Learning Module 9: Parametric and Non Parametric Tests of Independence

Q.4447 Assume a financial analyst, Alex Perez, is examining the significance of the correlation between the monthly returns of ETF 2 and the overall market index based on the table provided with 48 monthly observations. The sample correlation, $r_{ETF2,Market}$, is 0.9096. Perez wishes to use a t-test to check if the correlation is significant at a 0.01 level of significance. The sample t-table is given below:

df	p = 0.10	p = 0.05	p = 0.025	p = 0.01	p = 0.005
31	1.309	1.696	2.040	2.453	2.744
42	1.302	1.682	2.018	2.418	2.698
43	1.302	1.681	2.017	2.416	2.695
44	1.301	1.680	2.015	2.414	2.692
45	1.301	1.679	2.014	2.412	2.690
46	1.300	1.679	2.013	2.410	2.687
47	1.300	1.678	2.012	2.408	2.685
48	1.299	1.677	2.011	2.407	2.682

The conclusion about the significance of the correlation between ETF 2 and the market index is:

A. Significant because the calculated t-statistic is greater than the critical value.

B. Not significant because the calculated t-statistic is less than the critical value.

C. Not significant because the calculated t-statistic is greater than the critical value.

The correct answer is **A.**

Hypothesis test: Two-sided; $H_0 : \rho = 0$ versus $H_a : \rho \neq 0$. To test the significance of the correlation between ETF 2 and the market index, Alex Perez will calculate the t-statistic using the formula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Where: r = Sample correlation coefficient $= 0.9096$. n = Number of observations $= 48$. Substituting the given values into the t-test formula:

$$t = \frac{0.9096\sqrt{48-2}}{\sqrt{1-0.9096^2}} \approx \frac{0.9096 \cdot 6.7823}{0.41549} \approx 14.848$$

Since the critical t-value at the 0.01 level of significance is ±2.687, and the calculated t-statistic (14.848) is greater than the critical value, Perez should conclude that the correlation between

ETF 2 and the market index is significant. The test statistic is well outside the critical range, which leads to the rejection of the null hypothesis that there is no correlation.

B is incorrect. It suggests that the correlation is not significant due to the calculated t-statistic being less than the critical value. This is not the case here, as the calculated t-statistic (14.848) is indeed greater than the critical value at the 0.01 level of significance, indicating a significant correlation.

C is incorrect. The calculated t-statistic is greater than the critical value. In statistical hypothesis testing, a calculated t-statistic greater than the critical value at a given level of significance indicates that the correlation is significant, contrary to what option C suggests.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence. LOS (a): Explain parametric and non-parametric tests of the hypothesis that the population correlation coefficient equals zero and determine whether the hypothesis is rejected at a given level of significance.

Q.4451 The following contingency table shows the responses of two categories of investors (employed vs. self-employed) with regard to their risk tolerance levels (low, medium, or high). The total sample size is 200 investors.

	Low	Medium	High	Total
Employed	45	55	20	120
Self-Employed	30	40	10	80
Total	75	95	30	200

If we wish to test whether there is any significant difference between employed and self-employed investors concerning risk tolerance levels, the test statistic is *closest to:*

A. 0.222.

B. 0.333.

C. 0.730.

The correct answer is **C.**

Let H_0 be the hypothesis that there is no significant difference between employed and self-employed investors with regard to risk tolerance levels.

Let H_a be the hypothesis that there is a significant difference between employed and self-employed investors with regard to risk tolerance levels.

Step 1: Calculate the expected frequency of investors by their category and risk tolerance level using the formula:

	Low	Medium	High	Total
Employed	$\frac{(120\times75)}{200}=45$	$\frac{(120\times95)}{200}=57$	$\frac{(120\times30)}{200}=18$	120
Self-Employed	$\frac{(80\times75)}{200}=30$	$\frac{(80\times95)}{200}=38$	$\frac{(80\times30)}{200}=12$	80
Total	75	95	30	200

Step 2: Compute the scaled squared deviations for each category and risk tolerance level:

	Low	Medium	High
Employed	$\frac{(45-45)^2}{45}=0$	$\frac{(55-57)^2}{57}\approx0.070$	$\frac{(20-18)^2}{18}\approx0.222$
Self-Employed	$\frac{(30-30)^2}{30}=0$	$\frac{(40-38)^2}{38}\approx0.105$	$\frac{(10-12)^2}{12}\approx0.333$

Step 3: Calculate the total chi-square statistic:

$$\chi^2=\sum\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\approx0.070+0.222+0.105+0.333=0.730$$

A is incorrect. The option suggesting a test statistic of 0.222 does not accurately reflect the chi-square statistic calculated from the observed and expected frequencies. This value might represent a partial calculation for one of the categories but does not encompass the total variance observed across all categories and risk tolerance levels.

B is incorrect. The option indicating a test statistic of 0.333 similarly fails to capture the aggregate discrepancy between the observed and expected frequencies across all categories. While it might represent a calculation for a specific part of the table, it does not account for the total chi-square statistic which measures the overall difference.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence. LOS (b): Explain tests of independence based on contingency table data.

Q.4676 What type of test would you use to assess the correlation between excess risk-adjusted return (alpha) and mutual fund expense ratios for US large-cap growth funds?

A. Parametric test

B. Nonparametric test

C. Chi-square test

The correct answer is **B.**

Given that mutual fund expense ratios and excess risk-adjusted returns may not be normally distributed and may not meet the assumptions of a parametric test, a nonparametric test, such as the Spearman rank correlation coefficient, would be appropriate to assess the correlation between these variables.

A is incorrect. Parametric tests are based on specific distributional assumptions, which may not be met by the data in this scenario.

C is incorrect. The chi-square test is used to test for independence of categorical variables, not for assessing the correlation between continuous variables like excess risk-adjusted return and mutual fund expense ratios.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4695 What are the null and alternative hypotheses for assessing the correlation between excess risk-adjusted return (alpha) and mutual fund expense ratios for US large-cap growth funds?

A. $H0 : \rho = 0$ versus $Ha : \rho \neq 0$

B. $H0 : \rho \leq 0$ versus $Ha : \rho > 0$

C. $H0 : \rho \geq 0$ versus $Ha : \rho < 0$

The correct answer is **A.**

This formulation of hypotheses tests whether there is no correlation (null hypothesis) versus the alternative hypothesis that there is a correlation between excess risk-adjusted return and mutual fund expense ratios.

B and C are incorrect. These formulations of hypotheses are for one-sided tests, which do not align with the scenario described where we are interested in determining whether there is any correlation between the variables, regardless of direction.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4696 What is the *most* appropriate test statistic for conducting a test of correlation between excess risk-adjusted return (alpha) and mutual fund expense ratios for US large-cap growth funds using a nonparametric approach?

A. Spearman rank correlation coefficient

B. Pearson correlation coefficient

C. Chi-square test statistic

The correct answer is **A.**

When dealing with variables that may not meet distributional assumptions, such as mutual fund expense ratios and excess risk-adjusted returns, a nonparametric approach like the Spearman rank correlation coefficient is suitable for assessing correlation.

B is incorrect. The Pearson correlation coefficient assumes normality and linearity, which may not hold true for the variables in this scenario.

C is incorrect. The chi-square test statistic is used for testing independence between categorical variables, not for assessing correlation between continuous variables like excess risk-adjusted return and mutual fund expense ratios.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4697 If the calculated test statistic for assessing the correlation between excess risk-adjusted return (alpha) and mutual fund expense ratios is -0.55177, and the critical values at a 0.05 level of significance are ±2.306, what decision is *most likely* to be made?

A. Reject the null hypothesis

B. Fail to reject the null hypothesis

C. Cannot be determined from the information given

The correct answer is **B.**

Since the calculated test statistic falls within the range of the critical values, it does not provide sufficient evidence to reject the null hypothesis that the Spearman rank correlation coefficient is zero.

A is incorrect. Reject the null hypothesis: This decision would be made if the calculated test statistic falls outside the range of the critical values, indicating sufficient evidence to reject the null hypothesis.

C is incorrect. Cannot be determined from the information given: The decision can be determined based on whether the calculated test statistic falls within or outside the range of the critical values, so it can be determined from the information provided. Hence, this choice is incorrect.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4698 What is the significance level used in the test of correlation between excess risk-adjusted return (alpha) and mutual fund expense ratios for US large-cap growth funds?

A. 0.05

B. 0.01

C. 0.10

The correct answer is **A.**

The significance level, often denoted by α, is typically set at 0.05 in hypothesis testing unless otherwise specified.

B and C is incorrect. While these are common significance levels, the test of correlation between excess risk-adjusted return and mutual fund expense ratios in this scenario specifically mentions a 0.05 level of significance.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4699 What type of correlation coefficient is used when the assumptions for the parametric Pearson correlation is *least likely* to be met, such as when dealing with non-normally distributed variables?

A. Pearson correlation coefficient

B. Kendall's tau

C. Spearman rank correlation coefficient

The correct answer is **C.**

This coefficient is used in situations where the data may not meet the assumptions of the Pearson correlation coefficient, such as non-normally distributed variables or when dealing with ordinal data.

A is incorrect. This coefficient assumes normality and linearity in the data, which may not hold true in scenarios where non-normally distributed variables are involved.

B is incorrect. Kendall's tau: While Kendall's tau is another nonparametric measure of association, it specifically assesses the strength and direction of association between two variables based on the ranks of the data, rather than on the actual values of the variables. It is not mentioned in the context of the scenario provided.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4700 What are the null and alternative hypotheses to test whether the dividend and financial leverage groups are independent of one another?

A. H0: Dividend and financial leverage ratings are not related, Ha: Dividend and financial leverage ratings are related.

B. H0: Dividend and financial leverage ratings are related, Ha: Dividend and financial leverage ratings are not related.

C. H0: Dividend and financial leverage ratings are independent, Ha: Dividend and financial leverage ratings are dependent.

The correct answer is **A.**

This formulation of hypotheses correctly represents the null hypothesis that the two classifications are independent and the alternative hypothesis that they are related.

B is incorrect. This choice incorrectly states the null hypothesis as suggesting a relationship between dividend and financial leverage ratings, which is not the case.

C is incorrect. While this choice includes the concept of independence, it incorrectly formulates the alternative hypothesis as suggesting dependence rather than the two ratings being related.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (b) Explain tests of independence based on contingency table data.

Q.4703 What conclusion can be drawn if the p-value associated with the chi-square test statistic is 0.03 at a 5% level of significance?

A. Reject the null hypothesis.

B. Fail to reject the null hypothesis.

C. Accept the null hypothesis.

The correct answer is **A.**

When the p-value is less than the chosen significance level (0.05 in this case), we reject the null hypothesis and conclude that there is a significant relationship between the variables.

B is incorrect. This choice would be incorrect when the p-value is less than the significance level, indicating sufficient evidence to reject the null hypothesis.

C is incorrect. While technically correct in terms of statistical terminology, accepting the null hypothesis is not typically stated in practice; instead, we simply fail to reject it when there is insufficient evidence to do so. Therefore, this choice is less appropriate than "Reject the null hypothesis."

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (b) Explain tests of independence based on contingency table data.

Q.4704 What are the null and alternative hypotheses to test whether the dividend and financial leverage groups are independent of one another?

A. Null hypothesis: Dividend and financial leverage ratings are related; Alternative hypothesis: Dividend and financial leverage ratings are not related

B. Null hypothesis: Dividend and financial leverage ratings are not related; Alternative hypothesis: Dividend and financial leverage ratings are related

C. Null hypothesis: Dividend and financial leverage ratings are independent; Alternative hypothesis: Dividend and financial leverage ratings are dependent

The correct answer is **B.**

The null hypothesis assumes independence, while the alternative hypothesis suggests a relationship between dividend and financial leverage ratings.

A is Incorrect. This answer suggests that the null hypothesis assumes a relationship between dividend and financial leverage ratings, which is incorrect. The null hypothesis assumes independence, not a relationship.

C is incorrect. This answer incorrectly states that the null hypothesis assumes independence between the two classifications. However, the null hypothesis does not make an assumption; rather, it is a statement that is either rejected or failed to be rejected based on the evidence.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (b) Explain tests of independence based on contingency table data.

Q.4707 What is the significance of the critical value in the chi-square test of independence at a 5% level of significance?

A. The critical value indicates the probability of committing a Type I error.

B. The critical value signifies the threshold beyond which we reject the null hypothesis.

C. The critical value determines the strength of the relationship between the variables.

The correct answer is **B.**

In the chi-square test of independence, if the calculated chi-square value exceeds the critical value, we reject the null hypothesis, indicating a significant relationship between the variables.

A is incorrect. The critical value does not directly indicate the probability of committing a Type I error; it serves as a threshold for determining the significance level.

C is incorrect. The critical value does not determine the strength of the relationship between variables; it is a statistical parameter used for hypothesis testing.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (b) Explain tests of independence based on contingency table data.

Q.4708 Which statement accurately describes the degrees of freedom in a chi-square test of independence? The degrees of freedom is:

A. determined by the number of observations in the contingency table.

B. calculated as the difference between the total number of observations and the number of cells in the table.

C. equal to the product of one less than the number of rows and one less than the number of columns in the contingency table.

The correct answer is **C.**

This formula for degrees of freedom in a chi-square test of independence accounts for the constraints imposed by the row and column totals in the contingency table.

The expression $(r-1)(c-1)$ represents the degrees of freedom for a contingency table with r rows and c columns, calculated as the product of one less than the number of rows and one less than the number of columns.

A is incorrect. The degrees of freedom are not determined by the number of observations; they are calculated based on the structure of the contingency table.

B is incorrect. The degrees of freedom are not simply the difference between the total number of observations and the number of cells; they are determined by the number of rows and columns in the contingency table and are used to assess the variability in the data.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (b) Explain tests of independence based on contingency table data.

Q.4727 Which statement is *most likely* true regarding the significance level in testing the correlation coefficient between two variables' returns?

A. A lower significance level increases the likelihood of rejecting the null hypothesis.

B. A higher significance level increases the likelihood of Type I error.

C. The choice of significance level does not affect the rejection of the null hypothesis.

The correct answer is **B.**

A higher significance level (e.g., 10% instead of 5%) increases the likelihood of Type I error, which occurs when the null hypothesis is incorrectly rejected.

A is incorrect. A lower significance level (e.g., 1% instead of 5%) decreases the likelihood of rejecting the null hypothesis, which may result in a Type II error (failing to reject the null hypothesis when it is false).

C is incorrect. The choice of significance level directly influences the probability of Type I error and, consequently, affects the rejection of the null hypothesis.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4728 In testing the correlation coefficient between two variables' returns, what factor *most likely* determines the critical t-value used for hypothesis testing?

A. The level of significance chosen by the analyst.

B. The magnitude of the sample correlation coefficient.

C. The degree of freedom associated with the sample size.

The correct answer is **C.**

The critical t-value used for hypothesis testing is primarily determined by the degree of freedom associated with the sample size, which affects the shape of the t-distribution.

A is incorrect. While the level of significance influences the decision rule (i.e., whether to reject the null hypothesis), it does not directly determine the critical t-value.

B is incorrect. The magnitude of the sample correlation coefficient may influence the calculated test statistic but not the critical t-value used for hypothesis testing.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.

Q.4730 Which statement is *most likely* true regarding the t-test statistic used in testing the correlation coefficient between two variables' returns?

A. A higher t-test statistic indicates stronger evidence against the null hypothesis.

B. A lower t-test statistic suggests a higher probability of Type II error.

C. The t-test statistic is independent of the sample size.

The correct answer is **A.**

A higher t-test statistic indicates stronger evidence against the null hypothesis, suggesting that the sample correlation coefficient is significantly different from zero.

B is incorrect. A lower t-test statistic does not directly suggest a higher probability of Type II error; Type II error depends on factors such as the level of significance and the effect size.

C is incorrect. The t-test statistic depends on the sample size, with larger sample sizes typically resulting in larger t-values for the same correlation coefficient.

CFA Level I, Quantitative Methods, Learning Module 9: Parametric and Non-Parametric Tests of Independence, LOS (a) Explain parametric and nonparametric tests of the hypothesis that the population correlation coefficient equals zero, and determine whether the hypothesis is rejected at a given level of significance.
