From: **Dean Gladish** gladish.dean@gmail.com
Subject: Re: t-test for women and men subsets, promis_first_assessment_data_v1.1
Date: May 19, 2020 at 3:39 PM
To: Christopher Hanes cshanes@umich.edu, cshanes87@gmail.com, Dean Gladish gladishd@carleton.edu

DG

Chris,

- Is making their scores transformed to have the same mean and SD as HUI-3 a valid process?

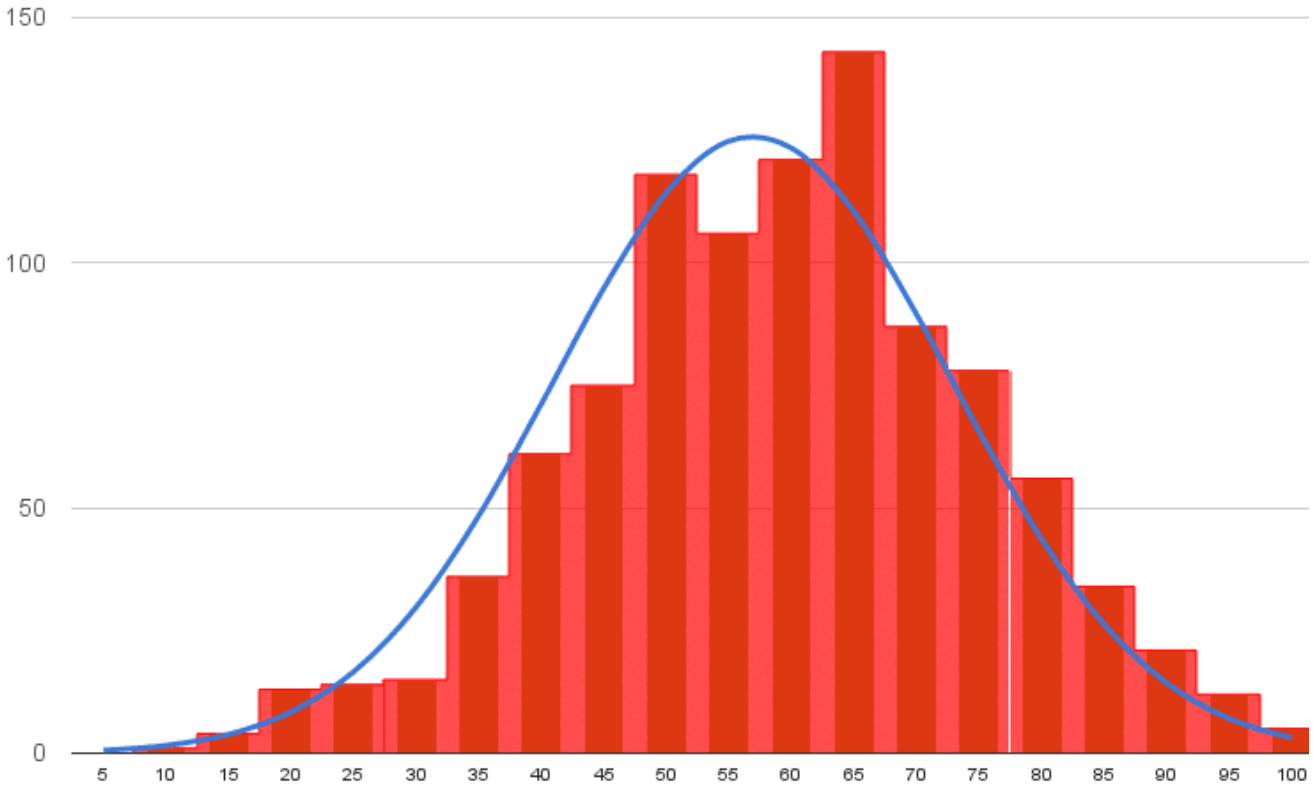
Yes. This is called linear equating. They borrowed from the article "Linking should replace regression when mapping from profile to preference measures":
To paraphrase, one is not interested in predicting the score a patient might have obtained on another examination. Rather, one wants to know what score is equivalent for the second exam, such that students of a particular ranking on one exam *are assigned the same ranking on the other exam*.

- Does this mean that PROMIS-29 is better at conveying outliers?

Yes, as long as we don't have real test scores from HUI-3 and we are using exclusively linear regression to map from PROMIS to HUI-3.

- What does this mean if we are showing physicians translated HUI-3 scores?

The validity of our translated HUI-3 scores depends on the validity of the underlying assumptions for our translation process. In the article about mapping studies, Fayers and Hays mention one simple approach known as equipercentile linking, which requires smoothing:



As an analogy, equipercentile linking can ensure that a student in the top 5% for PROMIS will also be in the top 5% when the PROMIS scores are converted to HUI-3 scores. Regression does not achieve this: the predicted HUI-3 scores for individuals assessed using PROMIS will be less extreme than the observed scores of similar individuals who were assessed using Y.
Suppose we have two tests, X and Y.

```
+---+-----+-------+
| X | Y   | Y=10X |
+---+-----+-------+
| 1 | 10  | 10    |
+---+-----+-------+
| 2 | 20  | 20    |
```

3	30	30
7	70	70
10	100	100

Here, we have no problem with regression to the mean because the correlation coefficient is 1 => 100(1-r) = 0 = % regression to the mean. However, consider this new data: the last column shows more realistic data for the hypothetical Y test and demonstrates how the observed values of 32 and 27 are both brought to the regression line value of 31.5, the process of which diminishes extreme values:

X	Y	Y_2=9.62X+2.64	Y_2
1	10	12.2	15
3	30	31.5	32
3	30	31.5	27
7	70	69.9	71
10	100	98.8	99

The predictions for rows 2 and 3 are closer to 30 than the observed. This shows how people get shoehorned onto the regression line; person 2 was dragged to a lower ranking, and person 3 was dragged to a higher ranking. Variance was removed; the very term "regression" is short for "regression to the mean". That is the theory behind linear regression: if Person 3 has a below average (for that group X) score on the second test, Y, then their score is likely to be partly the consequence of a below average underlying or "true" level and partly luck => their true score is likely to be closer to the mean.

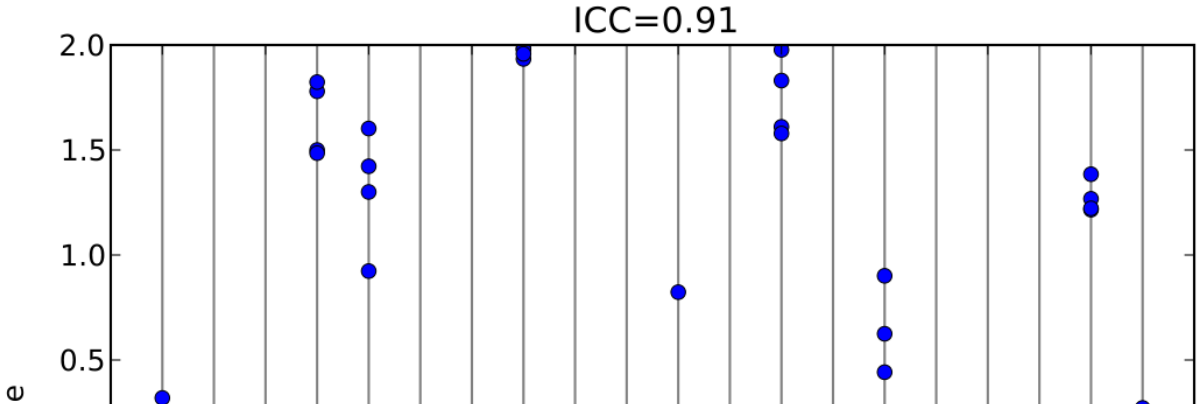
Furthermore. A 10 on test X should correspond to a 100 on test Y. However, the regression line does not reach 100. Direct equipercentile linking,

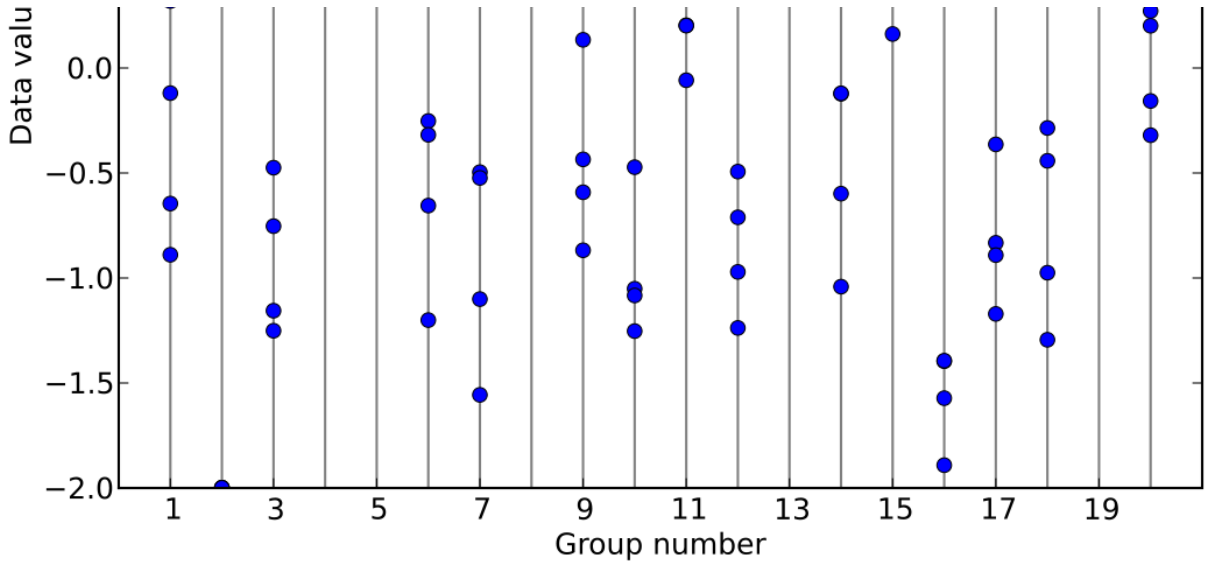
$$Y = \mu_Y + \frac{\sigma_Y}{\sigma_X}(X - \mu_X)$$

can certainly increase the validity of our mapping. In the original article, Hays et al. do both - linear regression and then linear equating.

- What does this mean: "their product-moment is equal to their intraclass correlation in that both values are at the relatively high value of 0.70, which indicates that between-group differences are just as important as across-group differences; that is why I think our demographic stuff (age and gender) is very valuable"?

So after standardizing, Hays et al. took one random half sample and estimated **product-moment** (or interclass) correlation, which measures the degree of scatter/the strength of correlation between predicted/observed HUI-3 preference scores, and **intraclass** correlation, which describes how strongly units in the same group resemble each other:





$$I.C.C. = \frac{\sigma^2_{weighted\ group\ means}}{\sigma^2_{weighted\ group\ means}\sigma^2_{all\ samples}}$$

Again, they measured these two types of correlation between the same categories - predicted and observed HUI-3 scores. Because intraclass correlation is designed to operate on data structured as groups, not as paired observations, it is an indication of consistency/reproducibility of quantitative measurements made by different observers measuring the same quantity; in our case, it shows that it doesn't matter who takes the test; once you decouple the PROMIS:HUI-3 pairings (anonymize them) we have disconnected separate patients from their scores. Because the intraclass correlation = interclass correlation = 0.70, their results are totally reproducible and have nothing to do with individual patient bias.

Now I was just thinking that we should do something which they *didn't do*, which is to calculate intraclass correlations to compare different demographic groups such as age and gender and see how similar values within those groups tend to be.

- Can you expand on the last part, how global pain is correlated with the error?

Global pain is a suppressor variable; this means that although it has a practically near-zero correlation with the dependent variables (HUI-3 preference scores), it paradoxically contributes to the predictive validity of the test. There's an interesting article about suppressor effects, which suppress irrelevant variance in the other predictor variables; they are able to do this because they do have a high correlation with other variables which do have correlation with the criterion/dependent variable; in essence, they provide more information about other variables, reducing the impact of variance/noise in those variables on the prediction. For these models, I think we should include predictive linear regression, yes, and also use linear equating to make it more of a direct translation.

Also, the phone call thing. It will be amazing. My phone number is the same as always, (224) 306-4486. Call me. I want to know. I can already see the formation of a great paper.

On Tue, May 19, 2020 at 3:00 PM Dean Gladish <gladish.dean@gmail.com> wrote:

> Chris,
> My previous response is outdated, I just revised (fixed the table in) my reply. I hope this clarifies things!
>
> - Is making their scores transformed to have the same mean and SD as HUI-3 a valid process?
>
> Yes. This is called linear equating. They borrowed from the article "Linking should replace regression when mapping from profile to preference measures":
> To paraphrase, one is not interested in predicting the score a patient might have obtained on another examination. Rather, one wants to know what score is equivalent for the second exam, such that students of a particular ranking on one exam *are assigned the same ranking on the other exam*.
>
> - Does this mean that PROMIS-29 is better at conveying outliers?
>
> Yes, as long as we don't have real test scores from HUI-3 and we are using exclusively linear regression to map from PROMIS to HUI-3.
>
> - What does this mean if we are showing physicians translated HUI-3 scores?
>
> The validity of our translated HUI-3 scores depends on the validity of the underlying assumptions for our translation process. In the article about mapping studies, Fayers and Hays mention one simple approach known as equipercentile linking, which requires smoothing: