missing completely at random (MCAR). The assumption of MCAR holds if the probability of missing data on Y is unrelated to Y or to the values of any other variable in the data set.

完全随机缺失（MCAR）。如果Y的数据缺失的概率与Y或数据集中的任何其他变量的值无关，那么MCAR的假设就成立。

The assumption of *missing at random* (MAR) is met when the probability that data are missing on Y may depend on the value of X, but is not related to the value of Y when holding X constant.

当数据在Y上缺失的概率可能取决于X的值，但在保持X不变的情况下与Y的值无关时，就符合随机缺失的假设（MAR）。

Missing data are said to be *nonignorable* if the missingness is related to values that would have been observed (i.e., the condition of MAR does not hold). Missing data are nonignorable if cases with missing data on a given variable would have higher or lower values on that variable than cases with data present, controlling for all other variables in the data set.

如果缺失的数据与本来可以观察到的数值有关（即MAR的条件不成立），则被称为不可忽略的数据。如果在控制数据集中的所有其他变量的情况下，某个变量的数据缺失的情况下，该变量的数值会比有数据的情况下更高或更低，那么缺失的数据就是不可忽略的。

In fact, when data are MCAR, listwise deletion will produce *consistent* (i.e., unbiased) parameter esti- mates, standard errors, and test statistics (e.g., model χ^2).

事实上，当数据是MCAR时，列表删除将产生一致的（即无偏的）参数估计、标准误差和测试统计（如模型χ2）。

However, estimates produced using listwise deletion are usually not *efficient*. This is because listwise deletion often results in loss of a considerable proportion of the original sample. Thus, standard errors will frequently be larger when listwise deletion is used, as compared with alternative methods (e.g., multiple imputation) that use all of the available data. The inflation in standard errors will thus decrease statistical power and lower the precision of the parameter esti- mates (wider confidence intervals).

然而，使用列表式删除法产生的估计值通常是不有效的。这是因为顺时针删除通常会导致损失相当一部分的原始样本。因此，与使用所有可用数据的替代方法（如多重归因法）相比，使用列表删除法时，标准误差往往会更大。因此，标准误差的膨胀将降低统计能力，并降低参数测定的精度（更宽的置信区间）。

listwise deletion can be a very effective (and straightforward) missing data strategy only when the MCAR assumption holds and only a very small part of the sample is removed.

只有当MCAR假设成立，并且只有很小一部分样本被删除时，列表删除才是一种非常有效（和直接）的缺失数据策略。

data are MAR, listwise deletion may produce results that are neither consistent nor efficient.

数据是MAR，列表式删除可能产生既不一致也不高效的结果。

Another common missing data strategy is *pairwise deletion*. A variety of statistical analyses (e.g., EFA, CFA, multiple regression, ANOVA) can be performed using means, variances, and covariances as input data (i.e., do not require raw data as input).

另一个常见的缺失数据策略是成对删除。各种统计分析（如EFA、CFA、多元回归、方差分析）可以使用平均值、方差和协方差作为输入数据（即不需要原始数据作为输入）。

For one, pairwise deletion produces biased standard errors.

其一，成对删除会产生有偏差的标准误差。

To illustrate: for three variables, X, Y, and Z, the correlation between X and Y ($r_{x,y}$) must be within a certain range, as determined by the following equation:

举例说明：对于三个变量X、Y和Z，X和Y之间的相关性（rx,y）必须在一定范围内，由以下公式决定。

$$r_{x,z}r_{y,z} \pm \text{SQRT}[(1 - r_{x,z}^2)(1 - r_{y,z}^2)]$$

if $r_{x,z} = .70$ and $r_{y,z} = .80$, then the value of $r_{x,y}$ must be with the range of .13 to .99 (i.e., .56 ± .43). If $r_{x,y} < .13$, then the input matrix would not be positive definite.

如果 $r_{x,z}$ =0.70， $r_{y,z}$ =0.80，那么$r_{x,y}$的值必须在0.13到0.99的范围内（即0.56±0.43）。如果 $r_{x,y}$ <.13，那么输入矩阵就不是正定的。

simple (or single) imputation (mean and regression imputation). For example, regression imputation (referred to by Allison, 2002, as “con- ditional mean imputation”) entails regressing the variable with missing data on other variables in the data set for cases with complete data.

简单（或单一）的归因法（平均值和回归归因法）。例如，回归归因法（Allison，2002年，称为“条件平均归因法"）需要将数据缺失的变量回归到数据集中的其他变量上，以获得完整的数据。

Recommended Missing Data Strategies

ML

Maximum likelihood (direct ML) and multiple impu- tation are the most widely preferred methods for handling missing data in SEM

最大似然法（直接ML）和多重暗示法是处理SEM中缺失数据的最广泛首选方法。

When ML is used in context of missing data, it is often referred to as *full information maximum likelihood*, or FIML.

当ML被用于缺失数据的情况下，它通常被称为完全信息最大似然，或FIML。

prefer the term *direct ML* or *raw ML*, because ML estimation with missing data requires that raw data be input to the analysis rather than a variance–covariance matrix (and means).

更喜欢直接ML或原始ML这个词，因为有缺失数据的ML估计需要将原始数据输入分析，而不是方差矩阵（和平均值）。

EM algorithm is a computational device for obtaining ML estimates of the means and the covariance matrix

EM算法是一种用于获得均值和协方差矩阵的ML估计的计算装置

limitation of using the EM algorithm to calculate input matrices for CFA/ SEM is that the resulting standard errors of the parameter estimates are not consistent. Thus, confidence intervals and significance tests may be com- promised. As with pairwise deletion, this is due in part to the problem of specifying the proper sample size

使用EM算法来计算CFA/SEM的输入矩阵的局限性在于，所得到的参数估计的标准误差是不一致的。因此，置信区间和显着性检验可能会出现问题。与成对删除一样，这部分是由于指定适当的样本量的问题造成的

Direct ML

Metho- dologists generally regard **direct ML** to be the best method for handling missing data in most CFA and SEM applications

计量学家普遍认为直接ML是处理大多数CFA和SEM应用中缺失数据的最佳方法。

Direct ML assumes that the data are MCAR or MAR and multivariate normal. However, when data are non-normal, direct ML can be imple- mented to provide standard errors and test statistics that are robust to non- normality using the MLR estimator

直接ML假设数据是MCAR或MAR和多变量正态的。然而，当数据为非正态时，直接ML可以通过MLR估计器来提供标准误差和测试统计数据，这些数据对非正态性是稳健的。

Multiple Imputation

multiple imputation is a useful approach to missing data when the researcher does not have access to a program capable of direct ML or wishes to estimate a CFA/SEM model with a fitting function other than ML.

当研究者无法使用能够直接进行ML的程序，或者希望用ML以外的拟合函数估计CFA/SEM模型时，多重归因是处理缺失数据的一种有用方法。

simple imputation procedures such as mean or regression imputation are problematic because they produce underesti- mates of variances and overestimates of correlations among the variables with imputed data. For instance, if regression imputation was used to sup- ply values on variable Y from data that are available on variable X (i.e., $Y^6 = a + bX$), the correlation between X and Y would be overestimated (i.e., for cases with imputed Y values, X is perfectly correlated with Y).

简单的归入程序，如平均数或回归归入，是有问题的，因为它们会产生方差的低估和被归入数据的变量之间的关联性的高估。例如，如果用回归归因法从变量X的数据中归纳出变量Y的值（即Y6=a+bX），那么X和Y之间的相关性就会被高估（即对于有归因Y值的案例，X与Y完全相关）。

Multiple imputation reconciles this problem by introducing random variation into the process. In other words, missing values for each case are imputed on the basis of observed values (as in regression imputation), but random noise is incorporated to preserve the proper degree of variability in the imputed data.

多重归因法通过在这个过程中引入随机变化来调和这个问题。换句话说，每个案例的缺失值都是在观察值的基础上进行推算的（如回归推算），但随机噪声被纳入其中以保持推算数据的适当变异程度。

$$\text{Y\_hat} = a + bX + S_{x,y}E$$

where $S_{x,y}$ is the estimated standard deviation of the regression's error term (root mean squared error), and E is a random draw (with replacement) from a standard normal distribution.

其中，Sx,y是回归误差项的估计标准差（均方根误差），E是从标准正态分布中随机抽取的（带替换）。

The first step is to impute multiple data sets

第一步是对多个数据集进行估算

In the second step, the *M* data sets are analyzed using standard analytic procedures. In the third step, the results from the *M* analyses are combined into a single set of parameter estimates, standard errors, and test statistics. Parameter estimates are combined by simply averaging the estimates across the *M* analyses. Standard errors are combined using the average of the standard errors over the set of analyses and the between-analysis parameter estimate variation

在第二步，使用标准的分析程序对M数据集进行分析。第三步，将M分析的结果合并为一组参数估计值、标准误差和检验统计。参数估计值是通过简单地对M分析中的估计值进行平均来合并的。标准误差是用一组分析的标准误差的平均值和分析间的参数估计变异来组合的。

problem with pairwise deletion is determining the appropriate sample size to specify in the analy- sis. For this illustration, two *N*s were used: (1) the number of nonmissing cases for the variable with the most missing data (N = 460); and (2) the size of the full sample (N = 650; cf. Allison, 2003). In this example, the results produced by the various missing data methods are similar. This can be attributed in large part to the size of the sample (elimination of 41% of the sample with at least one missing observation still results in an N = 385) and the fact that the data are MCAR. The biggest difference is seen in the standard errors produced in the analysis using listwise deletion.

成对删除的问题是确定在分析中指定适当的样本量。在这个例子中，我们使用了两个N。(1)缺失数据最多的变量的非缺失案例数(N=460)；(2)全样本的规模(N=650；参见Allison, 2003)。在这个例子中，各种缺失数据方法产生的结果是相似的。这在很大程度上可以归因于样本的规模（剔除41%的样本中至少有一个缺失的观察值，仍然可以得到N=385）以及数据是MCAR这一事实。最大的差异体现在使用列表式删除的分析中产生的标准误差。

similarity of the parameter estimates and standard errors produced by direct ML and multiple imputation.

通过直接ML和多重归因产生的参数估计和标准误差的相似性。

In summary, direct ML and multiple imputation are strong methodol- ogies for handling missing data when the data are either MCAR or MAR. If missing data are nonignorable

总之，当数据为MCAR或MAR时，直接ML和多重归因是处理缺失数据的有力方法。如果缺失的数据是不可忽视的

CFA WITH NON-NORMAL OR CATEGORICAL DATA

However, an alternative to ML for normal, continuous data is generalized least squares (GLS). GLS is a computationally simpler fitting function and produces approximately the same goodness of fit as ML (i.e., $F_{ML} = F_{GLS}$), especially when sample size is large. Nevertheless, ML (and GLS) are appropriate only for multivariate normal, interval-type data (i.e., the joint distribution of the continuous variables is distributed normally). When continuous data depart markedly from normality (i.e., marked skewness or kurtosis), or when some of the indicators are not interval level (i.e., binary, polytomous, ordinal), an esti- mator other than ML should be used.

然而，对于正常的、连续的数据，ML的一个替代方法是广义最小二乘法（GLS）。GLS是一个计算上比较简单的拟合函数，产生的拟合度与ML大致相同（即FML=FGLS），特别是当样本量很大时。然而，ML（和GLS）只适合于多变量正态、区间型数据（即连续变量的联合分布是正态分布）。当连续数据明显偏离正态性（即明显的偏度或峰度），或某些指标不是区间水平（即二元、多元、序数）时，应使用ML以外的测算方法。

Non-Normal, Continuous Data

Research has shown that ML (and GLS) is robust to minor departures in nor- mality (e.g., Chou & Bentler, 1995). However, when non-normality is more pronounced, an estimator other than ML should be used to obtain reliable statistical results

研究表明，ML（和GLS）对非正态性的轻微偏离是稳健的（例如，Chou & Bentler，1995）。然而，当非正态性比较明显的时候，应该使用ML以外的估计方法来获得可靠的统计结果。

ML is particularly sensitive to excessive kurtosis. The consequences of using ML under conditions of severe non-normality include (1) spuriously inflated model χ^2 values (i.e., overrejection of solu- tions); (2) modest underestimation of fit indices such as the TLI and CFI; and (3) moderate to severe underestimation of the standard errors of the parameter estimates

ML对过度的峰度特别敏感。在严重的非正态性条件下使用ML的后果包括：（1）虚假地夸大模型χ2值（即过度拒绝求解）；（2）适度低估适合指数，如TLI和CFI；以及（3）中度至严重低估参数估计的标准误差。

These deleterious effects are exacerbated as sample size decreases

这些有害的影响会随着样本量的减少而加剧

The two most commonly used estimators for non-normal continuous data are (1) robust ML (Bentler, 1995; Satorra & Bentler, 1994); and (2) weighted least squares

对非正态连续数据最常用的两种估计方法是：（1）稳健ML（Bentler, 1995; Satorra & Bentler, 1994）；（2）加权最小二乘法

The robust ML estimator (hereafter abbreviated, MLM) provides ML parameter estimates with standard errors and a mean-adjusted χ^2 test sta- tistic that are robust to non-normality. The mean-adjusted χ^2 test statistic is often referred to as the Satorra–Bentler scaled χ^2

稳健的ML估计器（以下简称MLM）提供了带有标准误差的ML参数估计值以及对非正态性具有稳健性的均值调整χ2检验统计量。均值调整后的χ2检验统计量通常被称为Satorra-Bentler尺度的χ2

As shown in this table, EQS and PRELIS produce very similar results. Some of the indicators evidence considerable non-normality (e.g., kurtosis of X5 = 9.4), and thus the assumption of multivariate normality does not hold (Table 9.6)

如该表所示，EQS和PRELIS的结果非常相似。一些指标显示出相当大的非正态性（例如，X5的峰度=9.4），因此，多变量正态性假设不成立（表9.6）

Specifically, the ML χ^2 (87.48) is considerably larger than the SB χ^2 (33.13), reflecting the tendency for ML to produce inflated χ^2 values when data are non-normal. In addition, the standard errors of the ML estimates are noticeably smaller than those based on MLM (e.g., .027 vs. .051 for the V2 indicator), illustrating the propensity for ML to under- estimate standard errors in this context. The underestimation of standard errors results in inflated test statistics (e.g., *z*s for V2 = 22.60 and 12.15 for ML and MLM, respectively), thus increasing the risk of Type I error. Note that the parameter estimates are not affected (e.g., $\lambda_{21} = .618$) by the type of estimator used (e.g., $\lambda_{21} = .618$ in both ML and MLM)

具体来说，ML的χ2（87.48）比SB的χ2（33.13）大得多，反映了当数据不正常时，ML有产生夸大χ2值的倾向。此外，ML估计值的标准误差明显小于基于MLM的估计值（例如，V2指标的标准误差为0.027比0.051），说明在这种情况下，ML有低估标准误差的倾向。对标准误差的低估导致了测试统计量的膨胀（例如，ML和MLM的V2的z=22.60和12.15），从而增加了I型错误的风险。请注意，参数估计值不受所使用的估计器类型的影响（例如，λ21=0.618）（在ML和MLM中λ21都=0.618）

Chi-square difference testing can be conducted using the SB χ^2 statis- tic. However, unlike ML-based analysis, this test cannot be conducted by simply calculating the difference in χ^2 values produced by the nested and comparison models

可以使用*SB χ2*统计量进行*Chi-square*差异检验。然而，与基于*ML*的分析不同，这种检验不能通过简单计算嵌套模型和比较模型产生的*χ2*值的差异来进行。

The reason is that a difference between two SB χ^2 values for nested models is not distributed as χ^2. Thus, a *scaled difference in* ^2s (SDCS) test should be used, SDCS test statistic, T_S,

原因是嵌套模型的两个 *SB χ2* 值之间的差值不是以 *χ2* 分布的。因此，应该使用*2s*内的比例差（*SDCS*）检验，*SDCS*检验统计量，*TS*。

$$T_S = (T_0 - T_1)/c_d$$

T_0 is the regular ML χ^2 for the nested model, T_1 is the regular ML χ^2 for the comparison (less restricted) model, and c_d is the difference test scaling correction. c_d is defined as

T0 是嵌套模型的常规 *ML χ2*，*T1* 是比较（限制较少）模型的常规 *ML χ2*，*cd* 是差异检验的比例校正。 *cd* 定义为

$$c_d = [(d_0 * c_0) - (d_1 * c_1)]/(d_0 - d_1)$$

d_0 is the degrees of freedom of the nested model, d_1 is the degrees of freedom of the comparison model, c_0 is the scaling correction factor for the nested model, and c_1 is the scaling correction factor for the comparison model.

*D0*是嵌套模型的自由度，*d1*是比较模型的自由度，*c0*是嵌套模型的比例校正系数，*c1*是比较模型的比例校正系数。

scaling correction factors can be readily computed by dividing the regular ML χ^2 by the SB χ^2: where T_0* is the SB χ^2 value.

缩放校正因子可以通过用常规的*ML χ2*除以*SB χ2*而轻易计算出来：其中*T0**为*SB χ2* 值。

$$c_0 = T_0 / T_0^*$$

First, the χ^2s must be obtained from ML and MLM for the nested and comparison models model with a correlated error with a single degree of freedom; $d_0 - d_1 = 5 - 4 = 1$). These four χ^2 values are then used to calculate the scaling correc- tion factors for the nested and comparison models; for example, $c_0 = T_0 / T_0^* = 87.478 / 33.128 = 2.641$

首先，必须从ML和MLM中获得嵌套模型和比较模型的χ2s，该模型具有单自由度的相关误差；d0-d1=5-4=1）。然后用这四个χ2值来计算嵌套模型和比较模型的比例相关系数；例如，c0 = T0 / T0* = 87.478 / 33.128 = 2.641

In the second step, the dif- ference test scaling correction (c_d) is computed using the scaling correc- tion factors and degrees of freedom from the nested and comparison models

第二步，利用嵌套模型和比较模型的比例校正因子和自由度，计算出差异检验的比例校正（cd）。

c_d in this example equals 3.013. In the third and final step, T_S is obtained by dividing the difference between the ML χ^2 val- ues of the nested and comparison models by c_d; that is, $T_S = (87.478 - 25.833) / 3.013 = 20.46$. T_S is interpreted in the same fashion as the regular χ^2 difference test. Because the T_S value is statistically significant ($df = 1, p < .001$), it can be concluded that the revised one-factor model provides a significantly better fit to the data than the original one-factor solution.

本例中cd等于3.013。在第三步，也是最后一步，TS是由嵌套模型和比较模型的ML χ2值之差除以cd得到的；也就是说，TS = (87.478 - 25.833) / 3.013 = 20.46。TS的解释与常规χ2差异检验相同。因为TS值具有统计学意义（df=1，p<0.001），所以可以得出结论，修订后的单因素模型对数据的拟合效果明显好于原来的单因素解决方案。

there are many situations where using the standard χ^2 difference test to compare nested models estimated by MLM will yield misleading results. Thus, the SCDS test should always be employed when comparing nested solutions estimated by MLM.

在很多情况下，使用标准χ2差异检验来比较由MLM估计的嵌套模型会产生误导性结果。因此，在比较由MLM估计的嵌套方案时，应始终采用SCDS检验。

Categorical Data

When at least one factor indicator is categorical (i.e., dichotomous, polytomous, ordinal), ordinary ML should not be used to estimate CFA models

当至少有一个因子指标是分类的（即二分法、多分法、顺序法），普通ML不应该被用来估计CFA模型

potential consequences of treating categorical variables as continuous variables in CFA are multifold, including that it can (1) pro- duce attenuated estimates of the relationships (correlations) among indi- cators, especially when there are floor or ceiling effects; (2) lead to “pseudofactors” that are artifacts of item difficulty or extremeness; and (3) produce incorrect test statistics and standard errors. ML can also produce incorrect parameter estimates,

在CFA中把分类变量当作连续变量的潜在后果是多方面的，包括：(1)产生指标间关系（相关）的减弱估计，特别是当存在底限或上限效应时；(2)导致 "伪因素"，是项目难度或极端性的假象；(3)产生不正确的测试统计和标准误差。ML也可以产生不正确的参数估计。

Thus, it is important that an estimator other than ML be used with categorical out- comes or severely non-normal data.

因此，对于分类数据或严重的非正态数据，使用ML以外的估计器是很重要的。

used with categorical indicators; for example, weighted least squares (WLS), robust weighted least squares (WLSMV), and unweighted least squares (ULS)

与分类指标一起使用；例如，加权最小二乘法（WLS）、稳健加权最小二乘法（WLSMV）和非加权最小二乘法（ULS）。

WLS is available in all of the major latent variable software programs (in Amos, WLS is referred to as asymp- totically distribution-free, ADF; in EQS, WLS is referred to as arbitrary generalized least squares, AGLS). WLS is closely related to the GLS estima- tor. Like ML, GLS minimizes the discrepancy between the observed ($\boldsymbol{S}$) and predicted (Σ) covariance matrices

WLS在所有主要的潜变量软件中都可以使用（在Amos中，WLS被称为ymp-totically distribution-free，ADF；在EQS中，WLS被称为arbitrary generalized least squares，AGLS）。WLS与GLS估计法密切相关。像ML一样，GLS最小化了观察（S）和预测（Σ）协方差矩阵之间的差异

WLS uses a different W; specifically, one that is based on estimates of the variances and covariances of each element of $\boldsymbol{S}$, and fourth-order moments based on multivariate kurtosis

WLS使用不同的W；具体来说，是基于对S中每个元素的方差和协方差的估计，以及基于多变量峰度的四阶矩。

WLS fit function is weighted by variances/ covariances and kurtosis to adjust for violations in multivariate normality; that is, if there is no kurtosis, WLS and GLS will produce the same minimum fit function value, $F_{\mathrm{WLS}} = F_{\mathrm{GLS}}$.

WLS的拟合函数由方差/协方差和峰度加权，以调整违反多变量正态性的情况；也就是说，如果没有峰度，WLS和GLS将产生相同的最小拟合函数值，FWLS=FGLS。

Consider a three-factor CFA model in which each latent factor is defined by 6 indicators ($p = 18$). Thus, there are 171 elements of $\boldsymbol{S}$; that is, $b = 18(19) / 2$ (see Eq. 3.14, Chapter 3). In this example, W is of the order $b \times b$ (171 × 171) and has 14,706 distinct elements; $b(b + 1) / 2 = 171(172) / 2 = 14{,}706$.

考虑一个三因素的CFA模型，其中每个潜在因素由6个指标（p=18）定义。因此，S有171个元素；也就是说，b=18(19)/2（见第3章公式3.14）。在这个例子中，W的阶数是b×b（171×171），有14,706个不同的元素；b（b+1）/2=171（172）/2=14,706。

In addition, WLS requires that sample size exceeds $b + p$ (number of elements of $\boldsymbol{S}$ plus number of indicators) to ensure that W is nonsingular

此外，WLS要求样本量超过b+p（S的元素数加上指标数），以确保W是非正交的。

Unless the sample size is quite large, very skewed items can make W not invertible; W will frequently be nonpositive definite in small to moderate samples with variables that evidence floor or ceiling effects. WLS behaves very poorly in small or moderately sized samples

除非样本量相当大，否则非常倾斜的项目会使W不能倒置；在小到中等的样本中，W经常是非正定的，而这些变量有底线或上限效应。WLS在小规模或中等规模的样本中表现得非常差。

WLS estimator with categorical outcomes is not favorable (e.g., over- sensitivity of χ^2 and considerable negative bias in standard errors as model complexity increases; Muthén & Kaplan, 1992). Thus, as with non-normal continuous data, WLS is not a good estimator choice with categorical out- comes

对于分类结果的WLS估计器是不利的（例如，随着模型复杂性的增加，χ2的过度敏感和标准误差的相当大的负偏差；Muthén & Kaplan, 1992）。因此，与非正态连续数据一样，WLS在分类结果中不是一个好的估计器选择。

Unlike WLS, WLSMV does not require W to be positive definite, because W is not inverted as part of the estimation procedure. In WLSMV, the number of elements in the diagonal W equals the number of sample correlations in $\boldsymbol{S}$, but this matrix is not inverted during estimation. Nevertheless, WLSMV estimation is fostered by N being larger than the number of rows in W. In the computation of the χ^2 test statistic and standard errors, the full W is used but not inverted.

与WLS不同，WLSMV不要求W是正定的，因为在估计过程中W没有被倒置。在WLSMV中，对角线W中的元素数等于S中样本相关性的数量，但是这个矩阵在估计过程中并没有被倒置。然而，WLSMV估计因N大于W中的行数而得到促进。在计算χ2检验统计量和标准误差时，使用完整的W，但不倒置。

Therefore, the measurement errors (θ) of the CFA model with categorical indicators are not free param- eters but instead reflect the remainder of 1 minus the product of the squared factor loading and factor variance; that is,

因此，带有分类指标的CFA模型的测量误差（θ）不是自由参数，而是反映了1的余数减去平方因子载荷和因子方差的乘积；也就是说。

$$\theta = 1 - \lambda^2\phi$$

Because these coefficients are based on the latent variable underlying the binary indicators, they differ in value from phi correlations, which are based on observed measures. The results indicate that the one-factor model fits the data well, χ^2 (9) = 9.54, p = .39, RMSEA = 0.009, TLI = 0.999, CFI =.999.

由于这些系数是基于二元指标的潜在变量，它们在价值上与基于观察到的措施的phi相关度不同。结果表明，单因素模型对数据的拟合很好，χ2（9）=9.54，P=0.39，RMSEA=0.009，TLI=0.999，CFI=0.999。

In the CFA, the y^* variances are standardized to 1.0, and thus parameter estimates should be interpreted accordingly. Squaring the completely standardized factor loadings yields the proportion of variance in y^* that is explained by the latent factor (e.g., Y1 = $.775^2$ = .601), not the proportion of variance explained in the observed measure (e.g., Y1), as in the interpretation of CFA with con- tinuous indicators. residual variances convey the proportion of y^* variance that is not accounted for by the latent factor; for example, for Y1: 1 – $.775^2$ = .399.

在CFA中，y*方差被标准化为1.0，因此参数估计值应作相应解释。将完全标准化的因子载荷平方化，可以得到潜伏因子所解释的y*方差的比例（例如，Y1 = .7752 = .601），而不是观察指标（例如，Y1）所解释的方差的比例，正如在解释连续指标的CFA时那样。 残差传达了潜伏因子所不占的y*方差的比例；例如，对于Y1：1 - .7752 = .399。

As with SB χ^2, the difference in χ^2 values for nested models estimated with WLSMV is not distributed as χ^2.

与SB的χ2一样，用WLSMV估计的嵌套模型的χ2值之差不以χ2分布。

that is, χ^2_{diff} (4) = 27.96, p < .001 indicates that the restriction of equal factor loadings significantly degrades the fit of the model.

也就是说，χ2diff (4) = 27.96, P < .001表明，因子载荷相等的限制大大降低了模型的拟合。

Comparison with Item Response Theory (IRT) Models

IRT, which has also been referred to as *latent trait theory*, relates characteristics of items (item parameters) and characteristics of individu- als (latent traits) to the probability of endorsing a particular response category

IRT也被称为潜在特质理论，它将项目的特征（项目参数）和个人的特征（潜在特质）与支持某一特定反应类别的概率联系起来。

In other words, an IRT model specifies how both the level of the latent trait and the item prop- erties are related to a person's item responses

换句话说，IRT模型规定了潜在特质的水平和项目参数是如何与一个人的项目反应相关的。

The probability of answering cor- rectly or endorsing a particular response category is graphically depicted by an *item response function*

回答正确或认可某一特定反应类别的概率用项目反应函数来描述。

IRFs reflect the nonlinear (logit) regression of a response probability on the latent trait.

IRF反映了反应概率对潜在特质的非线性（logit）回归。

For instance, IRT can be used to explore the latent dimensionality of categorical outcomes, to evaluate the psychometric properties of a test, and to conduct differ- ential item functioning analysis

例如，IRT可以用来探索分类结果的潜在维度，评估测试的心理测量特性，并进行不同项目功能分析。

In addition to the latent trait level (denoted θ in the IRT literature), either one, two, or three item parameters can be estimated in an IRT model. The choice of IRT model should be based on substantive and empirical considerations (e.g., model fit, although IRT currently pro- vides limited information in regard to goodness of model fit).

除了潜在的特质水平（在IRT文献中表示为θ），在IRT模型中可以估计一个、两个或三个项目参数。IRT模型的选择应该基于实质性的和经验性的考虑（例如，模型拟合度，尽管IRT目前提供的关于模型拟合度的信息很有限）。

The simplest model is the *one- parameter logistic model* (1PL),

最简单的模型是单参数逻辑模型（1PL）。

对一个项目作出积极反应的概率

probability of responding positively on an item is predicted by the latent trait (θ) and a single item parameter, *item difficulty*

对某一项目作出积极反应的概率是由潜在特征（θ）和单一项目参数（项目难度）预测的。

logistic function

$$P(y_{is}=1|\theta_s,b_i)=\exp(\theta_s-b_i)/[1+\exp(\theta_s-b_i)]$$

An item difficulty conveys the level of the latent trait (θ) where there is a 50% chance of a positive response on the item; for example, if b = .75, there is a .50 probability that a person with a trait level of .75 will respond positively to the item

一个项目的难度传达了潜在特质（θ）的水平，在这个水平上有50%的机会对项目作出积极反应；例如，如果b=0.75，那么特质水平为0.75的人对项目作出积极反应的概率是0.50。

rel- atively “easier” items have lower b values and are represented by curves closer to the horizontal axis. Accordingly, b is inversely related to a proportion-correct score

相关的 "更容易 "的项目有较低的b值，并以更接近横轴的曲线表示。因此，b与正确率分数成反比。

In a *two-parameter logistic model* (2PL), an *item discrimination* para- meter is included

在双参数逻辑模型(2PL)中，包括了一个项目判别参数

probability of a positive response is predicted by the logistic function

阳性反应的概率是由逻辑函数预测的

$$P(y_{is}=1\mid\theta_s,b_i,a_i)=\frac{\exp[a_i(\theta-b_i)]}{1+\exp[a_i(\theta_s-b_i)]}$$

Item discrimination parameters are analogous to factor loadings in CFA and EFA because they represent the relationship between the latent trait and the item responses

项目鉴别参数类似于CFA和EFA中的因子载荷，因为它们代表了潜在特质和项目反应之间的关系。

Thus, discrimination parameters influence the steepness of the *slope* of the IRF curves. items with rela- tively high a parameter values are more strongly related to the latent vari- able (θ) and have steeper IRF curves

因此，鉴别参数会影响IRF曲线的陡峭程度。参数值相对较高的项目与潜在变量（θ）的关系更强，IRF曲线更陡峭。

item discrimination (*a*) is a multiplier of the difference between trait level (θ) and item difficulty (*b*). This reflects the fact that the impact of the dif- ference between θ and *b* on the probability of a positive response (*P*) depends on the discriminating power of the item

项目辨别力（a）是特质水平（θ）和项目难度（b）之间差异的乘数。这反映了这样一个事实：θ和b之间的差异对阳性反应的概率（P）的影响取决于项目的鉴别力。

A *three-parameter logistic model* (3PL) can also be estimated in IRT, which includes a “guessing” parameter (denoted either as *c* or γ).

三参数逻辑模型（3PL）也可以在IRT中估计，它包括一个 "猜测 "参数（用c或γ表示）。

used to repre- sent IRF curves that do not fall to zero on the vertical axis (i.e., > .00 prob- ability of a positive response for persons with very low θ levels). In other words, if an item can be correctly answered by guessing (as in true/false or multiple-choice items on an aptitude test), the probability of a positive response is greater than zero even for persons with low levels of the latent trait characteristic

用来代表纵轴上不落到零的IRF曲线（即对于θ水平很低的人来说，积极反应的概率大于0.00）。换句话说，如果一个项目可以通过猜测来正确回答（如能力测试中的真/假或多选项目），即使对于潜质特征水平较低的人来说，积极反应的概率也大于零。

The second portion of the selected output presents the item calibra- tions (i.e., IRT parameter estimates). The item discrimination (*a*) and item difficulty (*b*) parameter estimates are provided under the “Slope” and “Threshold” columns, respectively. All asymptote parameters (*c*, “guess- ing”) equal zero because a 2PL model was specified. As in CFA, the item loadings are interpreted as the correlations between the items and the latent traits (θ). Loadings can be calculated by the equation

所选输出的第二部分显示了项目口径（即IRT参数估计）。项目区分度（a）和项目难度（b）的参数估计值分别在 "斜率 "和 "阈值 "栏中提供。所有的渐近参数（c，"猜测 "）都等于零，因为我们指定了一个2PL模型。如同在CFA中一样，项目载荷被解释为项目和潜在特征（θ）之间的相关关系。负荷率可以通过以下公式计算

$$a / \mathrm{SQRT}(1 + a^2)$$

loading for Y1 = 1.218 / SQRT(1 + 1.218^2) = .773. Item intercepts can be computed by the equation

Y1的载荷=1.218/SQRT（1+1.2182）=.773。项目截距可以通过以下公式计算：

$$-ab$$

-ab

Item thresholds (*b*) always have the opposite sign of item intercepts.

项目阈值（b）的符号总是与项目截距相反。

it appears that items Y1, Y4, and Y5 have the highest discrimina- tion; that is, are more strongly related to θ, such that the probability of a positive response changes most rapidly with a change in θ. The Y3 indica- tor has the lowest difficulty (b_3 = -2.309); that is, lower levels of the latent dimension of Alcohol Dependence (θ) are required for a positive response on the Y3 criterion

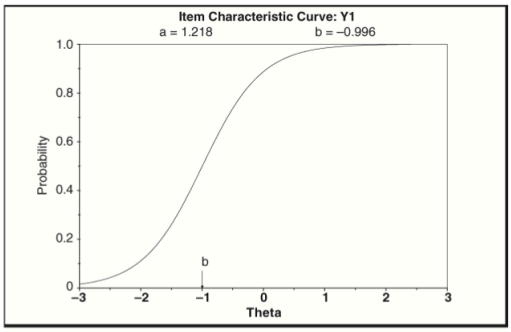
看来，项目Y1、Y4和Y5的辨别力最高；也就是说，与θ的关系更密切，因此，随着θ的变化，正面回答的概率变化最快。Y3指标的难度最低（b3=-2.309）；也就是说，Y3标准的正面回答需要较低水平的酒精依赖（θ）。

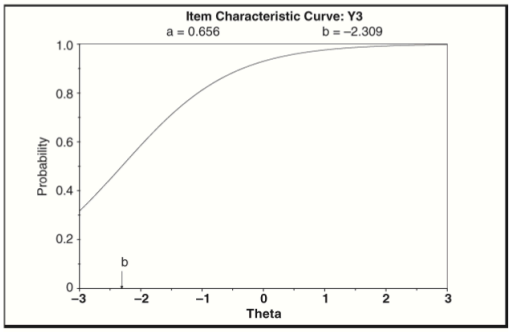
larger *b* values would be likely if a community sample rather than an outpatient sample was used).

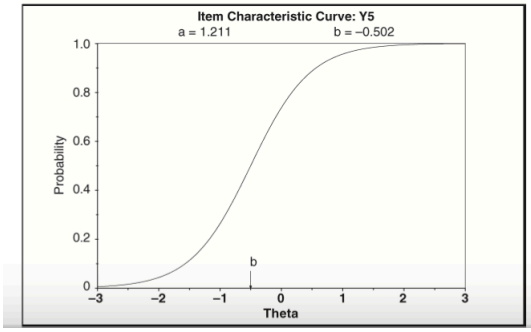
如果使用的是社区样本而不是门诊样本，可能会有更大的b值）。

For example, the steepest section of each is the middle of the curve, where small changes in θ are associated with the greatest increase in probability of a positive response

例如，每条曲线中最陡峭的部分是曲线的中部，θ的微小变化与阳性反应概率的最大增加有关。







each curve indicates that once a certain trait level is reached (i.e., $\theta \approx 1.5$), increases in the trait are not associated with appreciable change in item endorsement.

每条曲线都表明，一旦达到一定的特质水平（即θ≈1.5），特质的增加与项目认可的明显变化没有关系。

For instance, although items Y1 and Y5 have similar discrimination parameters (a = 1.218 and 1.211, respectively), Y5 is a "less difficult" item (b = –0.996 and –0.502, respectively)

例如，尽管项目Y1和Y5有相似的辨别参数（a=1.218和1.211，分别），但Y5是一个 "不太难 "的项目（b=-0.996和-0.502，分别）。

Thus, although items Y1 and Y5 have simi- lar a parameters, the shape of their curves differs somewhat because lower levels of θ are needed for a positive response to item Y5

因此，尽管项目Y1和Y5有相似的a参数，但它们的曲线形状却有些不同，因为对项目Y5作出积极反应需要较低的θ水平。

Item Y3 has the relatively weakest relationship with θ (a = .656); hence, the probability of a positive response on Y3 changes most slowly with a change in θ. Item 3 also has the lowest threshold (b = –2.309).

项目Y3与θ的关系相对最弱（a=0.656）；因此，Y3的积极反应的概率随着θ的变化而变化最慢。

Using CFA parameterization (and symbols), an IRT discrimination parameter can be calculated as

使用CFA参数化（和符号），IRT区分度参数可以计算为

$$a = \lambda / \mathrm{SQRT}(\theta)$$

$$a = \lambda / \mathrm{SQRT}(\theta)$$

where λ is the factor loading, and θ is the residual variance

其中λ是因子载荷，θ是残差方差。

Table 9.10, the item discrimination of Y1 would be calculated as a_1 = .775 / SQRT(.399) = 1.23 (cf. a_1 = 1.22 in Table 9.12).

在表9.10中，Y1的项目区分度将被计算为a1 = .775 / SQRT(.399) = 1.23（参见表9.12中a1 = 1.22）。

IRT difficulty parameter can be directly calculated as

IRT难度参数可以直接计算为

$$b=\tau/\lambda$$

$$b=\tau/\lambda$$

where τ is the CFA item threshold, and λ is the CFA factor loading. Thus, the item difficulty parameter of item Y1 is computed $b_1 = –0.759 / .775 = –0.979$ (cf. $b_1 = –0.996$ in Table 9.12).

其中，τ是CFA的项目阈值，λ是CFA的因子载荷。因此，项目Y1的项目难度参数计算为b1=-0.759/.775=-0.979（参见表9.12中b1=-0.996）。

Standard errors of IRT parameter estimates can also be obtained from CFA using the delta method

IRT参数估计的标准误差也可以用delta方法从CFA中得到。

MIMIC framework offers sev- eral potential advantages over IRT. These include the ability to (1) use either continuous covariates (e.g., age) or categorical background vari- ables (e.g., gender); (2) model a direct effect of the covariate on the latent factor (in addition to direct effects of the covariate on test items); (3) readily evaluate multidimensional models (i.e., measurement models with > one latent factor); and (4) incorporate an error theory (e.g., measure- ment error covariances). Indeed, a general advantage of the covariance structure analysis approach is that the IRT model can be embedded in a larger structural equation model (e.g., Lu, Thomas, & Zumbo, 2005).

与IRT相比，MIMIC框架提供了七种潜在的优势。这些优势包括：（1）使用连续协变量（如年龄）或分类背景变量（如性别）；（2）建立协变量对潜在因子的直接影响模型（除了协变量对测试项目的直接影响）；（3）容易评估多维模型（即具有>一个潜在因子的测量模型）；以及（4）纳入误差理论（如测量误差协方差）。事实上，协方差结构分析方法的一个普遍优势是，IRT模型可以被嵌入到一个更大的结构方程模型中（例如，Lu, Thomas, & Zumbo, 2005）。

Other Potential Remedies for Indicator Non-Normality

指标非正态性的其他潜在补救措施

Three other remedial strategies for non-normality are briefly presented: bootstrapping, item parceling, and data transformation

下面简要介绍其他三种针对非正态性的补救策略：引导、项目分割和数据转换。

Bootstrapping

引导法

Multiple samples (with the same *N* as the original sample) are ran- domly drawn from the original sample *with replacement*

从原始样本中随机抽取多个样本（与原始样本的N相同），并进行替换。

the CFA model is estimated in each data set, and the results are aver- aged over the data sets. The number of bootstrapped samples can be speci- fied by the researcher, but should be sufficiently large to foster the quality of the averaged estimates

在每个数据集上对CFA模型进行估计，并对数据集的结果进行平均化。研究者可以指定自举样本的数量，但应足够大以提高平均估计的质量。

The procedure is most appropri- ate for models with non-normal, continuous indicators

该程序最适合于具有非正态、连续指标的模型。

In Monte Carlo simulation, multiple samples (e.g., > 500) are randomly generated on the basis of population parameter values and other data aspects (e.g., sample size, amount of non-normality)

在蒙特卡洛模拟中，根据群体参数值和其他数据方面（如样本大小、非正态量），随机产生多个样本（如>500）。

As in bootstrapping, the results of models fitted in the simulated data sets are averaged to examine the behavior of the estimates (e.g., stability and precision of parameter estimates and test statistics).

与引导法一样，在模拟数据集中拟合的模型结果被平均化，以检查估计值的行为（例如，参数估计值和测试统计的稳定性和精确度）。

Bootstrapping is based on the notion that when the distributional assumptions of normal-theory statistics are violated, an *empirical sampling distribution* can be relied upon to describe the actual distribution of the population on which the parameter estimates are based

引导是基于这样的概念：当正态理论统计的分布假设被违反时，可以依靠经验抽样分布来描述参数估计所基于的人口的实际分布。

For the sake of brevity, only unstandardized factor load- ings are presented. These results are identical to the EQS output in Table 9.7. The next section of the output provides results from the bootstrapped samples. The first column, “SE,” is the bootstrap estimate of the standard errors of the factor loadings

为了简洁起见，只介绍了未标准化的因子负荷。这些结果与表9.7中的EQS输出相同。输出的下一节提供了自举样本的结果。第一列，"SE"，是对因子载荷的标准误差的引导估计

the bootstrapped stan- dard errors are considerably larger than the maximum likelihood esti- mates; for example, X2: the bootstrapped standard error of .0606 is 74% larger than the ML estimate of .0349

自举标准误差比最大似然估计值大得多；例如，X2：自举标准误差为0.0606，比ML估计值0.0349大74%。

bootstrapped standard errors are very similar in magnitude to the standard errors produced by MLM

引导的标准误差与MLM产生的标准误差的大小非常相似。

Indeed, the primary objective of bootstrapping is often to obtain better standard errors for the purpose of significance testing, calculation of confidence intervals, and so forth. The second column, “SE-SE,” presents standard errors of the bootstrapped standard error estimates. These should be low in magnitude, given the number of bootstrapped samples and original sample size. Values in the “Mean” column are the average unstandardized factor loadings across the 500 bootstrap samples

事实上，自举的主要目的通常是为了获得更好的标准误差，以便进行意义检验、计算置信区间等等。第二栏，"SE-SE"，显示了自举标准误差估计值的标准误差。考虑到被引导样本的数量和原始样本的大小，这些标准误差的大小应该是很低的。平均值 "一栏中的数值是500个引导样本中未标准化的因子载荷的平均值。

Values in the “Bias” column represent the difference between the original esti- mates and the averaged bootstrapped estimates (e.g., X2 = .7076 – .7027 = .0049). The last column, “SE-Bias” presents the standard errors of these bias estimates. The final section of the output lists the bias-corrected 90% confidence intervals of the unstandardized factor loadings; that is, confi- dence intervals of the original sample parameter estimates using standard errors that have been adjusted on the basis of bootstrapped results.

偏差 "一栏中的数值是原始估计值和平均引导估计值之间的差异（例如，X2=0.7076-0.7027=0.0049）。最后一栏，"SE-Bias "表示这些偏差估计值的标准误差。输出的最后一栏列出了未标准化因子载荷的偏差校正后的90%置信区间；也就是说，原始样本参数估计值的置信区间使用的是在自举结果基础上调整过的标准误差。

the bootstrap distribution will follow a noncentral χ^2 distribution rather than a central χ^2 distribution in accord with statistical theory

根据统计理论，自举分布将遵循非中心χ2分布，而不是中心χ2分布。

Item Parceling

Another remedial approach that has been used to address non-normality is *item parceling*. A parcel (also referred to as a "testlet") is a sum or average of several items that presumably measure the same construct.

The pri- mary potential advantages of using parcels are that (1) parcels may be more apt to approximate normality than individual items (thereby, the assumptions of ML are more likely to be met); (2) it offers improved reli- ability and relationships with other variables (3) models based on parcels may be considerably less complex

Data Transformation

A final potential remedial strategy for non-normality is to transform raw scores of a variable so they more closely approximate a normal distribution.

First, transformation is not always successful at reducing the skewness or kurtosis of a variable.Transformed variables must be reas- sessed in order to verify the success of the transformation in approximat- ing normality at the uni- and multivariate levels. Second, in addition to altering the distribution of variables, nonlinear transformation often changes the relationships a variable has with other variables in the analysis. Thus, the resulting fit statistics, parameter estimates, and standard errors of a CFA based on transformed indicators may differ markedly from an analysis based on the original variables This may be problematic for at least two reasons: (1) it "strains" reality if the true popu- lation distribution is not normally distributed; and (2) it makes the interpretability of the parameter estimates more complex.

项目打包

另一种用于解决非正态性的补救方法是项目包裹。一个包裹（也被称为 "测试单元"）是几个项目的总和或平均数，这些项目可能是测量同一结构的。

使用小包的主要潜在优势是：（1）小包可能比单个项目更容易接近正态性（因此，更有可能满足ML的假设）；（2）它提供了更好的可靠性和与其他变量的关系（3）基于小包的模型可能相当不复杂。

数据转换

对非正态性的最后一个潜在补救策略是转换变量的原始分数，使其更接近于正态分布。

首先，转换并不总是成功地减少变量的偏度或峰度。必须对转换后的变量进行重新评估，以验证转换在单变量和多变量水平上是否成功接近正态。第二，除了改变变量的分布外，非线性转换还经常改变一个变量与分析中其他变量的关系。因此，基于转化指标的CFA所产生的拟合统计、参数估计和标准误差可能与基于原始变量的分析有明显的不同，这至少有两个原因。(1) 如果真实的人口分布不是正态分布，它就会 "扭曲 "现实；(2) 它使参数估计值的可解释性更加复杂。