Introduction

Welcome to Advanced Statistics course. This is a major course for the Doctor of Philosophy in Nursing program aimed. The course reviews important concepts in statistics and presents their applications in nursing situations. This course equips you with the necessary skills and knowledge on data processing, data analysis, and data presentation. The focus is on advanced statistics that doctoral students would need in preparation for their dissertation.

The course is divided into three units. Unit 1, Review of Basic Statistics, is about descriptive statistics – how to organize and present data, such that they can be summarized and interpreted more meaningfully and it includes basic principles in inferential statistics. Unit 2, Common Inferential Statistical Methods, showcase statistical applications common in nursing, which includes tests of differences, and tests of relationships. Unit 3, Multivariate Analysis, is about statistical techniques that are useful in modeling techniques especially when testing or validating theories or models, which are required in the PhD in nursing program.

Course outcomes

At the end of the course, students should be able to:
1. Explain basic concepts and principles in statistics;
2. Demonstrate application of common inferential statistics used in nursing;
3. Perform advanced statistical methods and analysis given a data set; and
4. Critique data analysis and interpretations of select statistical applications in nursing studies.

Content outline

I.   Review of Basic Statistics
   1. Basic statistical concepts
   2. Frequency tables
   3. Measures of central tendency and variability
   4. Probability and sampling
   5. Estimation
   6. Testing of hypothesis

II.   Common Inferential Statistical Methods
   1. Test of independence
   2. Analysis of variance
   3. Correlation
   4. Regression

III. Multivariate Analysis
   1. Multivariate analysis of variance
   2. Multiple regression
   3. Logistic regression
   4. Factor analysis
   5. Path analysis
   6. Structural equation modeling

UNIT 2. Common Inferential Statistical Methods

Organizing Principle

Inferential statistics is used to draw inferences about a population given a sample. This builds on the previously learned concepts on descriptive statistics. From describing and summarizing data, you can now get more information from the sample by applying concepts, principles and processes from inferential statistics.

This Unit explores the most common inferential statistical methods, namely: t-test, chi-square test of independence, ANOVA, correlation and regression. Tests of independence are used when comparing groups of characteristics of a group. They can be categorized depending on the level of measurement of data or the number of groups (or characteristics) being compared, whether two or more than two. Correlation statistics are used to determine relationships between variables or characteristics of variables. Regression statistics are used to make predictions given a set of independent variables (or factors). This set of advanced statistics builds on the previously learned Unit where the focus is on test of differences, such as the use of t test and ANOVA.

Learning Objectives

After working on this module, you should be able to:
1. Apply tests of independence given parametric and statistic data;
2. Compare more than two groups (or characteristics of a group) using analysis of variance;
3. Infer significant relationships between variables using correlational techniques; and
4. Predict relationships between variables using linear regression;

Underlying Concepts

- o   Test of independence

- o   Analysis of variance

- o   Correlational statistics

- o   Regression statistics

Module 7. Tests of Independence

Experimental studies, whether true experiment or quasi-experimental, are commonly designed to compare means. These may involve comparing a sample mean with a specific value, comparing before and after means of the same group, or comparing means of two groups or more than two groups. In this module, you will learn about testing single mean when the variance is known and estimated; and difference between means test when there are independent groups and when there are correlated pairs. This also includes computing for independence when variables are measured in the nominal level.

Learning Objectives

After working on this module, you should be able to:
1. Describe the uses of t-test
2. State the assumptions required in testing single means;
3. State the assumptions required in testing differences between two means;
4. Compute t-test and the probability of a sample mean when sigma is known and when sigma is estimated;
5. Compute t-test and the probability value for the difference between two means with independent groups; and
6. Compute t-test and the probability value for the difference between two means with correlated pairs.
7. Compute independence of variables when their level of measurement is nominal
8.

7.1 Testing Single Means

How to test the null hypothesis that the population mean is equal to some hypothesized value has been discussed in previous modules on normal distribution and sampling distribution of the mean. This significance test known as student t-test or t-test consists of computing the probability of a sample mean differing from μ by one (the difference between the hypothesized population mean and the sample mean) or more.

In testing single means, the first step is to determine the sampling distribution of the mean and then computing the test statistic.   When population standard deviation σ is known, the probability calculations are made based on the standard normal distribution. This was done by computing Z based on the formula,

$$Z = \frac{M - \mu}{\sigma_M}$$

where Z is the value on the standard normal distribution, M is the sample mean, μ is the hypothesized value of the mean, and σ_M is the standard error of the mean.

When σ is not known, it is estimated using t-distribution with the following formula,

$$t = \frac{M - \mu}{S_M}$$

where t is the value we compute for the significance test, M is the sample mean, μ is the hypothesized value of the population mean, and s_M is the estimated standard error of the mean.

The assumptions for testing single means include the following:
- Each value is sampled independently from each other value.
- The values are sampled from a normal distribution.

Read "Testing a Single Mean" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/tests_of_means/single_meanM.html

Study Guide Questions
1. Which test statistic is used in testing single means when population standard deviation is known?
2. Which test statistic is used in testing single means when population standard deviation is unknown?

Activity 7-1.

The scores of a random sample of 10 students on a Statistics test are as follows: 60, 62, 67, 69, 70, 72, 75, 78, 82, 90. Test to see if the sample mean is significantly different from 65 (which is the usual passing score) at 0.05 level of significance. Use the table below to complete your answer.

Steps in Hypothesis Testing	*Application to scenario*
1. State the null and alternative hypotheses	
2. Determine the level of significance	
3. Identify the test statistic to be used	
4. Determine the critical region	
5. Compute the test statistic and probability value	
6. Decide whether the result is significant or not	
7. Make conclusions about the research hypothesis	

7.2 Difference Between Two Means with Independent Groups

In order to do t-test to check whether there is a difference between population means, three assumptions need to be met:
- The two populations have the same variance (homogeneity of variance)
- The populations are normally distributed (normal distribution)
- Each value is sampled independently from each other value (independence)

Comparing means of independent groups also require at least interval level data in the dependent (outcome) variable. The independent variable should be nominal level or categories and considered as the basis for grouping the data.

Computing t- test is somewhat similar to computing Z-score as shown in the previous section on testing single means.

Read "Difference Between Two Means (Independent Groups)" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/tests_of_means/difference_meansM.html

Study Guide Questions
1. Which of the assumptions in testing difference between population means should never be violated however slightly?
2. What is the formula in computing the t statistics?
3. How is the degree of freedom computed in t-test?
4. *Where do you get the probability value for one tailed t-test? two-tailed t-test?*

Activity 7-2.

A study measured the time spent by physicians with average-weight and overweight patients. The mean and standard deviation for the 33 samples from each group is shown below:

	Physicians' time spent with average-weight patients	*Physicians' time spent with overweight patients*
Mean	*23.94*	*31.36*
Standard deviation	*10.14*	*9.86*

Compute for t-test and do the hypothesis testing using *above data at 0.05 level of significance.*

Steps in Hypothesis Testing	*Application to scenario*
1. State the null and alternative hypotheses	

2. Determine the level of significance	
3. Identify the test statistic to be used	
4. Determine the critical region	
5. Compute the test statistic and probability value	
6. Decide whether the result is significant or not	
7. Make conclusions about the research hypothesis	

7.3 Difference Between Two Means with Correlated Pairs

When two means being compared are taken from the same group, e.g., scores in a pre-test and post-test, the test statistic used is a special formula for t-test called correlated t-test or paired t-test.

The correlated t test is computed by first computing the differences between the two scores for each subject. Then, a test of a single mean is computed on the mean of these difference scores.

Read "Difference Between Two Means (Correlated Pairs)" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/tests_of_means/specific_relatedM.html

Study Guide Questions
1. Which assumption for testing difference of means does not apply to correlated pairs?
2. What is the formula in computing the t test for correlated pairs?
3. How is the degree of freedom computed in t-test for correlated pairs?
4. What is the relationship between the correlation and the standard error of the mean in correlated pairs?

Activity 7-3.

A (hypothetical) experiment is conducted on the effect of alcohol on perceptual motor ability. Ten subjects are each tested twice, once after having two drinks and once after having two glasses of water. The two tests were on two different days to give the alcohol a chance to wear off. Half of the subjects were given alcohol first and half were given water first. The scores of the 10 subjects are shown below. The first number for each subject is the performance in the "water" condition. Higher scores reflect better performance. Test to see if alcohol had a significant effect.

Effect of water	*Effect of alcohol*	*Difference between water and alcohol*
16	13	-3
15	13	-2
11	10	-1
20	18	-2
19	17	-2
14	11	-3
13	10	-3
15	15	0
14	11	-3
16	16	0

Compute for t-test and do the hypothesis testing using *above data at 0.05 level of significance.*

Steps in Hypothesis Testing	*Application to scenario*
1. State the null and alternative hypotheses	
2. Determine the level of significance	
3. Identify the test statistic to be used	
4. Determine the critical region	
5. Compute the test statistic and probability value	

6. Decide whether the result is significant or not	
7. Make conclusions about the research hypothesis	

7.4.   Test of independence with nominal variables

When dealing with two nominal variables and you want to see whether the proportion of one variable is different from the other, use Chi-square test of independence. This Chi-square test of independence is used only when the sample size is large (greater than 1000). For small sample size, Fischer's exact test is more accurate. The chi-square test of independence, like other tests of independence, assumes that the individual observations are independent.

As an example, a study wanted to know if dengue vaccine can cause severe reaction in children aged 3-7 years old depending on injection site. The nominal variables are the injection site (arm vs. thigh) and presence of severe reaction (yes vs. no).

	Severe reaction		% severe reaction
Injection Site	Yes	No	
Arm	76	8840	0.85%
Thigh	30	4758	0.63%

This cross-tabulation shows that there is a higher proportion of severe reactions in children who received the vaccine in the arm.

Hypothesis testing can also be done with the null hypothesis stated as "the relative proportions of one variable are independent of the second variable". In the given example, this null hypothesis is stated as: "the proportion of children given vaccine in the arm who have severe reactions is equal to the proportion of children given vaccine in the thigh who have severe reactions.

For more readings on Chi-square test of independence, read
https://stats.libretexts.org/Bookshelves/Applied_Statistics/Book%3A_Biological_Statistics_(McDonald)/02%3A_Tests_for_Nominal_Variables/2.05%3A_Chi-square_Test_of_Independence

Module 8. Analysis of Variance

Comparing means of more than two groups require another set of test statistics other than t-test. This is Analysis of Variance or ANOVA. This module describes the uses, assumptions and computations of ANOVA and how these are applied in health and nursing studies.

Learning Objectives

After working on this module, you should be able to:
1. Describe the uses of ANOVA
2. Define the different types of ANOVA
3. State the assumptions of one-way ANOVA
4. Compute for F-statistics and the probability value in ANOVA.

8.1 Uses of ANOVA

Analysis of Variance (ANOVA) is a statistical method used to test differences between two or more means. Inferences about the means are made by analyzing variances between and within groups. ANOVA is used to test general rather than specific differences among means.

Read "Introduction to Analysis of Variance" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/analysis_of_variance/introM.html

Study Guide Questions
1. What is the null hypothesis tested by analysis of variance?
2. Why is ANOVA preferable to t-test (comparing two means at a time) when comparing means of more than two groups?
3. Why is Tukey HSD test is preferable to ANOVA in certain situations?
4. Why is ANOVA more commonly used than others considered as post hoc test such as Tukey test?

8.2 Different Types of ANOVA

There are many types of ANOVA, depending on the number of factors and/or levels. Hence, there is a one-way ANOVA if there is only one factor and two-way ANOVA if there are two factors. When different subjects are used for the levels of a factor, the factor is called a between-subjects factor or a between-subjects variable. When the same subjects are used for the levels of a factor, the factor is called a within-subjects factor or a within-subjects variable.

Read "Analysis of Variance Designs" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/analysis_of_variance/anova_designsM.html

Study Guide Questions
1. In describing an ANOVA design, the term factor is a synonym of dependent or independent

variable?
2. Within-subjects variables in ANOVA are sometimes referred to as _____ since these are _____ measurements of the same subjects.
3. What is the ANOVA design called when all combinations of levels and factors are included?

8.3 Assumptions of ANOVA

The assumptions for ANOVA are the same as for a t-test of differences between means except that they apply to two or more groups.

- The populations have the same variance (homogeneity of variance)
- The populations are normally distributed (normal distribution)
- Each value is sampled independently from each other value (independence)

Read "One Factor ANOVA (Between Subjects)" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/analysis_of_variance/one-wayM.html

8.4 Computing for One-way ANOVA

The standard method for determining the probability is based on the ratio of Mean Square Between (MSB) to Mean Square Error (MSE). This ratio is named after R. Fisher and is called the F ratio.

MSE is computed as the mean of the sample variances, whereas MSB is computed by multiplying the variance of the sample means by n (number of observations in each group).

The shape of the F distribution depends on the sample size, that is, it depends on the two degrees of freedom (df) parameters: one for the numerator (MSB) and one for the denominator (MSE).

MSB ➜ dfnumerator   = k-1
MSE ➜ dfdenominator = N-k

Read "One Factor ANOVA (Between Subjects)" and "Within-Subjects ANOVA" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the videos of this resource:
http://www.onlinestatbook.com/2/analysis_of_variance/one-wayM.html
http://www.onlinestatbook.com/2/analysis_of_variance/within-subjectsM.html

One of the important characteristics of ANOVA is that it partitions the variation into its various sources. In ANOVA, the term **Sum of Squares** (SSQ) is used to indicate variation. The total variation is defined as the sum of squared differences between each score and the mean of all subjects.

The **Analysis of Variance Summary Table** (an example shown below) is a convenient way to summarize the partitioning of the variance.

Source of Variation	Degrees of freedom	Sum of Squares	Mean Square	F ratio	p value
Condition	3	27.53	9.18	3.465	0.0182
Error	132	349.65	2.65		
Total	135	377.19			

ANOVA designs could be about **between-subjects factors** for which each subject's performance is measured only once and the comparisons are among different groups of subjects or **within-subjects factor** is sometimes referred to as a repeated-measures factor since repeated measurements are taken on each subject.

Study Guide Questions
1. What is the advantage of within-subjects design over between-subjects design?
2. What are the two components of the total sum of squares in one-factor between-subjects design?
3. What source of variation is found in an ANOVA summary table for a within-subjects design that is not in an ANOVA summary table for a between-subjects design?
4. When do MSB and MSE estimate the same quantity in ANOVA?

Activity 8 -1

Give the source and df columns of the ANOVA summary table for the following experiments:

Research scenario	Source of variation	Degrees of freedom
1. Forty nurses are each tested on a simple reaction time task and on a choice reaction time task.		
2. Twenty male and 20 female patients are each tested under three levels of drug dosage: 5 mg, 10 mg, and 20 mg.		
3. Thirty hypertensive patients are tested for blood pressure for three trials a day for two days after a lifestyle modification intervention		
4. An experiment is conducted in which post-operative patients are either assigned to a drug therapy group, a behavioral therapy group, or a control group for pain intervention. Ten subjects are assigned to each group. The level of measured once a month for four months.		

Module 9. Correlation

9.1 Basic Concepts in Correlation

In nursing research, we are often interested in describing relationships between two or more variables. The use of correlation analysis helps us establish statistical relationships between variables. Statistical relationships are based on two concepts: covariance and independence. **Covariance** means a change in the value of independent variable would entail similar change in the dependent variable. **Independence** is the opposite of covariance; i.e., there is no change in the dependent variable even if there is a change in the independent variable.

Read "Introduction to Bivariate Data" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource from:
http://www.onlinestatbook.com/2/describing_bivariate_data/introM.html

This section introduces describing and finding a statistical relationship between two quantitative variables, otherwise known as "bivariate data". Read the example on finding correlation between ages of husbands and wives.

Activity 9-1

Look for research studies in nursing that explores relationship between variables or determines correlation of variables between variables. Then try to answer the following questions:

1. What is/are the research objective/s?
2. What are the variables being explored?
3. What is/are the hypotheses of the study?
4. What are the correlational statistics used in the study?
5. What are the conclusions of the study in answer to the research hypotheses and objectives?

The type of correlation analysis to be used depends on the level of measurement of the variables. The table below shows the common measures of association/ correlation for each level of measurement. Correlations are either -1 or +1 with negative numbers indicating negative relationships, while positive numbers indicate positive relationships. The measure of association is equal to zero if the variables are independent of each other.

Correlational Measures	Measurement Level of Data	High Correlation	Independence
Pearson's product moment correlation	Interval/ Ratio	+1, -1	0
Spearman's rho	Ordinal	+1, -1	0
Chi-square test	Nominal	Infinity	0

A strong correlation, which shows that two variables are related does not mean that one variable caused the other. It is wrong to infer causation based on correlation analysis alone. There is the issue of **temporal ambiguity**, that is, whether the dependent variable really comes after the existence of the independent variables. Another reason might be the presence of **confounders** that are not controlled in the study. A confounder is a risk factor of the dependent variable at the same time related to the independent variable.

12

9.2 Pearson Product Moment Correlation

Finding statistical relationships between variables that are measured at the level of interval or ratio require Pearson product moment correlation or Pearson's r.

9.2.1   Properties of Pearson's r

Pearson's r allows us to define statistical relationships between interval or ratio variables. The correlation coefficient ranges from -1 to +1. The plus (+) or minus (-) sign indicate the type of relationship between the variables. The nearer to 1 is the correlation coefficient, the stronger the relationship.

Read "Properties of r" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource from:
http://www.onlinestatbook.com/2/describing_bivariate_data/propertiesM.html

After reading the resource, answer the following study guide questions.

Study Guide Questions
1.   What is Pearson product moment correlation?
2.   What does Pearson's r of 0 mean?
3.   What does Pearson's r of -1 mean?
4.   What does Pearson's r of +1 mean?

9.2.2   **Assumptions in Pearson's r**

The use of Pearson's r requires that certain assumptions must be met. In the following link, read these assumptions.

https://statistics.laerd.com/spss-tutorials/pearsons-product-moment-correlation-using-spss-statistics.php

These assumptions are very important. If the data that you have does not satisfy these assumptions, the use of Pearson's r would not be valid.

Activity 9-2

To check whether you understood the assumptions, fill in the table below.

Assumption	Example
1.	Requires variables measured in _____, such as weight, height, exam score
2.	Distribution of the variables must approximate the normal curve
3.	Single data points within the data that do not follow the usual pattern
4.	When variables are graphed, they should tend to form a straight line, not necessary lying on the line but scattered close to it

9.2.3   **Computing Pearson's r**

There are several formulas and ways of computing Pearson's r. You can do a manual computation or use statistical software. If you are doing a manual computation, refer to the following link:

Read "Computing Pearson's r" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/   or   access   the   video: http://www.onlinestatbook.com/2/describing_bivariate_data/calculationM.html

You can also use MS Excel in computing for Pearson's r.
https://www.youtube.com/watch?v=s2TVkYmmCAs

If you want to use SPSS in computing Pearson's r, you may find the following link useful:
https://statistics.laerd.com/spss-tutorials/pearsons-product-moment-correlation-using-spss-statistics.php

To test whether you were able to get the computations, try to do the following activity

Activity 9-3

Determine Pearson's r to find if there is a relationship between mother's age and infant's birth weight using the data below.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Mother's Age	4	2	6	7	8	9	12	11	8	2
Infant's Birthweight	6.5	4.8	7.8	8.2	8.8	9.0	10.2	9.4	8.4	4.9

9.2.4   **Interpreting Pearson's r**

As previously discussed, Pearson's r would range from -1 to +1. The sign refers to the direction of the relationship; whether positive (+) or negative (-). The number refers to the magnitude of the relationship; the nearer to one (1) to stronger the relationship and the nearer to zero (0), the weaker the relationship.

Refer to this resource to find out more about interpreting correlation coefficients:

http://www.psychology.emory.edu/clinical/bliwise/Tutorials/SCATTER/scatterplots/effect.htm

Activity 9-4

As an activity, try to interpret the following correlation coefficients by determining their direction and magnitude (strength).

- 0.38	0.23	-0.50
0.84	-1.00	0.11
- 0.06	0.62	-0.75

To check whether your answers are correct, click on the following link:
http://www.psychology.emory.edu/clinical/bliwise/Tutorials/SCATTER/scatterplots/corr.htm

9.2.5   **Finding statistical significance of Pearson's r**

Although Pearson's r or correlation coefficient can already give us meaningful information about the relationship of variables, it is still not complete to be able to infer whether there is a statistical relationship or not. In order to do this, you need to apply hypothesis testing. Refer to the following resource uploaded in your course site.

Kuan, L.G., & Bonito, S.R. & UPOU (2001). N298: Statistical methods applied in nursing. University of the Philippines Open University, Manila. pp 204-207.

Activity 9-4

Do hypothesis testing of the Pearson's r computed in Activity 14-2 to determine if there is a statistical significance of the correlation between mother's age and infant's birth weight. Use the following table to show the steps:

Steps in Hypothesis Testing	Application to scenario
1. State the null and alternative hypotheses	
2. Determine the level of significance	
3. Identify the test statistic to be used	
4. Determine the critical region	
5. Compute the test statistic and the probability value	
6. Decide whether the result is significant or not	
7. Make conclusions about the research hypothesis	

As a summary of this module, read the following link:
http://www.statstutor.ac.uk/resources/uploaded/pearsoncorrelation3.pdf

9.3 Spearman's rho

Finding statistical relationships between variables that are measured at the level of ordinal require Spearman's rho.

9.3.1   Use of Spearman's Rho

The Spearman's rho or coefficient of rank-order correlation is one of the nonparametric statistical analyses used when correlating two ordinal variables. Spearman's rho assumes that the variables under consideration were measured on at least an ordinal (rank order) scale. Variables measured at interval or ratio level can also sometimes transformed into ordinal variables. For example, exam scores can be transformed into ranks from highest to lowest or life satisfaction scales can be ranked from very satisfactory to unsatisfactory.

9.3.2   Assumptions in Spearman's rho

The ordinal measures for Spearman rho is Spearman's rho is appropriate whenever the relationship between X and Y is monotonically increasing or monotonically decreasing. In monotonically increasing function, as X increases, Y does not decrease but increases or remains constant. In monotonic decreasing function, as X increases, Y does not increase but decreases or remains constant. This property of monotonicity is appropriate in ordinal variables since they cannot always assume linearity. Ordinal variables can have linear or nonlinear relationships. Normality is not an assumption in Spearman's rho.

Activity 9-5

Compare the assumptions between Pearson's r and Spearman's rho by putting a check mark (✓) in the appropriate columns in the table.

Assumptions	Pearson's r	Spearman's rho
1. ordinal variables		
2. interval variables		
3. normal distribution		
4. monotonic function		
5. linearity		

9.3.3   Computing for Spearman's rho

Spearman's rho is not so different from Pearson's r. In fact, it is a derivation from Pearson's r applied to ranks rather than raw scores. Spearman's rho is computed from ranks. The individual observations (cases) can be ranked into two ordered series.

To do a manual computation of Spearman's rho, refer to the following resource uploaded in your course site. Kuan, L.G., & Bonito, S.R. & UPOU (2001). *N298: Statistical methods applied in nursing*. University of the Philippines Open University, Manila. pp 218-229.

You can also use MS Excel in computing for Spearman's rho.
https://www.youtube.com/watch?v=kt8kmqPDuK4

If you want to use SPSS in computing Spearman's rho, you may find the following link useful:
https://www.google.com/search?q=spearman%27s+rho+excel&ie=utf-8&oe=utf-8#q=spearman%27s+rho+spss

Activity 9-6.

Find out if there is a correlation between personality and intelligence ranking in a local beauty pageant. You can do a manual computation or use statistical software.

	Contestant 1	*Contestant 2*	*Contestant 3*	*Contestant 4*	*Contestant 5*
Personality	3	5	2	4	1
Intelligence	2	3	1	5	4

9.3.4 Interpreting Spearman's rho

Spearman's rho can be interpreted as the product moment correlation between the ranks on X and the ranks on Y. Just like Pearson's r, Spearman's rho may range from -1 to +1. Plus one (+1) indicates perfect agreement between the ranked variables while negative one (-1) indicates perfect disagreement.

Refer to the following resource to understand more about the interpretation of Spearman's rho
http://www.statstutor.ac.uk/resources/uploaded/spearmans.pdf

Study Guide Questions
1. What is similar in the interpretation of Pearson's r and Spearman's rho?
2. What is different in the interpretation of Pearson's r and Spearman's rho?

9.3.5 Finding statistical significance in Spearman's rho

Finding statistical significance in Spearman's rho is similar to the process of hypothesis testing in Pearson's r.

Refer again to the following resource uploaded in your course site to see the step-by-step process of hypothesis testing. Kuan, L.G., & Bonito, S.R. & UPOU (2001). *N298: Statistical methods applied in nursing*. University of the Philippines Open University, Manila. pp 218-229.

Activity 9-7.

Continuing the example given in Activity 16-3, find out if there is a statistical significance in the computed Spearman's rho.

Steps in Hypothesis Testing	Application to scenario
1. State the null and alternative hypotheses	
2. Determine the level of significance	
3. Identify the test statistic to be used	

4. Determine the critical region	
5. Compute the test statistic and the probability value	
6. Decide whether the result is significant or not	
7. Make conclusions about the research hypothesis	

9.4. Use of Chi-square test

Finding statistical relationships between variables that are at nominal level of measurement requires Chi-square test.

Chi-square test of association is a non-parametric technique used to measure the relationship between two nominal level variables. The variables may have two levels (dichotomous) or multiple levels (more than 2 levels or categories). Chi-square is based on the difference between the observed and the expected frequency. Large differences between the observed and expected frequency lead to the rejection of null hypothesis that there is no relationship between the variables being considered.

9.4.1   Assumptions in Chi-square test

The type of data required in Chi-square test is frequency data or variables measured at nominal level. Variables that are measured at higher-level scale, such as ordinal or interval, may be transformed into nominal provided that there is a theoretical basis for it. For example, actual blood cholesterol level can be categorized as high, normal, or low. Measures must be independent of each other, which mean the categories are mutually exclusive. There should be adequate sample size. Small sample size would lead to invalid results. The researcher needs to know the sample size necessary to attain the desired power for the specified significance level and hypothesized population effect size.

Activity 9-8

Complete the following sentences that give the assumptions for Chi-square test.
1. Only _____ data can be used in Chi-square test
2. There should be _____ basis for categorizing data
3. The categories should be _____ to satisfy independence.
4. _____ sample size is needed to ensure valid results.

9.4.2   Computing for Chi-square test

Computing for Chi-square can be done manually or with the use of statistical software.

Read "Contingency Tables" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/   or   access   the   video: http://www.onlinestatbook.com/2/chi_square/contingencyM.html

You can also refer to the following resource to do a manual computation of Chi-square test, refer to the following resource uploaded in your course site.

Kuan, L.G., & Bonito, S.R. & UPOU (2001). *N298: Statistical methods applied in nursing*. University of the Philippines Open University, Manila. pp 240-243.

You can also use MS Excel in computing for Chi-square.
https://www.youtube.com/watch?v=offtAejQm_c

If you want to use SPSS in computing Chi-square, you may find the following link useful:
https://www.youtube.com/watch?v=Ahs8jS5mJKk

9.4.3   Interpreting Chi-square test

Chi-square test result has no specific meaning unless you take a step further and use it in hypothesis testing. See below for the step-by-step process of finding statistical significance in Chi-square test.

9.4.4   Finding statistical significance in Chi-square test

Read the following resource (uploaded in the course site) to understand the hypothesis testing involving Chi-square test.

Kuan, L.G., & Bonito, S.R. & UPOU (2001). *N298: Statistical methods applied in nursing*. University of the Philippines Open University, Manila. pp 240-243.

You will also find in the above resource that there are correlational techniques for nominal variables that can be computed from the chi-square test of association. Find out more phi coefficient (ϕ), cramer's ν, and contingency coefficient.

Module 10. Simple Linear Regression Analysis

In correlation analysis, you were able to determine the magnitude and direction of relationship between independent and dependent variables. In regression analysis, the aim is to be able to predict outcomes and explain inter-relationships between variables. This module presents the concepts and application of simple linear regression.

Learning Objectives

After working on this module, you should be able to:
1. Describe the use of Simple Linear Regression Analysis;
2. Discuss the assumptions in using Simple Linear Regression Analysis;
3. Compute for estimates of Simple Linear Regression;
4. Test the statistical significance of regression equation; and
5. Interpret results of Simple Regression Analysis

10.1 Use of Simple Linear Regression Analysis

Regression analysis makes use of the correlation between variables to develop a prediction equation. The prediction equation allows you to predict the score of the dependent variables based on the scores of the independent variables.

Read "Introduction to Linear Regression" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/regression/introM.html

Study Guide Questions
3. What do you call the variable that we are predicting (referred to as Y)?
4. What do you call the variable from which you are basing your prediction (referred to as X)?
5. What do you call the best fitting line?
6. What do you call the value of the point minus the predicted value?
7. What is the most commonly used criterion for the best fitting line?

Activity 10-1.
Think of a research objective where you want to do a prediction. Identify what is the predictor variable and what is the criterion variable in this study.

    Research objective: _____
    Predictor variable: _____
    Criterion variable: _____

10.2 Assumptions in Simple Linear Regression Analysis

Computing the estimates for simple linear regression does not need assumptions; however, in doing statistical testing, the following assumptions must be met:
1. The sample must be representative of the population to which the inference will be made *(generalizability)*
2. The errors of prediction are distributed normally. This means that the distributions of deviations from the regression line are normally distributed *(normality)*
3. The variance of Y is the same as for any value of X *(homoscedasticity)*
4. The value of Y at one value of X does not depend on, and is not affected by the value of Y at another value of X *(independence)*
5. The relationship between X and Y must be linear, i.e., when the two scores for each individual are graphed, they should tend to form a straight line; not necessarily lying on the line but scattered close to it *(linearity)*

10.3 Computing for Regression Estimates

Linear regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a regression line. Computing for regression estimates can be done easily by statistical software. However, this could be done also manually.

Read "Introduction to Linear Regression" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/regression/introM.html

Study Guide Questions
5. What is the formula for a regression line?
6. What is the meaning of Y' (pronounced as Y prime) in the above formula?
7. What does b stands for?
8. What does A (or a) stands for?
9. What happens to the regression equation when the variables are standardized?

Another resource that you might find useful is the following. This will be uploaded in the course site. You can read through this resource and find out more about interpreting regression results.

Kuan, L.G., & Bonito, S.R. & UPOU (2001). *N298: Statistical methods applied in nursing*. University of the Philippines Open University, Manila. pp 252-257.

Activity 10-2.

Try to compute for the regression estimates of the following data (hypothetical data).
Find out whether the mother's weight during her last prenatal check-up predicts the weight of her baby during delivery. Fill in the table to do this manually. Then you can use statistical software to check your computations.

Mother's weight	Infant's birthweight	X^2	Y^2	XY
64	2.5			
58	2.8			
55	2.5			
62	3.8			
62	3.2			
70	3.1			
68	3.5			
66	3.7			
62	2.4			
62	3.3			
65	3.8			

10.4 Finding Statistical Significance of Regression Equation

This section shows how to conduct significance tests for the regression equation. Significance testing of the regression equation depends on the regression slope and correlation of the variables. If the regression slope is significantly different from zero, then the correlation coefficient is also significantly different from zero.

Read "Inferential Statistics for b and r" from Online Statistics Education: A Multimedia Course of Study from http://onlinestatbook.com/

You can also watch the video of this resource:
http://www.onlinestatbook.com/2/regression/inferentialM.html

Study Guide Questions
1. What is the null hypothesis for testing significance in regression?
2. What is the test statistic in hypothesis testing for regression?
3. How do you interpret the results of hypothesis testing in regression?

You can also refer to the following resource to see the step-by-step process in hypothesis testing.

Kuan, L.G., & Bonito, S.R. & UPOU (2001). *N298: Statistical methods applied in nursing*. University of the Philippines Open University, Manila. pp 252-257.

Activity 10-3.

Compute for t-test and do the hypothesis testing *at 0.05 level of significance* using the *data* in 10.2

Steps in Hypothesis Testing	Application to scenario
1. State the null and alternative hypotheses	
2. Determine the level of significance	
3. Identify the test statistic to be used	
4. Determine the critical region	
5. Compute the test statistic and probability value	
6. Decide whether the result is significant or not	
7. Make conclusions about the research hypothesis	

10.5 Interpreting Results of Regression Equation

Refer to the following resource to find out about interpreting results in regression.

Kuan, L.G., & Bonito, S.R. & UPOU (2001). *N298: Statistical methods applied in nursing*. University of the Philippines Open University, Manila. pp 252-257.

Study Guide Questions
1. What is the meaning of the computed regression of coefficient (b)?
2. What is the meaning of the intercept constant (a)?
3. What is the meaning of the coefficient of determination (R^2)?
4. What is the meaning of significant results in regression