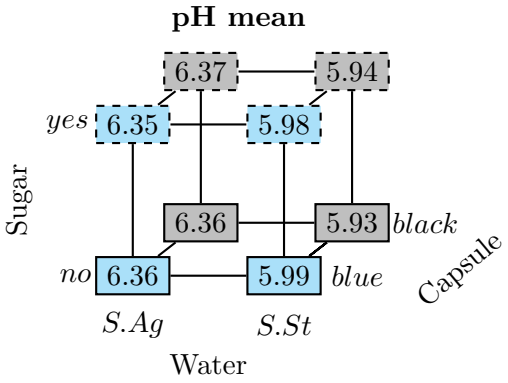
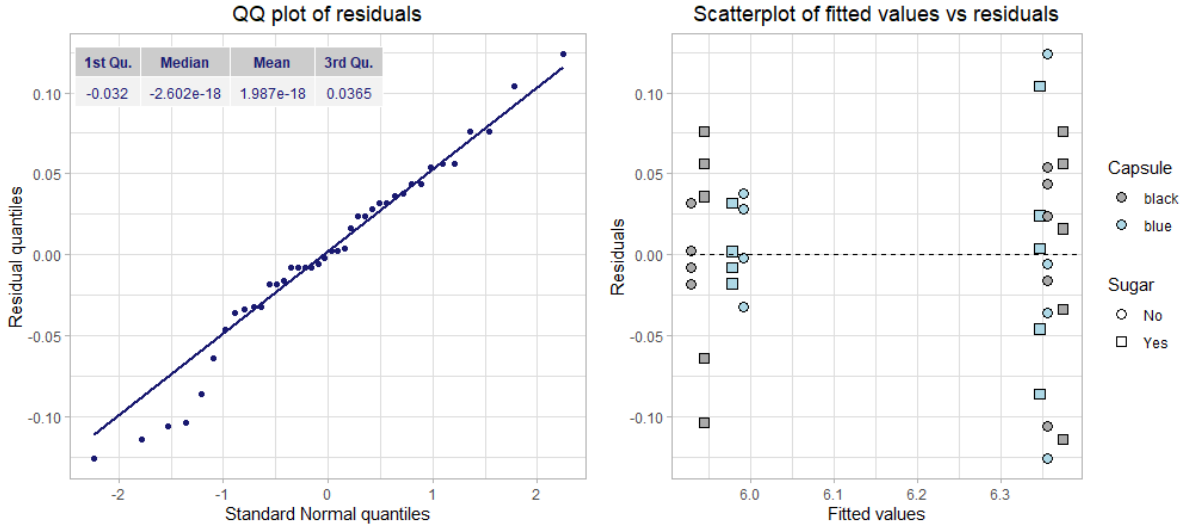


Figure 3.2: 2^3 factorial design with sample means

The null hypothesis of homogeneous variances also failed to be rejected, the modified Levene test returning a p-value of 0.272. There are possible signs of heteroscedasticity with higher variance for higher fitted values in Figure 3.3b, but the level of difference is not of major concern due to the robust, balanced design.

Inspection of the total and partial autocorrelation functions did not signify dependency among residuals. The ANOVA model is therefore considered appropriate for the analysis.



(a) **(b)**

Figure 3.3: ANOVA model adequacy checking: (a) Normal probability plot of residuals and (b) Residuals versus fitted values.

Table 3.1: Analysis of Variance Table

	DoF	Sum of Squares	Mean Square	F statistic	P Value
Water	1	1.580	1.580	394.892	1.443×10^{-19}
Capsule	1	0.003	0.003	0.765	0.388
Sugar	1	6.250×10^{-5}	6.250×10^{-5}	0.016	0.901
Water*Capsule	1	0.010	0.010	2.480	0.125
Water*Sugar	1	2.250×10^{-5}	2.250×10^{-5}	0.006	0.941
Capsule*Sugar	1	0.002	0.002	0.525	0.474
Water*Capsule*Sugar	1	2.500×10^{-6}	2.500×10^{-6}	0.001	0.980
Residuals	32	0.128	0.004		

Water is the only term where statistical significance was found. As no two- or three-way interactions were found to be significant, this main effect can be treated to be of practical value. Specifically, the main effect of water on pH of the brewed coffee, going from *Santo*

Stefano to *Sant'Agata*, is a 0.3975 increase (confidence interval: [0.3583, 0.4367]). Capsule intensity and addition of sugar were not found to have any significant effect.

4 Conclusions

The experiment has shown that different bottled waters can affect coffee acidity brewed with a capsule machine, providing a viable choice for those that suffer from heartburn. Despite carbonated *Sant'Agata* water having a lower pH, the coffee it brews is surprisingly $\approx 2.28 - 2.73\times$ less acidic than that brewed with still *Santo Stefano* water. Further research should be aimed at confirming whether this result is due to difference in carbonation using waters with matching mineral compositions.

Insufficient evidence is found that coffee intensity or sugar addition can be used to control coffee acidity. For those looking to brew lower acidity coffees for health reasons, these treatments – and others that rely on acidity perception – risk not having the intended effect.

The validity of these conclusions likely holds only in the restricted setting and materials of this experiment. Generalisation can be investigated further through, for example, extending to other coffee brewing methods with two-way ANOVAs, or comparing other sugar varieties in a one-way ANOVA. In particular, use of a home carbonation appliance would ensure the only difference between waters is the carbonation, eliminating any distortion in results due to mineral composition differences.

References

[1] Philip O Katz, Lauren B Gerson, and Marcelo F Vela. Guidelines for the diagnosis and management of gastroesophageal reflux disease. *American Journal of Gastroenterology*, 108(3):308–328, 2013.

[2] Sean D. Delshad, Christopher V. Almario, William D. Chey, and Brennan M.R. Spiegel. Prevalence of gastroesophageal reflux disease and proton pump inhibitor-refractory symptoms. *Gastroenterology*, 158(5):1250 – 1261.e2, 2020.

[3] National Health Service UK. Heartburn and acid reflux. https://www.nhs.uk/conditions/heartburn-and-acid-reflux/. Last accessed 12-01-2021.

[4] Mark Feldman and Cora Barnett. Relationships between the acidity and osmolality of popular beverages and reported postprandial heartburn. *Gastroenterology*, 108(1):125 – 131, 1995.

[5] Susana Andueza, María A Vila, M Paz de Peña, and Concepción Cid. Influence of coffee/water ratio on the final quality of espresso coffee. *Journal of the Science of Food and Agriculture*, 87(4):586–592, 2007.

[6] L. Navarini and D. Rivetti. Water quality for espresso coffee. *Food Chemistry*, 122(2):424 – 428, 2010. 5th Conference on Water in Food.

[7] Purdue University Chemical Education Division Groups. Solubility. http://chemed.chem.purdue.edu/genchem/topicreview/bp/ch18/soluble.php. Last accessed 10-02-2021.

[8] Ardo Van Den Hout. *Lecture Notes for STAT0029 - Statistical Design of Investigation.* 2020/2021.

[9] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611, 1965.

[10] Nornadiah Mohd Razali and Bee Yap. Power comparisons of some selected normality tests. *Statistics Faculty of Computer and Mathematical Sciences*, pages 126–138, 07 2010.

[11] Morton B. Brown and Alan B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, 69(346):364–367, 1974.

[12] Minitab 17, Statistical Software (2010). State College, PA: Minitab, Inc. www.minitab.com.

Word count : 1991