Power Analysis: Sample Size, Effect Size, Significance, and Power

16 Sep 2011
Dr. Sean Ho
CPSY501

cpsy501.seanho.com

Please download:
SpiderRM.sav

Trinity Western University

Outline for today

- Discussion of research article (Missirlian et al)
- Power analysis (the “Big 4”):
  - Statistical significance (p-value, α)
    - **What does the *p*-value mean, really?**
  - Power ($1-\beta$)
  - Effect size (Cohen's rules of thumb)
  - Sample size (n)
    - **Calculating min required sample size**
- Overview of linear models for analysis
- SPSS tips for data prep and analysis

Practice reading article

- Last time you were asked to read this journal article, focusing on the statistical methods:
- Missirlian, et al., "*Emotional Arousal, Client Perceptual Processing, and the Working Alliance in Experiential Psychotherapy for Depression*", Journal of Consulting and Clinical Psychology, Vol. 73, No. 5, pp. 861–871, 2005.
- How much were you able to understand?

Discussion:

- What research questions do the authors state that they are addressing?
- What analytical strategy was used, and how appropriate is it for addressing their questions?
- What were their main conclusions, and are these conclusions warranted from the actual results /statistics /analyses that were reported?
- What, if any, changes/additions need to be made to the methods to give a more complete picture of the phenomenon of interest (e.g., sampling, description of analysis process, effect sizes, dealing with multiple comparisons, etc.)?

Outline for today

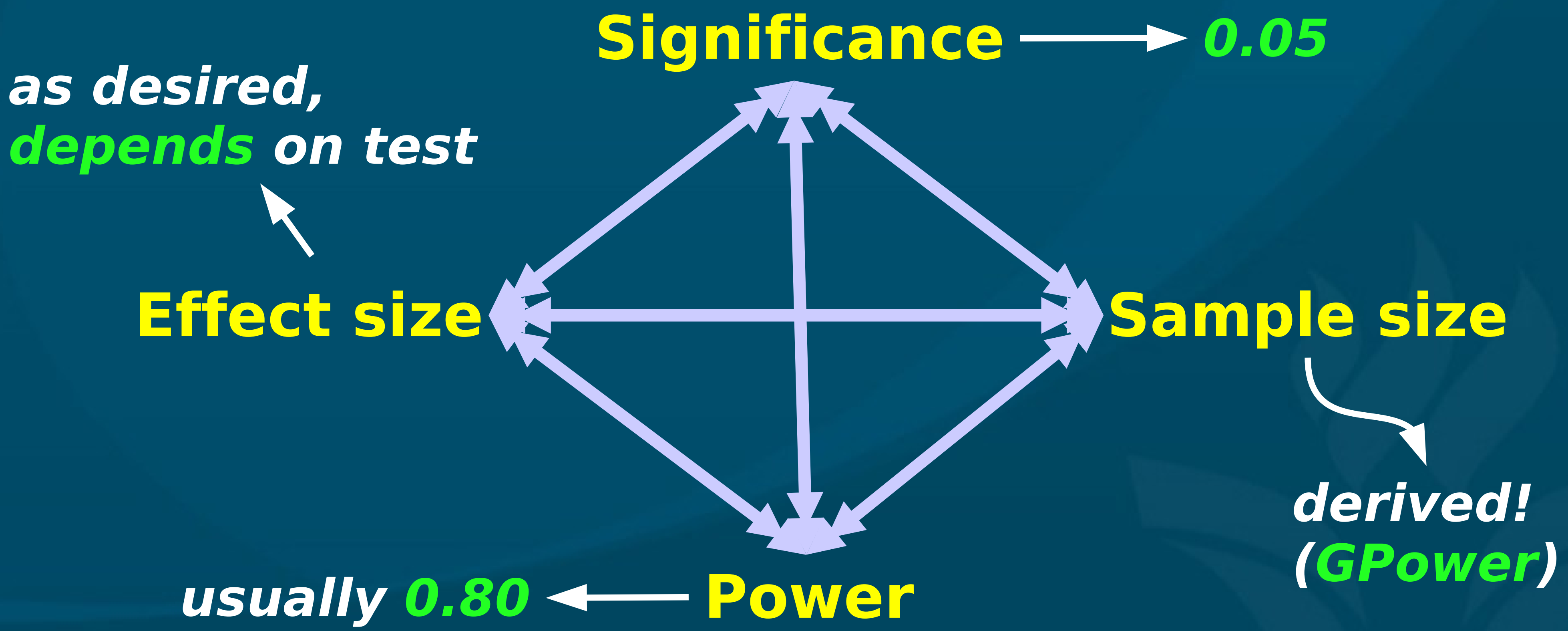
- Discussion of research article (Missirlian et al)
- Power analysis (the "Big 4"):
  - Statistical significance (*p*-value, α)
    - **What does the *p*-value mean, really?**
  - Power (1-β)
  - Effect size (Cohen's rules of thumb)
  - Sample size (n)
    - **Calculating min required sample size**
- Overview of linear models for analysis
- SPSS tips for data prep and analysis

Central Themes of Statistics

- Is there a real effect/relationship amongst the given variables?
  - How big is that effect?
- We evaluate these by looking at
  - Statistical significance (p-value) and
  - Effect size (r^2, R^2, η, d, etc.)
- Along with sample size (n) and statistical power ($1-\beta$), these form the “Big 4” of any test

The “Big 4” of every test

- Any statistical test has these 4 facets
- Set 3 as desired → derive required level of 4th



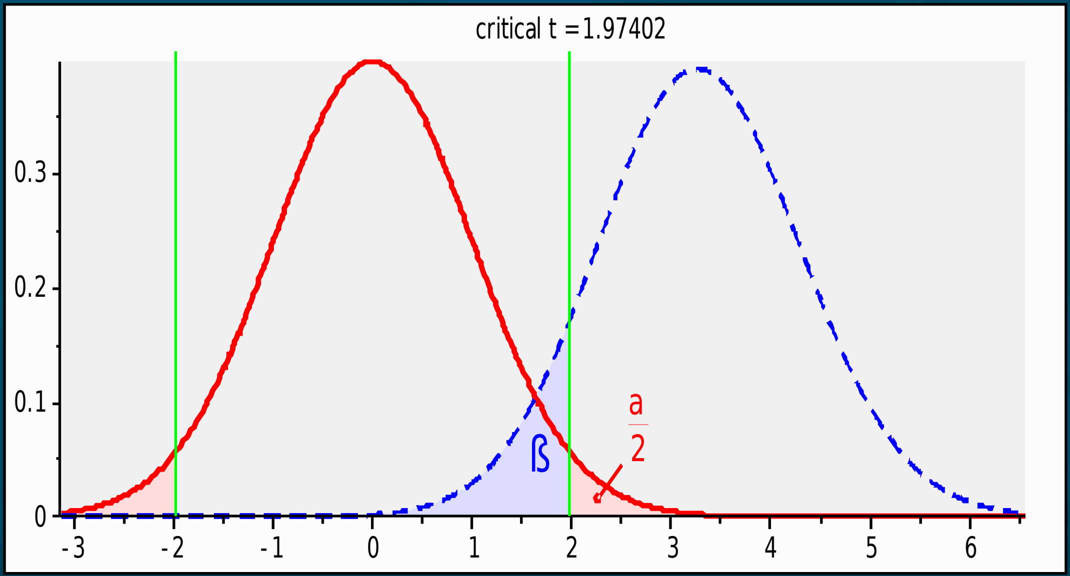
Significance (α) vs. power (1-β)

- α is our tolerance of Type-I error: incorrectly rejecting the null hypothesis (H_0)
- β is our tolerance of Type-II error: failing to reject H_0 when we should have rejected H_0.
- Set H_0 so that Type-I is the worse kind of error
- Power is 1-β
- e.g., parachute inspections (what should H_0 be?)

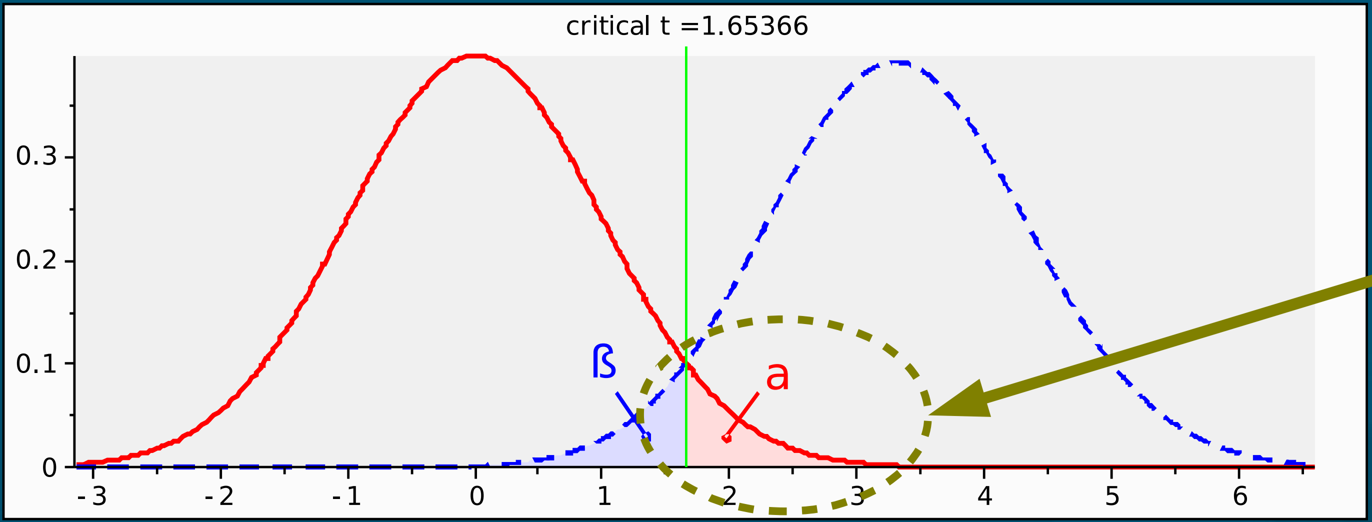
	Actual truth: H_0 true	Actual truth: H_0 false
Reject H_0	Type-I error (α)	Correct
Fail to reject H_0	Correct	Type-II error (β)

Significance: example

- RQ: are depression levels higher in males than in females?
  - IV: gender; DV: depression (e.g., DES)
  - What is H_0? (recall: H_0 means no effect!)
- Analysis: independent-groups *t*-test
- What does Type-I error mean?
- What does Type-II error mean?



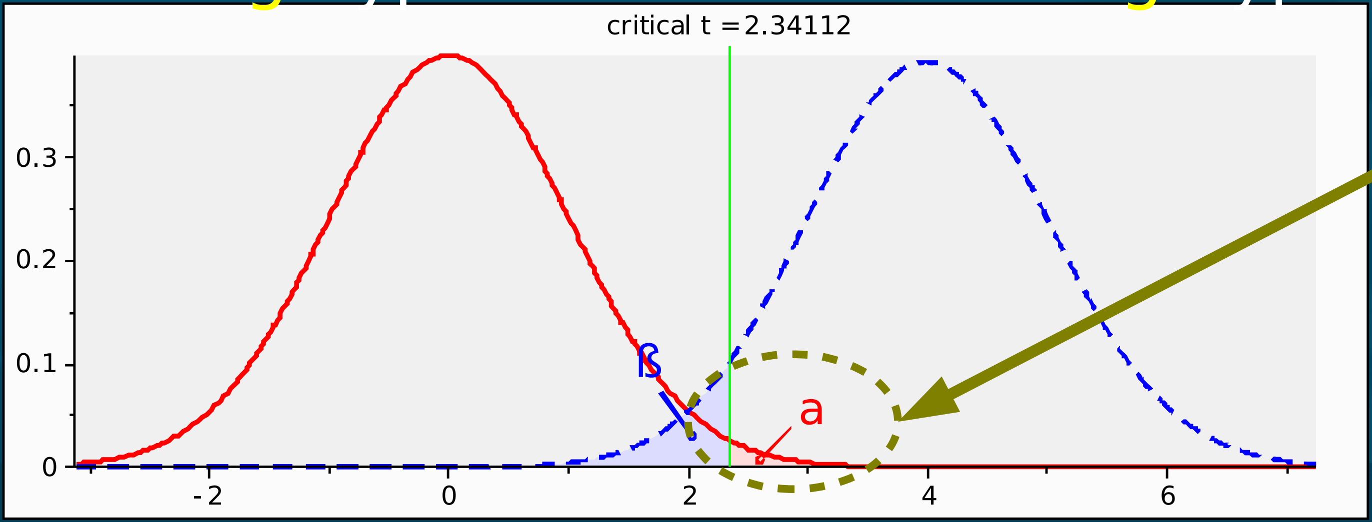
Impact of changing α



α=0.05

(without changing data)

decreasing Type-I means increasing Type-II!



α=0.01

What is the *p*-value?

- Given data and assuming a model, the *p*-value is the computed probability of Type-I error
  - Assuming model is true and H_0 is true, p is the probability of getting data that is at least as extreme as what we observed
  - Note: this does not tell us the probability that H_0 is true! $p(\text{data}|H_0)$ vs. $p(H_0|\text{data})$
- Set the significance level ($\alpha=0.05$) according to our tolerance for Type-I error: if $p < \alpha$, we are confident enough to say the effect is real

Myths about significance

- *(why are these all myths?)*

- Myth 1: "If a result is not significant, it proves there is no effect."
- Myth 2: "The obtained significance level indicates the reliability of the research finding."
- Myth 3: "The significance level tells you how big or important an effect is."
- Myth 4: "If an effect is statistically significant, it must be clinically significant."

Problems with relying on *p*-val

- When sample size is low, *p* is usually too big
  - → if effect size is big, try bigger sample
- When sample size is very big,
  p can easily be very small even for tiny effects
  - e.g., avg IQ of men is 0.8pts higher than IQ of women, in a sample of 10,000: statistically significant, but is it clinically significant?
- When many tests are run, one of them is bound to turn up significant by random chance
  - → multiple comparisons: inflated Type-I

Outline for today

- Discussion of research article (Missirlian et al)
- Power analysis (the “Big 4”):
  - Statistical significance (*p*-value, α)
    - **What does the *p*-value mean, really?**
  - Power (1-β)
  - Effect size (Cohen's rules of thumb)
  - Sample size (n)
    - **Calculating min required sample size**
- Overview of linear models for analysis
- SPSS tips for data prep and analysis

Effect size

- Historically, researchers only looked at significance, but what about the effect size?
- A small study might yield non-significance but a strong effect size
  - Could be spurious, but could also be real
  - Motivates meta-analysis – repeat the experiment, combine results: with more data, it might be significant!
- Current research standards require reporting both *p*-value as well as effect size

Measures of effect size

- For *t*-test and any between-groups comparison:
  - Difference of means: $d = (\mu_1 - \mu_2)/\sigma$
- For ANOVA: η^2 (eta-squared):
  - Overall effect of IV on DV
- For bivariate correlation (Pearson, Spearman):
  - r or r^2: r^2 is fraction of variability in one var explained by the other var
- For regression: R^2 and ΔR^2 ("R^2-change")
  - Fraction of variability in DV explained by overall model (R^2) or by each predictor (ΔR^2)

Interpreting effect size

- What constitutes a "big" effect size?
- Consult literature for the phenomenon of study
- Cohen '92, *"A Power Primer"*: rules of thumb
  - Somewhat arbitrary, though!
- For correlation-type *r* measures:
  - 0.10 → small effect (1% of var. explained)
  - 0.30 → medium effect (9% of variability)
  - 0.50 → large effect (25%)

Example: dependent *t*-test

- Dataset: SpiderRM.sav
- 12 individuals, first shown picture of spider, then shown real spider → measured anxiety
- Compare Means → Paired-Samples T Test
- SPSS results: $t(11)=-2.473$, $p<0.05$
- Calculate effect size: see text, p.332 (§9.4.6)
  - r ~ 0.5978 (big? Small?)
- Report sample size (*df*), test statistic (*t*), *p*-value, and effect size (*r*)

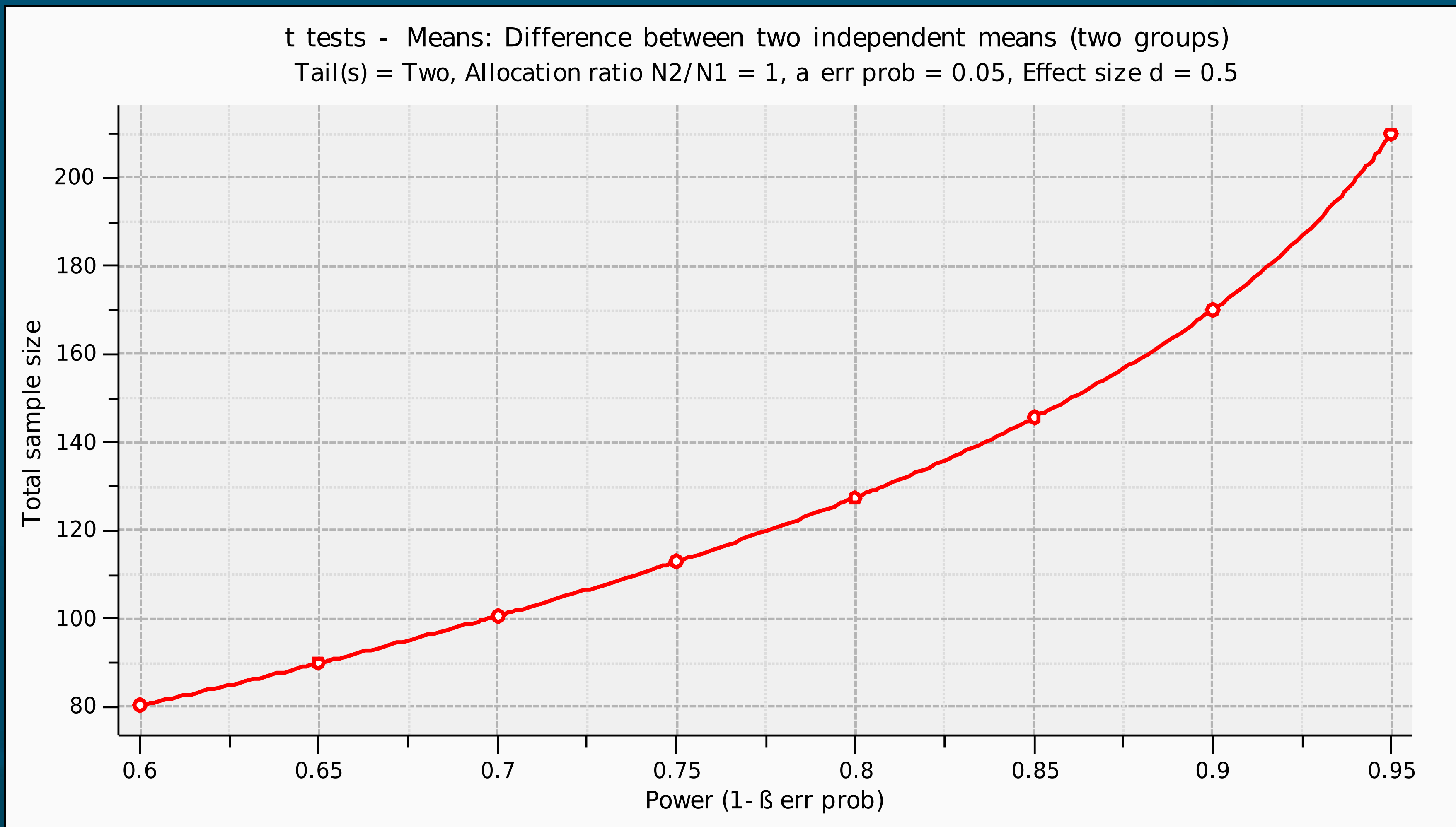
$$r=\sqrt{\frac{t^2}{t^2+df}}$$

Finding needed sample size

- Experimental design: what is the minimum sample size needed to attain the desired level of significance, power, and effect size (assuming there is a real relationship)?
- Choose level of significance: usually $\alpha = 0.05$
- Choose power: usually $1-\beta = 0.80$
- Choose desired effect size
  - From literature or Cohen's rules of thumb
- → use GPower or similar to calculate the required sample size (do this for your project!)

Power vs. sample size

- Choose α=0.05, effect size *d*=0.50, and vary β:



Outline for today

- Discussion of research article (Missirlian et al)
- Power analysis (the "Big 4"):
  - Statistical significance (*p*-value, α)
    - **What does the *p*-value mean, really?**
  - Power (1-β)
  - Effect size (Cohen's rules of thumb)
  - Sample size (n)
    - **Calculating min required sample size**
- Overview of linear models for analysis
- SPSS tips for data prep and analysis

Overview of Linear Methods

- For scale-level DV:
  - dichotomous IV → indep *t*-test
  - categorical IV → one-way ANOVA
  - multiple categorical IVs → factorial ANOVA
  - scale IV → correlation or regression
  - multiple scale IVs → multiple regression
- For categorical DV:
  - categorical IV → χ^2 or log-linear
  - scale IV → logistic regression
- Also: repeated measures and mixed-effects

SPSS tips!

- Plan out the characteristics of your variables before you start data-entry
- You can reorder variables: cluster related ones
- Create a var holding unique ID# for each case
- Variable names can't have spaces: try "_"
- Add descriptive labels to your variables
- Code missing data using values like '999'
  - Then tell SPSS about it in Variable View
- Clean up your output file (*.spv) before turn-in
  - Add text boxes / headers; delete junk