BASIC STATISTICS

Convenience Samples

Constant

If everyone in your study gives you the same answer to a specific question, then that column of data is a constant…it is constant across people.

Variable

If people in your study give you different answers to a particular question, then the column of data is a variable….it varies across people.

Classification of Variables

Nominal Variable – the numbers or categories have no inherent order. They are used only to distinguish one group from another.

Ordinal Variable – the numbers or categories can be ranked, but it would not be appropriate to average the numbers or categories. You could calculate the median.

Nominal and ordinal variables are considered **qualitative variables**.

Interval Variable – the numbers can be ranked AND it would be appropriate to average the numbers. The scale used to measure the interval variable can be changed without loss of information. The differences between numbers can be interpreted.

Ratio Variable – the numbers can be ranked AND it would be appropriate to average the numbers. The difference between numbers can be interpreted. The scale used to measure the ratio variable has a meaningful zero.

Interval and ratio variables are considered **quantitative variables**.

A mini questionnaire is given below. Indicate whether the responses from each question would be considered a nominal, ordinal, interval, or ratio variable.

What is your major?

What is your current gpa?

Have you completed all your general studies requirements? Yes No

What is your current classification? Freshman Sophomore Junior Senior

How many hours per week do you use the MTSU library?

How many MTSU football games have you attended this year? None 1 to 2 3 or more

How many MTSU softball games have you attended this year? None 1 to 2 3 or more

What year were you born? _____

What is your gender? Female Male

What is your age? _____

What is your age range?

1 to 18 19 to 25 26 to 49 50+

What is your annual salary? _____

What is your IQ score? _____

What is your current relationship status? Single Living as married Married Separated

What level of satisfaction best describes your current relationship status?

	1 Very Dissatisfied	2	3	4	5 Very Satisfied	

What level of satisfaction best describes your current relationship status?

	-10 Very Dissatisfied	-5	0	5	10 Very Satisfied	

What level of satisfaction best describes your current relationship status?

Very dissatisfied

Somewhat dissatisfied

Neither satisfied or dissatisfied

Somewhat satisfied

Very satisfied

Cleaning and Summarizing Data

Purpose – Data Cleaning, Graphs, and Descriptive Statistics for a Single Variable

Type of Variable	**Common Analysis Options**	
Nominal Variables	Frequency Table Bar Chart Pie Chart Mode	
Ordinal Variables	Frequency Table Bar Chart Pie Chart Box Plot Mode Median	Range Interquartile Range Semi-interquartile Range Maximum Minimum
Interval and Ratio Variables	Histogram Box Plot Mode Mean Variance Standard Deviation Skewness Kurtosis Minimum Maximum	**If the data are markedly skewed…** Median Range Interquartile Range Semi-interquartile Range

Purpose – Describing the Relationship Between Two Variables

Variables	**Common Analysis Options**
Two Nominal variables	Cross-tabulations Bar Chart by Group Pie Chart by Group Mode by Group Correlation
One nominal variable; one ordinal, interval, or ratio variable	Correlation For each NOMINAL group, consider the following…
Two variables that are ordinal, interval, or ratio	Correlation Scatterplot
Three or more quantitative variables	Correlation Matrix Scatter Matrix

Histogram Box Plot Mode Mean Variance Standard Deviation Skewness Kurtosis Minimum Maximum	**If the data are markedly skewed…** Median Range Interquartile Range Semi-interquartile Range

Descriptive Statistics

Mean: It is the mathematical average of the scores.

$$\text{Sample Mean} = \frac{\text{Sum of the Scores}}{\text{Number of Scores}} \qquad \bar{Y} = \frac{\sum Y}{n}$$

Faculty Salaries	Number of Children
35	0
33	2
42	1
55	3
65	5
46	1
46	
50	
42	
Calculate the sample mean, $\bar{Y}$.	Calculate the sample mean, $\bar{Y}$.
Interpret the sample mean, $\bar{Y}$.	Interpret the sample mean, $\bar{Y}$.

Characteristics of the Mean include:

- It makes sense to calculate a mean for quantitative variables, but not for qualitative variables.
- The mean can be distorted by outliers.
- $\sum(Y - \bar{Y}) = 0$. The average difference from the mean is zero. Some scores will be above the mean and some scores will be below the mean, but the differences will cancel each other out.

Variance: The variance measures the tendency for scores to depart from each other and from the mean. It is the "average" squared amount that the scores differ from the mean.

$$\text{Sample Variance} = \frac{\text{Sum}\left(\text{Score - Mean}\right)^{\text{Squared}}}{\text{Number of Scores - 1}}$$

$$S^2 = \frac{\sum\left(Y - \bar{Y}\right)^2}{n-1}$$

If we were calculating the variance for everyone in the population, then we would divide by *n* instead of *n – 1*.

Standard Deviation: The standard deviation measures the tendency for scores to depart from each other and from the mean. It is the "average" amount that the scores differ from the mean.

$$\text{Sample Standard Deviation} = \sqrt{\text{Sample Variance}}$$

Faculty Salaries (in thousands of dollars)		
Score Y	*Score – Mean* $(Y - \bar{Y})$	*(Score – Mean)*^{2} $(Y - \bar{Y})^2$
35 33 42 55 65		
Mean =		*SS* =

Calculate the variance, S^2.

Interpret the variance, S^2.

Calculate the standard deviation, *S*.

Interpret the standard deviation, *S*.

Number of Children		
Score Y	*Score – Mean* $(Y - \bar{Y})$	*(Score – Mean)*^{2} $(Y - \bar{Y})^2$
0 2 1 3 5 1		
Mean =		*SS* =

Calculate the variance, S^2.

Interpret the variance, S^2.

Calculate the standard deviation, *S*.

Interpret the standard deviation, *S*.

Variance	*Standard Deviation*
❑ The variance = 0 when all the scores are the same. ❑ The variance is larger than zero when the scores differ. ❑ The variance is not influenced by the mean. ❑ The variance is in ***squared units.***	❑ The SD = 0 when all the scores are the same. ❑ The SD is larger than 0 when the scores diffe ❑ The SD is not influenced by the mean. ❑ The SD is in the original ***units***.

SPSS Instructions
→Analyze→Compare Means→Means
Dependent List: FreshGPA
Independent List: Gender
→Options
Cell Statistics: Mean, Number of Cases, Standard Deviation,
Variance, Minimum, Maximum

Means

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
freshgpa * gender	30	69.8%	13	30.2%	43	100.0%

Report

freshgpa

gender	Mean	N	Std. Deviation	Variance	Minimum	Maximum
Female	2.1066	15	.60818	.370	1.13	3.01
Male	2.3530	15	.40882	.167	1.68	3.08
Total	2.2298	30	.52436	.275	1.13	3.08

SPSS Instructions
→Analyze→Descriptive Statistics→Descriptives
Variables: FreshGPA
→Options
√ Mean √ Standard Deviation √ Variance √ Minimum √ Maximum

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
freshgpa	43	1.13	3.63	2.4571	.57991	.336
Valid N (listwise)	43					

A random sample is expected to differ from the population because of "chance".

Sampling Error. The discrepancy, or amount of error, that exists between a sample statistic and the corresponding population parameter.

Age

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Sample Mean	Sample Mean	Sample Mean	Sample Mean	Sample Mean

Population Mean =

Samples will differ from each other because of random selection (i.e., chance).

The sample means will differ from the population mean because of the random selection used to obtain the samples.

More generally, statistics (sample mean, sample standard deviation, etc.) will differ from the population parameters (population mean, population standard deviation, etc.) because of the random variability among the samples.

Statistical methods are used to determine whether observed differences among groups are likely due to the random selection process (i.e., there are no *population* differences) or due to actual differences in the populations.

Analyzing Data

Commonly Used Methods for the Behavioral Sciences

Purpose – Compare a sample of data to a theoretical or known population

Dependent Variable	*Analysis*
Nominal or Ordinal	Chi-square Goodness of Fit
Interval or Ratio	One Sample *Z* Test (if the population mean and population variance is known) One Sample *t* Test (if the population mean is known, but the population variance is unknown)

Purpose – Predict the Dependent Variable

Dependent Variable	*Predictor(s)*	*Analysis*
Nominal	Nominal	Chi-square Test of Independence *Z* Test of Proportions Logistic Regression
Nominal	Nominal, Ordinal, Interval, and Ratio	Logistic Regression
Ordinal	Nominal, Ordinal, Interval, and Ratio	Logistic Regression
Interval or Ratio	Nominal, Ordinal, Interval, and Ratio	Linear Regression (a.k.a., multiple regression, multiple linear regression)

Purpose – Compare the Group Means[1, 2, 3]

Dependent Variable	*Predictor(s)*	*Analysis*
Interval or Ratio	Nominal	T-test if only two groups Analysis of Variance (T-TEST, ANOVA or GLM)
Interval or Ratio	Nominal while controlling for an extraneous variable that is nominal	Analysis of Variance (GLM)
Interval or Ratio	Nominal while controlling for an extraneous variable that is ordinal, interval, or ratio	Analysis of Covariance (GLM)

[1]The analysis is often done via the general linear model, but an overachieving stat person could use the linear regression procedure to obtain *identical* mathematical results as the GLM procedure when there is only one dependent variable.

[2]If the dependent variable is measured more than once (e.g., pre-test/post-test) or if the group conditions reflect subtest scores then the analysis becomes a repeated measures t-test, anova, or ancova.

[3]If the researcher wants to compare groups on more than one dependent variable (e.g., Anxiety and Depression) at the same time, then it becomes a multivariate t-test, manova, or mancova.