Item and Test Statistics (Chapter 3)
Algebra of Expectations (Appendix A)

Measurement Methods

Lecture 4

2/7/06

Today’s Class

- Item Statistics
  - Binary item score statistics
  - Quantitative item score statistics
- Test Statistics
- Algebra of Expectations

Prerequisite Conditions

- In the examples we will discuss today, we assume several things about the test we administered:
  - We have a finite sample with which we wish to characterize the population.
  - Each respondent has been drawn independently and randomly.
  - It is our hope that our sample (if not truly random) behaves as our population of interest.

Binary Item Statistics: Initial Example

- To begin our class example, imagine if a single examinee is drawn from a well-defined population of interest.
  - The examinee takes a single binary item (0/1).
- Now imagine repeating this process multiple times.
  - Multiple examinees are drawn randomly and produce similar responses (0/1).

The Result

TABLE 3.1
A Constructed Binary Variable

Observations				*Cumulative Proportion*
01101	00001	01001	00110	8/20 = .40
00100	00010	01111	10111	18/40 = .45
00000	00100	00100	10000	21/60 = .35
00101	01100	00000	10000	26/80 = .325
00001	10101	00000	00000	30/100 = .30
00100	00000	00101	00001	34/120 = .283
11000	00011	00010	00011	41/140 = .293
00000	10000	10001	10010	46/160 = .287
00101	00000	00101	00001	51/180 = .283
00010	10000	00100	10110	57/200 = .285
. . .				

Population proportion = .3 (probability)

- The result of our sample process can be shown in the table above.
- Note here that as N gets large, the proportion correct approaches 0.3 – a nice feature of consistent statistics (from the omnipresent law of large numbers).

Why Does This Example Matter?

- The example we just talked about demonstrates the process we like to think we are performing when we build tests.
- This is similar to the examples you may have seen in your basic stats courses with people drawing black and red balls from an urn (with replacement).

Expanding Our Example

- Now imagine we have a total of m cognitive items ($j = 1,\ldots,m$).
  - Each is binary valued, or scored 0/1.

- We will use the following notation throughout the lecture (and the book).
  - X_j is the j^{th} item score (a random variable).
  - x_{ji} is the score of the i^{th} examinee ($i=1,\ldots,n$).
  - x_{ji} is the realization of random variable X_j

Binary Item Statistics

- We represent the **population** proportion of examinees passing item *j* by:

$$\pi_j$$

- In our **sample**, the number of examinees who pass item *j* is thus represented by:

$$n_j$$

Binary Item Statistics

- The relative frequency of passes for item *j* in our sample can be represented by:

$$p_{\mathrm{j}}$$

- And also, recall the sample mean:

$$\overline{X}_j = \sum_{i=1}^{n} \frac{x_{ji}}{n}$$

Proportion Correct

- What we discover is that with binary scoring, in the sample:

$$p_j = \bar{X}_j$$

- And in the population:

$$\pi_j = \mu_j$$

- It is comforting to note that as the sample size increases, p_j becomes an increasingly precise estimate of μ_j

Proportion Correct

- We note something that may be obvious, but is important to remember: both p_j and π_j are proportions – confined between zero and one.
- An item that everyone passes ($\pi_j = 1$) is too easy.
- An item that everyone fails ($\pi_j = 0$) is too hard.

Item Difficulty

- Building off of our example of easy and hard items, in *classical test theory* (CTT), we define **item difficulty**.

- In CTT, the item difficulty parameter is π_j.
  - Thus, the estimate of item difficulty is p_j, which is sometimes noted by p^+.

Terminology

- Ironically, π_j is an inverse measure of item difficulty.
  - Items are more difficult as π_j gets smaller.
  - Items are easier as π_j gets bigger.

- You will discover there are a few people who refer to π_j as item easiness.
  - Although this is not often the case.

Terminology (continued)

- Another problem with the term "item difficulty" arises when we consider non-cognitive items.
- How would you term the π_j for an item such as:
  - I often have hallucinations.
- I would hope this would be an "difficult" item (with a low p_j)

Algebra of Expectations

- What we are working toward is characterizing the distribution of a variable by a set of summary numbers.
- The topics presented today should be really basic.
- The underlying statistical premise of these topics, however, is a bit more complicated.
- The numbers we will use to characterize a distribution are derived from using expectations in statistics (this is shown in Appendix A).

Expectations

- The mean of a binary variable can be represented by the following notation:

$$\mu_j = \pi_j = \mathrm{E}(X_j) = \sum_{x \in \Omega} p_{x_j} x_j$$

- For our item (with π_j = 0.3), this would be represented by:
  - (1-0.3)*0 + (0.3)*1 = 0.3

Proportion	0	1
x	0.7	0.3

Variance

- The variance of a variable can be described by the following expectation:

$$\sigma_{jj} = Var(X_j) = E\left[(X_j - \mu_j)^2\right]$$

- For binary values, the variance is a function of the population mean.

Variance

X_j	$P(X_j)$	$X_j - \mu_j$	$(X_j - \mu_j)^2$
0	$1 - \pi_j$	$-\pi_j$	π_j^2
1	π_j	$1 - \pi_j$	$(1 - \pi_j)^2$

- Here, the variance is formed by taking:

$$\sigma_{jj} = \sum_{j=0}^{1} P(X_j)\left[\left(X_j - \mu_j\right)^2\right]$$

Variance

X_j	$P(X_j)$	$X_j - \mu_j$	$(X_j - \mu_j)^2$
0	$1 - \pi_j$	$-\pi_j$	π_j^2
1	π_j	$1 - \pi_j$	$(1 - \pi_j)^2$

- Or…

$$
\begin{aligned}
\sigma_{jj} &= \pi_j(1-\pi_j)^2 + (1-\pi_j)(-\pi_j)^2 \\
&= \pi_j(1-\pi_j)[1-\pi_j+\pi_j] \\
&= \pi_j(1-\pi_j)
\end{aligned}
$$

Sample Variance

- Our calculation of the variance formula for binary items was done for the population.
- For the sample, the formula is found by inserting the sample proportion p_j:

$$s_j^2 = p_j (1 - p_j)$$

Mean and Variance of Binary Items

- For binary items, the variance is a function of the mean.

TABLE 3.2
Binary Item Variance and Difficulty

π_j	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
σ_j^2	.0	.09	.16	.21	.24	.25	.24	.21	.16	.09	.0

Item Association

- Up to this point, we have talked about characterizing the distribution of a single binary item.

- We will now consider measure of item association.
  - These measures characterize the joint distribution of two binary items.

- You will find that for binary items, there are a multitude of measures of item association.

Two Items

- For two binary items, *j* and *k*, there are four different possible response combinations:

$X_j = 0$ (fail)	$X_k = 0$ (fail)
$X_j = 1$ (pass)	$X_k = 0$ (fail)
$X_j = 0$ (fail)	$X_k = 1$ (pass)
$X_j = 1$ (pass)	$X_k = 1$ (pass)

Independence

- From elementary probability, recall that if two events A and B were independent, then:

$$P(\text{A and B}) = P(\text{A})*P(\text{B})$$

- If we imagine that responses to our two items are independent, we have the same result:

$$P(X_{\text{j}} = 1 \text{ and } X_{\text{k}} = 1) = P(X_j = 1)*P(X_{\text{k}} = 1)$$

Item Independence

- On cognitive tests, however, we do not think items are necessarily independent.
  - We like to think that examinees getting one item correct would tend to have a higher chance of getting the other item correct.
- The idea is that we should take the difference:

$$P(X_j = 1 \text{ and } X_k = 1) - P(Xj = 1)*P(X_k = 1)$$

$$\pi_j - \pi_k\pi_j$$

Correlation

- Recall from basic statistics the idea of correlation as a measure of association between two variables.
- Here, we define the Pearson product-moment correlation:

$$r_{xy} = \frac{(1/n)\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(1/n)\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{(1/n)\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

More Correlation

- The Pearson correlation takes values from -1 to 1, with 0 indicating no association.
- Because of this uniform range, the correlation is thought of as being “dimensionless” or “scale-free.”

Covariance

- The numerator of the Pearson correlation is the sample covariance:

$$s_{xy} = (1/n)\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xy} = (1/n)\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

- The scale of the sample covariance is often difficult to understand (it is in the product of the scales of *x* and *y*).

Covariance of Binary Items

- The covariance of binary items can be defined similarly.
  – In the population – $\sigma_{jk} = \pi_{jk} - \pi_k\pi_j$
  – In the sample – $s_{jk} = p_{jk} - p_j p_k$

- Here, $p_{jk} = P(X_j = 1 \text{ and } X_k = 1)$

Covariance Example

TABLE 3.4
Computation of Binary Item Covariances

	Items	
Ex.	*1*	*2*
1	1	1
2	1	0
3	1	1
4	0	1
5	0	0
6	1	0
7	0	0
8	0	0
9	1	1
10	0	0

$\bar{x}_1 = 5/10 = .5$
$\bar{x}_2 = 4/10 = .4$

$1/10(\Sigma\ x_{1i}x_{2i}) = 3/10 = .3$
$s_{12} = .3 - .5 \times .4 = .1$

$$p_{12} = \frac{\text{count of (1,1) pairs}}{10} = .3$$

$$p_1 = \frac{\text{count of 1s in item 1}}{10} = .5$$

$$p_2 = \frac{\text{count of 1s in item 2}}{10} = .4$$

$p_{12} - p_1 p_2 = .3 - .5 \times .4 = s_{12}$

Correlation Example

- From the covariance and variance of *j* and *k*, we can derive our correlation:

$$r_{\mathrm{jk}} = s_{\mathrm{jk}} /(s_{\mathrm{j}}s_{\mathrm{k}})^{.5} = 0.1/(.25*.24)^{.5} = 0.408.$$

Correlation in Binary Items

- Correlation in binary items represents a problem in that the range is bounded.
  - This bound occurs when the marginal proportion correct for each item is different.
- If two items *j* and *k* have different difficulty parameters, but otherwise were perfectly associated:
  - We could perfect predict the easier item from the more difficult one.
  - But, the correlation is not 1.0 in this case.

Correlation Bounds

- In our thought example, imagine $\pi_j > \pi_k$.

- The maximum possible covariance between the two items would then be:

$$\sigma_{jk} = \pi_j(1 - \pi_k)$$

- The maximum possible correlation would then be:

$$\rho_{jk} = \sqrt{\frac{\pi_j(1-\pi_k)}{\pi_k(1-\pi_j)}}$$

More Correlations

- Because of the bounds on the Pearson correlation, people have developed other measures of **binary** item association:
  - Cohen's Kappa
  - Tetrachoric Correlation
- We will cover tetrachoric correlation once we revisit binary items in Chapter 11.

Correlation/Covariance Matrices

- If one has multiple items, a matrix of correlations or covariances can be constructed.
- This matrix is symmetric – everything in lower triangle is repeated exactly in the upper triangle:

TABLE 3.6
Binary Covariance and Correlation Matrices

(a) Covariance Matrix

	1	2	3	4	5
1	.25	.10	.15	.10	−.05
2	.10	.24	.00	.08	−.06
3	.15	.00	.25	.10	−.05
4	.10	.08	.10	.16	−.02
5	−.05	−.06	−.05	−.02	.09

(b) Correlation Matrix

	1	2	3	4	5
1	1	.408	.600	.500	−.333
2	.408	1	.000	.408	−.408
3	.600	.000	1	.500	−.333
4	.500	.408	.500	1	−.167
5	−.333	−.408	−.333	−.167	1

Correlation Matrix Inspection

- As we progress through this class, we will come to understand that a lot of information can be culled from a correlation matrix.
- Often times, groups of items are highly intercorrelated within, but relatively uncorrelated between.
- These ideas will serve as a base to answer one of our fundamental questions about a test:
  - Does one score suffice or should two scores be used?

Next Time

- Continuous/quantitative item statistics.
- Test scores.
- Algebra of expectations for continuous distributions.