# STAT2203/7203: Assignment 3

Alexander White

Student ID: 43218307

## Question 1: Rejection Sampling and Benford's Law

### (a) Verify that $\frac{f_D(x)}{\log_{10}(2)} \leq 1$ for all $x \in \{1, 2, \ldots, 9\}$

**Solution:**
For x = 1:

$$f_D(x) = log_{10}(\frac{x+1}{x})$$

$$\frac{f_D(1)}{\log_{10}(2)} = \frac{log_{10}(\frac{2}{1})}{\log_{10}(2)} = 1$$

For x = 2:

$$\frac{f_D(2)}{\log_{10}(2)} = \frac{log_{10}(\frac{3}{2})}{\log_{10}(2)} \approx 0.585$$

We know $0 < a \leq b$, then $log_{10}(a) \leq log_{10}(b)$.
Therefore, as the numerator is getting smaller as x gets larger towards 9, $\frac{f_D(x)}{\log_{10}(2)} \leq 1$ for all $x \in \{1, 2, \ldots, 9\}$

### (b) Joint probability mass function of $(X, Y)$

**Solution:**
pmf of X is given by: $f_X(x) = \frac{1}{9}, for x \in 1, 2, \ldots, 9$
$Y \mid X = x \sim Bernoulli(\frac{f_D(x)}{\log_{10}(2)})$

Joint pmf of (X,Y) is given by $P(X = x, Y = y) = P(X = x) \cdot P(Y = y \mid X = x)$

For Y = 1:
$P(X = x, Y = 1) = \frac{1}{9} \cdot \frac{f_D(x)}{\log_{10}(2)})$

For Y = 0:
$P(X = x, Y = 0) = \frac{1}{9} \cdot \left(1 - \frac{f_D(x)}{\log_{10}(2)})\right)$

### (c) Probability $P(Y = 1)$

**Solution:**

$$P(Y = 1) = \sum_{x=1}^{9} P(X = x, Y = 1)$$

Substituting the expression for $P(X = x, Y = 1)$ from part 1.b:

$$P(Y = 1) = \sum_{x=1}^{9} \frac{1}{9} \cdot \frac{f_D(x)}{\log_{10}(2)}$$

$$= \frac{1}{9 \cdot \log_{10}(2)} \sum_{x=1}^{9} f_D(x)$$

Since $f_D(x)$ is the probability mass function of Benford's Law, it satisfies the property that the sum of probabilities over all possible values equals 1:

$$\sum_{x=1}^{9} f_D(x) = 1$$

Therefore:

$$P(Y = 1) = \frac{1}{\log_{10}(2)}$$

### (d) Conditional probability mass function of $X$ given $Y = 1$

**Solution:**
Can use Bayes' theorem:

$$P(X = x|Y = 1) = \frac{P(X = x, Y = 1)}{P(Y = 1)}$$

Can substitute the values from part 1.b and 1.c:

$$P(X = x|Y = 1) = \frac{\frac{1}{9} \cdot \frac{f_D(x)}{\log_{10}(2)}}{\frac{1}{\log_{10}(2)}} = \frac{f_D(x)}{9}$$

Therefore, the conditional probability mass function is:

$$P(X = x|Y = 1) = f_D(x)$$

### (e) R Code for simulating a random variable from Benford's law

**Solution:**
R Code:

```r
benford_func <- function() {
  Y <- 0
  while (Y == 0) {
    X <- sample(1:9, 1)  # Simulate X
    Y <- rbinom(1, 1, log10((X + 1) / X) / log10(2))  # Simulate Y
  }
  return(X)
}
```

### (f) Distribution of the number of pairs of $(X, Y)$

**Solution:**
The number of times we need to simulate the pairs relies on the process of the independent Bernoulli trials with success probability of P(Y = 1)
Each pair is independent of any previous simulations.
This follows a geometric distribution. The probability mass function of a geometric distribution is:

$$P(T = k) = (1 - p)^{k-1} p$$

where $p = P(Y = 1)$ and $T$ is the number of trials needed to get the first success. Therefore, the number of pairs follows:

$$T \sim \text{Geometric}\left(\frac{1}{\log_{10}(2)}\right)$$

# Question 2: Confidence Intervals and Hypothesis Tests

## (a) 95% confidence interval of the population mean CAPS-IV score at commencement of the study

**Solution:**
The formula for the confidence interval for the population mean is:

$$\text{CI} = \bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Substituting the values:

$$\bar{x} = 81.35, \quad s = 17.54, \quad n = 48, \quad z_{\alpha/2} = 1.96 \text{ (for 95\%)}$$

$$\text{CI} = 81.35 \pm 1.96 \cdot \frac{17.54}{\sqrt{48}} = 81.35 \pm 1.96 \cdot 2.53 = 81.35 \pm 4.96$$

Therefore, the 95% confidence interval is:

$$\text{CI} = [76.39, 86.31]$$

## (b) Does the data provide evidence that MDMA treatment is associated with decrease in mean CAPS-IV score?

**Solution:**

- Null hypothesis $H_0$: $\mu = 0$ (no decrease in mean CAPS-IV score).
- Alternative hypothesis $H_a$: $\mu > 0$ (there is a decrease in mean CAPS-IV score).

The test statistic is calculated as:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{24.2 - 0}{\frac{23.1}{\sqrt{24}}} = \frac{24.2}{4.71} \approx 5.14$$

- Degrees of freedom: $df = n - 1 = 23$.
- Using $\alpha = 0.05$ for moderate to strong evidence.

Therefore, the critical value for a one-sided t-test at $\alpha = 0.05$ with 23 degrees of freedom is approximately 1.714.

Since the t-statistic ($t = 5.14$) is greater than the critical value, the null hypothesis is rejected. We can conclude that the high dose MDMA treatment is associated with a significant decrease in mean CAPS-IV score.

## (c) Hypothesis test for greater decrease in CAPS-IV score (High vs Low dose)

**Solution:**
Can use a two-sample t-test to test if the decrease in CAPS-IV score is greater for the high dose group or low dose group.

- Null hypothesis $H_0$: $\mu_{\text{high}} = \mu_{\text{low}}$ (no difference in decrease between groups).
- Alternative hypothesis $H_a$: $\mu_{\text{high}} > \mu_{\text{low}}$ (greater decrease in the high dose group).

The test statistic is:

$$t = \frac{\bar{x}_{\text{high}} - \bar{x}_{\text{low}}}{\sqrt{\frac{s_{\text{high}}^2}{n_{\text{high}}} + \frac{s_{\text{low}}^2}{n_{\text{low}}}}} = \frac{24.2 - 12.7}{\sqrt{\frac{23.1^2}{24} + \frac{19.4^2}{24}}} = \frac{11.5}{\sqrt{37.92}} \approx 1.87$$

- Degrees of freedom: $df = 2(24 - 1) = 46$.
- Using $\alpha = 0.05$ for moderate to strong evidence.

Therefore, the critical value for a one-sided t-test at $\alpha = 0.05$ with 46 degrees of freedom is $\approx 1.68$.

Since $t = 1.87$ is greater than the critical value, we reject the null hypothesis and conclude that the high dose MDMA treatment results in a greater reduction in the CAPS-IV score compared to the low dose treatment.

### (d) 95% confidence interval for population proportion (High dose, 20% drop)

**Solution:**
We can use the formula CI formula:

$$\text{CI} = \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Using the values:

$$\hat{p} = \frac{11}{24} = 0.458, \quad n = 24, \quad z_{\alpha/2} = 1.96$$

Therefore:

$$\text{CI} = 0.458 \pm 1.96 \cdot \sqrt{\frac{0.458(1-0.458)}{24}} = 0.458 \pm 1.96 \cdot 0.102 = 0.458 \pm 0.2$$

Thus, the 95% confidence interval for the population proportion is:

$$\text{CI} = [0.258, 0.658]$$

### (e) Hypothesis test for difference in population proportions (High vs Low dose, 20% drop)

**Solution:**

- Null hypothesis $H_0$: $p_{\text{high}} = p_{\text{low}}$.
- Alternative hypothesis $H_a$: $p_{\text{high}} > p_{\text{low}}$.

The test statistic for comparing two proportions is:

$$z = \frac{\hat{p}_{\text{high}} - \hat{p}_{\text{low}}}{\sqrt{\hat{p}_{\text{pool}}(1 - \hat{p}_{\text{pool}}) \left(\frac{1}{n_{\text{high}}} + \frac{1}{n_{\text{low}}}\right)}}$$

The pooled sample proportion $\hat{p}_{\text{pool}}$ is:

$$\hat{p}_{\text{high}} = \frac{11}{24} = 0.458$$

$$\hat{p}_{\text{low}} = \frac{6}{24} = 0.25$$

$$\hat{p}_{\text{pool}} = \frac{11+6}{24+24} = \frac{17}{48} \approx 0.354$$

Sub in values:

$$z = \frac{0.458 - 0.25}{\sqrt{0.354(1-0.354)\left(\frac{1}{24}+\frac{1}{24}\right)}} = \frac{0.208}{0.138} \approx 1.51$$

- Degrees of freedom: $df = 2(24-1) = 46$.
- Using $\alpha = 0.05$ for moderate to strong evidence.

The critical value for a one-sided z-test at $\alpha = 0.05$ is $z_\alpha = 1.645$, for a standard normal distribution.
As $z = 1.51$ is less than the critical value, we do not reject the null hypothesis.

Therefore, there is not enough evidence to conclude that the proportion of patients with a 20% or more drop in CAPS-IV score is greater in the high dose group than in the low dose group.

### (f) Validity of assumptions/approximations in part (e)

**Solution:**
Assumptions for the z-test for proportions used in part 2.e:

- **The data follows a binomial distribution.** Each patient either experienced a 20% or more drop in CAPS-IV score (success) or did not (failure). Therefore, the data is appropriately modeled by a binomial distribution.
- **Large enough sample size for the z-test to be valid.** The z-test for proportions assumes that the sample sizes in each group are large enough for the normal approximation to be valid.
  - In the high dose group: 11 successes and 13 failures.
  - In the low dose group: 6 successes and 18 failures.

  Since both the number of successes and failures in each group are greater than 5, the sample sizes are large enough, and the normal approximation is valid.
- **Independent samples.** The two groups are independent of each other, and each patient's outcome is independent of the others.

Given these points, the assumptions for the z-test in part (e) are valid, and the conclusions drawn from the test are reliable.

## Question 3: Linear Regression on FEV and FVC Data

### (a) Linear regression and diagnostic plots

**Solution:**
R Code:

```r
ozone <- read.csv('ozone.csv')
head(ozone)

lm_ozone <- lm(FEV ~ FVC, data = ozone)
summary(lm_ozone)

plot(lm_ozone)
```

Output:

```
Call:
lm(formula = FEV ~ FVC, data = ozone)

Residuals:
    Min      1Q  Median      3Q     Max
-17.701  -7.880  -1.301   8.132  20.558

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.7526     1.4309  -1.225    0.226
FVC           0.8392     0.1146   7.324 8.37e-10 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1

Residual standard error: 10.32 on 58 degrees of freedom
Multiple R-squared:  0.4805,    Adjusted R-squared:  0.4715
F-statistic: 53.64 on 1 and 58 DF,  p-value: 8.374e-10
```

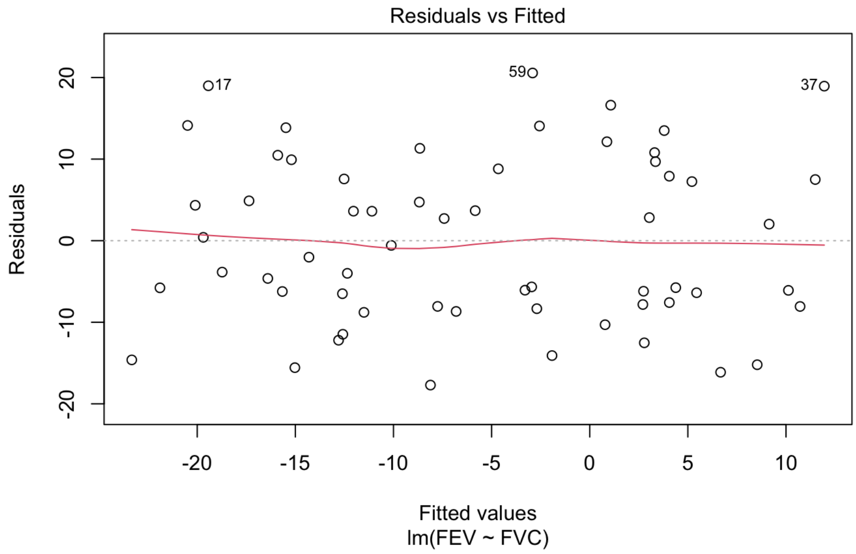


Figure 1: Linear Model Residuals vs Fitted Values

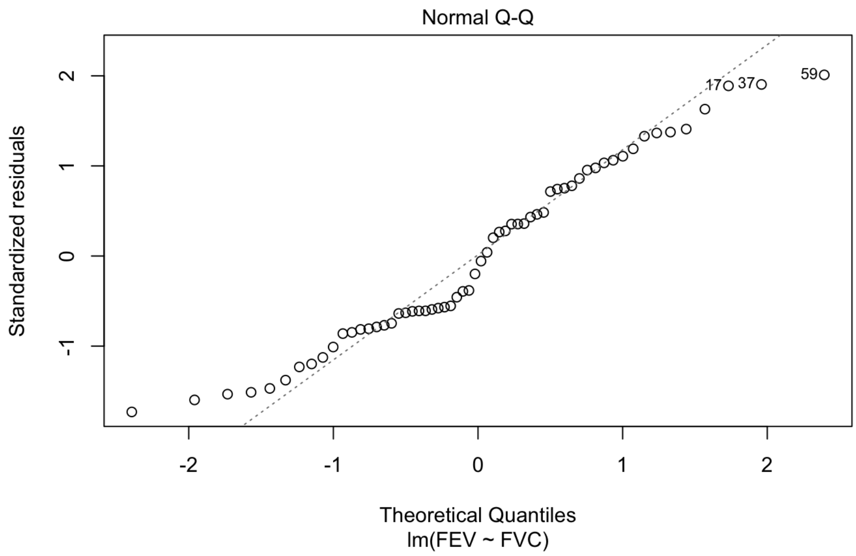


Figure 2: Normal QQ Plot of Residuals

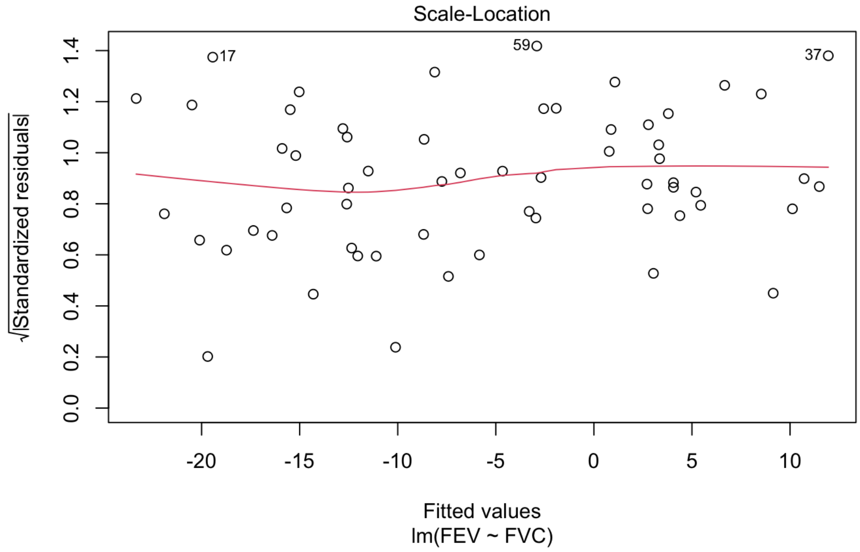


Figure 3: Scale-Location Plot

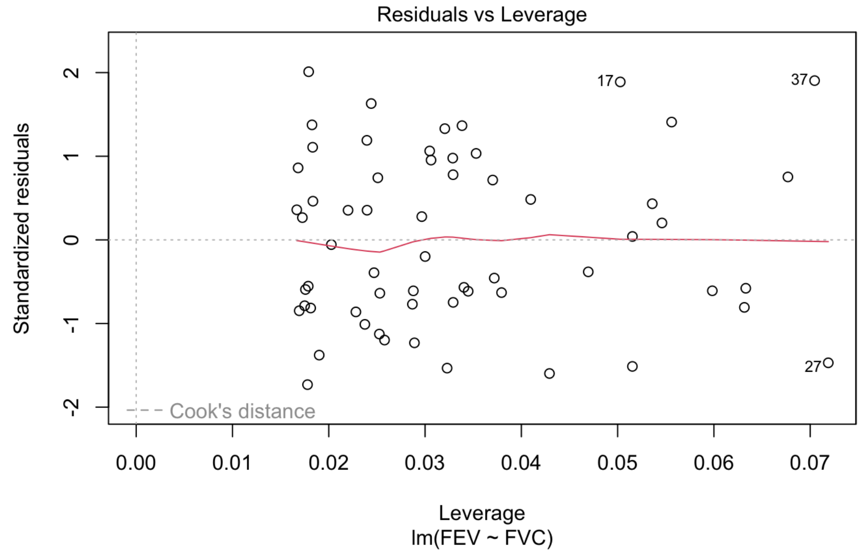


Figure 4: Residuals v Leverage Plot

## (b) Assumptions of the Linear Regression Model

**Solution:**

Assumptions of the Linear Regression model:

- **Normality**: The residuals should be approximately normally distributed.
- **Independence**: The residuals should be independent of each other.
- **Homoscedasticity**: The residuals should have constant variance.
- **Linearity**: There should be a linear relationship between FVC and FEV.

**Checking the assumptions:**

- **Normality**: As seen in the Normal Q-Q plot (Figure 2), the residuals should follow the diagonal line if they are normally distributed. There is some evidence that this assumption is violated, as the points deviate from the line, especially in the middle and at the start and end. This suggests that the residuals may not be perfectly normal, and there may be some outliers affecting the shape of the distribution.
- **Independence**: The assumption of independence is generally based on the fact that there was no clear indication that the FVC or FEV measurements were dependent on each other. The study doesn't mention repeated measures or time-dependent data, thus we can we assume the independence assumption is valid.
- **Homoscedasticity**: Using the Residuals vs. Fitted plot (Figure 1), we can check for homoscedasticity. Since there is no clear pattern or shape of the plot, and the residuals are spread relatively evenly, the assumption of homoscedasticity is valid.
- **Linearity**: Figure 1 also helps to assess linearity. Since there is no clear non-linear pattern in the residuals, we can assume that the linearity assumption is valid, meaning that the relationship between FEV and FVC can be modeled with a linear regression.

## (c) Normal distribution of FVC%

**Solution:**

The linear regression model of FEV  FVC does not require the independent variable (FVC%) to be normally distributed, but rather we assume that the residuals will be normally distributed. The validity of the model relies on this assumption being valid.
The residuals of the model can be checked using the Normal Q-Q plot (Figure 2). If they follow the plotted line, then we can say that the residuals are normally distributed.

## (d) 99% confidence interval for the slope

**Solution:**

Using the linear model represented as $y = \hat{\beta}_0 + \hat{\beta}_1 x$, where $\hat{\beta}_1$ is the slope. The 99% confidence interval for the slope $\hat{\beta}_1$ of the regression model is calculated as:

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} \cdot \mathrm{SE}(\hat{\beta}_1)$$

From the regression output:

- $\hat{\beta}_1 = 0.8392$ (estimated slope)
- $\mathrm{SE}(\hat{\beta}_1) = 0.1146$
- For a 99% confidence interval with $df = 58$, $t_{0.005,58} \approx 2.663$

Subbing these values into the formula:

$$0.8392 \pm 2.663 \times 0.1146$$

Therefore, the 99% confidence interval for the slope $\hat{\beta}_1$ of the linear model is:

$$(0.5342, 1.1442)$$

### (e) Prediction and confidence intervals for FEV%

**Solution:**
The formula for the prediction interval is given by:

$$\hat{y}_0 \pm t_{\alpha/2,n-2} \cdot \sqrt{\text{MSE}\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}$$

Using the following values from the regression output:

- $\hat{\beta}_0 = -1.7526$ (intercept),
- $\hat{\beta}_1 = 0.8392$ (slope),
- $\text{MSE} = 10.32^2 = 106.6624$,
- $n = 60$ (sample size),
- $\bar{x} = -4.558$ (mean of FVC values),
- $x_0 = 10$ (FVC = 10),
- $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 8109.645$,
- $t_{0.05/2,58} \approx 1.672$ (critical value for 90% prediction interval).

The predicted FEV change for $x_0 = 10$ is:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 = -1.7526 + 0.8392 \times 10 = 6.6394$$

The standard error for the prediction interval is calculated as:

$$\text{SE}_{\text{pred}} = \sqrt{\text{MSE}\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}$$

Subbing in the known values:

$$\text{SE}_{\text{pred}} = \sqrt{106.5024\left(1 + \frac{1}{60} + \frac{(10 - (-4.557956))^2}{8109.645}\right)}$$

$$\approx 10.538$$

The 90% prediction interval is:

$$\hat{y}_0 \pm t_{\alpha/2,n-2} \cdot \text{SE}_{\text{pred}} = 6.6394 \pm 1.672 \times 10.538$$

This gives:

$$6.6394 \pm 17.62$$

Thus, the prediction interval is:

$$(-10.98, 24.26)$$

The formula for the confidence interval is given by:

$$\hat{y}_0 \pm t_{\alpha/2,n-2} \cdot \sqrt{\text{MSE}\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)}$$

Subbing in the known values:

$$\mathrm{SE}_{\mathrm{pred}} = \sqrt{106.5024\left(\frac{1}{60} + \frac{(10-(-4.557956))^2}{8109.645}\right)}$$

$$\approx 2.135$$

The 90% confidence interval is:

$$\hat{y}_0 \pm t_{\alpha/2,n-2} \cdot \mathrm{SE}_{\mathrm{pred}} = 6.6394 \pm 1.672 \times 2.135$$

This gives:

$$6.6394 \pm 3.57$$

Thus, the confidence interval is:

$$(3.07, 10.21)$$

### (f) Hypothesis test for zero intercept

**Solution:**
The hypotheses for the intercept are:

- Null hypothesis $H_0$: $\beta_0 = 0$ (the intercept is zero).
- Alternative hypothesis $H_a$: $\beta_0 \neq 0$ (the intercept is not zero).

The t-statistic for the intercept is:

$$t = \frac{\hat{\beta}_0}{\mathrm{SE}(\hat{\beta}_0)}$$

From the regression output:

- $\hat{\beta}_0 = -1.7526$
- $\mathrm{SE}(\hat{\beta}_0) = 1.4309$

Therefore, the test statistic is:

$$t = \frac{-1.7526}{1.4309} \approx -1.225$$

Using a significance level of $\alpha = 0.05$ and degrees of freedom $n - 2 = 58$, the critical value for a two-sided test is $t_{0.025,58} \approx 2.002$.
Since $|t| = 1.225$ ¡ the critical value (2.002), we do not reject the null hypothesis. Therefore, there is not significant evidence to conclude that the intercept is different from zero.

### (g) Interpretation of R-squared value

**Solution:**
The $R^2$ value from the regression output is 0.4805. This means that approximately 48.05% of the variability in the FEV change can be explained by the change in FVC.
Therefore, the linear model explains 48.05% of the total variation in FEV. This indicates a moderate fit, but not a particularly strong model. The remaining 51.95% of the variation is due to other factors not captured by the model.

### (a) Deriving Var($\hat{\beta}_1$)

**Solution:**
We need to show that:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

The least squares estimator for the slope $\hat{\beta}_1$ is given by:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

We know the linear regression model is represented as:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Therefore, we can sub this into the equation for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta_0 + \beta_1 x_i + \epsilon_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Cancelling out the $\beta_0$ and $\beta_1$ terms, after centering at $\bar{x}$ and $\bar{y}$, we can simplify further:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\epsilon_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

To find the variance of $\hat{\beta}_1$:

$$\text{Var}(\hat{\beta}_1) = \text{Var}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})\epsilon_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

Since the denominator is a constant, it can be factored out:

$$\text{Var}(\hat{\beta}_1) = \frac{\text{Var}\left(\sum_{i=1}^{n}(x_i - \bar{x})\epsilon_i\right)}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

Using the fact that the residuals are normally distributed, $\epsilon_i \sim N(0, \sigma^2)$, and applying the variance to a sum of independent random variables:

$$\text{Var}\left(\sum_{i=1}^{n}(x_i - \bar{x})\epsilon_i\right) = \sigma^2 \sum_{i=1}^{n}(x_i - \bar{x})^2$$

We can then sub this back into the equation:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2}$$

Simplifying further we have the solution:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

## (b) Deriving Var($\hat{\beta}_0$)

**Solution:**
Need to show that:

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)$$

The least squares estimator for the intercept is:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

To find the variance of $\hat{\beta}_0$:

$$\text{Var}(\hat{\beta}_0) = \text{Var}(\bar{y} - \hat{\beta}_1 \bar{x}) = \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\hat{\beta}_1)$$

To find $\text{Var}(\bar{y})$:
$\bar{y}$ is given by:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

Since $\bar{y}$ is the average of the independent random variables $Y_i$, the variance of $\bar{y}$ is:

$$\text{Var}(\bar{y}) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(Y_i)$$

With $Y_i$ having a variance $\sigma^2$, we get:

$$\text{Var}(\bar{y}) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

From part (a), we know:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

We can then sub these values into the variance formula for $\hat{\beta}_0$:

$$\text{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\bar{x}^2 \sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Factoring out $\sigma^2$, we get:

$$\text{Var}(\hat{\beta}_0) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right)$$

## (c) Showing that $\sum_{i=1}^{n} x_i e_i = 0$

**Solution:**
Need to show that the sum of the product of $x_i$ and the residuals $e_i$ is zero:

$$\sum_{i=1}^{n} x_i e_i = 0$$

We are given:

- $e_i = y_i - \hat{y}_i$ as the residuals.

- $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ are the predicted values.

Starting with:

$$e_i = y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

Subbing this into the sum:

$$\sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} x_i \left( y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \right)$$

Expanding the sum:

$$\sum_{i=1}^{n} x_i e_i = \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2$$

We know that *Least Squares* seeks to minimize the function:

$$SSE(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

To find the optimal values of $\hat{\beta}_0$ and $\hat{\beta}_1$, we take the partial derivatives of $SSE$ with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$, and set them to zero.

The partial derivative with respect to $\hat{\beta}_0$ is:

$$\frac{\partial SSE}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

Setting this equal to zero gives:

$$\sum_{i=1}^{n} e_i = 0$$

The partial derivative with respect to $\hat{\beta}_1$ is:

$$\frac{\partial SSE}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

Setting this equal to zero gives:

$$\sum_{i=1}^{n} x_i e_i = 0$$

Thus, we have shown that, to satisfy the requirements of Least Squares Estimators:

$$\sum_{i=1}^{n} x_i e_i = 0$$