Scale Building with Confirmatory Factor Analysis

Latent Trait Measurement and
Structural Equation Models
Lecture #7

February 27, 2013

Today's Class

- Scale building with confirmatory factor analysis
  - Item information
  - Use of factor scores
  - Reliability for factor scores
    - General concept of reliability

- Additional Psychometric Issues:
  - Construct Maps
  - Item Design
  - Model Fit
  - Model Modification
  - Scale Interpretation
  - Item Information
  - Factor Scores
  - Reliability for Factor Scores
  - Test Information
  - Validity

Data for Today's Class

- Data were collected from two sources:
  - 144 "experienced" gamblers
    - Many from an actual casino
  - 1192 college students from a "rectangular" midwestern state
    - Many never gambled before

- Today, we will combine both samples and treat them as homogenous – one sample of 1346 subjects
  - Later we will test this assumption – measurement invariance (called differential item functioning in item response
    theory literature)

- We will build a scale of gambling tendencies using the first 24 items of the GRI
  - Focused on long-term gambling tendencies

Pathological Gambling: DSM Definition

- To be diagnosed as a pathological gambler, an individual must meet 5 of 10 defined criteria:

1. Is preoccupied with gambling
2. Needs to gamble with increasing amounts of money in order to achieve the desired excitement
3. Has repeated unsuccessful efforts to control, cut back, or stop gambling
4. Is restless or irritable when attempting to cut down or stop gambling
5. Gambles as a way of escaping from problems or relieving a dysphoric mood
6. After losing money gambling, often returns another day to get even
7. Lies to family members, therapist, or others to conceal the extent of involvement with gambling
8. Has committed illegal acts such as forgery, fraud, theft, or embezzlement to finance gambling
9. Has jeopardized or lost a significant relationship, job, educational, or career opportunity because of gambling
10. Relies on others to provide money to relieve a desperate financial situation caused by gambling

Our 24 GRI Items

- The first 24 items of the GRI were written to represent the 10 DSM criteria in the gambling tendencies construct:

Criterion	Item Count
1	3
2	2
3	4
4	1
5	4
6	4
7	2
8	2
9	1
10	1

BUILDING OUR GAMBLING SCALE

Building our Gambling Scale

- The 41 items of the GRI represent the full set of items generated to study the tendency to gamble as a construct
  - The 10 DSM criteria were the basis for the items of the GRI
  - We will use the first 24 items as our item pool (the remaining 17 you will use in your homework assignment)

- Our goal: to create a scale that accurately measures one overall gambling factor

- The key: make sure the one-factor model fits the data
  - If the model does not fit, inferences cannot be made

- The problem: balancing model fit with the nature of the construct
  - Not all items will be retained – so the final construct will likely be different from the original construct

First Step: Analysis of 24 Items

- The first step in our analysis is to examine how the one-factor model fits the entire item pool

```
TITLE:
    Gambling Research Instrument Items
    Data from 1192 College Students/144 Gamblers
    41 Likert Items (1-6): GRI1-GRI41
    12 SOGS items (SOGS4-SOGS15), mostly dichotomous
    ================================================================
    Identification: Marker Item Factor Variance, Zero Factor Mean
    ================================================================
    One-Factor GAMBLING tendencies model with first 24 GRI items

DATA:
    FILE = gamblingdata.csv;

ANALYSIS:
    ESTIMATOR = MLR;

VARIABLE:
    NAMES = GRI1-GRI41 SOGS4-SOGS15 Student ID;
    USEVARIABLES = GRI1-GRI24;
    IDVARIABLE = ID;
    MISSING = ALL(99);

MODEL:
    GAMBLING by GRI1-GRI24;

OUTPUT:
    STANDARDIZED MODINDICES(ALL 0) RESIDUAL SAMPSTAT;
```

Step #1: Assessment of Model Fit

- Our assessment of model fit begins with our global indices of model fit:
  - Model χ^2, RMSEA, CFI, TLI, and SRMR

```
Chi-Square Test of Model Fit

          Value                              2697.493*
          Degrees of Freedom                      252
          P-Value                              0.0000
          Scaling Correction Factor            1.3825
            for MLR
```

The model χ^2 indicated the model did not fit better than the saturated model – but this statistic can be overly sensitive

```
RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.086
          90 Percent C.I.                    0.083  0.089
          Probability RMSEA <= .05           0.000
```

The model RMSEA indicated the model did not fit well (want this to be < .05)

```
CFI/TLI

          CFI                                0.684
          TLI                                0.654
```

The model CFI and TLI indicated the model did not fit well (want these to be > .95)

```
SRMR (Standardized Root Mean Square Residual)

          Value                              0.087
```

The SRMR indicated the model did not fit well (want this to be < .08)

Assessment of Global Model Fit

- **Assessment of global model fit:**
  - Recall that item intercepts, factor means, and variances are just-identified
    - ***<u>Therefore, misfit comes from inaccurately modeled covariances</u>***
  - χ^2 is sensitive to large sample size
  - Pick at least one global fit index from each class; hope they agree (e.g., CFI, RMSEA)
  - If model fit is not good, you should NOT be interpreting the model estimates
    - They will change as the model changes
    - All models are approximations – close approximations are best
  - If model fit is not good, it's your job to find out WHY
  - If model fit is good, it does not mean you are done, however…
    - You can have good fit and poorly functioning items

Step #2: Assessment of Model Misfit

- Our one-factor model ended up not fitting well
  - We must make modifications before we can conclude we have built a good scale to measure gambling tendencies

- Because the model did not fit well, we cannot look at any model-based parameters to give us indications of misfit
  - These are likely to be biased (= wrong or misleading)

- What we must examine is the **residual covariance matrix** using the **normalized residual covariances**
  - Normalized residual covariances are like z-scores (bigger than +/- 2 indicate significant misfit)

Normalized Residual Covariances

- Normalized residual covariances are like z-scores
  - Values bigger than +/- 2 indicate significant misfit

- Positive residuals: items are more related than your model predicts them to be
  - Something other than the factor created the relationship

- Negative residuals: items are less related than your model predicts them to be
  - The overall model causes these to be off

- Evidence of model misfit tells you where the problems with the model lie, but not what to do about fixing them

GRI 24 Item Analysis Normalized Residuals

- The largest normalized residuals:
  - -15.947 (Covariance of Item 20 and Item 22)
    - Negative value: model causes misfit
  - 14.643 (Covariance of Item 20 and Item 12)
    - Positive value: additional features causing extra item covariances
  - 14.083 (Covariance of Item 20 and Item 4)
    - Positive value: additional features causing extra item covariances

- Often, we examine the wording and content of the items for clues as to why they do not fit the model:
  - Item 20: When gambling, I have an amount of money in mind that I am willing to lose, and I stop if I reach that point. (6R)
  - Item 22: Gambling has hurt my financial situation. (10)
  - Item 4: I enjoy talking with my family and friends about my past gambling experiences. (7R)
  - Item 12: When I lose money gambling, it is a long time before I gamble again. (6R)

Ways of Fixing The Model #1: Adding Parameters – <u>Increasing Complexity</u>

- A common source of misfit is due to items that have a significant residual covariance
  - Said another way: items are still correlated after accounting for the common factor
  - For a one-factor model – this indicates that one factor does not fit data (so theory of one-factor is incorrect)

- Solutions that increase model complexity:
  - **Add additional factors** (recommended solution)
    - Factors are additional dimensions (constructs) that are measured by the items
  - **Add a residual covariance between items** (dangerous solution)
    - Use modification indices to determine which to add
    - Error covariances are unaccounted for multi-dimensionality
      - This means you have measured your factor **and** something else that those items have in common (e.g. stem, valence, specific content, additional factors)

Solution #1: Adding Factors

- Adding a factor is equivalent to stating that the hypothesized one-factor model does not fit the data
  - The evidence suggests a one-factor model is not adequate

- The GRI was created to measure each of the 10 criteria of the DSM, not one general gambling factor
  - Likely, gambling tendencies have more than one factor

- We will revisit adding an additional factor in a future class
  - When we look at multidimensional factor models

- For now, our focus will be on building a one-factor scale

Solution #2: Examining Modification Indices for Residual Covariances

- Note: this solution is not recommended as it weakens the argument that a single factor underlies a scale
  - It also is seen as a trick to improve model fit

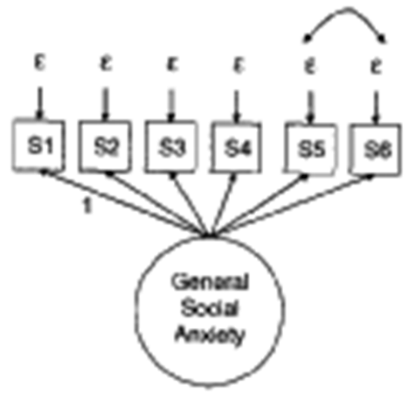
- The largest modification indices for residual covariances:
  - Item 20 with Item 4; MI = 294.968
  - Item 22 with Item 20; MI = 247.625
  - Item 20 with Item 12; MI = 226.539

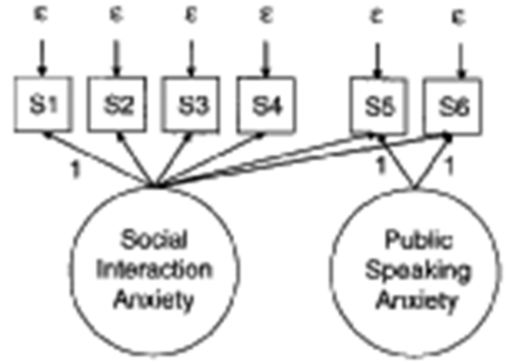
- Each of these items is reverse coded
  - Indication that wording of items elicits different response
  - Potential for a reverse-worded method factor

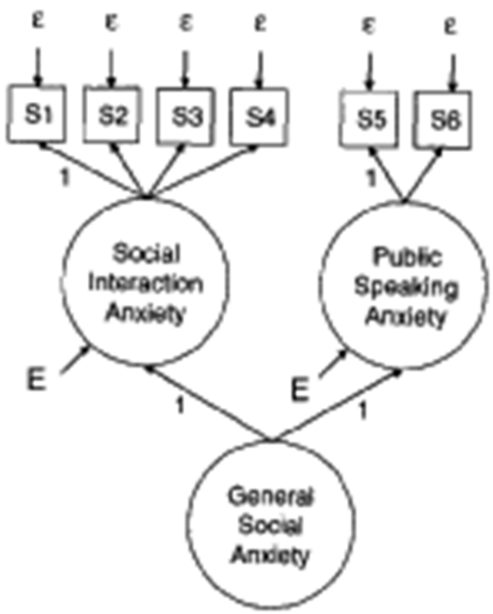
- We will not add these to our model – we want our single factor to be only thing that "explains" items

Error Covariances Actually Represent Multidimensionality

From Brown (2006)







Here there is a general factor of "Social Interaction Anxiety" that includes two items dealing with public speaking specifically.

The extra relationship between the two public speaking items can be modeled in many different, yet statistically equivalent ways… error covariances really represent another factor (which is why you should be able to explain and predict them if you include them).

Ways of Fixing The Model #2: Removing Terms – <u>Decreasing Complexity</u>

- Solution #1: When multiple factors are correlated > .85 may suggest a simpler structure – remove factors
  - Nested model comparison: fix factor variances to 1 so factor covariance becomes factor correlation, then test correlation ≠1 at $p < .10$

- Solution #2: Dropping Items; Drop items with:
  - <u>**Non-significant loadings:**</u> If the item isn't related, it isn't measuring the construct, and you most likely don't need it
  - <u>**Negative loadings**</u>: Make sure to reverse-coded as needed ahead of time – otherwise, this indicates a big problem!
  - <u>**Problematic leftover positive covariances between two items**</u> – such redundancy implies you may not need both items

- However – models with differing items are NOT COMPARABLE AT ALL because their Log-Likelihood values are based on different data!
  - No model comparisons of any kind (including AIC and BIC)
  - To do a true comparison, you'd need to leave the item in the model but remove its loading (≈ original test of its loading)

List of Items to Remove

- Because our model fit is terrible we will modify our model by dropping items that do not fit well
  - This will change our gambling construct but will allow us to (hopefully) have one factor measured by the test
- There are 9 items with 10 or more significant normalized residual covariances:

GRI22	16
GRI4	15
GRI20	15
GRI2	13
GRI8	13
GRI19	13
GRI24	13
GRI12	12
GRI17	12

Item Removal Logic and Details

- By dropping items with a number of significant normalized residual covariances we will reduce the number of items in our analysis thereby reducing the number of terms in the saturated covariance matrix

- This can make it easier to achieve a reasonable approximation as the number of covariances increases exponentially with each additional item, but the number of statistical parameters increases by two (factor loading and unique variance)
  - We would be trying to approximate a lot of covariances terms with only a few parameters

- We will remove the 9 items from the list on the previous page, leaving us with a 15-item analysis

GRI 15 Item Analysis

- The 15 item analysis gave this model fit information:

```
Chi-Square Test of Model Fit

      Value                          420.480*
      Degrees of Freedom                   90
      P-Value                          0.0000
      Scaling Correction Factor        1.6453
        for MLR

RMSEA (Root Mean Square Error Of Approximation)

      Estimate                         0.053
      90 Percent C.I.                  0.048  0.058
      Probability RMSEA <= .05         0.157

CFI/TLI

      CFI                              0.909
      TLI                              0.894

SRMR (Standardized Root Mean Square Residual)

      Value                            0.042
```

The model χ^2 indicated the model did not fit better than the saturated model – but this statistic can be overly sensitive

The model RMSEA indicated acceptable model fit (want this to be < .05; .06-.08 is acceptable)

The model CFI and TLI indicated the model adequately (want these to be > .95)

The SRMR indicated the fit well (want this to be < .08)

Examining the Normalized Residuals

- The normalized residuals from the analysis indicated that several items had questionable fit:
  - Items 7, 16, and 18 had four significant normalized residuals
  - The rest had 2 or fewer (4 items had none)

- At this point, the choice of removal of additional items is ultimately up to theory
  - The fit of the model is adequate – removal of items may make the model fit better
  - The construct may be significantly altered by removing items measuring certain features

- We will choose to omit items 7, 16, and 18 from the scale, and rerun the analysis with 12 items

GRI 12 Item Analysis

- The 12 item analysis gave this model fit information:

```
Chi-Square Test of Model Fit

          Value                            185.178*
          Degrees of Freedom                     54
          P-Value                            0.0000
          Scaling Correction Factor          1.6034
            for MLR

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.043
          90 Percent C.I.                     0.036  0.050
          Probability RMSEA <= .05            0.949

CFI/TLI

          CFI                                 0.952
          TLI                                 0.941

SRMR (Standardized Root Mean Square Residual)

          Value                               0.032
```

The model χ^2 indicated the model did not fit better than the saturated model – but this statistic can be overly sensitive

The model RMSEA indicated good model fit (want this to be < .05)

The model CFI and TLI indicated the model fit well (want these to be > .95)

The SRMR indicated the fit well (want this to be < .08)

- Additionally, only 4 normalized residuals were significant
  - Item 15 with item 1
  - Item 14 with item 3
  - Item 21 with item 3
  - Item 15 with item 6

Interpreting Parameters

- The one-factor model seems to fit the 12 GRI items so we will interpret the parameters

Item	Unstandardized Loading	Residual Variance	Standardized Loading (STDYX)	
GRI 1	1.000 (0.000)	0.697 (0.066)	.567 (.037)	.322
GRI 3	0.785 (0.073)	0.545 (0.044)	.522 (.036)	.272
GRI 5	1.118 (0.096)	0.499 (0.047)	.673 (.032)	.453
GRI 6	0.815 (0.070)	0.309 (0.024)	.645 (.033)	.416
GRI 9	0.960 (0.059)	0.215 (0.017)	.766 (.023)	.587
GRI 10	1.068 (0.081)	0.369 (0.040)	.711 (.028)	.506
GRI 11	1.012 (0.075)	0.854 (0.078)	.533 (.033)	.284
GRI 13	1.172 (0.073)	0.462 (0.047)	.704 (.028)	.496
GRI 14	1.023 (0.095)	1.837 (0.078)	.398 (.026)	.159
GRI 15	0.857 (0.071)	1.219 (0.078)	.408 (.029)	.166
GRI 21	0.967 (0.071)	0.377 (0.036)	.672 (.028)	.451
GRI 23	1.086 (0.080)	0.514 (0.044)	.657 (.026)	.432

Interpreting Parameters

- Each item had a statistically significant factor loading
  - The item measures the factor/is correlated with the factor

- The standardized factor loadings ranged from .766 (item 9) to .398 (item 14)

- The R^2 is the squared standardized loading
  - Item 9 had an R^2 of .587
    - Items with high R^2 are better for measuring the factor
  - Item 14 had an R^2 of .159
    - Items with low R^2 are not contributing much to the factor

Interpreting the Scale

- The final gambling scale had a different set of items than the original gambling scale
  - As such, the construct measured by the final scale is different that the construct that would be measured if the full set of items were used (and fit a one-factor model)

Criterion	24-Item Count	12-Item Count
1	3	2
2	2	2
3	4	2
4	1	1
5	4	2
6	4	1
7	2	1
8	2	1
9	1	0
10	1	0

Pathological Gambling: DSM Definition

- To be diagnosed as a pathological gambler, an individual must meet 5 of 10 defined criteria:

1. Is preoccupied with gambling
2. Needs to gamble with increasing amounts of money in order to achieve the desired excitement
3. Has repeated unsuccessful efforts to control, cut back, or stop gambling
4. Is restless or irritable when attempting to cut down or stop gambling
5. Gambles as a way of escaping from problems or relieving a dysphoric mood
6. After losing money gambling, often returns another day to get even
7. Lies to family members, therapist, or others to conceal the extent of involvement with gambling
8. Has committed illegal acts such as forgery, fraud, theft, or embezzlement to finance gambling
9. Has jeopardized or lost a significant relationship, job, educational, or career opportunity because of gambling
10. Relies on others to provide money to relieve a desperate financial situation caused by gambling

Final 12 Items on the Scale

Item	Criterion	Question
GRI1	3	I would like to cut back on my gambling.
GRI3	6	If I lost a lot of money gambling one day, I would be more likely to want to play again the following day.
GRI5	2	I find it necessary to gamble with larger amounts of money (than when I first gambled) for gambling to be exciting.
GRI6	8	I have gone to great lengths to obtain money for gambling.
GRI9	4	I feel restless when I try to cut down or stop gambling.
GRI10	1	It bothers me when I have no money to gamble.
GRI11	5	I gamble to take my mind off my worries.
GRI13	3	I find it difficult to stop gambling.
GRI14	2	I am drawn more by the thrill of gambling than by the money I could win.
GRI15	7	I am private about my gambling experiences.
GRI21	1	It is hard to get my mind off gambling.
GRI23	5	I gamble to improve my mood.

QUANTIFYING AN ITEM’S CONTRIBUTION: ITEM INFORMATION

Item Information

- The amount of information an item provides about a factor is a combination of:
  - The size of the factor loading
  - The size of the error variance
- The index of the statistical information that item *i* provides for factor f is $\frac{\lambda_{if}^2}{\psi_i^2}$
  - The unstandardized loadings provide the information for a factor with a variance fixed at σ_f^2
  - The standardized loadings provide the information for a factor with a variance of 1
  - The rank order of the information will be the same using either unstandardized or standardized loadings
    - So choice is arbitrary as both work the same way
- Note: as information depends on the model parameters (loadings and unique variances) a model must fit well to have an accurate sense of the information provided by each item

Information of Our Items

- Given that the one-factor model seems to fit the 12 GRI items, we use the parameters to calculate item information

Item	Unstandardized Loading	Residual Variance	Item Information	Info Rank
GRI 1	1.000 (0.000)	0.697 (0.066)	1.434	8
GRI 3	0.785 (0.073)	0.545 (0.044)	1.130	10
GRI 5	1.118 (0.096)	0.499 (0.047)	2.504	4
GRI 6	0.815 (0.070)	0.309 (0.024)	2.148	7
GRI 9	0.960 (0.059)	0.215 (0.017)	4.296	1
GRI 10	1.068 (0.081)	0.369 (0.040)	3.092	2
GRI 11	1.012 (0.075)	0.854 (0.078)	1.201	9
GRI 13	1.172 (0.073)	0.462 (0.047)	2.971	3
GRI 14	1.023 (0.095)	1.837 (0.078)	0.570	12
GRI 15	0.857 (0.071)	1.219 (0.078)	0.602	11
GRI 21	0.967 (0.071)	0.377 (0.036)	2.482	5
GRI 23	1.086 (0.080)	0.514 (0.044)	2.297	6

Using the Mplus NEW Command to Calculate Item Information

```
MODEL:
    GAMBLING by GRI1
                GRI3   (L3)
                GRI5   (L5)
                GRI6   (L6)
                GRI9   (L9)
                GRI10 (L10)
                GRI11 (L11)
                GRI13 (L13)
                GRI14 (L14)
                GRI15 (L15)
                GRI21 (L21)
                GRI23 (L23);

    GRI1 (U1); GRI3 (U3); GRI5 (U5); GRI6 (U6); GRI9 (U9); GRI10 (U10);
    GRI11 (U11); GRI13 (U13); GRI14 (U14); GRI15 (U15); GRI21 (U21); GRI23 (U23);

MODEL CONSTRAINT:
NEW(INFO1 INFO3 INFO5 INFO6 INFO9 INFO10 INFO11 INFO13 INFO14 INFO15 INFO21 INFO23);

INFO1 = 1/U1;
INFO3 = L3^2/U3;
INFO5 = L5^2/U5;
INFO6 = L6^2/U6;
INFO9 = L9^2/U9;
INFO10 = L10^2/U10;
INFO11 = L11^2/U11;
INFO13 = L13^2/U13;
INFO14 = L14^2/U14;
INFO15 = L15^2/U15;
INFO21 = L21^2/U21;
INFO23 = L23^2/U23;
```

NEW Command Output:

New/Additional Parameters				
INFO1	1.434	0.135	10.638	0.000
INFO3	1.130	0.241	4.687	0.000
INFO5	2.504	0.543	4.613	0.000
INFO6	2.148	0.419	5.131	0.000
INFO9	4.296	0.635	6.763	0.000
INFO10	3.092	0.610	5.070	0.000
INFO11	1.201	0.233	5.143	0.000
INFO13	2.971	0.507	5.858	0.000
INFO14	0.570	0.116	4.909	0.000
INFO15	0.602	0.116	5.203	0.000
INFO21	2.482	0.435	5.708	0.000
INFO23	2.297	0.411	5.582	0.000

GAMBLING BY				
GRI1	1.000	0.000	999.000	999.000
GRI3	0.785	0.073	10.791	0.000
GRI5	1.118	0.096	11.661	0.000
GRI6	0.815	0.070	11.589	0.000
GRI9	0.960	0.059	16.173	0.000
GRI10	1.068	0.081	13.181	0.000
GRI11	1.012	0.075	13.584	0.000
GRI13	1.172	0.073	16.131	0.000
GRI14	1.023	0.095	10.791	0.000
GRI15	0.857	0.071	12.000	0.000
GRI21	0.967	0.071	13.590	0.000
GRI23	1.086	0.080	13.635	0.000

Residual Variances				
GRI1	0.697	0.066	10.638	0.000
GRI3	0.545	0.044	12.360	0.000
GRI5	0.499	0.047	10.559	0.000
GRI6	0.309	0.024	12.902	0.000
GRI9	0.215	0.017	12.732	0.000
GRI10	0.369	0.040	9.315	0.000
GRI11	0.854	0.078	10.927	0.000
GRI13	0.462	0.047	9.912	0.000
GRI14	1.837	0.078	23.406	0.000
GRI15	1.219	0.078	15.615	0.000
GRI21	0.377	0.036	10.560	0.000
GRI23	0.514	0.044	11.618	0.000

Notes About Item Information

- Under CFA (which assumes items that are continuous/normally distributed), information is constant for each item
  - This differs when you have categorical items: Information differs by values of the latent trait (some items measure certain trait levels better than others)

- Because item information is constant under CFA, it gets very little attention
  - By extension, certain types of tests aren't possible with CFA – such as computer adaptive tests (as the best item at each point would be the best item overall)

- Item information can be used to select a set of items to be used to measure a construct
  - A shorter yet efficient test
  - Remember – changing the items changes the construct

Building A Shorter Form

- Using item information, if we wanted to build a shorter form of our test, we would pick the top five items
  - We may stratify by content if that is an important facet for the nature of the construct

- Neglecting content, a five-item test would be based on:
  - Item 9 (Information = 4.296)
  - Item 10 (Information = 3.092)
  - Item 13 (Information = 2.971)
  - Item 5 (Information = 2.504)
  - Item 21 (Information = 2.482)

- The 5-item test would still have less precision (reliability) than the 12 item test (more on this shortly)

THE USE OF FACTOR SCORES

Factor Scores

- Once a scale has been constructed and is shown to fit some type of analysis model (here, we used a one-factor model) the next step in an analysis is typically to assess each subject's score on the scale
  - Representing their level of the construct measured by the scale

- In SEM/CFA ("continuous" items), factor scores are not typically used, instead:
  - If a scale "fits" a unidimensional model, sum scores are used (less ideal – not preferred in SEM)
    - More on this next week
  - A path model with the latent factor as a predictor (IV) or outcome (DV) is used (the preferred approach)
    - More on this after the CFA/measurement models section

More on Factor Scores

- Factor scores (by other names) are used in many domains
  - Item response theory (CFA with categorical items) – i.e. the GRE
- Because the historical relationship between CFA and exploratory factor analysis, factor scores are widely avoided and deadpanned
  - In exploratory factor analysis, factor scores are indeterminate
    - No single "best" score
  - Due to too many loadings
- The use of factor scores from CFA is possible and is justified as they are uniquely determined (like IRT/unlike EFA)
  - I seek to describe why in the next few slides

Factor Scores: The Big Picture

- A factor score is the estimate of a subject's level of an unobserved latent variable

- Because this latent variable is not measured directly, it is essentially a piece of missing data

- It is difficult to pin down what the missing data value (factor score value) should be precisely
  - Each factor score has a distribution of possible values
  - The mean of the distribution is the "factor score" – it is the most likely value
  - Depending on the test, there may be a lot of error (variability) in the distribution

- Therefore, the use of factor scores must reflect that the score is not known and is represented by a distribution

Factor Scores: Empirical Bayes Estimates

- For most (if not all) latent variable techniques (e.g., CFA, IRT, mixed/multilevel/hierarchical models), the estimate of a factor score is a so-called empirical Bayes estimate
  - Empirical = some or all of the parameters of the distribution of the latent variable are estimated (i.e., factor mean and variance)
  - Bayes = the score comes about from the use of Bayes' Theorem

- Bayes' Theorem states the conditional distribution of a variable A (soon to be our factor score) given values of a variable B (soon to be our data) is:

$$f(A|B) = \frac{f(B|A)f(A)}{f(B)} = \frac{f(B|A)f(A)}{\int_{a \in A} f(B|A=a)f(A=a)da}$$

Factor Scores through Empirical Bayes

- In the empirical Bayes formula:
  - The variable A is our factor score (F_{sf} for subject s and factor f)
  - The variable B is the data (the vector $\mathbf{Y}_s$ for a subject s)

$$f\left(F_{sf}\middle|\mathbf{Y}_s\right) = \frac{f\left(\mathbf{Y}_s\middle|F_{sf}\right)f(F_{sf})}{f(\mathbf{Y}_s)} = \frac{f\left(\mathbf{Y}_s\middle|F_{sf}\right)f(F_{sf})}{\int_{a\in A} f\left(\mathbf{Y}_S\middle|F_{sf}=a\right)f\left(F_{sf}=a\right)da}$$

  - $f\left(F_{sf}\middle|\mathbf{Y}_s\right)$ is the **distribution** of the factor score given the data
    - **This is the posterior distribution (we are doing Bayesian stats!)**
  - $f\left(\mathbf{Y}_s\middle|F_{sf}\right)$ is the CFA model (think multivariate normal density), evaluated at factor score F_{sf}
  - $f(F_{sf})$ is the distribution of the factor score, evaluated at F_{sf}
    - This comes from our CFA assumption of normality of factor score: $F_{sf} \sim N(\mu_f, \sigma_f^2)$ where μ_f and σ_f^2 are set by identifiability constraints

Moving from Distributions to Estimates

- The previous slide provides the distribution of the factor score - <u>understanding a factor score has a distribution is the key to its use</u>

- For the CFA model ("continuous"/normal data and normal factors), this distribution will be normal (for a single factor) or multivariate normal (for more than one factor)
  - For our example with one factor: univariate normal
  - Univariate normal has two parameters: mean and variance
    - Mean = factor score (<u>**called Expected A Posteriori or EAP estimate**</u>)

$$\hat{F}_{sf} = E\left(F_{sf}\right) = \int_{a \in F} a\, f\left(F_{sf} | \mathbf{Y}_s\right) da$$

    - Variance = variance of factor (<u>**directly related to reliability**</u>)

$$\hat{\sigma}^2_{F_f} = E\left(F_{sf} - E\left(F_{sf}\right)\right)^2 = \int_{a \in F} \left(a - E\left(F_{sf}\right)\right)^2 f\left(F_{sf} | \mathbf{Y}_s\right) da$$

Factor Scores from Mplus

- The SAVEDATA command is used to have Mplus output the factor scores for each subject:

```
VARIABLE:
    NAMES = ID GRI1-GRI41;
    USEVARIABLES = GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23
                   GRI12sum GRI12avg;
    IDVARIABLE = ID;
    MISSING = ALL(99);
    AUXILIARY = GRI12sum GRI12avg;
```

```
MODEL:
    GAMBLING by GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23;
```

```
DEFINE:
    GRI12avg = MEAN(GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23);
    GRI12sum = SUM(GRI1 GRI3 GRI5-GRI6 GRI9-GRI11 GRI13-GRI15 GRI21 GRI23);
```

```
SAVEDATA:
    !saves latent trait estimates
    SAVE = FSCORES;

    !puts latent trait estimates into file named *.dat
    FILE = gri10item_fscores.dat;
```

ID Variable

Variables in Analysis

Auxiliary Variables (not in analysis)

Latent Variable Person Estimate: Mean and SE

```
Order and format of variables

GRI1           F10.3
GRI3           F10.3
GRI5           F10.3
GRI6           F10.3
GRI9           F10.3
GRI10          F10.3
GRI11          F10.3
GRI13          F10.3
GRI14          F10.3
GRI15          F10.3
GRI21          F10.3
GRI23          F10.3
ID             I5
GRI12SUM       F10.3
GRI12AVG       F10.3
GAMBLING       F10.3
GAMBLING_SE    F10.3
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	GRI1	GRI3	GRI5	GRI6	GRI9	GRI10	GRI11	GRI13	GRI14	GRI15	GRI21	GRI23	ID	GRI12SUM	GRI12AVG	GAMBLING	GAMBLING_SE
2	3	1	4	1	4	5	1	1	1	2	1	1	1	25	2.083	0.776	0.19
3	2	1	1	1	4	2	3	5	6	4	3	4	2	36	3	1.185	0.19
4	4	4	1	2	4	6	5	5	2	2	2	4	3	41	3.417	1.818	0.19
5	4	2	5	5	5	4	2	5	2	5	3	2	4	44	3.667	2.163	0.19
6	4	2	4	1	3	4	2	3	4	3	3	2	5	35	2.917	1.186	0.19

Interpreting Factor Score Output

- From Mplus:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	GRI1	GRI3	GRI5	GRI6	GRI9	GRI10	GRI11	GRI13	GRI14	GRI15	GRI21	GRI23	ID	GRI12SUM	GRI12AVG	GAMBLING	GAMBLING_SE
2	3	1	4	1	4	5	1	1	1	2	1	1	1	25	2.083	0.776	0.19
3	2	1	1	1	4	2	3	5	6	4	3	4	2	36	3	1.185	0.19
4	4	4	1	2	4	6	5	5	2	2	2	4	3	41	3.417	1.818	0.19
5	4	2	5	5	5	4	2	5	2	5	3	2	4	44	3.667	2.163	0.19
6	4	2	4	1	3	4	2	3	4	3	3	2	5	35	2.917	1.186	0.19

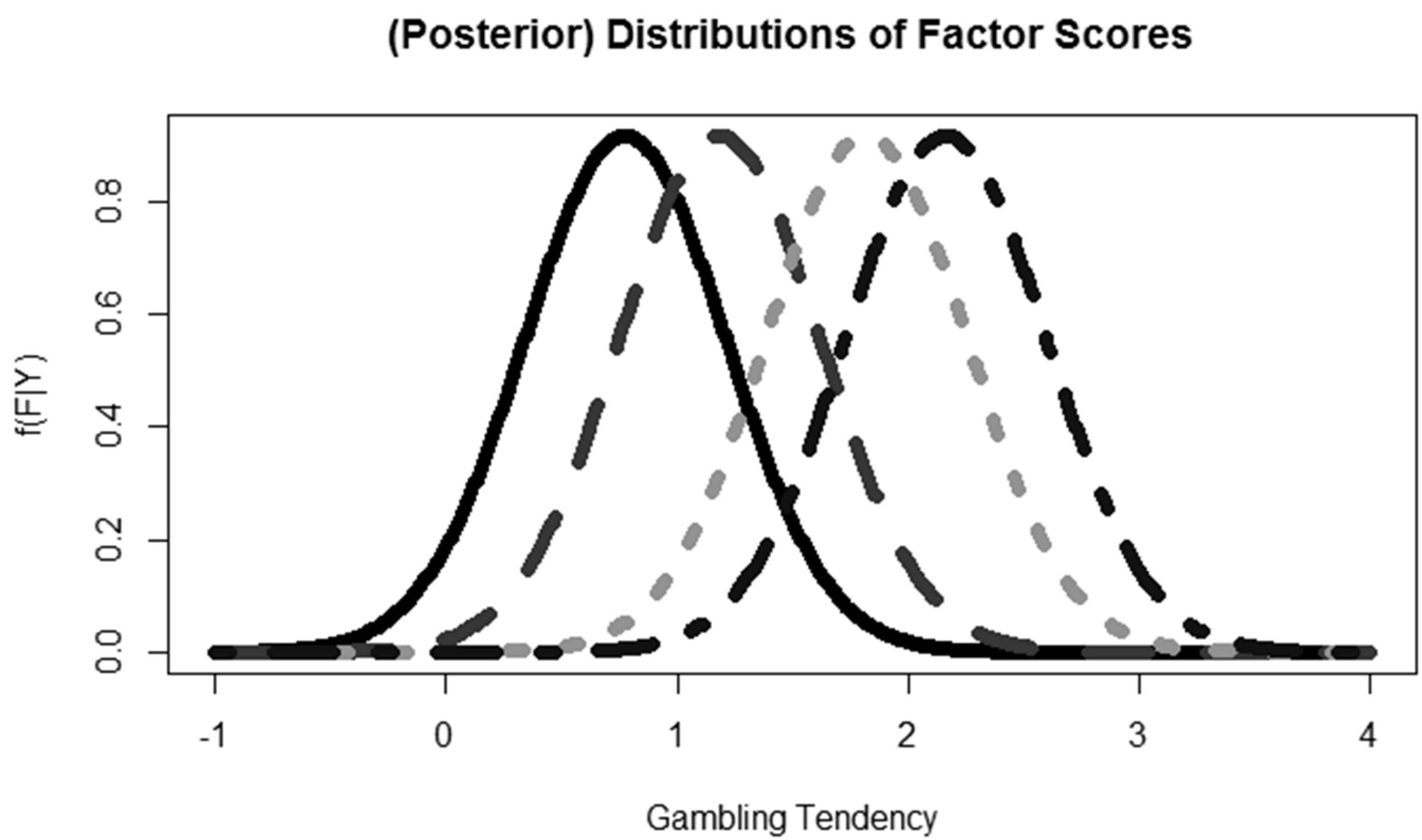
- Subject #1 – Factor Score: $\hat{F}_{11} = 0.776$; Standard Error of Factor Score: $\hat{\sigma}_{F_1} = 0.19$
- Subject #2 - $\hat{F}_{21} = 1.185; \hat{\sigma}_{F_1} = 0.19$
- Subject #3 - $\hat{F}_{31} = 1.818; \hat{\sigma}_{F_1} = 0.19$
- Subject #4 - $\hat{F}_{41} = 2.163; \hat{\sigma}_{F_1} = 0.19$

- Note: σ_{F_1} is the same for all subjects with complete data
  - This is due to us using CFA (item information is constant)

- For subjects with missing data, the standard errors will be different:

GRI1	GRI3	GRI5	GRI6	GRI9	GRI10	GRI11	GRI13	GRI14	GRI15	GRI21	GRI23	ID	GRI12SUM	GRI12AVG	GAMBLING	GAMBLING_SE
3	1	2	1	1	1	4	1	3	*	1	2	75	*	1.818	-0.045	0.192
2	1	4	1	2	2	1	4	*	3	2	2	76	*	2.182	0.596	0.192

Distributions of Factor Scores for Subjects 1-4



Secondary Analyses with Factor Scores

- If using CFA, factor scores are not widely accepted
- BUT you want to use results from a survey in a new analysis

- Solutions (which we will get to next class):
  - <u>Best:</u> Use SEM – error in factor scores is already partitioned variance
  - <u>Similarly good:</u> Use "plausible values" (repeated draws from posterior distribution of each person's factor score) – essentially what SEM does – but with factor scores that vary within a person
    - Used in National Assessment of Educational Progress (NAEP)
  - <u>Possibly Okay (but widespread)</u>: for scales that are unidimensional (and verified in CFA), use **sum scores**
    - Assumes unidimensionality and "high" reliability
    - Should also be a distribution
  - <u>Okay:</u> Use SEM with "single indicator" factors using sum scores
    - More on this later
    - Make error variance = (1-reliability)*Variance (Sum score); factor loading = 1
  - <u>Not Cool:</u> Use factor scores only
    - Considered bad because of EFA (but CFA has different scores)
    - Widespread in other measurement models (like IRT)
    - Should give better results than sum-scores, but also neglects reliability and model fit issue

RELIABILITY OF FACTOR SCORES

Reliability of Factor Scores

- The factor score itself is not measured perfectly - the error of the factor score distribution, $\hat{\sigma}_F^2$, indicates how much each factor score estimate varies
  - If $\hat{\sigma}_F^2$ were to be 0, then we would have perfect measurement
  - Perfect measurement = reliability of 1.0

- Our goal is to quantify the ***reliability*** of the factor score

- The classical notion of reliability still holds:

$$\rho_F = \frac{Var(True\ Score)}{Var(True\ Score) + Var(Error)} = \frac{\sigma_F^2}{\sigma_F^2 + \hat{\sigma}_F^2}$$

  - σ_F^2 = variance of the factor (from the CFA model – the true score)
  - $\hat{\sigma}_F^2$ = variance of the estimated factor score (error)

Reliability of Our Factor Score

- Using the CFA model estimate, we found that: $\sigma_F^2 = 0.331$

- From our estimated factor scores, $\hat{\sigma}_F = 0.19; \hat{\sigma}_F^2 = 0.036$

- The reliability for the factor score is:

$$\rho_F = \frac{\sigma_F^2}{\sigma_F^2 + \hat{\sigma}_F^2} = \frac{0.331}{0.331 + 0.036} = .902$$

  - This reliability is above a minimal standard of .8 – meaning our factor is fairly well-measured

Test Information

- The information about a factor each item contributes is an indirect index of that item's ability to measure the factor

- The information a TEST has about a factor is the sum of the item information for each item of the test

- For our test (from item information slide):

$$I_{test} = \sum_{i=1}^{12} I_i$$
$$= 1.434 + 1.13 + 2.504 + 2.148 + 4.296 + 3.092 + 1.201 + 2.971 + 0.570 + 0.602 + 2.482 + 2.297 = 24.727$$

Test Information and Factor Score Variance

- The factor score variance is directly related to the test information function:

$$\hat{\sigma}_F^2 = \frac{1}{I_{test}}$$

- For our test: $\hat{\sigma}_F^2 = \frac{1}{I_{test}} = \frac{1}{24.727} \approx 0.04$
  - And, consequently, $\hat{\sigma}_F = 0.19$

- To increase the reliability of the measurement of a factor, you must increase the test information
  - To do that, you must add **informative** items

CONCLUDING REMARKS

Wrapping Up

- Today, we built a scale using confirmatory factor analysis

- In the process, we touched on many important topics in psychometrics:
    - Construct Maps
    - Item Design
    - Model Fit
    - Model Modification (by removing items)
    - Scale Interpretation
    - Item Information
    - Factor Scores
    - Reliability for Factor Scores
    - Test Information
    - Validity

Where We Are Heading

- Next week, we will continue our discussion of confirmatory factor analysis with one factor
- We will compare classical test theory to confirmatory factor analysis (and show how we can test the assumptions of the ASU model)
  - This will bring about a comparison of reliability coefficients
- We will discuss how to use factor scores, sum scores, item parcels, and single-indicator models in analyses