AEA Intro to Quant Methods eStudy

RunDown:

Time	Slides/Notes	Materials
Intro, Warm Up & Review	**WARM UP POLL** – No. 1 (experience with t-test)– *Start session with this poll while people are joining and getting materials ready.*	Session 1 pre-work: 1. "Homework" reading & sampling distribution demo 2. PDF of online survey so they can refer to it 3. Polls & answers 4. NELS Excel data set 5. Survey Excel data set
	Slides 2– 3 Intro to topic, agenda, objectives *Review participants, noting how this seminar will NOT make them data analysts! Discuss helix model of learning.*	
	Slides 5 -13: Review of last week's discussion: *Review of the inferential stats part and how it applies to t-tests.* **Slide 6. Why do these kids vary?** This slide demonstrates that there is indeed variance in an outcome we're interested in. So why is that? What could explain the high scores vs. the low scores? Brainstorm some explanatory variables. **Slide 7. Graph** *Describe* **Slide 8. Back to variance…** *Note that this is a population estimate, N-1 in the denominator. Discuss the expectation part of the formula.* **Slide 9. Our model** Linear model for predicting 12^{th} grade math achievement: mean math achievement plus whether or not student has taken advanced math in 8^{th} grade. How does that variable add to our explanation of variance? Is it adding anything beyond knowing the mean? Talk about this as a "model." It's linear. What does that mean with	

	respect to relationships between variables? Steady, additive…what do those features mean? We'll return to this concept of a linear model later… **Slide 10. Histogram of 12th grade math achievement** See the mean, or "expectation." Illustrate how variance is conceived--how much the other scores vary from this. Eyeballing the histogram, is there a lot of variance about 56.9? (YES!). What would a histogram look like with LITTLE variance? What would it look like with GREAT variance? **Slide 11. Independent Samples T-test** **Slide 12. Explaining variance scatterplot** This slide shows how statistically we test to see if group membership is a good predictor of the outcome: we look at how much the means of the groups vary from each other, or the mathematical equivalent: how much each varies from the grand mean. This provides good discussion for the use of means for group comparisons to begin with—does knowing which group a person belongs to help to predict their achievement score? If so, that grouping variable (taking Adv Math in 8th grade) is a predictor worth considering. **Slide 13. T-test: difference between means** This is our evaluation question put in statistical (inferential stats) terms.	
Independent Samples T-test	**Slide 14: Divider Slide** *Here we will discuss the Independent Samples T-test in detail. This will set us up for Dependent (Paired) Samples T-test and ANOVA.* **Slide 15. T-Test** Remember, this is an inferential procedure so we have to test our hypothesis against the Null (Sampling error or chance hypothesis). P = program group, e.g., those taking Advanced Math in 8th grade C = comparison group, e.g., those not taking Advanced Math in 8th grade **Slide 16. T-test distribution of differences** *Clarify that we're looking at the difference between 2 means. If there is none, it should be or equal to 0. Each row represents a different sample from the same population, as if we ran multiple studies, one for each row. We're setting up the sampling distribution on which t is based.*	

Slide 17. Sampling distribution of differences
The distribution of differences between the means of 2 independent samples, repeatedly sampled from the same population(s). **The mean of this distribution can be shown to equal the population mean difference** (see the mu's in green). The variance of the distribution is given by The **Variance Sum Law**: the variance of the sum or difference of two INDEPENDENT variables is equal to the sum of their variances. This variance is called the standard error of differences between means.

Also, an important theorem in statistics states that the sum or difference of 2 independent, normally distributed variables is itself normally distributed. Therefore **this distribution of differences is normally distributed.** For the NULL HYPOTHESIS, the mean of this distribution is ZERO – i.e. there is NO DIFFERENCE between the two group means.

Slide 18. Std Error
Remind audience of similarity between mean & std deviation of a variable, and mean & std error of a sampling distribution of a statistic: like SD, size matters! Stability is a measure of certainty – if unstable (i.e. large std error), we're less certain that the obtained statistic is an accurate representation of the relationship we're testing in our sample to the population

Slide 19. Sampling distribution of the difference
Just as we discussed with std normal distribution, the frequency distribution of a statistic has a mean (presumed to be the population mean) and variance, i.e. the std error of the statistic. In this example, the statistic is the difference between means. For a sample, we can calculate the difference, but we have to ESTIMATE it for the population; our sample statistic is an estimate, with error bars around it, based on the std error of the difference between means.

Slide 20-22. Homework Review
Demo this website to review sampling distribution of a statistic, noting how each one has a mean and variance, as we know regarding the distribution of an outcome variable.

Slide 23. Check in…
Answer a few questions before moving on

Slide 24. Inferential stats: stats vs. parameters
Objective of inferential statistics: to infer the extent to which the observed sample statistic reflects the TRUE population parameter; we test against the NULL HYPOTHESIS that the observed statistic is due to SAMPLING ERROR (vs. it is a true reflection of the population

	parameter). **Slide 25. T-test: test of the difference between means** Talk about the numerator & denominator & how as t increases, it becomes more apparent that the groups truly are different from each other (i.e. mean difference = big, pooled variance = small). **Slide 26. The t-distribution** For practical purposes, the shape of the t-distribution is identical to the normal distribution when sample size is large. However, when sample sizes are small (below 30 subjects), the shape of the t-distribution is flatter than that of the normal distribution, and the t-distribution has greater area under the tails. The reason that the distribution is not normal is because the standard error is estimated using the sample standard deviation instead of the population standard deviation (because the population standard deviation is not known). This creates some uncertainty that is reflected in the t-distribution having greater area under the tails than the normal distribution, especially when the sample size is below 30 subjects. **Slide 27. Demonstration** *Go to http://simon.cs.vt.edu/SoSci/converted/T-Dist/* *Show how t-distribution approaches the normal distribution as the sample size increases, specifically to 30 and higher.* **Slide 28. T-distribution: A family of sampling distributions** The t-distribution is actually a family of sampling distributions, because every sample size produces a slightly different t-distribution, i.e. t-distributions change as a function of degrees of freedom. Because of this, evaluators must always report the sample size used in the t-test. Image & description from http://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm Note that the probability density function for t helps us to know the probability of an obtained t, just like how we know with a Gaussian normal distribution the probability associated with a specified number of std deviations from the mean. **Slide 29. T-distribution and critical values DEMO** Go to http://simon.cs.vt.edu/SoSci/converted/T-Dist/ and show how critical value of t changes as you change alpha (represented by % in the center box) for N = 20. The % value in the middle is the likelihood of making a Type I error. It determines the critical values for t, e.g., for N = 20, critical t = +/-2.093 for a 2 tailed test.	

	Slide 30. Divider slide – SPSS DEMO *Use NELS data set with Adv8Math as the IV and 12th grade math achievement as the DV. Focus on the output table with t and it's significance value, that's it because we get into a full interpretation of results later.* **Slide 31 & 32. Test of Mean Difference** *Show where the std error of the difference is for this example, with the 95% confidence intervals. If the interval were to include 0, the t-test would be non-sig, because that would mean that there really is no statistically reliable difference between the 2 group means.* **Slide 33. Summary so far…** *Review then ask for questions* **QUESTION BREAK – ALLOW 3 MINS** **Slide 34 QUESTION BREAK – ALLOW 3 MINS**	
Homogeneity of Variance	**Slide 35. You may have noticed…** *Point out Levene's Test* **Slide 36. Divider Slide** *We'll talk about this assumption for the t-test next…* **Slide 37. (Graphs – unequal variances)** The assumption of homogeneity of variance is that the variability in the POPULATION of interest is homogeneous – the variance in math achievement of the population of students taking advanced math in 8th grade is equal to the variance of the population NOT taking advanced math. Sample variances not expected to be equal due to sampling error. How do you think the math test scores would be distributed for the non-Advanced math POPULATION (in pink)? How about the Advanced Math POPULATION (in blue)? Do you think they'd be equally distributed? Why or why not? **Slide 38. VOTE** *Show of hands for A or B, twice. Stephanie to clear the voting after each vote so we can determine % choosing A vs. B.* **Slide 39. Std. Error** *Don't forget that 's' is based on sample size (it's in the denominator), and sample size is in the denominator of this statistic as well. Given that the magnitude of the std error is related to sample size, B is the*	

	correct answer to the previous questions. Discuss why larger 'n' = smaller SEM. **Slide 40. Pooling variance** Pooled variance is a weighted average (weighted by sample size) of the separate sample variances. This is done to correct for different sample sizes in the two groups, which is very often the case. When the sample sizes are unequal, they do not provide an equal estimate of the population variance; larger samples provide better estimates, so we weight by sample size (technically, degrees of freedom or n-1). **Slide 41. Homogeneity of variance** Note that this assumption does not apply to the sample – we know that sample variances will be different just due to sampling error. However, based on the sample variance, we estimate the population variance and then test whether they are statistically different from each other. **Slide 42. Heterogeneity of variance** Correction occurs for degrees of freedom, which influences obtained t. It also influences whether t will be statistically significant (with low sample sizes, i.e. under 30). **Slide 43 - 45. Levene's Test** *Examples—show how to interpret.* **Slide 46 & 47. Degrees of Freedom** Remind folks of their pre-session reading for DFs: dfs means the number of pieces of info that are free to vary. In this situation, the group mean is not free to vary (that's the -1 part of the equation). Since there are two group means, we subtract 2 dfs from the total sample size—the bits of info that are free to vary.	
Hypothesis testing	**Slide 48. Divider Slide and SPSS DEMO** *Review the AdvMath8 example, and one from the survey comparing males & females on satisfaction with work and outside of work. Talk through the results, then intro the notion of directional tests.* **Slides 49-50. One- and Two-Tailed tests** *Intro the concept, then have them apply it to the AdvMath8 example evaluation.* **Slide 51 & 52. One-tailed test** *Directions for one-tailed Independent Samples t-test in SPSS can be found at http://www.ats.ucla.edu/stat/spss/faq/pvalue.html* *(No option for one-tailed test in current version). You divide the obtained p-value for the 2-tailed test by 2, and that will tell you the*	

	significance of the observed t-value for a one-tailed test. (The one-tailed test doesn't change the t-value, only the probability that it is not due to sampling error). Report that value if you want an exact p-value; otherwise indicate if it is < .05 or whatever the chosen nominal alpha. **Slide 53. Reporting results** *Refer back to SPSS output and show how that fits into reporting.* **Slide 54 - 56. Sampling error** Recall ping pong ball experiment, apply to hypothesis testing with respect to observed t vs. critical t: When we compare group means using a T-test, we'll get an obtained t-statistic with a p-value, which is the probability that the obtained t-statistic was obtained by chance, given the sample size. The probability is obtained because there is a probability density function that tells us how likely it is to obtain a specific t-value for a given sample size, just like the probability density function for obtaining 0 heads in 10 coin flips. The t-distribution has a mean of 0 in its center because this is the distribution for the null hypothesis – that there is no difference between the group means. The probability of any given obtained t value can be obtained from the sampling distribution of the t-statistic, given a specific sample size up to 30; beyond 30 the values are virtually identical. **Slide 57. Hypotheses** Remember, inferential statistics involved hypothesis testing. Descriptive stats do not. **Slide 58. Type I & II errors** Our t-test result tells us to reject the null, but we do not know if in this instance we made a Type I error or correctly rejected the null. The probability that we've made a Type I error is < .001 (it's the reported p-value), but that means if we ran this study repeatedly and each time obtained a t-statistic of about the same value, the likelihood that we're finding this due to sampling error is nearly 0. **Slide 59. Effect size and t-tests** If sample sizes are large enough, std errors will be small and therefore two means can look statistically different than each other (i.e. the t-test will be significant). However, the difference may not be meaningful. A statistical test of the size of the difference is the effect size. It communicates how many standard deviation units one group mean is from the other. **Slide 60 & 61. Cohen's D** *Show 2 formulas, and note there are more (e.g., converting 'r' into 'd', etc.). Note that d is in units of std deviation (see the denominator),*	

	therefore communicates how many SDs the two group means are from each other. Also note that this is another SIGNAL to NOISE ratio! **Slide 62. Effect size calculator DEMO** *Go to http://www.uccs.edu/~faculty/lbecker/* For our AdvMath8 example, the "No" group is group 1, so the sign is negative because the mean is smaller than the program group's mean. **Slide 63. Interpretation** *Cohen was reluctant to assign these value judgments due to literal translation and strict adherence to cut-off values (e.g., .49 is small?). They are to be used heuristically, and depending on the phenomenon of interest. (E.g., a "small" effect size could mean that 1,000s of lives are saved).* **Slide 64. Exercise 2.2 Back to Our Survey CHAT BOX** *Attendees select a grouping variable & outcome variable from the survey & enter in Chat Box. We select one or two to demo.*	
Wrap Up	**Slide 65. Wrap Up – Divider Slide – QUESTIONS?** **Give 3 – 5 mins if there is time** **Slide 66. Objectives** *Review* **Slide 67. Poll 2** If there's time.	

Notes:

1. T-distribution demos: http://simon.cs.vt.edu/SoSci/converted/T-Dist/
2. Effect size calculator demo: http://www.uccs.edu/~faculty/lbecker/
3. Have copy of survey from last week (PDF)
4. Two data sets if you would like to analyze along with me: NELS (student achievement data set) and our Survey data set