Education Transfer Plan
ABSTRACT

ETP TITLE: Introducing New Topics in Elementary Statistics
FELLOW NAME: Alyson M. Clark
FELLOW PRIMARY E-MAIL: alyson_clark@westvalley.edu
SPONSOR COMPANY: Intel
MENTOR NAME: Lynette Caubin

ABSTRACT (50 words or less:)

After learning how statistics is used in manufacturing at Intel, and incorporating statistical methodology into the collection, analysis, and reporting of IP&M data, I realized I needed to re-evaluate what I teach in my Elementary Statistics class. Thus, I decided to extend a previous ETP ("Elementary Statistics Workbook," Synopsys, 2004) to include statistical methods widely used at Intel that are currently *not* part of our curriculum, and to teach these new methods using the data I worked with this summer.

ETP TYPE: Veteran—Extended ETP **GRADE LEVEL:** Community College
Subject: Math **Document Format(s):** ☒ Word doc ☐ PowerPoint ☐ Excel ☐ Other

Describe how your ETP aligns with NBC or State standard as stated in your proposal:

My ETP aligns with two of the National Board standards. First, it aligns with Standard X: Reflection and Growth – as part of my work this summer, I reflected on what I teach and how I teach it based on discussions with Intel statisticians. Second, my ETP aligns with Standard XII: Contributing to the Professional Community – I have collaborated with statisticians at Intel to gain their insight on strengthening my school's program and improving the practice of statistics in an industry setting.

Describe the connection between your ETP and the Summer Fellowship.

My summer fellowship project involved reviewing, researching, and evaluating how the IP&M team gathers and presents data. My objective was to identify gaps and recommend new ways to statistically gather, analyze, and present data. The objective of my ETP was to review the current curriculum for Elementary Statistics at my school, identified gaps in the topics taught, and created ways to present new material and practice it with real-life data from my summer fellowship.

Checklist for sections contained in ETP:

☒ Clearly Stated Outcomes & Standards	☒ List of Materials/Resources	☒ Hard Copy Turned In to Peer Coach
☒ Specific Procedure/Plan	☒ Rubric or plan for evaluating outcomes.	☒ Soft Copy Turned into Peer Coach

I. FELLOW SIGNATURE-- Required Before August 12, 2005

I, the IISME **Fellow** named above, affirm that the ETP I am submitting is my own work, that I acknowledged sources where appropriate, and that I avoided including any proprietary information of the Sponsor Company. By my submission I am assigning to IISME my entire copyright in the ETP. I understand IISME is simultaneously granting me a license to use the ETP for pedagogical purposes.

Signature: ______________________________ Date: August 3, 2005

II. MENTOR SIGNATURE-- Required Before August 12, 2005

I, the **Mentor** named below [please select one of the following],

☐ have read the attached ETP, and my comments, if any, appear below.

☐ have read the attached ETP, and, as outlined in the IISME-Company Fellowship Agreement, have reviewed it on behalf of the Sponsor Company, and have determined that the ETP does not contain any Sponsor-proprietary information. My additional comments, if any, appear below.

Mentor Comments:

Signature: ______________________________ Date: ______________

Printed Name: ______________________________

III. ADMINISTRATOR SIGNATURE-- Required Before October 3, 2005, submit to IISME on or before October 3 to be eligible for $300 grant.

I, the **Administrator** named below have read the attached ETP and my comments, if any, appear below.

Administrator comments:

Signature: ______________________________ Date: ______________

Printed Name: ______________________________

Introducing New Topics in Elementary Statistics

An Educational Transfer Plan
by
Alyson M. Clark

School:
West Valley College
Mathematics Department
14000 Fruitvale Ave.
Saratoga, CA 95070
(408) 741-2590
alyson_clark@westvalley.edu

Sponsor:
Lynette Caubin
Worldwide Operations Manager
Issue Prevention and Management

Intel Corporation
2200 Mission College Blvd.
Santa Clara, CA 95052

Table of Contents

Introduction

During my 2004 summer fellowship at Synopsys, I chose to create an Elementary Statistics workbook that would supplement lecture and the textbook on key concepts taught in a statistics course. This workbook compiled years of individual activities and handouts I had into a single, semester-long tool for the students to use both during and outside of class. Motivated by the work I completed as part of my ETP last summer, I also put my lecture notes into PowerPoint presentation form. With these new tools I now teach Elementary Statistics using a computer and projector, while easily referring to practice problems in the workbook. All of this work has proven to be a much more effective and dynamic way of teaching than my usual chalk-and-chalkboard method of teaching.

During my fellowship this summer at Intel, I spent time interviewing statisticians about the specific methodologies they would like employees to know, and the technology skills they would like employees to have with respect to statistics. As a result of these conversations, I was forced to re-think the current curriculum I teach so that it would better serve my students, especially if they were to work for a technical company like Intel. Thus, I decided to extend my previous ETP work to include the topics in statistics highlighted by the Intel statisticians that are currently not part of our curriculum. These topics included levels of measurement, control charts, normal probability plots, and the chi-square test.

The goal of my ETP this summer was to develop the lecture notes and corresponding practice problems for the workbook to teach these new topics, with the practice problems coming specifically from the data sets I worked with as part of my fellowship. This aligned with two National Board standards: (1) Standard X: Reflection and Growth – I am reflecting on what I teach and how I teach it; and (2) Standard XII: Contributing to the Professional Community – I am strengthening my school's math program and improving the practice of statistics in an industry setting through my collaboration with Intel's statisticians.

Implementing this ETP will introduce several challenges. First, I will need to find a way to modify an already-full curriculum to allow time for these new topics. This will involve reducing or omitting topics that are currently taught, or finding ways to save time during class without hurting the students. Second, I will need to convince my department that these are the right topics to add to the curriculum. This will involve justifying that these topics are more useful to our students and might even be more interesting, especially when tied to data from my summer fellowship. Finally, I will need the resources necessary to carry out my ETP. This involves a lot of paper and photocopying for the workbook, and a room with a computer and projector for lecture. Ultimately, however, I believe I can face these challenges, deal with them, and move on to create a statistics curriculum that will best serve the students we teach.

Elementary Statistics – Curriculum Modification

Omissions or Modifications to Curriculum

- Omit survey methods (Section 1.3)
- Omit experimental design (Section 1.3)
- Spend very little time on basic graphs (Section 2.1)
- Use calculator for most, if not all, summary statistics (Chapters 3, 5)
  - → Mean, median, standard deviation, probability distribution parameters

Additions to Curriculum

- Levels of measurement (Section 1.1, pgs. 7 – 10)
- Control charts (Section 6.1, pgs. 298 – 299)
- Normal probability plots ("Using Technology" at end of Chapter 6, pgs. 353 – 354)
- Chi-square test (Section 11.1, pgs. 671 – 685)

Sections to Cover Under Changes to Curriculum

- 1.1 – 1.2
- 2.1 (brief), 2.2 – 2.3 (full)
- 3.1 – 3.2, 3.4 (all sections use calculator)
- 4.1
- 5.1 (use calculator)
- 6.1 – 6.3, Ch. 6 "Using Technology"
- 7.1 – 7.2
- 8.1 – 8.5
- 9.1 – 9.5
- 10.1 – 10.3
- 11.1

Classroom Modifications

Put lecture notes into handout form for the students; however, leave pieces blank (like the definition of a term, or the formula of a statistic) so that the students still have to attend class and follow along during lecture.

Lecture Notes

<u>Levels of Measurement</u>
(Section 1.1 – What is Statistics?)

- One way to categorized data is qualitative or quantitative
- Another way to classify data is by *levels of measurement*
  - → Nominal, ordinal, interval, ratio
- Nominal – data that consists of names, labels, or categories
- Ordinal – data that may be arranged in order (but differences in data values are meaningless)
- Interval – data that can be arranged in order (and differences are meaningful)
- Ratio – data that can be arranged in order (both differences between and ratios of data are meaningful – zero is starting point for all measurements)
- To determine the level of measurement, state the highest level that can be justified

Lowest	⟶		**Highest**
Nominal	Ordinal	Interval	Ratio

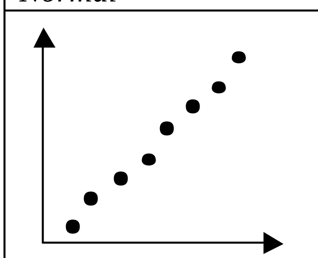
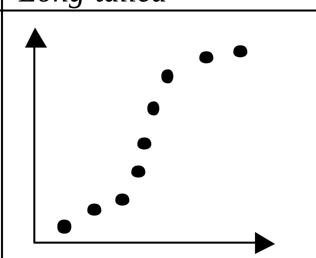
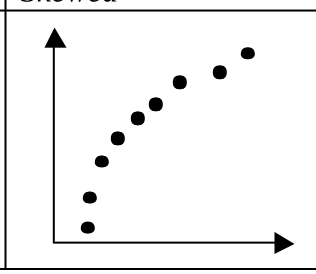
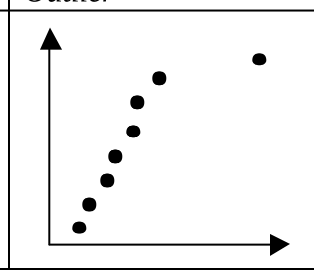
<u>Control Charts</u>
(Section 6.1 – Graphs of Normal Probability Distributions)

- Recall
  - → Time-series graph shows data values over time to look for trends ("trend chart")
- There exists variability in any sequential set of data, or data collected over time
- To separate special cause variation from normal cause variation, we create *control charts*
- A control chart is a trend chart with statistically determined control limits
- Data points that fall outside the control limits, or that exhibit unusual or nonrandom behavior, indicate the process has gone out of control
- A control chart has three components:
  - → A center line (CL) for average process performance when the process is in control
  - → A lower control limit (LCL) and a upper control limit (UCL), which are the limits of the range of process performance when the process is in control
- To create a control chart:
  1. Create a trend chart of the data
  2. Using data from when the process was in control, or using specified "target" values, find the mean, μ, and the standard deviation, σ
  3. Draw the CL at μ, the UCL at $\mu + 3\sigma$, and the LCL at $\mu - 3\sigma$
  4. Lines at $\mu \pm 1\sigma$ and $\mu \pm 2\sigma$ may also be added to the chart to help look for out-of-control warning signals in the existing data
- Analyzing a control chart involves looking for specific warning signals
  - → One or more points fall beyond the 3σ level (process is out of control)
  - → A run of 9 consecutive points on one side of the CL (small process shift or trend)
  - → At least 2 of 3 consecutive points lie between the 2σ and 3σ levels on the same side of the CL (medium shift or fast-moving trend)

Normal Probability Plots

(Chapter 6 – Using Technology)

- Recall
  - → Histograms help visualize the distribution of a data set (normal, skewed, bimodal)
- Unfortunately, a histogram's shape can be influenced by the categories used to create it
- Since much of statistical methodology requires the data be normally distributed, we need a more reliable graph for checking that a distribution is normal
- Such a graph is called a *normal probability plot*
- Normal probability plots are very difficult to create by hand
  - → Order data and determine percentile occupied by each value
  - → Find the *z*-score corresponding to each percentile
  - → Plot each data value against the corresponding percentile *z*-score
- Luckily we have our calculators to do the work! Here are the steps to follow:
  1. STAT 1 then enter data set in one of the lists
  2. 2nd Y= to go into STAT PLOTS menu
  3. 1 to go into Plot1
  4. ENTER to turn plot On
  5. Press down arrow key once, then right arrow key five times so curser is over normal probability plot, and hit ENTER to select
  6. Press down arrow key once, then type 2nd and the key corresponding to the list where the data is stored
  7. Choose Data Axis and Mark (or leave as X and a square)
  8. Hit GRAPH then ZOOM 9 to see the normal probability plot
- Once you have the normal probability plot, you can determine if the distribution is normal or not
  - → If the points lie close to a straight line, then the plot indicates the data is normal
  - → Systematic deviations from a straight line, or bulges in the plot, indicate the data is not normal
  - → Individual points off the line may be outliers
- Possible normal probability plots and distributional patterns

Normal	*Long-tailed*	*Skewed*	*Outlier*
			

Chi-Square Test

(Section 11.1 – Chi-Square: Tests of Independence)

- Suppose you have two qualitative variables you want to compare
- Every analysis until now has required quantitative data, so how do we deal with qualitative

data?

- Recall that qualitative data can be broken down into categories
- Thus, we can conduct tests on the number of observations that fall into each category instead of the observations themselves
- One such test is called the *chi-square test* (denoted χ^2)
- The χ^2 distribution is a non-symmetrical distribution that changes shape like the *t*-distribution depending on the degrees of freedom (*d.f.*)
  - → All χ^2 values are greater than or equal to zero
  - → Peak of distribution occurs over *d.f.* – 2
  - → As the degrees of freedom increase, the χ^2 distribution becomes more bell-shaped and symmetric
- The χ^2 test evaluates the following hypotheses:
  - → H_0: The two variables are independent (i.e. distribution across categories of one variable stays the same for each category of the other variable)
  - → H_1: The two variables are not independent (i.e. the distribution changes)
- To set up a χ^2 test, you must first create a contingency table for the observations, filling in the counts for each cell

		Variable 2			
		Category 1	Category 2	…	Category *c*
Variable 1	Category 1				
	Category 2				
	…				
	Category *r*				

- Once you have the contingency table, conduct the χ^2 test on your calculator:
  1. Edit matrix A: [2nd] [MATRIX] [←] [1] to edit
  2. Type the number of rows in the contingency table and hit [ENTER], then the number of columns and hit [ENTER]
  3. Enter the data in matrix A exactly as it appears in the contingency table
  4. Edit matrix B: [2nd] [MATRIX] [←] [2] to edit
  5. Type in the same number of rows and columns as matrix A (number of rows, [ENTER], number of columns, [ENTER]), but don't enter any data in the actual matrix
  6. Go into the χ^2 test: [STAT] [←] [ALPHA] [C]
  7. Leave Observed as [A] and Expected as [B], and go down to Calculate and hit [ENTER]
  8. Calculator will report *p*-value, use to make decision and state conclusion as part of our usual four steps of hypothesis testing
- The expected matrix contains numbers that we would expect to see if the null hypothesis were true, and is used with the observed data to calculate the *p*-value
  - → For the test to be accurate, the expected values should be greater than or equal to 5
- Good news! Unlike the *z*-tests and *t*-tests, the χ^2 test does <u>not</u> require that the data come from a normal distribution – it is what we call a *nonparametric test*

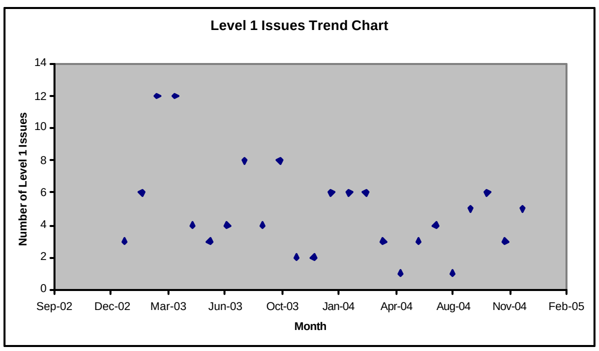
Workbook Problems

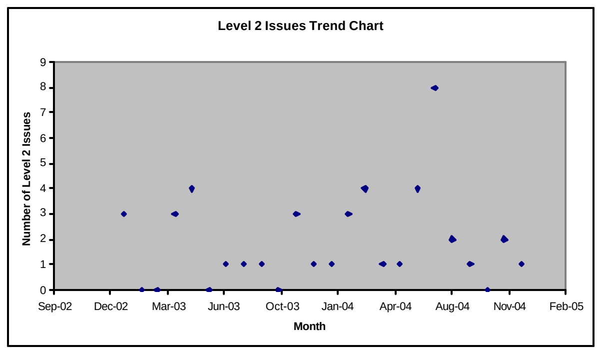
Levels of Measurement
(Section 1.1 – What is Statistics?)

1. Determine the level of measurement for each data set.
    - (a) Number of calls into Asia/Pacific call center each day
    - (b) Temperature of room in fabrication factory
    - (c) Type of fraud committed: damaged/defaced, counterfeit, warranty, other
    - (d) Months in which an issue with a product came up

Control Charts
(Section 6.1 – Graphs of Normal Probability Distributions)

1. Every month, the managers at a large corporation record the number of issues that came up. The issues are categorized level 1 or level 2, depending on the severity. Trend charts of the number of issues for both levels from 2003 to 2004 are shown below.





   - (a) For level 1 issues, the target mean is $\mu = 4.88$ issues and the target standard deviation is $\sigma = 2.91$ issues. Use this information to first draw the CL, LCL, and UCL. Then add two more lines at $\mu \pm 2\sigma$.

   - (b) For level 2 issues, the target mean is $\mu = 1.88$ issues and the target standard deviation is $\sigma = 1.87$ issues. Use this information to first draw the CL, LCL, and UCL. Then add two more lines at $\mu \pm 2\sigma$.

   - (c) Do the data indicate that the number of level 1 and level 2 issues is in control or out of control? Identify any warning signals that you observe and what that warning signal is indicating about shifts or trends in the data.

Normal Probability Plots
(Chapter 6 – Using Technology)

1. To measure the use of a company's website, the number of web hits are recorded each month (see data below). Before performing certain types of statistical analysis on this data, the company must first determine if the data set follows a normal distribution. Create a normal

probability plot for number of website his each month and comment on what the graph tells you about the shape of this distribution.

Month	Number of Web Hits
October 2004	299
November 2004	1870
December 2004	441
January 2005	170
February 2005	170
March 2005	1258
April 2005	1012
May 2005	1326
June 2005	5829

<u>Chi-Square Test</u>
(Section 11.1 – Chi-Square: Tests of Independence)

1. The worldwide equipment launch team was formed to prepare for and handle any issues that may arise during the launch of new equipment. Every quarter, each member of the team must report how ready they are to launch the newest system. Each team member has a set of categories they are responsible for, and they score each category on its level of readiness – red, yellow, or green. The total counts for the red, yellow, and green categories are then rolled-up so the whole team can determine their overall readiness (see data set below).

Readiness	Red	Yellow	Green
Q3 ‘02	30	21	64
Q4 ‘02	30	18	67
Q1 ‘03	27	21	67
Q2 ‘03	32	14	69
Q3 ‘03	32	15	68
Q4 ‘03	29	12	74
Q1 ‘04	32	12	64
Q2 ‘04	27	18	62
Q3 ‘04	39	21	48
Q4 ‘04	29	16	60
Q1 ‘05	11	18	63
Q2 ‘05	13	16	61

(a) To look for significant improvements in readiness, the team wants to see if the distribution of red, yellow, and green stays the same from quarter to quarter or if there is a change. Conduct the chi-square test to determine if levels of readiness and quarter are independent at the 0.10 level of significance. Clearly show the four steps of hypothesis testing.

(b) Suppose the Q3 ’05 data just came in, and we can now add the following information to the contingency table above. Conduct the chi-square test once again, with the new data added, to determine if levels of readiness and quarter are still independent at the 0.10 level of significance. Clearly show the four steps of hypothesis testing.

Readiness	**Red**	**Yellow**	**Green**
Q3 ‘05	4	5	81

Assessment

These surveys would be distributed at the beginning of the semester. The results of would be used to evaluate my decision to include the new topics outlined in this ETP, while removing or reducing the topics outlined in the curriculum modification section.

<u>Student Survey</u>

Please complete the following survey regarding your Elementary Statistics class.

1. Why are you taking Elementary Statistics? Because you want to, because it's required?

2. If this class is required, what is your major or what are you planning to study?

3. Do you have any idea how statistics will be used as part of your degree program? Will you actually be performing research as part of your studies?

4. Outside of this class, have you ever used statistics in a professional setting?

5. How do you think a company like Intel would use statistics?

<u>Math Faculty Survey</u>

Please complete the following survey indicating you thoughts and feelings regarding Elementary Statistics.

1. What do you think of the current Elementary Statistics curriculum? Do we teach too much, too little? Do we teach the right topics, the wrong topics?

2. If you could <u>add</u> topics to the current curriculum, what would they be?

3. If you could <u>omit</u> topics from the current curriculum, what would they be?

4. How do you feel about the time allotted for teaching Elementary Statistics (3 units)? Do you usually have enough time to cover the required topics? Do you ever need or want more time?

5. If you could teach your students statistical methods on a computer instead of a calculator, would you? If you could choose any statistical software package to teach from, which would you choose and why?