Welcome SSPG Grad Students! – Use this guide for preparation for the stats screening exam.

Part 1 – Introduction

As you know, the introductory graduate methods course PUBP/GOVT/PUAD 511 has a screening examination. This is to ensure that all students are prepared to begin on day 1 without having to reteach extensively many of the topics covered in a typical undergraduate stats or social sciences methods requirement course. In order to help students prepare for the screening exam, the Schar School of Policy and Government offers two paths to help you pass the exam:

1. Use this brief guide to help you study for the exam itself, and write the exams on one of the days available to you during the semester prior to your intention to register in your desired section. (Note: successful completion of the screening exam does not guarantee you a spot, so the earlier you take the exam successfully, the greater the likelihood that you will be able to register in your desired section.) *This path is recommended for those students who feel confident in their ability and knowledge that they have retained from their undergraduate career coursework.*

2. Attend one of the 5 scheduled stats workshops held throughout the year and write the screening exam at the conclusion of the workshop. The workshop is more than a refresher, it assumes no prior statistical training whatsoever, so *this path is recommended for those who have not taken statistics and/or a social science research methods class in a long time or not at all, and those who may wish to refresh themselves before undertaking more challenging material in GOVT/PUAD/PUBP 511.*

Part 2 – General Principles

What follows is a brief list of terms, concepts, ideas, calculations, etc. for which you will be expected to have knowledge of to successfully pass the screening exam.

Research Design	Graphics	Frequency Distribution Tables	Univariate Measures	Statistical Inference	Confidence Intervals
Dependent Variables	Histograms	Percent	Level of Measurement	The Law of Large Numbers	Margin of Error
Independent Variables	Box-Plots	Proportion	Central Tendency	Central Limit Theorem	Intervals for Interval/Ratio Level Data
Research Motivations	Bar Graphs	Cumulative Tables	Dispersion	Standard Errors (Theoretical)	Intervals for Proportions
Identifying Audiences	Pie Charts		Z-Scores & Probability	Standard Errors of Ratio and Interval level data	Basic Sample Size Estimation
Hypotheses Development	Scatter Plots		Outliers	Standard Errors of Proportions	The Ignorance Assumption

Part 3 – Some Example Screening Exam Questions

Digesting research – read the abstracts below. Identify the following and justify your response:

a) Who you believe the audience to be (sector/motivations, etc.)
b) Who the producers of the argument are (sector/motivations, etc.)
c) 2 independent variables
d) 2 dependent variables
e) Identify the data the authors are using
f) 2 hypotheses the authors *may* be testing, given the abstract

Stephanie Riegg Cellini
The George Washington University

"Financial Aid and For-Profit Colleges: Does Aid Encourage Entry?" ***Journal of Policy Analysis and Management*****. Summer 2010. Vol. 29 no. 3 pp. 526-552.**

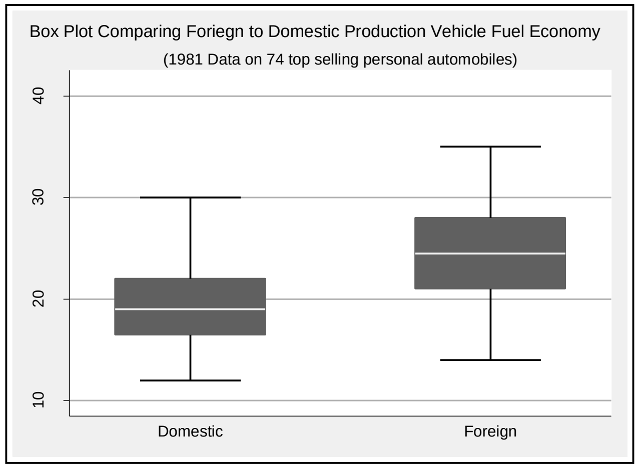
Concerns over rising college tuition and slow economic growth have brought renewed attention to the role of federal and state financial aid programs in opening access to education. Despite a large body of literature examining the effects of grant aid on four-year and public two-year college enrollment, for-profit colleges — particularly the vast majority that offer two-year degrees and certificates — have largely been ignored. Using panel data methods and a new administrative data set of for-profit colleges operating in California between 1989 and 2003, I assess the impact of the federal Pell Grant program, the G.I. Bill, and California's Cal Grant program on the net number of for-profit colleges per county. The results suggest that for both Pell and Cal Grants, increases in the per-student maximum award encourage for-profit entry. This relationship is particularly strong in counties with high adult poverty levels, where more students are eligible for aid. Further, these gains in the private sector do not appear to come at the expense of the public sector. Rather, public community colleges also experience enrollment gains as the generosity of Pell and Cal Grants increases, although this reaction appears to be weaker than the reaction of for-profits.

Arijit Chatterlee & Donald C. Hambrick
Pennsylvania State University

"It's all about me: Narcissistic Chief Executive Officers and their efforts on Company Strategy and Performance" ***Administrative Science Quarterly*****. September 2007. Vol. 52 no. 3 pp. 351-386.**

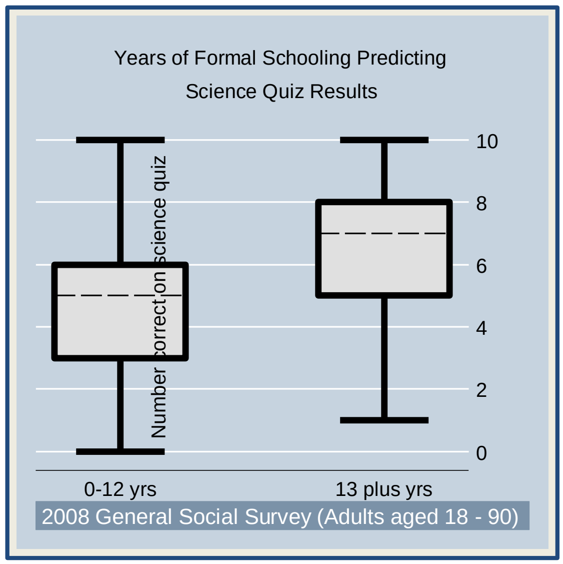
This study uses unobtrusive measures of the narcissism of chief executive officers (CEOs)—the prominence of the CEO's photograph in annual reports, the CEO's prominence in press releases, the CEO's use of first-person singular pronouns in interviews, and compensation relative to the second-highest-paid firm executive—to examine the effect of CEO narcissism on a firm's strategy and performance. Results of an empirical study of 111 CEOs in the computer hardware and software industries in 1992–2004 show that narcissism in CEOs is positively related to strategic dynamism and grandiosity, as well as the number and size of acquisitions, and it engenders extreme and fluctuating organizational performance. The results suggest that narcissistic CEOs favor bold actions that attract attention, resulting in big wins or big losses, but that, in these industries, their firms' performance is generally no better or worse than firms with non-narcissistic CEOs.

The questions below refer to the Box Plot to the right.



a. What can you infer about fuel consumption from this graphic comparison between domestic and foreign vehicles? Discuss both **location** and **variability** in your answer.
b. What is the level of measurement for both the **independent** and **dependent** variable?
c. Estimate a **5 number summary** for domestic automobile fuel consumption, given what you can see in the graph.

The following questions refer to the Box Plot (on the right) taken from the 2008 General Social Survey (GSS).



a. What can you infer about the relationship between the years of schooling that an adult has and the number of correct responses to a general science quiz as administered as a component of the 2008 GSS? Discuss both **location** and **variability** in your answer.

b. What is the <u>level of measurement</u> for both the **independent** and **dependent** variable?

c. Estimate a **5 number summary** for the science quiz among the category of survey respondents with the greater number of years of formal education, given what you can see in the graph.

Answer the following questions in short essay form:

- Why does probability sampling allow us to use statistics to make estimates of population parameters?
- Explain the benefits of the scientific method as a "knowledge acquisition system".
- What does it mean to "operationalize concepts"?
- How does our choice of instrumentation constrain our ability to run statistical tests?
- Explain what a confidence interval is.
- Why can we use an ignorance assumption when estimating sample sizes from proportions?

Below are listed a random sample of jury deliberation times (the time it takes a jury to determine a suspect's guilt or innocence) taken from a sample of trials conducted from May to July of 2012 in the Fairfax County Circuit Court. Complete the questions that follow about this distribution:

Fairfax County Circuit Court - Jury Deliberation Times (In Hours)

4	6	3	4
8	4	5	7
3	4	5	8
7	5	5	9
6	4	6	

a. Build an appropriate cumulative percent frequency table for jury deliberation times in Fairfax County.

Find each of the following:

b. Mode
c. Median
d. Mean
e. Variation ratio
f. Q1
g. Q3
h. IQR
i. Range
j. High value
k. Low value
l. Variance
m. Standard deviation
n. Build a +/- 2s interval around the mean and make a claim about outliers.
o. The Z-score of the highest value
p. **Probability** of finding a score as high as the highest value or higher.

Confidence Intervals.

a. Below are some summary statistics for the "Feeling Thermometer" score respondents to the 2008 National Election Study gave to the US Supreme Court. The thermometer scores range from 0 "Very Cold" to 100 "Very Warm". ***Calculate the standard error of the estimate of the mean.***

```
    Variable |        N        Mean      Std. Dev.       Min        Max
-------------+---------------------------------------------------------
Supreme Court|      2039    61.98087    19.46994          0        100
```

b. Using the standard error you calculate above, create a 90% confidence interval around the estimate and make a claim about the location of the population parameter.

c. A recent poll of 261 government administrators found that 44% worked uncompensated overtime during July of 2012. ***Calculate the standard error of the estimate of the proportion.***

d. Using the standard error of the proportion you calculated above, build a 95% confidence interval around the estimate and make a claim about the location of the population proportion.

e. Using the original proportion used in question c, determine how much sample you would need to make an estimate of the proportion within 3% with 90% confidence.

Part 4 – Formulas to know

$$M_{Median(position)} = (N+1)*0.5$$

$$v_{ratio} = \frac{\sum f\,(\text{nonmodal})}{N}$$

$$IQR = Q_3 - Q_1$$

Populations

$$\mu = \frac{\sum X_i}{N}$$

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N}$$

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}}$$

$$z = \frac{X - \mu}{\sigma}$$

Samples

$$\bar{x} = \frac{\sum x_i}{n} \qquad s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \qquad MOE = Z_\alpha \frac{s}{\sqrt{n}}$$

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \qquad n = \frac{{Z_{CL}}^2 s^2}{MOE^2}$$

$$z = \frac{x - \bar{x}}{s} \qquad s_{\bar{p}} = \sqrt{\frac{pq}{n}} \qquad s_p = \sqrt{pq}$$

Confidence Intervals

$$\bar{x} \pm z_\alpha \left(\frac{s}{\sqrt{n}} \right)$$

or $\bar{x} \pm z_\alpha s_{\bar{x}}$

$$p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$$

$90\%\ CI = \bar{X} \pm 1.645 s_{\bar{X}}$

$95\%\ CI = \bar{X} \pm 1.96 s_{\bar{X}}$

$99\%\ CI = \bar{X} \pm 2.575 s_{\bar{X}}$

Part 5 – The Normal Distribution Table

Normal Distribution Probability of <u>not</u> being in the RH Tail

z-score + column (X.X)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990