Individual Assessment for Applied Statistics (Takuya Otsuki)

Q1.

(a)

Scatter plot was created to plot the graph between Logistics Bill and GDP as a scatter plot is effective for the analysis of the relationship between two continuous variables. The result is shown in the figure below:

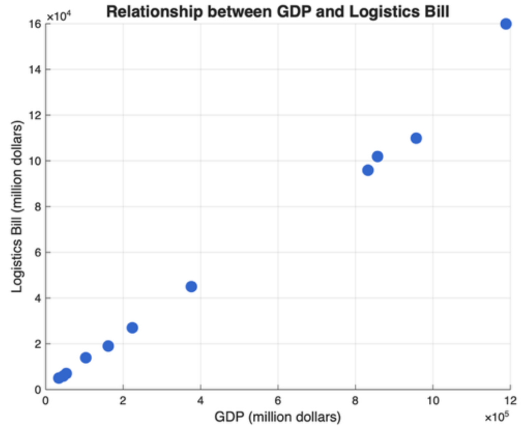

Figure 1. relationship between GDP and Logistic Bill

(b)

To explore the dependency between the Logistics Bill and GDP, linear regression was selected as the first choice of analytical method because from the scatter plot of Figure 1 the plot pattern shows a strong positive linear relationship with only one potential outlier between the two variables. Furthermore, question (b) asks whether the Logistics Bill is dependent on the GDP, and it matches the purpose of the linear regression, which is to figure out the relationship between a dependent variable and one or more independent variables to predict.

The fitlm estimated the regression equation in Formula 1.

$$Logistics\ Bill = b_0 + b_1 * GDP + e = -727.56 + 0.1239 * GDP \qquad (1)$$

, where e represents random error.

Before using this regression equation, it is needed to analyse whether applying the simple linear regression is suitable to solve this question. To evaluate this model of simple linear regression, predictive power (R square), significance of model (P-Value), normally distributed residuals, and whether residuals have constant variances and are independent were checked.

To begin with the result, using the linear regression with the raw data did not meet some of these assumptions. Thus, log transformation was conducted to the raw data to make the distribution more normal. In the following section, the results for each index in the model test are presented for both before and after applying the log transformation.

Firstly, R squares and p-values for significance of the model are shown in Table 1.

Table 1. R square and p-value

	Before log transformation	After log transformation
R-squared value	0.9896 (=98.96%)	0.998 (=99.8%)
Adjusted R-Squared value	0.988 (=98.8%)	0.997 (=99.7%)
p-value	3.17e-10	3.83e-13

The result of R-squared value and adjusted R-squared value were 0.9896, 0.988 for before the log transformation and 0.998, 0.997 respectively, which means around 99% of variances are explained by the formula 1 and the rest of just 1% is unexplainable. So, predictive power of the model is strong enough regardless of the log transformation.

For testing the significance of the model, hypothesis testing is essential. In this case, null hypothesis (H0) is that the line slope is flat (b1 = 0, Logistics Bill = b0) and there is no significance while alternative hypothesis (H1) is that the line slope is not flat (b1 != 0) and there is a decent significance. Since both p-values, 3.17e-10 for before and

3.83e-13 for after, are significantly smaller than 0.05, H0 is rejected, indicating the model has significant regression power regardless of the log transformation.

Normality test for residuals is also needed to meet the assumption of the simple linear regression to ensure that residuals are normally distributed. The results of chi-square test including the descriptive statistics are below:

Table 2. Descriptive statistics residuals

	Before log transformation	After log transformation
Count	11	11
Mean	-2.0186e-17	-2.0186e-17
Std	1.0000	1.0000
Min	-1.4292	-1.3351
25%	-0.5094	-0.5580
50%	0.0028	-0.2717
75%	0.2585	0.4243
Max	2.4268	2.1649

Table 3. Boundaries for outliers

	Before log transformation	After log transformation
Lower bound for outliers	-1.6614	-2.0315
Upper bound for outliers	1.4105	1.8978

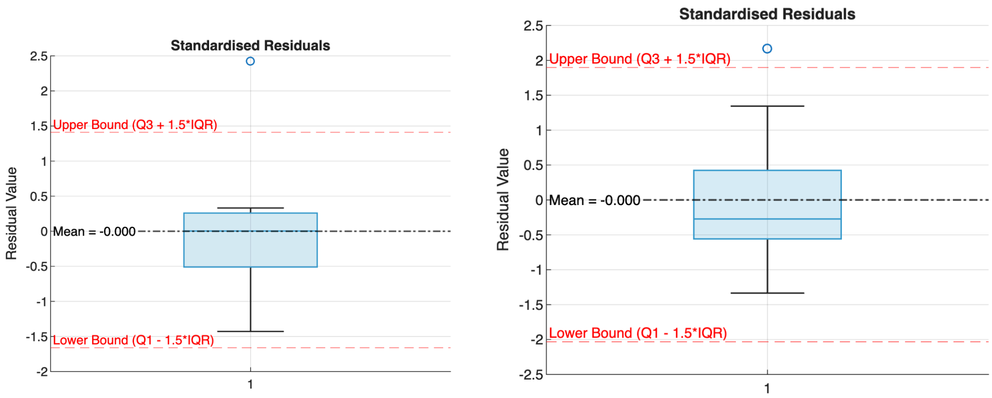

Figure 2. Residuals Boxplot with IQR Bounds (left: before, right: after log transformation)

Table 4. Develop the classes

	Before log transformation	After log transformation
No. of classes	4.46	4.46
Rounded No. of classes	4	4
Class range	0.96	0.87

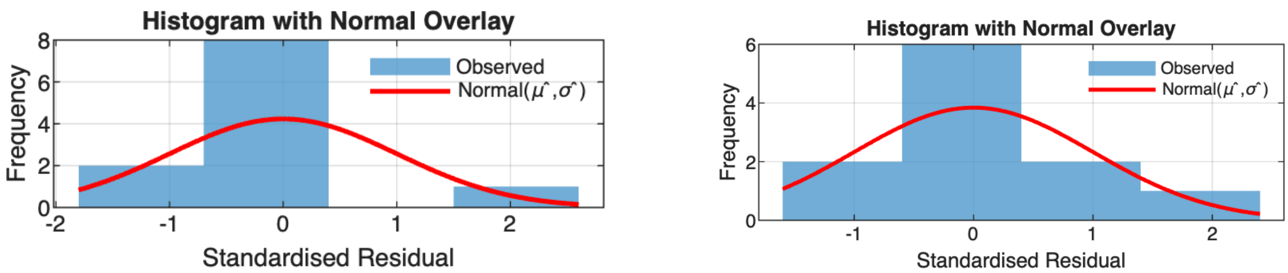

Figure 3. Histogram with normal overlay (left: before, right: after log transformation)

Table 5. Goodness-of-fit test (Normality via Chi-square)

	Before log transformation	After log transformation
Chi-square	5.8533	1.1813
p-value	0.015548	0.277088

In Figure 2 and 3, it is visually clear that the skewness shrank and the distribution after the log transformation shows better alignment with normal distribution (bell-shaped) than one before the log transformation, respectively. Looking at Table 5, there are p-values for both before and after. However, p-value before the log transformation is smaller than 0.05 while one after the transformation is greater than 0.05. In this case, the null hypothesis is that residuals are consistent with normality, and the alternative hypothesis is opposite to it. Therefore, only the residuals after the transformation fail to reject H0 and is normally distributed. Figure 4 support this thing by the sample data being align with the line.

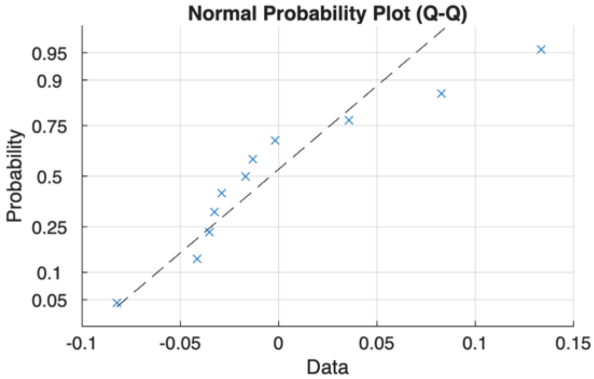

Figure 4. Q-Q plot

Lastly, the relationship between the standardised residuals and fitted values and one between standardised residuals and order for both before and after the transformation are provided below.

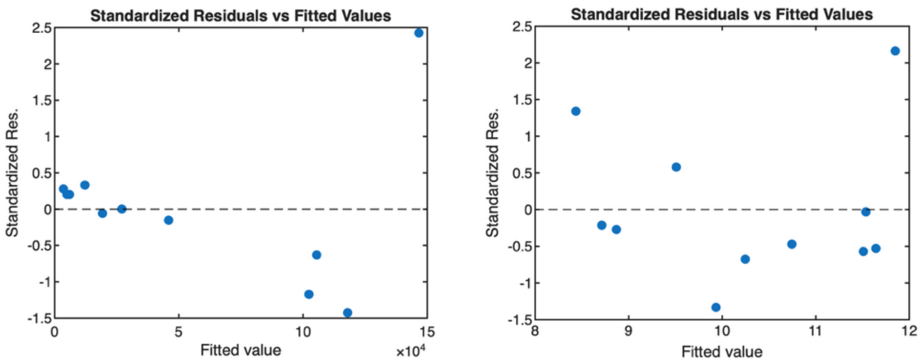

Figure 5. Standardised residuals vs fitted values (left: before, right: after log transformation)

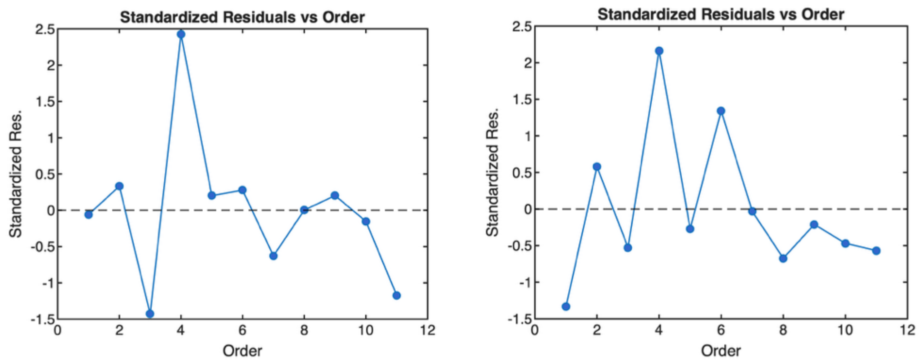

Figure 6. Standardised residuals vs order (left: before, right: after log transformation)

In Figure 5, the scatter distribution after the transformation is more symmetric than before, suggesting the variance is more constant in the residuals after the transformation. Figure 6 reveals that after-log-transformation is more random in the data variance. Then, it can be said that the transformation improved the residuals independence.

As a final note about testing the model, the mean values were both almost zero and standard deviations were also 1, which is a sign of correct standardisation. Furthermore, Germany is an outlier from Figure 2 and 5. But it is not an extreme value because z-score is less than 3.

Having verified that the log-transformed linear regression model satisfied all the assumptions, now it is possible to interpret the relationship between Logistics Bill and GDP confidently. Since the estimated coefficients from the log-transformed data are -1.5792 for the intercept b0 and 0.9600 for the slope b1, the relationship is described by Formula 2 and Figure 7.

$$\log(Logistics\ Bill) = -1.5792 + 0.9600 * \log(GDP) \quad (2)$$

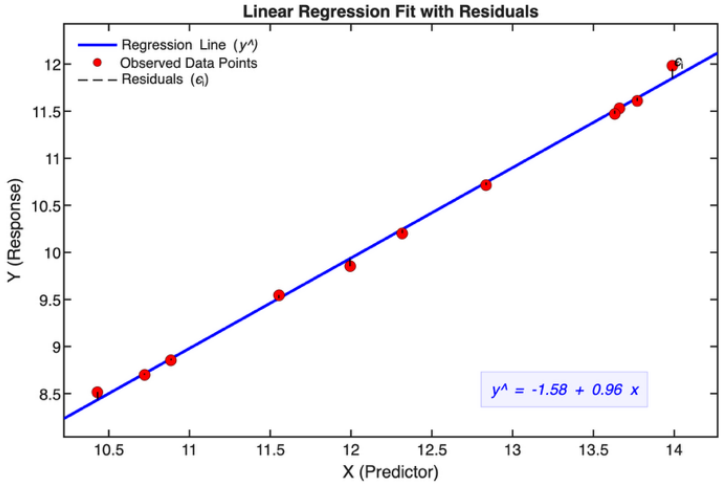

Figure 7. The linear regression model

The 95% confidence interval for the regression coefficient is also below:

Table 6. The 95% confidence interval for the regression coefficient

Lower boundary	0.924824
Upper boundary	0.995088

The slope coefficient b1 =0.96 is very close to 1, with a 95% confidence interval of [0.924824, 0.995088]. Thereby, it can be said that Bill is dependent on the GDP and they are an almost completely proportional relationship.

(c)

To make a prediction of the Logistics Bill for a country with a GDP of $600,000 million, Formula 2 was used, and it is equal to Formula 3.

$$Logistics\ Bill = e^{-1.58} * GDP^{0.96} \approx 0.206 * GDP^{0.96} \quad (3)$$

This model was selected as it was justified in Q1(b). The prediction is shown below:

Table 7. Prediction of the Logistics Bill

	Values (million dollars)
Logistics Bill	72597.12
Lower bound for the 95% prediction interval	62026.07
Upper Bound for the 95% prediction interval	84969.79

Table 7 shows that the predicted Logistics Bill for a country with a GDP of $600,000 million is 72597.12 million dollars. This is roughly 12% of the GDP, which is within the typical range. Additionally, the 95% prediction interval is [$62026.07 million, $84969.79 million], indicating that the actual Logistics Bills are expected to fall into this range with 95% confidence.

(d)

Looking at Figure 5, there is an outlier (Germany) whose z-score is over two and it may strongly affect the regression line and inflate the apparent R square.

Also, an eleven is small for a sample size. It likely causes the unbalanced data distribution, and it may make the model less reliable. In this case, Looking at Figure 1, there is a cluster in the lower-left area, which might cause the poor prediction ability of the model in the middle area. The log transformation may make the distribution look better. However, it also means the data is complexly manipulated and it could violate the reliability of the model.

(e)

The estimated intercept b0 in the log transformed model is -1.579 with a 95% confidential interval of [-2.0152, -1.1433]. As it is shown in Figure 8, the regression line is unlikely to pass the origin. If GDP is equal to zero, the Logistics Bill should also be zero. Furthermore, Logistics Bill cannot be a negative value. However, this phenomenon happens because GDP is not the only factor of the variation in the Logistics Bill while the Logistics Bill is strongly dependent on GDP.

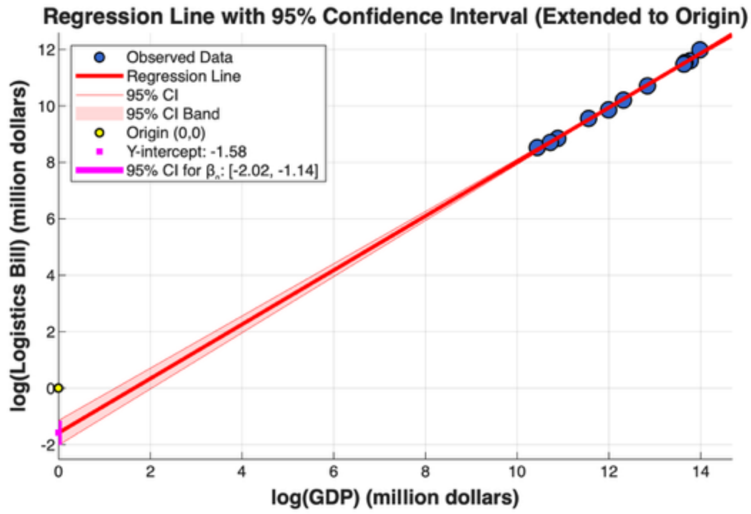

Figure 8. Line with a 95% CI

Q2

(i)

(a)

There are two estimated factors – the line speed and carbonation – that may cause deviation from the target fill level. Therefore, a two-way ANOVA was selected as an analytical method for this question. This method can allow exploration of the effect of each factor and the interaction between the two factors. Since the data was collected five times, the two-way ANOVA was conducted with a replication of five. In order to apply the two-way ANOVA, the five assumptions need to be satisfied: independent observations, balanced design, normality of errors in each cell and equal variance across cells.

It can be said that the first assumption is satisfied as any sample is not used several times.

The second one is also met because each cell has the same number of samples which is five.

For the normality check, Shapiro-Wilk test was selected as a main analytical method as the data is not categorical but continuous. Q-Q plot was also used to help the visual understanding. The results are shown in Table 8 and Figure 9 respectively.

Table 8. The result of Shapiro-Wilk test

Line speed / Carbonation	H	p-value	W
210 bpm / 10 %	0	0.6671	0.9402
240 bpm / 10 %	0	0.0867	0.8036
270 bpm / 10 %	0	0.1508	0.9232
300 bpm / 10 %	0	0.3838	0.9281
210 bpm / 12 %	0	0.4504	0.9088
240 bpm / 12 %	0	0.6201	0.9774
270 bpm / 12 %	0	0.4795	0.9419
300 bpm / 12 %	0	0.5994	0.9290

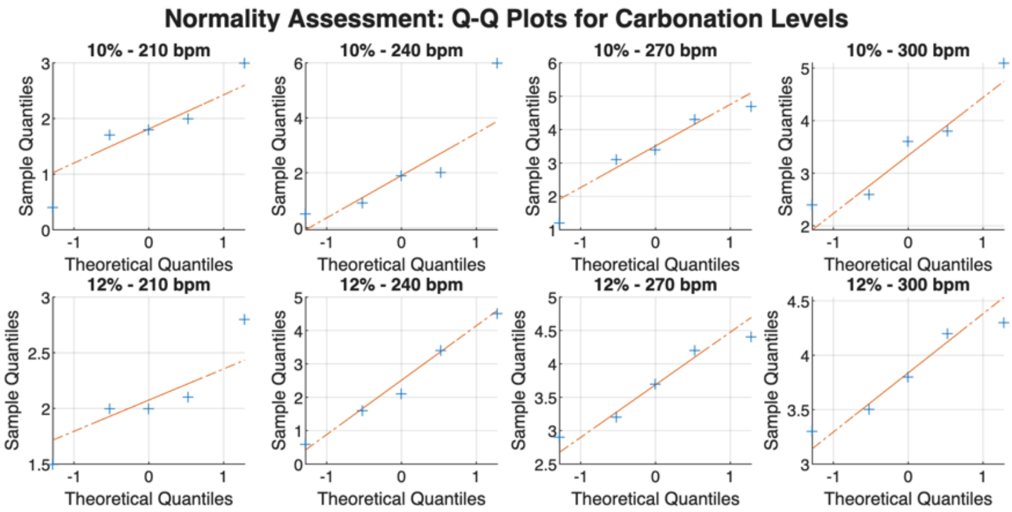

Figure 9. Q-Q plot

In Tables 8, all the H indicate zero, which means the null hypothesis of normality was not rejected in every combination between the line speed and carbonation. This is also proved with W values being close to 1 and p-values being much greater than 0.05. The sample data in Figure 9 also show alignment with each line, indicating normality in each group.

Lastly, Levene's test was conducted to check whether the variances are equal across cells. The result is shown in Figure10.

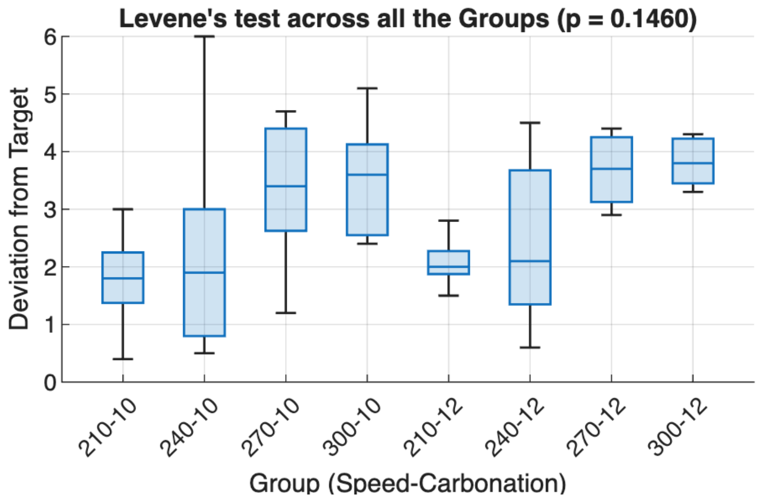

Figure 10. Levene's test

Figure 10 reveals that the variability in the 240 bpm –10% group is considerably large. However, Levene's test yielded p = 0.1460, which is greater than 0.05. This indicates that the assumption of the equal variances was satisfied despite a large variance in one group as the null hypothesis that the variances are equal across all the groups was not rejected.

As it was verified that all the assumptions were satisfied, it is possible to confidently use the two-way ANOVA for this question. The p-values from the two-way ANOVA are shown in Table 9.

Table 9. P-value for each factor

Factor	P-value
Carbonation	0.4639
Line speed	0.0063
Interaction	0.9989

Table 9 illustrates that the line speed has notable significance to the deviation from the target fill level, since the figure of the p-value is smaller than 0.05 and the result rejected the null hypothesis that the line speed has no effect on the deviation in the fill level.

(b)

This question is also solved by using Table 9. The p-value for the interaction factor is shown as 0.9989, which is greater than 0.05 and considerably close to one. This indicates that there is no interaction between the line speed and carbonation because the hypothesis that there is no interaction between the two effects was not rejected.

(ii)

(a)

The Poisson distribution is suitable for this case because all the assumptions were met as:

1. The eight-hour shift is a fixed interval for time
2. Events of breakdown cannot happen at the same time
3. Each breakdown does not affect each other, which mean they are independent
4. The potential number of breakdowns happening is theoretically infinite
5. Breakdowns occur with a constant probability 1.5 per eight-hour shift for all three shifts

(b)

As it was mentioned, the Poisson distribution can be applied for this case. Thus, the probability of two breakdown during the night shift is calculated as:

$$P(X=2) = \frac{1.5^2 * e^{-1.5}}{2!} = 0.2510 \tag{3}$$

and the answer is 25.10%

(c)

The probability of less than two breakdowns during the afternoon shift is equal to the sum of the probability of zero breakdown and that of one breakdown during the shift. Therefore, the answer is 55.78% by the following calculation.

$$P(X<2) = P(X=0) + P(X=1) = \frac{1.5^0 * e^{-1.5}}{0!} + \frac{1.5^2 * e^{-1.5}}{2!} = 0.5578 \tag{4}$$

(d)

Since each shift does not happen at the same time, the probabilities of no breakdown during one shift can be multiplied. The probability of no breakdown during one shift is calculated as:

$$P(X=0\ during\ one\ shift) = \frac{1.5^0 * e^{-1.5}}{0!} = e^{-1.5}$$

Therefore, the probability of no breakdowns during three successive shifts is 1.11% as the following calculation:

$$P(X=0\ thruogh\ three\ shifts) = (e^{-1.5})^3 = 0.0111 \tag{5}$$

Q3

(i)

According to the question, there are two groups, and it is asked to compare the outcome of the methods. Additionally, the term “significantly” implies the application of the hypothesis testing. Therefore, one of the two sample tests comparing the averages of the variables should be selected. In order to check whether two sample tests can be applied, normality check was conducted at the firs stage. For the normality check, Shapiro-Wilk test was conducted for each method, and the result is shown in Table 10.

Table 10. Result of Shapiro-Wilk test

	H	P-value	W
Computer-assisted	0	0.7193	0.91556
Group-based	0	0.5633	0.9448

It can be checked that H is equal to 0 in both groups, meaning that the data in both methods are normally distributed. This is supported by each p-value being greater than a significance of 0.05.

After the normality check, F-test was conducted since the two variables were not in pairs. The outcome of the F-test is below:

Table 11. Result of F-test

F-statistic	0.2813
Degree of freedom for Computer-assisted	11
Degree of freedom for Group-based	11
Two-tailed p-value	0.0461

As p-value is smaller than 0.05, the null hypothesis that the variances are equal was rejected. Thereby, Welsh's t-test was selected instead of two-sample t-test. As the goal of this question is to investigate whether the time of the computer-assisted group is smaller than that of the grouped-based group, the test is one tailed. The null hypothesis of this test is that the time of the computer-assisted group is greater or equal to that of the group-based one while the alternative hypothesis is defined as the time of the computer-assisted one being shorter than that of the group-based one. The result of the Welsh's one tailed t-test below:

Table 12. the result of Welsh's t-test

	Computer-Assisted	Group-Based
Mean	17.85	20.217
Variance	5.4973	19.542
P-value for one-tail	0.06	

The mean value of the computer-assisted was smaller than that of the group-base. However, it can be concluded that there is no evidence that it is significantly true because the p-value is larger than 0.05 and the alternative hypothesis is not rejected. 95% CI for the difference in mean values (computer-assisted – grouped-based) was [-5.418, 0685] seconds.

(ii)

(a)

The specification was 44.0 ± 2.5 N. The data calculated from the question sentence is shown in Table 13.

Table 13. Calculated data

Sample size (n)	100
Sample mean (x)	44.175 N
Sample variance (s^2)	1.017
Sample standard deviation (s)	1.008 N

To calculate the cumulative probability at the upper and lower boundary of the specification range, the two boundary has to be transformed to z-score. The z-score for the lower boundary was -2.6525 and upper was 2.3055. After that, since the distribution is considered being normal, the cumulative probability is calculated as:

$$P(outside\ spec) = 1 - P(within\ spec) = 1 - (P(X < 46.5) - P(X < 41.5)) = 0.014564 \qquad (6)$$

Thus, approximately 1.46% of the springs would lie outside the specification.

(b)

Since data is normally distributed and measurements are independent, one sample t-test can be applied. the hypothesis is defined as

Ho: population mean is equal to 44.0

H1: population mean is not equal to 44.0

The significance level is set to 0.05 in this question. The outcome of the one sample t-test is shown in Table 14 and Figure11.

Table 14. Outcomes of one sample t-test

Degree of freedom	99
p-value	0.0858
t-statistics	1.7353
t-critical	1.984

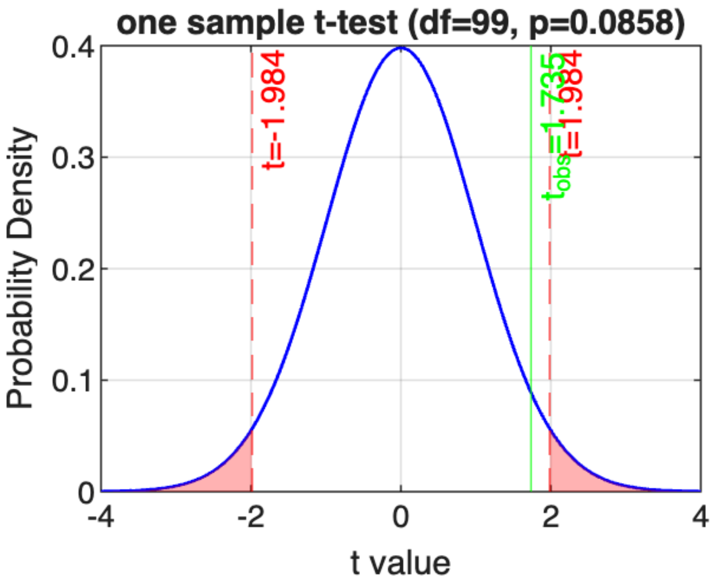

Figure 11. Relationship between t-statics and t-criticals

The p-value is equal to 0.0858, which is greater than 0.05. Therefore, the null hypothesis was not rejected, and the population mean is likely to be on target. This can be explained visually in Figure 11 as well. T-statistics is not in the shaded area, indicates the null hypothesis not being rejected.

(c)

This question is about the population mean as the part (b) was. Thus, the same data and parameters from part (b) can be used as well as from part (a). The confidential interval is calculated by using sample values and margin of error. The margin of error and 95% confidential interval can be calculated as:

$$Margin\ of\ Error = t_{critical} \times \frac{s}{\sqrt{n}} = 1.984 \times \frac{1.008}{\sqrt{100}} = 0.200 \quad (7)$$

$$95\%\ CI = x \pm Margin\ of\ Error = [43.975, 44.375] \quad (8)$$

Therefore, the population mean with the 95 confidential interval is [43.975, 44.375] N. Since the target value is 44 N, this interval is significantly close to the target, which is good alignment with part (b).

Appendix

Q1

(A)

```
scatter(GDP, Logistics_Bill, 100, 'filled', 'MarkerFaceColor', [0.2 0.4 0.8]);
```

(B)

```
X = log(GDP);
Y = log(Logistics_Bill);

% Fit linear model using fitlm
mdl = fitlm(X, Y);

% display the result
disp('========== result of the regression analysis ==========');
disp(mdl);

% Extract stats
a = mdl.Coefficients.Estimate(1); % intercept
b = mdl.Coefficients.Estimate(2); % slope
Y_pred = predict(mdl, X);
R2 = mdl.Rsquared.Ordinary;
RMSE = mdl.RMSE;
```

```
% 95% CI
CI_b = coefCI(mdl, 0.05); % alpha = 0.05 (95%CI)
CI_lower = CI_b(2,1);
CI_upper = CI_b(2,2);

% pvalue
p_value = mdl.Coefficients.pValue(2);
```

(C)

```
% Target GDP value (in million dollars)
X_new = log(600000); % log scale

% Point prediction (on log scale)
Y_pred = predict(mdl, X_new);

% Calculate 95% prediction interval (on log scale)
n = length(X);
x_bar = mean(X);
s_x = std(X);
alpha = 0.05;
df = n - 2;
t_value = tinv(1 - alpha/2, df);

% Residual standard error (RSE)
residuals = mdl.Residuals.Raw;
RSE = sqrt(sum(residuals.^2) / df);

% Standard error of prediction (on log scale)
SE_pred = RSE * sqrt(1 + (1/n) + ((X_new - x_bar)^2) / ((s_x^2) * (n - 1)));

% Prediction interval (on log scale)
PI_lower = Y_pred - t_value * SE_pred;
PI_upper = Y_pred + t_value * SE_pred;

% Convert all values back to the original scale
Y_pred_nonlog = exp(Y_pred);
PI_lower_nonlog = exp(PI_lower);
PI_upper_nonlog = exp(PI_upper);
```

(E)

```
% 95% confidence interval for intercept
CI_a = CI_b(1,:); % 95% CI for intercept
p_value_intercept = mdl.Coefficients.pValue(1);

fprintf('Intercept (β₀) = %.4f (in log scale)\n', a);
fprintf('95%% Confidence Interval: [%.4f, %.4f]\n', CI_a(1), CI_a(2));
fprintf('p-value: %.4f\n', p_value_intercept);

% Statistical conclusion
if CI_a(1) <= 0 && CI_a(2) >= 0
    fprintf('\nConclusion: The 95%% CI includes zero.\n');
    fprintf('→ We cannot reject H₀: β₀ = 0 (p = %.4f > 0.05)\n', p_value_intercept);
    fprintf('→ The data are consistent with the line passing through the origin.\n');
else
    fprintf('\nConclusion: The 95%% CI does NOT include zero.\n');
    fprintf('→ We reject H₀: β₀ = 0 (p = %.4f < 0.05)\n', p_value_intercept);
    fprintf('→ The line does not pass through the origin.\n');
end
```

Q2

(i)

(A)

```
% --- Shapiro-Wilk ---
alpha = 0.05;
[H210_10, pValue210_10, W210_10] = swtest(speed210_10, alpha);
[H240_10, pValue240_10, W240_10] = swtest(speed240_10, alpha);
[H270_10, pValue270_10, W270_10] = swtest(speed270_10, alpha);
[H300_10, pValue300_10, W300_10] = swtest(speed300_10, alpha);
[H210_12, pValue210_12, W210_12] = swtest(speed210_12, alpha);
[H240_12, pValue240_12, W240_12] = swtest(speed240_12, alpha);
[H270_12, pValue270_12, W270_12] = swtest(speed270_12, alpha);
[H300_12, pValue300_12, W300_12] = swtest(speed300_12, alpha);
```

(B)

```
data10 = [speed210_10'; speed240_10'; speed270_10'; speed300_10'];
group10 = [ones(n210_10,1); 2*ones(n240_10,1); 3*ones(n270_10,1); 4*ones(n300_10,1)];

figure('Name','Equal Variance Check (10%)','Color','w');
boxchart(categorical(group10), data10);
xlabel('Line Speed (Group)');
ylabel('Measurement Value');
title('Levene Test: 10% Carbonation','FontSize',13);

[p10, tbl10] = vartestn(data10, group10, 'TestType','LeveneAbsolute','Display','off');
```

```
Y = reshape(permute(Fill, [3 1 2]), [], 2);

% Step 4. Perform Two-Way ANOVA with replication
[p, tbl, stats] = anova2(Y, 5, 'off');  % 5 replicates per cell
```

Q3

(i)

```
%% === CAL ===
chi_contrib_cal = (Frequency_cal - expected_cal').^2 ./ expected_cal';

Chi2Table_cal = table(LengthBins_cal, Frequency_cal, bin_prob_cal', expected_cal', chi_contrib_cal, ...
    'VariableNames', {'BinCenter','Observed','BinProb','Expected','Chi2Contribution'});
disp('=== Chi-square Table: CAL ===');
disp(Chi2Table_cal);

Chi2_Stat_cal = sum(chi_contrib_cal);        % χ² statistic
df_cal = k_cal - 1 - 2;                      % Subtract 2 estimated parameters (μ, σ)
p_val_cal = 1 - chi2cdf(Chi2_Stat_cal, df_cal);  % Right-tail p-value

fprintf('\n[CAL] Chi-square Statistic = %.3f\nDegrees of Freedom = %d\np-value = %.3f\n', ...
    Chi2_Stat_cal, df_cal, p_val_cal);


%% === GBL ===
chi_contrib_gbl = (Frequency_gbl - expected_gbl').^2 ./ expected_gbl';

Chi2Table_gbl = table(LengthBins_gbl, Frequency_gbl, bin_prob_gbl', expected_gbl', chi_contrib_gbl, ...
    'VariableNames', {'BinCenter','Observed','BinProb','Expected','Chi2Contribution'});
disp('=== Chi-square Table: GBL ===');
disp(Chi2Table_gbl);

Chi2_Stat_gbl = sum(chi_contrib_gbl);
df_gbl = k_gbl - 1 - 2;
p_val_gbl = 1 - chi2cdf(Chi2_Stat_gbl, df_gbl);

fprintf('\n[GBL] Chi-square Statistic = %.3f\nDegrees of Freedom = %d\np-value = %.3f\n', ...
    Chi2_Stat_gbl, df_gbl, p_val_gbl);

alpha = 0.05;
```

```
%% === F-test ===
% degree of freedom
df1 = n_cal - 1;   % for CAL
df2 = n_gbl - 1;   % for GBL

F_stat = VarVal_cal / VarVal_gbl;     % ratio of variances (Male / Female)

% One-tail p-value: probability that an F random variable ≤ F_obs
p_one = fcdf(F_stat, df1, df2);

% Two-tailed p-value = 2 × smaller tail
p_two = 2 * min(p_one, 1 - p_one);

% Critical F values (for α = 0.05, two-tailed test → 0.025 each tail)
Fcrit_low2 = finv(0.025, df1, df2);    % lower critical
Fcrit_hi2  = finv(0.975, df1, df2);    % upper critical

fprintf('\n--- F-TEST RESULTS ---\n');
fprintf('F-statistic = %.4f\n', F_stat);
fprintf('df_M = %d, df_F = %d\n', df1, df2);
fprintf('One-tailed p (P(F<=f)) = %.4f\n', p_one);
fprintf('Two-tailed p = %.4f\n', p_two);
fprintf('F-critical (lower 2.5%%) = %.4f | (upper 97.5%%) = %.4f\n', Fcrit_low2, Fcrit_hi2);

% Explanation:
% - fcdf(x,df1,df2) gives the probability that F ≤ x
% - finv(p,df1,df2) returns the value x where P(F ≤ x) = p
% - p_two doubles the smaller tail to make it two-sided


% 95% Confidence Interval for Variance Ratio (σ₁²/σ₂²) ---
CI_var_low  = F_stat / Fcrit_hi2;
CI_var_high = F_stat / Fcrit_low2;
```

```
% Convert to CI for Standard Deviation Ratio (σ₁/σ₂) ---
CI_std_low  = sqrt(CI_var_low);
CI_std_high = sqrt(CI_var_high);
```

```
%% === Welch two-sample t test (CAL vs GBL) ===
% H0: μ_CAL ≥ μ_GBL   vs   H1: μ_CAL < μ_GBL  (left-tailed)

% map to original "Excel-style" symbols while reusing the stats above
meanA = MeanVal_cal;   varA = VarVal_cal;   nA = n_cal;   % CAL
meanB = MeanVal_gbl;   varB = VarVal_gbl;   nB = n_gbl;   % GBL

SE_welch   = sqrt(varA/nA + varB/nB);          % separate-variances SE
diff_means = meanA - meanB;                    % (μ_CAL - μ_GBL)
t_stat     = diff_means / SE_welch;            % test statistic

% Satterthwaite approximation for degrees of freedom
num = (varA/nA + varB/nB)^2;
den = (varA^2 / (nA^2*(nA - 1))) + (varB^2 / (nB^2*(nB - 1)));
df  = num / den;

% p-values
p_two_t   = 2 * (1 - tcdf(abs(t_stat), df));  % two-tailed p
p_one_lt  = tcdf(t_stat, df);                 % left-tailed p for H1: diff < 0

% critical t-values for α = 0.05
tcrit_two = tinv(1 - 0.025, df);              % two-tail
tcrit_one_left = tinv(0.05, df);              % left-tail (negative)

% 95% CI for (μ_CAL - μ_GBL)
CI_low  = diff_means - tcrit_two * SE_welch;
CI_high = diff_means + tcrit_two * SE_welch;
```

(ii)

(A)

```
% Calculate z-scores
z_lower = (lower_spec - mu) / sigma;
z_upper = (upper_spec - mu) / sigma;
```

```
% Probability within specification
p_below_lower = normcdf(z_lower);
p_below_upper = normcdf(z_upper);
p_within = p_below_upper - p_below_lower;
p_outside = 1 - p_within;
```

(B)

```
% Standard error
SE = sample_std / sqrt(n);

% t-statistic
t_stat = (sample_mean - spec_mean) / SE;

% Degrees of freedom
df = n - 1;
```

```
% Critical value (two-tailed, α = 0.05)
alpha = 0.05;
t_critical = tinv(1 - alpha/2, df);

% p-value (two-tailed)
p_value = 2 * (1 - tcdf(abs(t_stat), df));
```

(C)

```
% Margin of error
margin_error = t_critical * SE;

% Confidence interval
CI_lower = sample_mean - margin_error;
CI_upper = sample_mean + margin_error;
```