Data analysis Assignment: Physical Geography option.

Testing species richness and abundance at different sites

Key :

Site 1 Cabecera IA : Invertebrate abundance

Site 2 Villacarli ISR: Invertebrate species richness

Site 3 Puebla SS: Stored Sediment

Site 4 Carresquero

1. Produce diagrams showing site mean and 99% CI values for invertebrate abundance, species richness and stored sediment. Comment on the general patterns evident in these diagrams (3 marks)

A graph with lines and dots

Description automatically generated with medium confidence

*Figure 1: Mean invertebrate abundance (individuals per square meter) across sites 1-4. Error bars represent the 99% confidence intervals (CI) for each site's mean.*

Cabecera exhibits the highest mean abundance, while Carresquero shows the lowest. Differences in mean abundance and variability among the sites are visually evident

A graph with blue and black dots

Description automatically generated

*Figure 2: Mean invertebrate species richness (taxa per square meter) across sites 1-4. Error bars represent the 99% confidence intervals (CI) for the mean’s richness at each site.*

Cabecera demonstrates the highest species richness, while Carresquero exhibits the lowest. The differences in species richness among the sites highlight varying ecological conditions or biodiversity levels.

A graph of bar graphs

Description automatically generated with medium confidence

*Figure 3: Mean stored sediment (grams per square meter) across sites 1-4. Error bars represent the 99% confidence intervals (CI) for the mean stored sediment at each site.*

Villacarli shows the highest mean sediment storage, followed by Carresquero and Puebla, while Cabecera has negligible sediment storage. The variation in sediment storage highlights differences in sediment dynamics among the sites.



*Figure 4: Summary statistics (4dp) for invertebrate abundance (ind/m²), species richness (taxa/m²), and stored sediment (g/m²) across sites 1-4. The table provides the lower 99% confidence interval (CI), mean, and upper 99% CI for each metric.*

Commentary on the general patterns:

**Graph 1; Invertebrate Abundance:**There is no overlap in the 99% confidence intervals (CIs) for invertebrate abundance between Site 1 (Cabecera, 562.73–667.81 ind/m²), Site 2 (Villacarli, 139.59–320.54 ind/m²), and Site 3 (Puebla, 386.92–471.75 ind/m²), indicating clear differences in abundance among these sites. However, the CIs for Site 2 and Site 4 (Carresquero, 211.62–300.65 ind/m²) overlap, making it impossible to determine a significant difference in abundance between these two sites. Site 1 has the highest mean abundance, while Sites 2 and 4 show similar, much lower values.

**Graph 2; Species Richness:**There is little similarity among the four sites for invertebrate species richness, as most CIs show minimal overlap. Site 1 (12.44–15.43 taxa/m²) stands out with the highest richness, while Site 2 (3.14–5.66 taxa/m²) has the lowest. Comparisons are more challenging between Site 3 (8.67–11.06 taxa/m²) and Site 4 (5.77–8.63 taxa/m²) due to some overlap in their CIs, but Site 3 generally shows a higher mean richness than Site 4.

**Graph 3; Stored Sediment:**  
Sites 2 (880.63–1158.44 g/m²) and 4 (801.09–1051.69 g/m²) have overlapping CIs for stored sediment, suggesting similarity in sediment storage between these sites. However, Site 1 (48.78–72.19 g/m²) and Site 3 (283.07–506.13 g/m²) are starkly different from Sites 2 and 4, with significantly lower sediment storage values. Site 1 has negligible sediment storage compared to the other sites.

These patterns highlight significant ecological differences among the sites for invertebrate abundance, species richness, and sediment storage, with overlap suggesting some similarities in specific cases.

1. Explore the relationship between invertebrate abundance and invertebrate species richness. Do sites that have greater invertebrate abundance also have more invertebrate species? (3 marks)

To estimate: Scatter plot + regression line.

To test: Pearsons or Spearman’s.

CAN use a Pearson correlation analysis test to explore relationship between IA and ISR, as both data sets are continuous and normally distributed (see below)

Shapiro-Wilk test: IA test statistic: Significance >0.05 (.107) ISR test statistic: Significance >0.05 (.063)  
The histograms also show a normal distribution; therefore, the data sets meet the parametric assumptions to do Pearsons.

A graph of a number of blue dots

Description automatically generated with medium confidenceScatter Plot:

*Figure 5: Scatter plot showing the relationship between invertebrate abundance (ind/m²) and iinvertebrate species richness (taxa/m²). Each point represents a site observation.*

The plot reveals a strong positive relationship, with a linear trendline (R² = 0.693) indicating that approximately 69.3% of the variation in species richness is explained by invertebrate abundance.

Pearsons Correlation analysis:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Invertebrate species richness: (taxa/m2)** | | |
| **Invertebrate abundance : (ind/m2)** | | | Pearson Correlation | 0.832 |
|  | | | Sig. (1-tailed) | <0.001 |
|  | | | N | 60 |

*Figure 6: Table of results for the Pearsons correlation analysis between IA and ISR.*

The Pearson correlation coefficient is 0.832, indicating a strong positive relationship between the two variables. This suggests that as IA increases, the ISR also tends to increase. The p-value is < 0.001, which is highly significant. This means there is a very low probability (less than 0.1%) that this strong correlation occurred by chance.

1. Do patterns of stored sediment, invertebrate abundance and species richness differ significantly between sites, and if so, which sites differ from which? (6 marks)

Initial plan: 3 separate one way Anova tests to test for differences and post-hoc tests to identify where those differences are. But not all 3 variables are normally distributed. See below.

|  |  |  |
| --- | --- | --- |
| **Variable** | **Normality Test (Shapiro-Wilk)** | **Levene’s Test for Homogeneity** |
| **Invertebrate Abundance** | **Normality: Yes** (Shapiro-Wilk p > 0.05 for all sites) | **Variance violated:** Levene’s p = 0.011 |
| **Invertebrate Species Richness** | **Normality: Yes** (Shapiro-Wilk p > 0.05 for all sites) | **Variance not violated:** Levene’s p = 0.925 |
| **Stored Sediment** | **Normality: No** (Shapiro-Wilk p < 0.05 for Cabecera; p > 0.05 for others) | **Variance violated:** Levene’s p < 0.001 |

*Figure 7: Summary table of normality and variance assumptions for invertebrate abundance, invertebrate species richness, and stored sediment).*

*Normality was assessed using the Shapiro-Wilk test, and homogeneity of variances was tested with Levene’s test.*

Explanation of the Statistics:

**Invertebrate Abundance:**

Passed normality for all sites (p > 0.05), but variance assumption violated (p = 0.011), so standard ANOVA is not suitable. I will therefore use Kruskal-Wallace as a non-parametric alternative.

**Invertebrate Species Richness:**

Passed normality for all sites (p > 0.05), and variance assumption met (p = 0.925). I therefore can use One-Way ANOVA since all assumptions are satisfied.

**Stored Sediment:**

Normality (Shapiro-Wilk): Failed for Cabecera (p = 0.037), though others passed (e.g., Villacarli p = 0.389). However, as the data set is relatively small (15 data points per site) and the variance assumption is not met (Levene’s: p < 0.001), we cannot get away with using one-way ANOVA. Will therefore use Kruskal-Wallis.

|  |  |  |
| --- | --- | --- |
| **Kruskal-Wallis** | Invertebrate abundance (ind/m2) | Stored sediment (g/m^2) |
| Chi-Square statistic | 47.175 | 50.317 |
| Sig. | <0.001 | <0.001 |

*Figure 8: Summary table of relevant Kruskal-Wallis test results for Invertebrate abundance and stored sediment.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Invertebrate species richness (taxa/m2) | | | | |
| **ANOVA** | Sum of Squares | Mean Square | F | Sig. |
| Between Groups | 740.983 | 246.994 | 80.106 | <0.001 |
| Within Groups | 172.667 | 3.083 |  |  |

*Figure 9: Summary table of one-way ANOVA results for Invertebrate species richness (taxa/m2).*

Conclusion: Patterns of stored sediment, invertebrate abundance and species richness differ significantly between sites.

For Kruskal-Walis, the very low p-value (p<0.05 on both accounts) indicates that that there are statistically significant differences in invertebrate abundance and stored sediment between sites.

For ANOVA: p is also less than 0.05 ( p<0.001) therefore, there are some sites that differ significantly.

Identifying which sites differ from which:

|  |  |  |  |
| --- | --- | --- | --- |
| Invertebrate species richness: | | Mean Difference (I-J) | Sig. |
|  |
| Cabecera | Villacarli | **9.**53333\* | <0.001 |  |
| Puebla | **4**.06667\* | <0.001 |  |
| Carresquero | **6.**73333\* | <0.001 |  |
| Villacarli | Cabecera | -9.53333\* | <0.001 |  |
| Puebla | -5.46667\* | <0.001 |  |
| Carresquero | -2.80000\* | <0.001 |  |
| Puebla | Cabecera | -4.06667\* | <0.001 |  |
| Villacarli | 5.46667\* | <0.001 |  |
| Carresquero | 2.66667\* | <0.001 |  |
| Carresquero | Cabecera | -6.73333\* | <0.001 |  |
| Villacarli | 2.80000\* | <0.001 |  |
| Puebla | -2.66667\* | <0.001 |  |

ISR: Tukey post hoc test.

*Figure 10: Table summarising Tukey post-hoc test results. For Invertebrate species richness at different sites.*

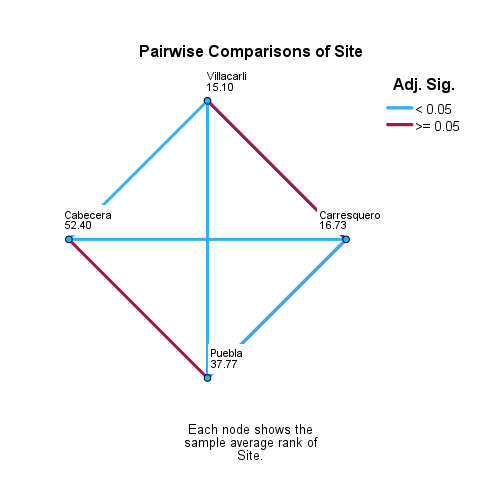
Summary of results:

Cabecera consistently had the highest ISR, significantly greater than all other sites. Villacarli consistently had the lowest ISR, significantly lower than all other sites.Puebla and Carresquero exhibited intermediate ISR values, with Puebla significantly higher than Carresquero.

This analysis confirms clear and significant differences in ISR between sites, with the strongest contrast between Cabecera (highest ISR) and Villacarli (lowest ISR).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pairwise Comparisons of Site for IA | | | | |
| Sample 1-Sample 2 | Test Statistic | Sig. | Adjusted p value | Significant? |
| Villacarli-Carresquero | -1.633 | 0.798 | 1.000 | No |
| Villacarli-Puebla | -22.667 | 0.000 | 0.002 | Yes |
| Villacarli-Cabecera | 37.300 | 0.000 | 0.000 | Yes |
| Carresquero-Puebla | 21.033 | 0.001 | 0.006 | Yes |
| Carresquero-Cabecera | 35.667 | 0.000 | 0.000 | Yes |
| Puebla-Cabecera | 14.633 | 0.022 | 0.130 | No |

*Figure 11: table summarising the pairwise comparisons of Invertebrate abundance at the different sites test results with adjusted p values. Adjusted by the Bonferroni correction for multiple tests.*

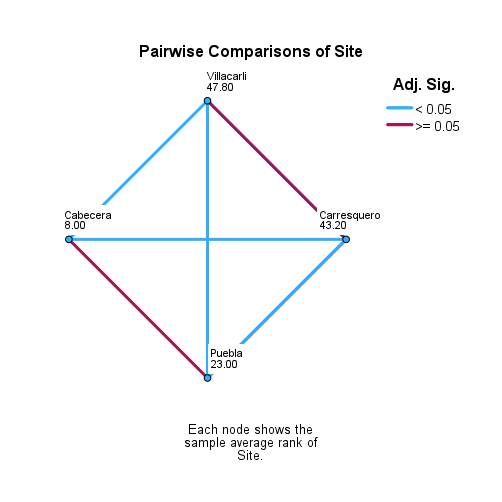
**

*Figure 12:* *Visualisation of figure 11’s results, showing where there are and aren’t significant differences between sites regarding invertebrate abundance.*

The test was 2-sided, so the significance level is therefore 0.05, as we can see in Figure 12, only two values were deemed insignificant, meaning that there were multiple sites that showed significant patterns of difference regarding IA. For instance, Cabecera consistently displayed significant differences in IA compared to other sites. Similarly, Villacarli exhibited significant variation when compared with Puebla and Cabecera. Whereas, the observed lack of difference between Villacarli – Carresquero, and Puebla – Cabecra suggests that these sites may have similar properties that result in less variation of abundance levels, which we can confirm to be true for the site 2 – site 4 pairing if we go back and look at figure 1, showing the means of IA for both these sites to be relatively similar.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pairwise Comparisons of Site for SS** | | | | |
| Sample 1-Sample 2 | Test Statistic | Sig. | Adjusted P value | Significant? |
| Cabecera-Puebla | -15.000 | 0.019 | 0.112 | No |
| Cabecera-Carresquero | -35.200 | 0.000 | 0.000 | Yes |
| Cabecera-Villacarli | -39.800 | 0.000 | 0.000 | Yes |
| Puebla-Carresquero | -20.200 | 0.002 | 0.009 | Yes |
| Puebla-Villacarli | 24.800 | 0.000 | 0.001 | Yes |
| Carresquero-Villacarli | 4.600 | 0.471 | 1.000 | No |

*Figure 13: stored sediment pairwise comparison test results with adjusted significance for Bonferroni significance.*

**

*Figure 14: Visualisation of figure 13’s results, showing where there are and aren’t significant differences between sites regarding invertebrate abundance.*

The Kruskal-Wallis pairwise comparison results for IA and SS demonstrate some interesting patterns regarding site differences. For SS, Villacarli and Carresquero show similar high values (47.50 and 43.20 respectively) with no significant difference (adjusted p=1.000), while Cabecera shows the lowest values (8.00). For IA, the pattern differs with Cabecera showing the highest values (52.40) and Villacarli-Carresquero showing the lowest similar values (15.10 and 16.73 respectively). Despite these contrasting high-low patterns, both analyses maintain some similarities in their non-significant pairings: Villacarli-Carresquero (adjusted p=1.000 for both) and a pairing involving Puebla (IA: Puebla-Cabecera adjusted p=0.130; SS: Cabecera-Puebla adjusted p=0.112).

When examining ISR, notably different trends emerge through the Tukey post-hoc analysis. Unlike IA and SS, the ISR analysis shows significant differences between all sites (all p<0.001), with no shared groupings. A clear hierarchical pattern emerges where Cabecera maintains the highest ISR values, followed sequentially by Puebla, then Carresquero, with Villacarli showing the lowest values. This distinct pattern, showing no site similarities, contrasts with the grouping patterns seen in both IA and SS analyses, suggesting that ISR may be influenced by different environmental factors than those affecting abundance and sediment storage.

Question 4: Multiple Regression Analysis

Initial approach: Multiple regression models using stepwise method to identify best predictors for:

* Invertebrate abundance (Model 1)
* Species richness (Model 2)

(i) Invertebrate Abundance Model:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model Summary | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of Estimate |
| 1 | .893 | .798 | .795 | 78.79344 |
| 2 | .901 | .812 | .805 | 76.71344 |

*Figure 15: Model summary statistics showing R² values and goodness of fit for invertebrate abundance regression model. Model 1 includes only stored sediment, while Model 2 includes both stored sediment and bed shear stress.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Coefficients | | | | | | |
|  | Unstandardised Coefficients | | Standardised Coefficients | |  |  |
| Variable | B | SE | Beta | t | Sig. | VIF |
| (Constant) | 557.426 | 29.068 |  | 19.176 | <.001 |  |
| Stored sediment (g/m²) | -.356 | .025 | -.859 | -14.364 | <.001 | 1.084 |
| Bed shear stress (N/m²) | .569 | .278 | .122 | 2.046 | .045 | 1.084 |

*Figure 16: Multiple regression coefficients for invertebrate abundance model, showing significant predictors and their relative importance.*

A graph with blue dots

Description automatically generated

*Figure 17: Scatterplot of standardised residuals showing model fit and assumptions.*

A graph of a normal growth

Description automatically generated with medium confidence

*Figure 18: Normal P-P plot demonstrating normality of residuals for abundance model.*

The stepwise regression for invertebrate abundance identified a two-predictor model. The final model achieved an R² of 0.812 (adjusted R² = 0.805), explaining 81.2% of the variance. Stored sediment emerged as the primary predictor (β = -0.859, p < 0.001), showing a strong negative relationship. Bed shear stress contributed a smaller but significant positive effect (β = 0.122, p = 0.045). Both velocity and nitrate were excluded due to non-significance in the stepwise process.

(ii) Species Richness Model:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model Summary | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of Estimate |
| 1 | .844 | .712 | .707 | 2.13006 |

*Figure 19: Model summary statistics for species richness regression analysis, showing final model with stored sediment as predictor.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Coefficients | | | | | | |
|  | Unstandardised Coefficients | | Standardised Coefficients | |  |  |
| Variable | B | SE | Beta | t | Sig. | VIF |
| (Constant) | 13.605 | .483 |  | 28.165 | <.001 |  |
| Stored sediment (g/m²) | -.008 | .001 | -.844 | -11.974 | <.001 | 1.000 |

*Figure 20: Multiple regression coefficients for species richness model, showing stored sediment as the only significant predictor.*

A graph with blue dots

Description automatically generated

*Figure 21: Scatterplot of standardised residuals for species richness model.*

A graph of a normal growth

Description automatically generated with medium confidence*Figure 22: Normal P-P plot showing distribution of residuals for species richness model.*

For species richness, the stepwise procedure retained only stored sediment in the final model (R² = 0.712, adjusted R² = 0.707). This single predictor explained 71.2% of the variance in species richness, demonstrating a strong negative relationship (β = -0.844, p < 0.001). All other variables were excluded as non-significant predictors during the stepwise process.

Model Critique:

Strengths:

* High explanatory power evidenced by strong R² values (0.812 and 0.712)
* Normal distribution of residuals confirmed by P-P plots
* Low VIF values (<1.1) indicating absence of multicollinearity
* Statistical significance achieved for all retained predictors (p < 0.05)

Limitations:

* Species richness model's reliance on a single predictor suggests possible oversimplification
* Potential interaction effects between variables remain unexplored
* Evidence of mild heteroscedasticity in residual plots
* Limited sample size may affect model robustness
* Possible omission of other influential variables

The analysis reveals stored sediment as the dominant factor in both models, showing consistent negative relationships with both abundance and richness. The additional influence of bed shear stress on abundance, but not on richness, suggests differing mechanisms affecting these two aspects of community structure. These findings align with the patterns observed in the earlier site-specific analyses (Questions 1-3), particularly regarding the inverse relationship between sediment levels and biodiversity metrics.

Question 5: Badlands Location Effect

To test the influence of badlands location, a dummy variable was created (0 = outside badlands [Sites

1 and 3], 1 = inside badlands [Sites 2 and 4]) and added to the previous regression models. For both

models, the stepwise regression retained stored sediment as the primary predictor, with badlands

location showing negative but relatively modest effects on both invertebrate abundance (β = -0.045)

and species richness (β = -0.127). Though these results suggest lower biodiversity metrics within

badland areas, the effect was not strong enough to be retained in the final models when accounting

for other environmental variables, particularly stored sediment.