**Developmental Reading Study**

**COM 968-32 : Statistics for Social Research II**

**(Fall 2024, Sub-term A)**

**Assignment No. 11**

**Dr. Peter Abraham Airewele, DSL**

**Omega Graduate School**

**Professor**

**Dr. Sean Taladay, Edd**

**November 7, 2024**

**Assignment #2: Developmental Reading Log**

1. Create Developmental Readings from seminal sources and scholarly peer-reviewed

journal articles. Review instructions for Assignment #3, the course essential elements,

and course readings to identify selections of books and journals to create entries.

a. Refer to the "Student Guide to Developmental Readings" in the General

Helps folder for updated information on sample comments, the grading rubric,

and key definitions related to developmental readings.

**Introduction**

The developmental reading study focuses on statistical assumption tests, their normality in linear regressions (Shatz, 2024), including statistical simulations and their constituent elements in form of relatable precepts, concepts and models. Simulation has become an essential tool in psychological research (Kelter, 2024), offering unique insights into statistical concepts, optimizing research project planning and modelling human behavior and cognition (Auger & Normand, 2024). Abstract. Attention is given to scientific research studies which are often associated with complex statistical algorithms and methodologies. Attempts will be made to see the possibility of demystifying statistical inferences, inductions and deductive reasonings and conclusions for easy understanding, interpretation and perhaps application as may be required. Statistical visualization takes a central stage (Waples, 2024), to create an enabling environment of study research and dissertation that involves the use of statistical models, charts, and graphs to convey the purpose and objectives of a given project in very simple and professional terms, to avoid statistical complexities.

**Source One:** Shatz, I. (2024). Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods*, *56*(2), 826-845.

**Comment 1:** Statistical Assumptions tests are used to validate whether a given data sample meets the necessary conditions, and the assumptions required to accurately apply a specific statistical test.

**Quote/Paraphrase:** Statistical methods generally have assumptions (e.g., normality in linear regression models). Violations of these assumptions can cause various issues, like statistical errors and biased estimates, whose impact can range from inconsequential to critical (Shatz, 2024). Abstract.

**Essential Elements:** Statistical methods, assumptions (e.g., normality in linear regression models). Violations, statistical errors, biased estimates, range from inconsequential to critical.

**Additive/Variant Analysis:** The additive is “violations of these assumptions can cause statistical errors and biased estimates.” In every assumption test, efforts are taken to present some levels of validity and reliability in the data analysis. It is cost and time effective to work with the right tools to assess and evaluate the data ensuring  a given data sample meets the necessary conditions and the assumptions required to accurately apply a specific statistical test.

**Contextualization:**  The reminder is that assumptions are checked to ensure that the chosen statistical test is appropriate for the data. The ultimate goal is to engage in statistical findings and results (data or information) which can be confidently interpreted.  Such statistical data interpretation enables easy comprehension of the tests conducted.

**Comment 2:** Establishing some levels of rationality to curtail statistical inaccuracies are crucial.

**Quote/Paraphrase:** Practical problems with missing data are common-On a central focus, there have been longstanding interests on the mechanism governing data missingness, -The foundational idea is to develop appropriate discrepancy measures -to demonstrate the feasibility, validity, and efficacy of the new test by theoretical analysis, simulation studies, and a real data analysis (Duan, 2024).

**Essential Elements:** Practical problems with missing data, central focus, longstanding interests, the mechanism governing data missingness, foundational ideas, appropriate discrepancy measures, feasibility, validity, and efficacy, theoretical analysis, simulation studies, real data analysis.

**Additive/Variant Analysis:** The additive is, “The foundational idea is to develop appropriate discrepancy measures. To demonstrate the feasibility by theoretical analysis, simulation studies, and a real data analysis”. Here, Duan emphasizes the demonstration of “statistical simulations.” These are vital in statistical analysis.

**Contextualization:** Statistical simulations are products or models fashioned out of theoretical, statistical analysis. They are practical to demonstrate. They enhance research visualizations making it to comprehend the data being analyzed.

**Source Two:** Auger, V., & Normand, A. (2024). Data simulations for advancing psychological research: Insights, preparations and investigations. *International Journal of Psychology*.

**Comment 3:**  Modern academic institutions and organizations embrace data simulations to accelerate projects and program data analysis.

**Quote/Paraphrase:** Simulation has become an essential tool in psychological research, offering unique insights into statistical concepts, optimizing research project planning and modelling human behavior and cognition (Auger & Normand, 2024). Abstract.

**Essential Elements:** Simulation, tool in psychological research, unique insights into statistical concepts, optimizing research project planning, modelling human behavior and cognition.

**Additive/Variant Analysis:** The additive is “optimizing research project planning.” This optimization is designed to investigate the activities associated

with research conduct and reasoning in the process of research design, planning and execution. “By forcing researchers to translate verbal theory into formal models, simulation can help specify the assumptions” (Auger & Normand, 2024). Abstract.

**Contextualization:** Statistical simulations are designed to create easily applicable models. They translate into perhaps simplified examples, and visualize patterns of analyzed data to apply in each experiment. Simulations can assist in explaining the correlation between data, and are fundamental in discovering social behavior and reasoning, as well as statistical techniques.

**Comment 4:** Sometimes, Scientific research studies are often associated with statistical algorithms and methodologies which compound the statistical analysis process.

**Quote/Paraphrase:** Biomedical researchers are regularly faced with an array of algorithms and statistical procedures they could potentially use to analyze or design observational studies and clinical trials. As stressed by Boulesteix et al. ([2018](https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.202200095#bimj2423-bib-0012)), the abundance of available methods makes it difficult to retain an overview of competing methods and their performance (Sauerbrei et al., [2014](https://onlinelibrary.wiley.com/doi/full/10.1002/bimj.202200095#bimj2423-bib-0104)). Neutral comparison studies can be considered an essential brick in the improvement of the reliability of the research (Kelter, 2024, para. 1 and 2). Introduction.

**Essential Elements:** Biomedical researchers, an array of algorithms and statistical procedures, design observational studies and clinical trials, competing methods and their performance, Neutral comparison studies, reliability of the research.

**Additive/Variant Analysis:** The variant is, “faced with arrays of algorithms.”. This indicates some levels of complex data or design methodologies that are perhaps intricate to unravel. However, the additive is “Neutral comparison studies can be considered an essential brick in reliability of research.” Neutral reviews are designed to avoid bias findings and outcomes.

**Contextualization:** When statistical data are outlined in rounds of difficult algorithms and models, the purpose and content of the research is problematic to comprehend. Neutrality offers room for comparative studies. This lends validity and reliability to the data and research conducted. Comparative reviews are crucial.

**Source Three:** Pirani, S. (2024). Simplifying statistical Decision Making: A Research Scholar’s Guide to parametric and Non-Parametric Methods. *International Journal of Multidisciplinary Research & Reviews*, *3*(03), 184-192.

**Comment 5:**  In statistical parametric procedures research students or officials are required to work in harmony as much as they can to prevent bias and curtail errors.

**Quote/Paraphrase:** The purpose is to ease the confusion or lack of knowledge among the researchers in application of various parametric and non-parametric techniques- to simplify the statistical decision for future research scholars- converts the complex theoretical concepts of parametric and nonparametric techniques into simplified summarized content. This will eventually lead to the research fraternity fostering effective solutions to the social issues (Pirani, 2024).

**Essential Elements:** Parametric and non-parametric techniques, statistical decision, complex theoretical concepts of parametric and nonparametric techniques, research fraternity.

**Additive/Variant Analysis:** The additive is “research fraternity fostering effective solutions to the social issues.” Pirani attempts to simplify parametric and non-parametric statistical formulars and applications. According to OGS.edu (2024) Using parametric procedures in statistics means you can generalize your findings from the sample to the population. You can use parametric procedures when you have a large sample size or when your sample meets certain assumptions about normality. Using a nonparametric procedure means your findings only apply to the sample, not the population.

**Contextualization:** The essence of a collective adoption of the parametric and non-parametric t-test is to justify their relevant use. The “research fraternity” imply a collective approach or perhaps a consensus of t-test adoption and applications. Their simplification enhance decision making in certain research studies.

**Comment 6:** Statistical research is not a function of performing a calculus stunt, it is more of practical, comprehensible interpretations, which can also be visualized.

**Quote/Paraphrase:** Advanced Statistics in Research does not show how to perform statistical procedures--it shows how to read, understand, and interpret them, as they are typically presented in journal articles and research reports. It demystifies the sophisticated statistics that stop most readers cold: with multiple regression, logistic regression, bias analysis, ANOVA, ANCOVA, MANOVA, factor analysis, path analysis, structural equation modeling, meta-analysis. Advanced Statistics in Research assumes that you have never had a course in statistics. It begins at the beginning, with research design, central tendency, variability, z scores, and the normal curve. (Hatcher, 2013). Book Introduction.

**Essential Elements:** Advanced Statistics in Research, sophisticated multiple regression, logistic regression, discriminant analysis, ANOVA, ANCOVA, MANOVA, factor analysis, path analysis, structural equation modeling, meta-analysis.

**Additive/Variant Analysis:** The additive, is “demystifies the sophisticated statistics” and the variant is “stop most readers cold.” Hatcher attempts to convey statistics as a study that has to start from the fundamentals, bottom up and not from the top-down approach. He is right in his attempt to demystify the seeming tacky, complex approach usually associated with some statistical studies.

**Contextualization:**  Most times, the study of modern statistics takes a three-prong approach, i) utilizing the qualitative, descriptive, narrative, ii) and quantitative mathematical approach. iii) The interpretation of the inductive and the deductive are simplified through the use of statistical visualizations, through models, charts and graphs.

**Point of Note:**

* **Tables 3.1 to 3.4** are culled from: International Journal of Multidisciplinary Research & Reviews 2024 /Pirani, 2024) Introduction.
* The comparative differences in the parametric and non-parametric statistics are presented in charts and tables.
* These visualizations enable easy reading or study, understanding, and interpretation of the statistical formular, calculated equations and findings.

**Simplified Comparison of the application of Parametric and Non-Parametric Test.**

**Table 3.1**

|  |  |  |
| --- | --- | --- |
|  | **Parametric** | **Non-Parametric** |
| Conclusions | Conclusions are applicable to population and thus can generalize the Result. | Conclusions are limited to the sample and thus cannot Generalize the Result. |
| Distribution | Can apply parametric Test only if the data is Normally Distributed (Can examine with the help of Kolmogrov-smirnov, Shapiro Wilcoxon or by computing Skewness and Kurtosis) | Distribution free test (Can also apply in case of Positively skewed and Negatively Skewed data). |
| Homogeneity of variance | Can apply parametric Test only if there is Homogeneity of variances of various Groups (Can examine with the help of Levene’s Test, FTest). | Not Required. |
| Sampling | Random Sampling is preferred. | Probability or non-probability sampling methods can be used. |
| Variables | At least one scale variable is needed (Interval or Ratio scale variable) | Nominal and Ordinal Data set is used to apply Non-parametric Test. |
| Sample Size | Large Sample size is Required to apply parametric Test. | Can apply in case of small sample size (n < 30). |
| Outliers | There must be no outliers in the data set. | No assumption of No outliers to be Fulfilled. |
| Observation | Independent observation must be fulfilled. | No assumption of independent observation to be Fulfilled. |

**Types of Parametric T-Tests Table 3.2**

|  |  |  |
| --- | --- | --- |
| **Purpose** | **Types of T-Tests** | **Formular** |
| To Test the difference between Sample, mean and population mean | One Sample T-Test | T-test: Definition, Formula, Types, Applications |
| To test the Difference between two independent sample means | Independent sample T-test formular | T-test: Definition, Formula, Types, Applications |
| To test the Difference between pre and post event outcomes. | Paired/Repeated Sample  t-test | Paired t Test MTB_00 |

**Table 3.3 One and Two Sample T-Tests’ Illustrations**

|  |  |
| --- | --- |
| **The formula for a one sample t-test is:**  T-test: Definition, Formula, Types, Applications  **where:** | **The formular of two-sample t-test:**  T-test: Definition, Formula, Types, Applications  **where:** |
| * **x̄**: is the sample mean. * **μ**: is the population mean (the   value you are comparing your  sample to).   * **s**: is the sample standard deviation. * **n**: is the sample size. | T-test: Definition, Formula, Types, Applications |

**Table 3.4: Comparison between Z-test and T-test**

|  |  |
| --- | --- |
| **Z-Test** | **T-Test** |
| Test the differences of mean of two groups, which may be sample mean and population mean, two sample means, proportions of sample and population, two sample proportions, etc. | Test the difference of mean of 2 groups which may be sample mean and population mean, two sample means, repeated sample means, paired sample means, etc. |
| Finite population. | Infinite population. |
| Population mean and variance are known with certainty. | Population Mean and Variance is unknown. |
| Large sample and its variance are same as population variance. | Sample is small compared to population size. |

**Source Four:** Fernandez, K. (2014, September 26). What is sampling error [Video]. YouTube.

https://www.youtube.com/watch?v=uGuWrPFStdg [Seminal] [Time = 2:49].

**Comment 7:** A sampling error occurs when there is a selection of samples that do not represent the results obtained from the entire population.

**Quote/Paraphrase:** Taking probability samples of large populations is considered common practice in the social sciences, the best way of getting a representative sample from a population, it doesn’t guarantee a perfect sample. We must acknowledge that even the best random samples will always be a little different from the true population. We call that “sampling error.” We can’t avoid sampling error, but we can estimate the size of sampling error and there are ways of reducing sampling error. The margin of error that you commonly see with survey results is an estimate of sampling usually 5% or less, that the margin of error is actually larger than stated in a report. As your sample size increases, your sampling error decreases (Fernandez, 2014).

**Essential Elements:** Probability samples of large populations, common practice in the social sciences, a representative sample from a population, sampling error. margin of error and estimate of sampling usually 5% .

**Additive/Variant Analysis:** The variant “The margin of error that you commonly see with survey results is an estimate of sampling usually 5% or less. That margin of error is larger than stated in the report. As your sample size increases, your sampling error decreases”. However, the additive, is “We can’t avoid sampling error, but we can estimate the size of sampling error and there are ways of reducing sampling error.” Most times, sensitive statistical surveys demand third party reviews to minimize the type 1 and type 2 errors.

**Contextualization:** Some scrupulous measures in planning the data sampling are required to contain sampling errors. Estimates of the projected errors are made ahead, every turn in the process of the survey of the population is well recorded and compared to the other. A decrease in the data sampled is likely to create more sampling errors. Increase in survey sampling perhaps offers a sampling error of 5% or less.

**Comment 8:**  The controversy about the effective format or way to determine Type 1 and Type 2 errors in statistical, hypothetical surveys still exist but require planning and alternative statistical interventions, perhaps aligned to the Frequentist (objective) and the Bayesian (subjective) model approach.

**Quote/Paraphrase:** In clinical studies upon which decisions are based there are two types of errors that can be made: a type I error arises when the decision is taken to declare a positive outcome when the truth is in fact negative, and a type II error arises when the decision is taken to declare a negative outcome when the truth is in fact positive. Commonly the primary analysis of such a study entails a two-sided hypothesis test with a type I error rate of 5% and the study is designed to have a sufficiently low type II error rate, for example 10% or 20%.

This simplistic approach has recently been challenged by numerous authors both from a frequentist and Bayesian perspective since when resources are constrained there will be a need to consider a trade-off between type I and type II errors (Grieve, 2024).

**Essential Elements:** Clinical studies, type I error, a type II error, a two-sided hypothesis test with a type I error rate of 5%, type II error rate for 10% or 20%., simplistic, a frequentist and Bayesian perspective, the trade-off, the costs to optimize the choice of error rates for simple, point null and alternative hypotheses .and extend the results to composite, or interval hypotheses, showing links to the Probability of Success of a clinical study.

**Additive/Variant Analysis:** The variant is,“This simplistic approach has recently been challenged by numerous authors both from a frequentist and Bayesian perspective.” The **Frequentist** perspective assumes that probabilities are fixed and objective and are based on long-term frequencies. The methods are data-driven and don't assign probabilities to hypotheses or parameters. The **Bayesian** perspective embraces subjectivity and the idea that probabilities change based on new information. And the methods can calculate the probability that a hypothesis is true.

**Contextualization:**  Both the Frequentist and the Bayesian models are still embraced by many researchers in their statistical survey approach and decision-making process. The former is objective, and data driven perhaps a drift to the null hypothesis, “Ho.”. The latter is subjective, with a probability change based on new information, perhaps a drift toward the alternative hypothesis, “Ha.” In essence the Ho and the Ha alternative can be applied to derive some reasonable statistical outcomes.

**Source Five:** Okoye, K., & Hosseini, S. (2024). Chi-Squared (X2) Statistical Test in R. In *R Programming: Statistical Data Analysis in Research* (pp. 211-223). Singapore: Springer Nature Singapore.

**Comment 9:** The Chi-squared (X2) analysis in R comes as one of the statistical alternative tests also popular in comparative statistical inferences in research study.

**Quote/Paraphrase:**  This chapter explains and practically illustrates to the readers how to apply a Chi-squared (X2) analysis in R. The Chi-squared test is used to compare how expectations are linked or related with the actual observed (frequency, fact, behavior, relationship, fitting, distribution) datasets or experimental data. Two main types of tests or analyses are usually applied by the researchers using the Chi-squared analysis. This includes the i) *Independence test* which is defined as a test of “relationship” that allows the researcher or data analyst to compare two (categorical) variables to determine whether they are related or not, and ii) *Goodness of fit* *test* which allows the user to determine whether a proportion of a data sample matches the larger population. Thus, *if* the analyzed data does not match or fit the assumed (expected) characteristics of the intended population, usually determined through the p-value (p ≤ 0.05), *then* the users or researcher may not consequentially want to use the drawn data or sample to make any conclusion about the studied (larger) population in question (Okoye & Hosseini, 2024).

**Essential Elements:**  Chi-squared (X2) analysis in R. actual observed (frequency, fact, behavior, relationship, fitting, distribution) datasets or experimental data, (i) *Independence test,* a test of “relationship, (ii) *Goodness of fit* *test*, the p-value (p ≤ 0.05).

**Additive/Variant Analysis:** The additive is “Chi-squared analysis includes the i) *Independence test* which is defined as a test of “relationship” that allows the researcher or data analyst to compare two (categorical), ii) *Goodness of fit* *test*

to determine whether a proportion of a data sample matches the larger population.

Many will employ the use of Chi Squared as a precursor tool to contain the type 1 and 2 errors. The comparative “independence test”, and the relativity of the goodness of fit are crucial and embraced by academia in dissertation study presentations.

**Contextualization:** The Chi Squared is required in organizational and academic statistical reports, they convey independent comparative test assessment and evaluation. The goodness of fit assessments determines how relevant a sample taken from a given population fits or represents the characteristics of the population. So, either approach is applicable based on the assigned task.

**Comment**  **10:**  Power, which is the probability of rejecting a false null hypothesis, is calculated as 1-β (also expressed as “1 - Type II error probability”). For a Type II error of 0.15, the power is 0.85.

**Quote/Paraphrase:** Although null hypothesis testing as originally conceived (and as is still used in applied research) is a powerful tool for decision-making. Power analysis of a specific desired effect size may then be carried out before an experiment in which a null hypothesis is tested against an alternative hypothesis that is based on this desired effect size. As a result, hypotheses in applied research are highly testable (Ashton, 2013, para. 2).

**Essential Elements:**  Null hypothesis testing, applied research, powerful tool for decision-making, Power analysis, alternative hypothesis that is based on this desired effect size.

**Additive/Variant Analysis:** The additive is, “Power analysis of a specific desired effect size may then be carried out before an experiment in which a null hypothesis is tested against an alternative hypothesis.” Power analysis of a “specific desired size” is another proactive process of intervention to test the veracity of a null hypothesis. The essence is to curtail study survey errors to barest minimum.

**Contextualization:**  It is believed that statistical surveys based on data extrapolation and analysis cannot be free from errors. But such errors must be envisaged ahead of time, and minimized during the experiment requiring a null hypothesis to be tested. Such a statistical process helps to put in place hypothetical testing with less errors.

**Source Six:** Waples, J. (2024) Simple Linear Regression: Everything You Need to Know. Data

Camp. https://www.datacamp.com/tutorial/simple-linear-regression.

**Comment 11:** Simple Linear Regression model show dotted lines of dependent and independent variables in slopes and intersections, and the equations usually follow.

**Quote/Paraphrase:**  Simple linear regression is a linear regression with one independent variable, also called the explanatory variable, and one dependent variable, also called the response variable. In simple linear regression, the dependent variable is continuous (Waples, 2024).

**Essential Elements:** Simple linear regression, independent variable, the explanatory variable, one dependent variable, response variable, the dependent variable. The slope and intersections, simple equations, linear regression equations.

**Additive/Variant Analysis:** The additive is “Linear regression with one independent variable, also called the explanatory variable, and one dependent variable, also called the response variable.” Analogy implies the variables either exhibit a dynamic flow in slope or intersect. The intersections could be levels of comparative data like “gender vs schooling”, School enrollment vs Fees, etc.

**Contextualization:** The assumptions below show series of simple linear regression in both slope intersections and linear equations. They depict both the dependent and independent variables either in slopes or at varied intersection points: See assumptions and details below:

**Assumptions:**

**Simple linear regression model coefficients:**

If we were using only the slope-intercept equation, we would find the values of m (slope) and b (y-intercept) by first calculating the slope as ‘rise over run,’ by which we mean measuring the change in y over the change in x between two points on the line. Then, once we have found the slope, we would find the y-intercept b by substituting the coordinates of one point on the line into the equation and solving for b. This final step gives you the point where the line crosses the y-axis.

**Simple linear regression equation:**

Let's look at the simple linear regression equation. We can start by first looking at the slope-intercept form of a straight-line and start at the beginning:

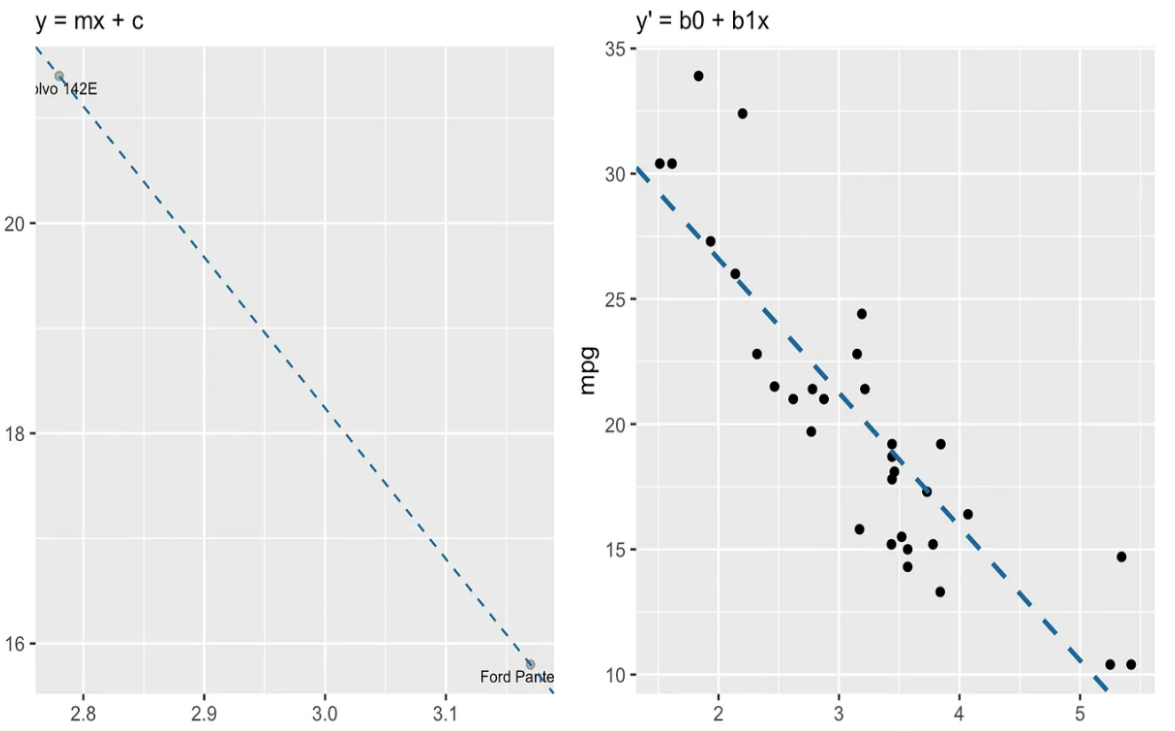
|  |
| --- |
| * ***m*** is the slope of the line. * ***b*** is the interception.   ***Y = mx + b*** |
| **Here:**   * In the context of data science, you are more likely to see this equation instead:   simple linear regression equation  **Where:**   * *b0* is the y-intercept * *b1* is the slope.   **The Streamlined Simple Linear Regression Equation is:**  simple linear regression slope and intercept equations | | |

The following visualization shows the conceptual difference between the slope-intercept form of the line, on the left, and the regression equation, on the right shown below:

**Figure 6.1: The Simple linear regression diagnostic plots:**

Diagnostic plots show how a simple linear regression model fits well and doesn’t violate the assumptions. Any patterns or deviations in these plots suggest model issues that need attention or information that wasn’t captured. One diagnostic plot that is unique to simple linear regression is the x values versus residuals plot, as you can see below. Additional plots include the Q-Q plot, scale-location plot, observation number vs. cook’s distance, and others.

|  |  |
| --- | --- |
| **The Slope Intercept**  (is Left) | **Linear Regression Equation**  (is Right) |



Slope-intercept form vs. simple linear regression equation. Image by Author

**Comment 12:** Modern digital, statistical software like the “R” Programming provide less complex statistical results, customizable, and user friendly with other programs like, IBM-SPSS, SAS, Tableau, Python, JMP, Scilab, Origin Pro, etc.

**Quote/Paraphrase:** The “R” Programming Language is an open software and programming language for computing statistics and graphical displays through methods such as data manipulations, modeling and calculations. (Ihaka & Gentleman). R is a functional and objective oriented programming, and in comparison, to other programs (Douglas et al., 2020; Maloof, 2011) includes the capacity to write more efficient codes, methods of vectorizations, artful- illustrative graphs, Tableau, Python, differentiate packages, image manipulation and data analysis (Okoye, 2024, p. 3 and 4, para. 1 and 2).

**Essential Elements:** The “R” Programming Language, open software, data manipulations, modeling and calculations, functional, objective oriented programming, capacity, efficient codes, methods of vectorizations, artful- illustrative graphs, Tableau, Python, differentiate packages, image manipulation and data analysis.

**Additive/Variant Analysis:** The additive is, the “R” Programming Language is an open software and programming language for computing statistics and graphical displays through methods such as data manipulations, modeling and calculations.”. The R-Program is customizable and partly user friendly and can work with other statistical programs. However, some skills in computer and software programs and data entry are required to operate the R-program effectively.

**Contextualization:**  Simplifying the use of R-Program will make it easily accessible and applicable. The shortlist of the 10 Best Statistical Analysis Software Options are 1) Tableau — Best for Salesforce users, 2. Domo — Best dashboards, 3. IBM SPSS — Best for large teams with different skill sets, 4. SAS Viya — Best cross-department tool. Others are 5. JMP — Best for Mac users, 6. MATLAB — Best for cleaning large volumes of data, 7. Scilab — Best open-source statistical analysis software, 8. Origin Pro — Best for graphing, 9. Minitab — Best for Six Sigma professionals, 10. Stata — Best for advanced programming.

**Steps to Using R Programming for Statistical Data Analysis**

**Table 6.1**

|  |  |
| --- | --- |
| Program | Supports different sets of packages and programs. |
| Transform | Implement collections of libraries, statistical analysis, data analytics, data science tasks. |
| Discover | Investigates, define, refine hypothesis and methods. |
| Model | Wide arrays of tools, and methods to capture the right tests and models. |
| Communicate | Compute, compile and run the codes, viewed as graphs and reports with R-markdown to share with the world. |

**Table 6.1 is culled from:** Okoye, K. (2024). *R Programming: Statistical Data Analysis in Research*. Springer Nature.

**Point to Note**

|  |
| --- |
| * Table 6.1 highlightsstepsto using R program. * It shows a wide spectrum of functions it avails. |

**Conclusion**

There is a broad sense of statistical education and skill sets realized from the development study, from statistical assumption testing, statistical simulations (Kelter, 2024; Hatcher, 2013) to the strategies to curtail Type 1 and 2 Errors associated with statistical analysis surveys. The Chi-squared (X2) analysis in R comes as one of the statistical alternative tests (Okoye & Hosseini, 2024) also popular in comparative statistical inferences and its application in research study. There are extensive studies in the use, and application of various parametric and non-parametric techniques, the T-test and Z-test variations, (Pirani, 2024; Shatz, 2024), the simple linear regressions, the dynamics of the dependent and independent variables on the line slope and the intersections, and the streamlined linear regression equations that follow. The glaring capacities of the R-Programming software and other customizable statistical software facilitate statistical research studies. Other statistical software that share similar application, and strategic relevance are “Tableau (best for salesforce users), IBM SPSS (best for large teams with different skill sets), SAS Viya (best cross-department tool), and Scilab (best open-source statistical analysis software) (Litt, 2024). Academia and modern organizations embrace these series of statistical software, because they possess fast track performance capacities in data analytics, storage and retrieval processes. Moreover, they demonstrate added advantages in time, error minimization, and risk management.

**Works Cited**

Ashton, J. C. (2013). Experimental power comes from powerful theories, the real problem in

null hypothesis testing. *Nature Reviews Neuroscience*, *14*(8), 585-585.

Auger, V., & Normand, A. (2024). Data simulations for advancing psychological research:

Insights, preparations and investigations. *International Journal of Psychology*.

Duan, R., Liang, C. J., Shaw, P. A., Tang, C. Y., & Chen, Y. (2024). Testing the missing at

random assumption in generalized linear models in the presence of instrumental variables. *Scandinavian Journal of Statistics*, *51*(1), 334-354.

Fernandez, K. (2014, September 26). What is sampling error [Video]. YouTube.

https://www.youtube.com/watch?v=uGuWrPFStdg [Seminal] [Time = 2:49]

Grieve, A. P. (2024). Optimizing the Trade-Off Between Type I and Type II Errors: A Review

and Extensions. *arXiv preprint arXiv:2409.12081*.

Hatcher, L. (2013). *Advanced statistics in research: Reading, understanding, and writing up*

*data analysis results.*

Kelter, R. (2024). The Bayesian simulation study (BASIS) framework for simulation studies in

statistical and methodological research. *Biometrical Journal*, *66*(1), 2200095.

Litt, S. (2024) Shortlist of the 10 Best Statistical Analysis Software Options. The CFO Club.

Financial Planning & Analysis.

Okoye, K., & Hosseini, S. (2024). Chi-Squared (X2) Statistical Test in R. In *R Programming: Statistical Data Analysis in Research* (pp. 211-223). Singapore: Springer Nature Singapore.

Okoye, K. (2024). *R Programming: Statistical Data Analysis in Research*. Springer Nature.

Pirani, S. (2024). Simplifying statistical Decision Making: A Research Scholar’s Guide to

parametric and Non-Parametric Methods. *International Journal of Multidisciplinary Research & Reviews*, *3*(03), 184-192.

Shatz, I. (2024). Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods*, *56*(2), 826-845.

Waples, J. (2024) Simple Linear Regression: Everything You Need to Know. Data

Camp. https://www.datacamp.com/tutorial/simple-linear-regression.