

# Introduction to Structural Equation Modeling using lavaan

## Model Selection

R. M. Kuiper

Department of Methodology & Statistics  
Utrecht University

## Previously: Model Comparisons: AIC / BIC

Suitable for comparing

- both nested and non-nested models/hypotheses.
- more than two models/hypotheses.

Choose the model/hypothesis with the **lowest** AIC and/or BIC.

To be discussed in the Model Selection lecture

- Quantify relative support via IC weights.

```
library(devtools) # Make sure you have Rtools
install_github("rebeccakuiper/ICweights")
library(ICweights)
?IC.weights
```

- Extension AIC for order-restricted hypotheses (GORIC and GORICA).

```
library(restriktor)
library(goric)
```

# Outline of this lecture

IC

AIC in R

IC weights

GORICA in R

Complement

Extra

The end

# Table of Contents

IC

## AIC in R

IC weights

GORICA in R

## Complement

Extra

The end

# Why information criteria?

Why do we need information criteria?

- to compare **non-nested** models/hypotheses.
- to compare **more than two** models at the time.
- because there are **different ideas** about what poses a good model.

# Model selection using information criteria

Describe data as good as possible (fit)  
with fewest number of parameters (simplicity / non-complexity).

General form of information criteria:

**model misfit + model complexity**

Model/Hypothesis with the **smallest value** is considered the best model.

# AIC

$$\text{IC} = \text{model misfit} + \text{model complexity}$$

**AIC:** minimize the **Kullback-Leibler distance**

$$\text{AIC} = -2 \log f(y|\hat{\theta}_y) + 2p$$

where  $\hat{\theta}_y$  is the maximum likelihood estimate.

# Intermezzo: Kullback-Leibler (K-L) distance

The K-L distance quantifies the distance between

- the **truth**:  $p(\cdot)$
- the **model under consideration**:  $f(\cdot|\theta)$

Extra:

$$\text{K-L} = E_{p(y)} \left[ \log p(y) - \log f(y|\theta) \right]$$



# Alternative to AIC: BIC

$$IC = \text{model misfit} + \text{model complexity}$$

**BIC**: maximize the **marginal model probability** (cf. Bayes Factor)

$$\text{BIC} = -2 \log f(y|\hat{\theta}_y) + \log(n)p$$

where  $\hat{\theta}_y$  is the maximum likelihood estimate.

## Note:

Difference in penalty AIC vs BIC:  $2p$  vs  $\log(n)p$ .  
(due to different derivation)

# Order-restricted generalization of AIC: GORIC

$$\text{IC} = \text{model misfit} + \text{model complexity}$$

**GORIC:** minimize the **K-L distance**

$$\text{GORIC} = -2 \log f(y|\tilde{\theta}_y) + 2PT$$

where  $\tilde{\theta}_y$  is the order-restricted maximum likelihood estimate and  $PT$  can be seen as the expected number of distinct parameters.

Kuiper, R.M., Hoijtink, H. and Silvapulle, M.J. (2011). An Akaike type information criterion for model selection under inequality constraints. *Biometrika*, 98, 495-501.

# Intermezzo: Idea penalty (PT)

loose interpretation

$$H_1 : \mu_1 > \mu_2 > \mu_3$$

contains 1 ordering of three means, 1-2-3.

Thus, not complex (i.e., parsimonious).

$$H_2 : \mu_1 > \mu_2, \mu_3$$

contains 2 orderings of three means: 1-2-3 and 1-3-2.

Thus, more complex (less parsimonious).

$$H_u : \mu_1, \mu_2, \mu_3$$

contains all six possible orderings of three means.

Thus, is most complex one (not parsimonious).

# GORIC Approximation: GORICA

$$IC = \text{model misfit} + \text{model complexity}$$

**GORICA:** Approximate GORIC, thus minimize the **K-L distance**

$$\text{GORICA} = -2 \log f(\hat{\theta}, \hat{\Sigma}_{\theta} | \tilde{\theta}) + 2PT$$

where

- $\hat{\theta}$  are the parameters of the unconstrained model,
- $\hat{\Sigma}_{\theta}$  is the variance matrix of  $\hat{\theta}$ ,
- $\tilde{\theta}$  is the order-restricted maximum likelihood estimate, and
- $PT$  can be seen as the expected number of distinct parameters.

Altınışık, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., and Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods*, 26(5), 599-621.

# GORICA vs GORIC

## Similarities with GORIC:

- Form:  $-2 \text{ fit} + 2 \text{ complexity}$ .
- Broad type of restrictions.

## Differences compared to GORIC:

- Uses asymptotic expression of the likelihood (is a normal):  
can therefore be easily applied to all types of statistical models.  
Disadvantage: might work less well in case of small samples.
- Does not need data set, but mle's and their covariance matrix.
- Can leave out nuisance parameters (i.e., not part of hypotheses).

## Note:

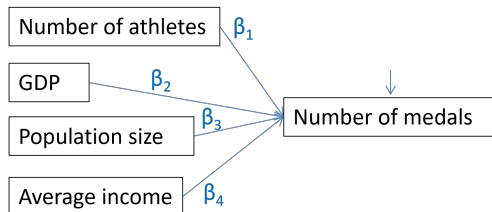
In case of normal linear models and/or not too small samples:  
GORICA weights = GORIC weights.

# Why GORIC and GORICA?

Most researchers have a-priori “order-restricted” / “informative” / “theory-based” hypotheses, e.g.:

$H_1 : \mu_1 > \mu_2 > \mu_3$  (in an ANOVA model) or

$H_1 : \beta_1 > \beta_2 > \beta_3 > \beta_4$  (in a regression/SEM model).



# Note on comparable estimates

## Continuous predictors

If compare relative strength/importance of parameters (e.g.,  $\beta_1 > \beta_2$ ), then make sure comparable:  
e.g., standardize continuous predictors.

## Multiple outcomes

If compare parameters across outcomes, then (also) standardize outcomes.

# Note on including “unconstrained” hypothesis

If set of hypotheses does not contain a reasonable/good one:  
Select the best of set of weak hypotheses.

Prevent choosing a weak hypothesis  
by including unconstrained hypothesis  $H_u$  (or  $H_a$ ):

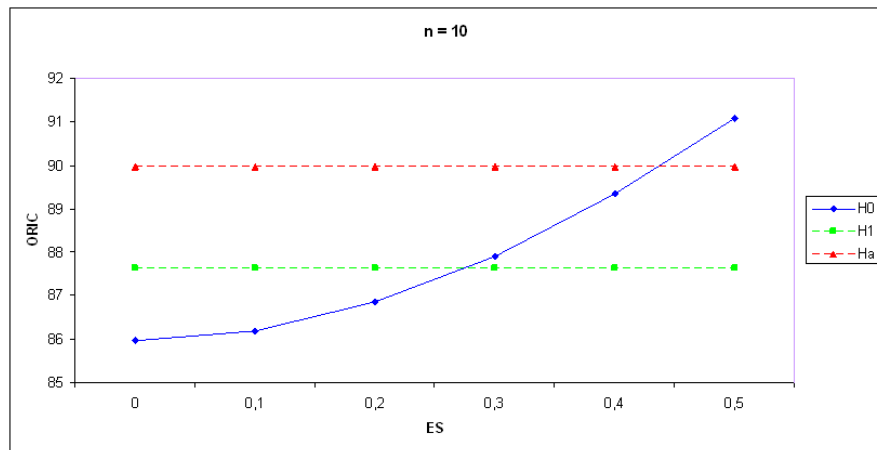
$$\begin{aligned} H_0 : & \quad \mu_1 = \mu_2 = \mu_3, \\ H_1 : & \quad \mu_1 > \mu_2 > \mu_3, \\ H_u : & \quad \mu_1, \mu_2, \mu_3, \\ & \quad \text{(i.e., no restrictions).} \end{aligned}$$

$H_u$  highest fit but also most complex, thus failsafe.

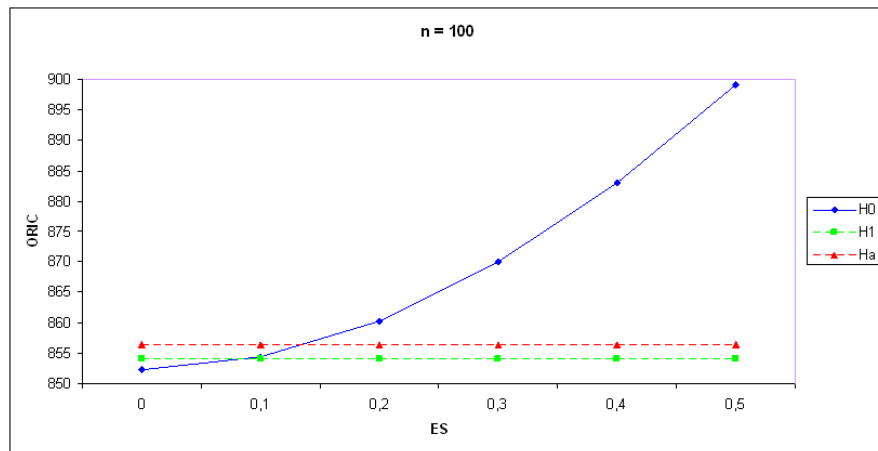
Note: GORIC for  $H_0$  and  $H_u$  equals AIC value.



Confirmation more power: 1 data set. GORIC values for 3 groups, effect size  $ES$ , and  $n = 10$  observations per group



Confirmation more power: 1 data set. GORIC values for 3 groups, effect size  $ES$ , and  $n = 100$  observations per group



# Note on hypotheses

1. Only include hypotheses with sound theoretical and/or empirical basis.

Note: Often a null hypothesis is not of interest.

2. Keep the number of hypotheses included as small as possible.
3. This is indeed a subjective endeavor, aim for inter-peer / inter-subjective agreement.

# Table of Contents

IC

AIC in R

IC weights

GORICA in R

Complement

Extra

The end

# Example: Multigroup regression

Regression example from lab meeting 'Intro and regression' and 'Multigroup analysis' lecture:

- Outcome: sw
- Predictors: overt and covert
- Group: gender (males and females)

# Example: Multigroup regression - data

```
# Data
data_regr <- read.table("popular_regr.txt", header = T)

data_regr[sapply(data_regr, function(x)
  as.character(x) %in% c("-99", "-999") )] <- NA

data_regr$gender <- factor(data_regr$gender,
  labels = c("male", "female"))
```

# Example: Multigroup regression - model specification

```
# Model specification
```

```
# Model 1
```

```
model.MG regr <- '  
  sw ~ overt + covert # regression  
  sw ~~ sw             # residual variance  
  sw ~ 1               # intercept  
'
```

```
# Model 2
```

```
model.MG regr_equal <- '  
  # model with labeled parameters  
  sw ~ c(b1,b1)*overt + c(b2,b2)*covert  
  sw ~~ sw  
  sw ~ 1  
'
```

# Example: Multigroup regression - fitting model

```
# Fit model
```

```
# Model 1
```

```
fit_MGregr <- lavaan(model = model.MGregr,  
                      data = data_regr,  
                      group = "gender") # multigroup
```

```
# Model 2
```

```
fit_MGregr_equal <- lavaan(model = model.MGregr_equal,  
                            data = data_regr,  
                            group = "gender") # multigroup
```



# Example: Multigroup regression - Alternative specification

```

model.MGregr_unequal <- '
  # model with labeled parameters
  sw ~ c(b1_m,b1_f)*overt + c(b2_m,b2_f)*covert
  sw ~~ sw
  sw ~ 1
'

# Model 1
fit_MGregr <- lavaan(model = model.MGregr_unequal,
                     data = data_regr,
                     group = "gender") # multigroup

# Model 2
fit_MGregr_equal <- lavaan(model = model.MGregr_unequal,
                           data = data_regr,
                           group = "gender", # multigroup
                           constraints = c("b1_m == b1_f;
                                             b2_m == b2_f"))

```

# Example: Multigroup regression - model comparison

```
# Compare models,
# i.e., evaluate equality of regression coefficients,
# using AIC

#anova(fit_MG regr, fit_MG regr_equal)[1:2]
# or
AIC(fit_MG regr, fit_MG regr_equal)

##                df        AIC
## fit_MG regr      8 2319.833
## fit_MG regr_equal 6 2316.868
```

Model with equality constraints has lowest AIC, so is preferred.  
But how much?

# Example: Multigroup regression - model comparison

AIC for  $n = 100$  vs  $n = 1491$ :

```
round(AIC_n, 3)
```

##	AIC_n100	AIC_n1491
## fit_MGregar	125.170	2319.833
## fit_MGregar_equal	124.309	2316.868
## abs. diff.	0.861	2.965

AIC increases with sample size.

Absolute difference in AIC values also increases.

Model with equality constraints has lowest AIC, so is preferred.

But how much?

# Table of Contents

IC

AIC in R

IC weights

GORICA in R

Complement

Extra

The end

# IC vs IC weights

IC values:

Cannot be interpreted, only compared: smallest is best.

IC weights ( $w_m$ ):

- $w_m$ : quantifies how much more  $H_m$  is supported than others in set.
- $w_m/w_{m'}$ : quantifies relative support of  $H_m$  vs  $H_{m'}$ .

Note: The bigger, the better.

Extra:

$$w_m = \frac{\exp(-0.5 \text{ IC}_m)}{\sum_{m'=1}^M \exp(-0.5 \text{ IC}_{m'})}.$$

Reference:

Kuiper, R.M., Hoijtink, H. and Silvapulle, M.J. (2012). Generalization of the order restricted information criterion for multivariate normal linear models. *Journal of Statistical Planning and Inference*, 142, 2454-2463.

# IC weights: Interpretation

Say,  $w_m/w_{m'} = .8/.2 = 4$ .

Possible interpretation:

- $H_m$  is 4 times more likely than  $H_{m'}$ .
- $H_m$  is 4 times more supported than  $H_{m'}$ .

If  $H_{m'}$  is complement of  $H_m$  (i.e., denoting all other possibilities):  
There is  $w_{m'} * 100\% = 20\%$  that  $H_m$  is not the best.

# GORIC weights vs BF and PMPs

For those familiar with or interested in Bayesian model selection:

$w_m/w_{m'} = \text{relative weight} \sim \text{Bayes factor } (BF_{mm'})$ .

$w_m \sim \text{posterior model probability (PMP)}$ .

# IC weights: Illustration in R

```
library(devtools) # Make sure you have Rtools
install_github("rebeccakuiper/ICweights")
#
library(ICweights)
#?IC.weights
```

```
# AIC weights
AICvalues <- AIC(fit_MGregar, fit_MGregar_equal)$AIC
Hypo.names <- c("Unc", "Equal")
#summary(IC.weights(AICvalues, Hypo.names))
IC.weights(AICvalues, Hypo.names)

##
## Per hypothesis/model, the information criterion value (IC), its weight (IC.w
##
##      IC    IC.weights ratio.IC.weights.vs.Unc ratio.IC.weights.vs.Equal
## Unc   2320 0.185      1.0              0.227
## Equal 2317 0.815      4.4              1.000
```



# IC weights: Illustration in R Ctd.

```
# AIC weights
IC.weights(AICvalues, Hypo.names)$IC.weights

##          Unc      Equal
## 0.1850568 0.8149432

IC.weights(AICvalues, Hypo.names)$ratio.IC.weights

##          vs Unc  vs Equal
## Unc      1.000000 0.2270794
## Equal    4.403746 1.0000000
```

$H_{Equal}$  is  $.815/.185 \approx 4.4 > 1$  times more supported than  $H_{Unc}$ .  
Thus, there is quite some evidence that  $H_{Equal}$  is the best of this set.

## *Repetition:* Note on comparable estimates

### Continuous predictors

If compare relative strength/importance of parameters (e.g.,  $\beta_1 > \beta_2$ ), then make sure comparable:  
e.g., standardize continuous predictors.

### Multiple outcomes

If compare parameters across outcomes, then (also) standardize outcomes.

# Note on comparable estimates in lavaan

## Default lavaan (and Mplus)

**U**nstandardized estimates.

## Standardized estimates

Can obtain them, e.g.: `summary(fit, standardized = TRUE)`  
but cannot do tests or model selection on them.

## Solutions

- Standardize your data first.
- Use GORICA weights.



# GORICA in R

```

# Option 1 - used in this course
if (!require("restriktor")) install.packages("restriktor")
library(restriktor)
#
#goric(fitted.lavaan.model,
#      hypotheses = list(H_1, H_2, ..., H_M),
#      standardized = TRUE)
#
#type = "gorica" # default in case of fitted lavaan model
#comparison = "complement" # will be discussed later

# Option 2
if (!require("gorica")) install.packages("gorica")
library(gorica)
#?gorica

```

# Example: Multigroup regression - model

```
# Model specification: Unconstrained model
model.MGRegr_unequal <- '
  # model with labeled parameters
  sw ~ c(b1_m,b1_f)*overt + c(b2_m,b2_f)*covert
  sw ~~ sw
  sw ~ 1
'

# Fit model: Unconstrained model
fit_MGRegr <- lavaan(model = model.MGRegr_unequal,
                      data = data_regr,
                      group = "gender") # multigroup
```

## Extra: GORICA as AIC proxy

```
H_AIC <- "abs(b1_m) = abs(b1_f); abs(b2_m) = abs(b2_f)"
```

```
#On unstandardized estimates, as done by default in lavaan:
set.seed(100)
AIC_Unstand <- gorica(fit_MGregar, hypotheses = list(H_AIC=H_AIC))
AIC_Unstand$result$gorica.weights

## [1] 0.8149104 0.1850896

# Same as for AIC obtained via lavaan
```

```
#On standardized estimates
set.seed(100)
AIC_Stand <- gorica(fit_MGregar, hypotheses = list(H_AIC=H_AIC),
                    standardized = T)
AIC_Stand$result$gorica.weights

## [1] 0.8102574 0.1897426
```

# Example: Multigroup regression - hypothesis

```
H1 <- "abs(b1_m) > abs(b1_f); abs(b2_m) < abs(b2_f)"
```

Remark: I assume we want to compare the absolute strength.



# Example: Multigroup regression - model comparison

```
set.seed(100)
results_MGregar <- goric(fit_MGregar, hypotheses = list(H1),
                        standardized = T)
#results_MGregar$result # Output too long for the slide
results_MGregar$result$gorica.weights

## [1] 0.6908713 0.3091287

results_MGregar$ratio.gw

##                vs. H1 vs. unconstrained
## H1                1.0000000          2.234899
## unconstrained 0.4474475          1.000000
```

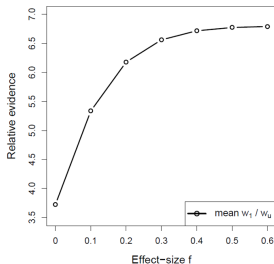
$H_1$  is  $.691/.309 \approx 2.235 > 1$  times more supported than the unc.  
 Thus, there is some evidence that  $H_1$  is the best of this set.  
 Note that the unconstrained always includes the true hypothesis...



# Alternative safeguard: Complement of $H_m$

## Until now

Prevent choosing weak hypothesis, include unconstrained hypothesis ( $H_u$ ).



**Figure:** IC weights can have an upper bound, when informative hypothesis has maximum fit (i.e., is fully in agreement with the data).

# Alternative safeguard: Complement of $H_m$

## Alternatively (in case of one hypothesis of interest)

Evaluate hypothesis of interest against its complement.

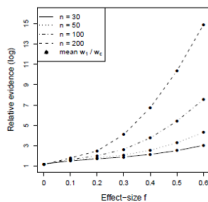
More powerful than against the unconstrained  
if  $H_m$  has maximum fit (which can lead to an upper bound for the  
GORIC weights).

Reference:

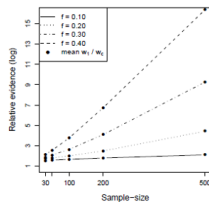
Vanbrabant, L., Van Loey, N., & Kuiper, R. M. (2020). Evaluating a Theory-Based Hypothesis Against Its Complement Using an AIC-Type Information Criterion With an Application to Facial Burn Injury. *Psychological Methods*, 25(2), 129-142.  
<https://doi.org/10.1037/met0000238>

# Alternative safeguard: Complement of $H_m$

vs complement

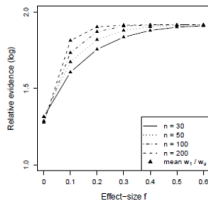


(a)

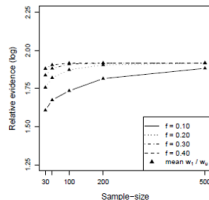


(b)

vs unconstrained



(c)



(d)

## Note: complement not always higher weight

In case  $H_m$  is almost true, but not true:

1.  $H_m$  does not have maximum fit (but much lower penalty).
2. Support for  $H_m$  is less when evaluating it against its complement (than  $H_u$ ).

This is of course a good thing, since  $H_m$  is not true.

Explanation:

Penalty of  $H_c$  is smaller than that of  $H_u$ .

Against  $H_u$ , you choose  $H_m$  'sooner' because of low penalty for  $H_m$ .

## Example: Multigroup regression - model comparison

```
set.seed(100)
results_MGRegr_compl <- goric(fit_MGRegr,
                              hypotheses = list(H1),
                              comparison = "complement",
                              standardized = T)
#results_MGRegr_compl$result # Output too long for the slide
results_MGRegr_compl$result$gorica.weights

## [1] 0.6219892 0.3780108

results_MGRegr_compl$ratio.gw[1,2]

## [1] 1.645427
```

$H_1$  is  $.622/.378 \approx 1.645$  times more supported than its complement.  
 Thus, there is only little evidence that  $H_1$  is the best of this set.  
 I would now explore for a competing hypothesis for next research...

# Use of unconstrained and complement

In case of 1 hypothesis of interest:

Use the complement.

It acts as another competing (non-overlapping) hypothesis.

In case of 2 or more hypotheses of interest:

Use the complement of the set, but this is not yet possible in the software.

So, use the unconstrained as failsafe.

This is used to check whether at least one of the hypothesis is not weak.

Then, the non-weak hypotheses can be compared to all other.





# Contact

## Websites

<https://github.com/rebeccakuiper/Tutorials>

[www.uu.nl/staff/RMKuiper/Software](http://www.uu.nl/staff/RMKuiper/Software)

[www.uu.nl/staff/RMKuiper/Websites%20%2F%20Shiny%20apps](http://www.uu.nl/staff/RMKuiper/Websites%20%2F%20Shiny%20apps)  
[informative-hypotheses.sites.uu.nl/software/goric/](http://informative-hypotheses.sites.uu.nl/software/goric/)

## E-mail

[r.m.kuiper@uu.nl](mailto:r.m.kuiper@uu.nl)

# Confirmatory methods - some of my references

- Kuiper, R. M., and Hoijtink, H. (2010). Comparisons of Means Using Exploratory and Confirmatory Approaches. *Psychological Methods*, 15(1), 69–86.
- Kuiper, R. M., Klugkist, I., and Hoijtink, H. (2010). A Fortran 90 Program for Confirmatory Analysis of Variance. *Journal of Statistical Software*, 34(8), 1–31.
- Kuiper, R.M., Hoijtink, H. and Silvapulle, M.J. (2011). An Akaike type information criterion for model selection under inequality constraints. *Biometrika*, 98, 495-501. (GORIC)
- Kuiper, R.M., Nederhof, T., and Klugkist, I. (2015). Properties of hypothesis testing techniques and (Bayesian) model selection for exploration-based and theory-based (order-restricted) hypotheses. *British Journal of Mathematical and Statistical Psychology*, 68(2), 220 – 245.
- Altınışık, Y., Van Lissa, C. J., Hoijtink, H., Oldehinkel, A. J., and Kuiper, R. M. (2021). Evaluation of inequality constrained hypotheses using a generalization of the AIC. *Psychological Methods*, 26(5), 599-621.  
<https://doi.org/10.1037/met0000406> (GORICA)

## Note on possibilities multiple studies

- Update hypotheses.  
First data set (or a part of it) generates one or more hypotheses.  
Then, other(s) used to determine evidence for that/those.
- Aggregate evidence for hypotheses.  
Aggregate the support for theories (diverse designs allowed).  
Note: Meta-analysis aggregates parameter estimates or effect sizes which need to be comparable.

# Table of Contents

IC

AIC in R

IC weights

GORICA in R

Complement

Extra

The end

# Summary

- IC
- AIC in R
- IC weights
- GORICA in R
- Complement

Exploratory vs confirmatory.

# Thanks & How to proceed

Thanks for listening!

Are there any questions?

- Ask fellow participant on course platform.
- Ask teacher during Q&A (or via course platform).
- See if making the lab exercises help.
- Check a GORICA article or a GORICA tutorial: e.g., <https://github.com/rebeccakuiper/Tutorials>.

You can start working on the lab exercises.