University of
PITTSBURGH

Statistical Tools to Support Implementation: Variable Selection and Post-Selection Inference in Genomic Nursing Research

Lacey W. Heinsberg, PhD, RN (she/her)

Assistant Professor of Nursing and Human Genetics,
University of Pittsburgh

Conflicts of Interest

Beyond my grant support acknowledged at the end of this presentation, I have nothing to disclose.

Pitt

Objectives

Understand the need and methods for **variable selection** in multivariable modeling

Describe the issues of **drawing inference** from data used for variable selection

Review a **strategy** to counter such issues

Pitt

Road map

- Introduction of the **problem**

- **Solution primer**

- Practical **application**

Pitt

Road map

Introduction of the **problem**

Introduction to a **solution**

Practical **application**

Pitt

Complex phenotypes

Most **health conditions/outcomes** are determined by **many** competing **genetic and non-genetic factors**.

Body composition

Research question: What factors are associated with fat mass? (And which of these are potentially modifiable?)

Understand a complex outcome

Describe

Explain

Predict

Adjust for study design

Ascertainment

Pitt

Modeling approaches

A priori (model is predefined)

Data-driven approach to select variables

Pitt

Modeling approaches

A priori (model is predefined)

Data-driven approach to select variables

Leisman et al, 2020. Development and Reporting of Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and Critical Care Journals. *Critical Care Medicine*.

Pitt

Modeling approaches

A priori (model is predefined)

Data-driven approach to select variables

Heinze, Wallisch, and Dunkler 2017. Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal.*

Pitt

We are seeing an increasing reliance on data-driven approaches

Pitt
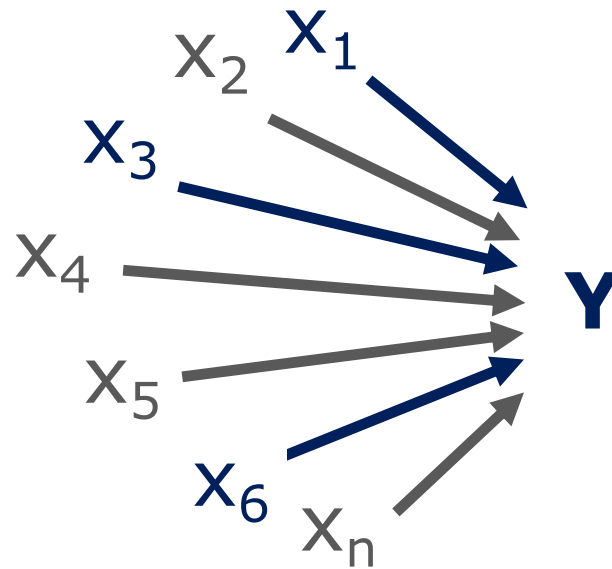
Data-driven approaches introduce uncertainty and invalidate classical inference.

Pitt

Data-driven variable-selection approaches

Bivariable p-values

Best subset

Forward selection

Backward elimination

Pitt

Data-driven variable-selection approaches

Bivariable p-values

Best subset

Forward selection

Backward elimination

Pitt

Bivariable p-values

Age

Weight

Height

Hair color

Diet

Physical activity

Eye color

Fat mass

Pitt

Bivariable p-values

Age

Weight

Height

Hair color

Diet

Physical activity

Eye color

p=0.02

Fat mass

Pitt
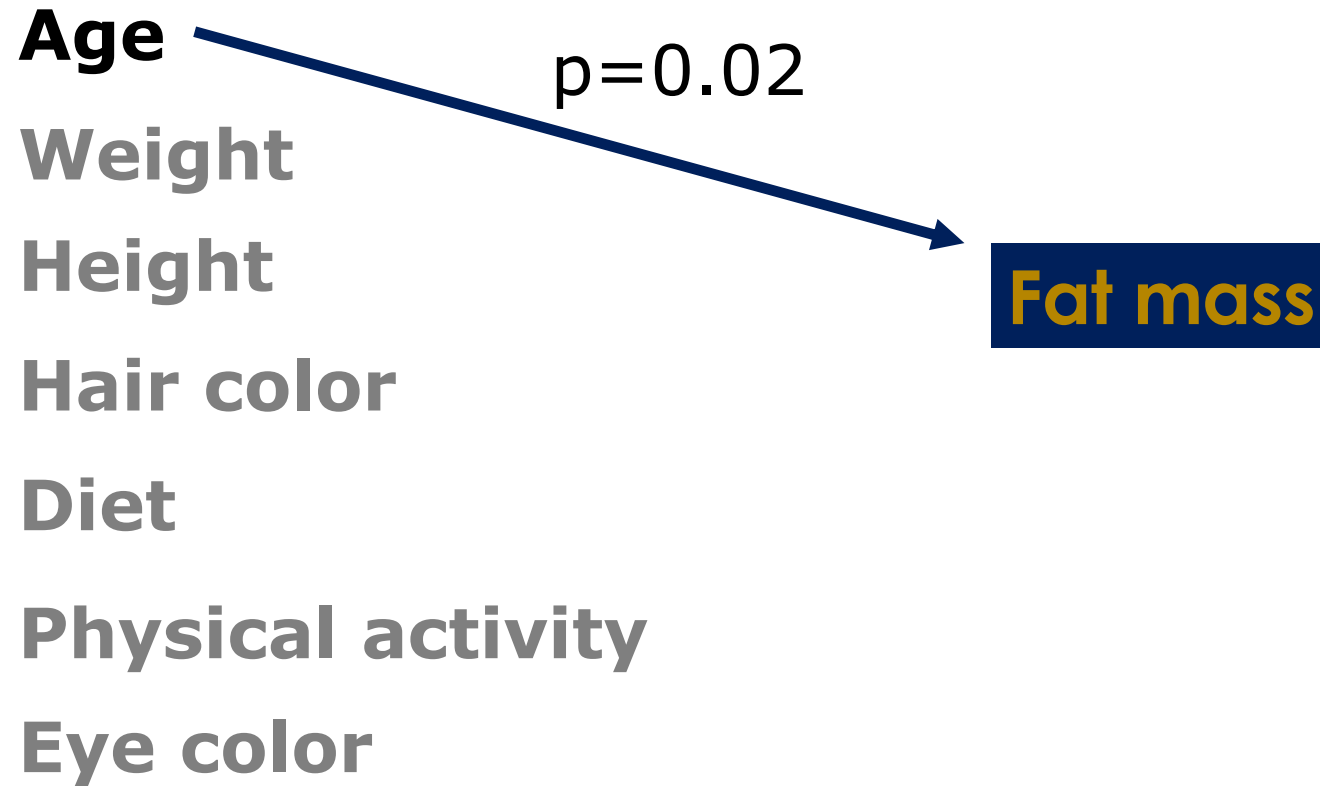
Bivariable p-values

Age

Weight   p=0.01

Height

Hair color

Diet

Physical activity

Eye color

Fat mass

Pitt

Bivariable p-values

Age

Weight

Height   p=0.57   **Fat mass**

Hair color

Diet

Physical activity

Eye color

Pitt

Bivariable p-values

Age

Weight

Height
Diet

Physical activity

Fat mass

Fat mass ~ Age + Weight + Height + Diet + Physical activity

Bivariable p-values

Pros: Simple/straightforward; reproducible

Cons: Leads to spurious conclusions
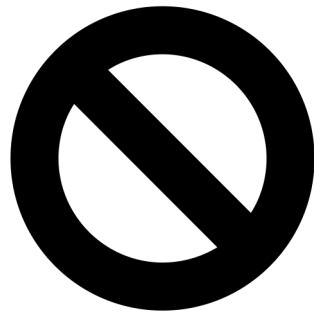
DO NOT USE

Pitt

Data-driven variable-selection approaches

Bivariable p-values

Best subset

Forward selection

Backward elimination

Pitt

Best subset selection

Weight

Height

Hair color

Fat mass

Best subset selection (Fat mass ~ …)

Models with 1 variable
- (1) Weight
- (2) Height
- (3) Hair color

Models with 2 variables
- (4) Weight + Height
- (5) Weight + Hair color
- (6) Height + Hair color

Model with 3 variables
- (7) Weight + Height + Hair color

Pitt

Best subset selection (Fat mass ~ …)

Models with 1 variable
1. **Weight**
2. Height
3. Hair color

Models with 2 variables
4. **Weight + Height**
5. Weight + Hair color
6. Height + Hair color

Model with 3 variables
7. **Weight + Height + Hair color**

Pitt

Best subset selection (Fat mass ~ …)

Models with 1 variable
- (1) Weight
- (2) Height
- (3) Hair color

Models with 2 variables
- (4) **Weight + Height**
- (5) Weight + Hair color
- (6) Height + Hair color

Model with 3 variables
- (7) Weight + Height + Hair color

Pitt

Best subset selection

Pros: Simple; easily interpreted; objective/reproducible

Cons: Computational limitations, theoretical limitations

Pitt

Data-driven variable-selection approaches

Bivariable p-values

Best subset

Forward selection

Backward elimination

Pitt

Forward selection

Age
Height
Weight
Hair color
Diet
Physical activity
Eye color

Fat mass

Null model
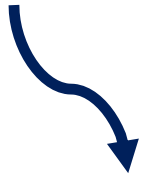
Fat mass ~ _____

Forward selection

Age

Height

Hair color

Diet

Physical activity

Eye color

Fat mass

Fat mass ~   Weight

Forward selection

Age

Hair color

Diet

Physical activity

Eye color

Fat mass

Fat mass ~ <u>Weight + Height</u>

Forward selection

Hair color

Fat mass

Eye color

Fat mass ~ <u>Weight + Height + Age + Diet + Physical activity</u>

Backward elimination

Fat mass

Full model
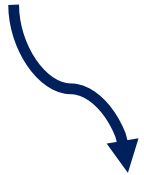
Fat mass ~ Weight + Height + Age + Diet + Physical activity + Hair color + Eye color

Backward elimination

Hair color

Fat mass

Fat mass ~ Weight + Height + Age + Diet + Physical activity + Eye color

Backward elimination

Hair color

Fat mass

Eye color

Fat mass ~ <u>Weight + Height + Age + Diet + Physical activity</u>

Forward selection/backwards elimination approaches

Pros: Simple; interpretable; reproducible; commonly used

Cons: Does not consider all possible combinations; collinearity problems can arise; biased regression coefficients/confidence intervals/p-values; unstable selection of variables; does not support valid inference

Pitt

Using data to select a model introduces uncertainty and invalidates classical inference

Pitt

Age
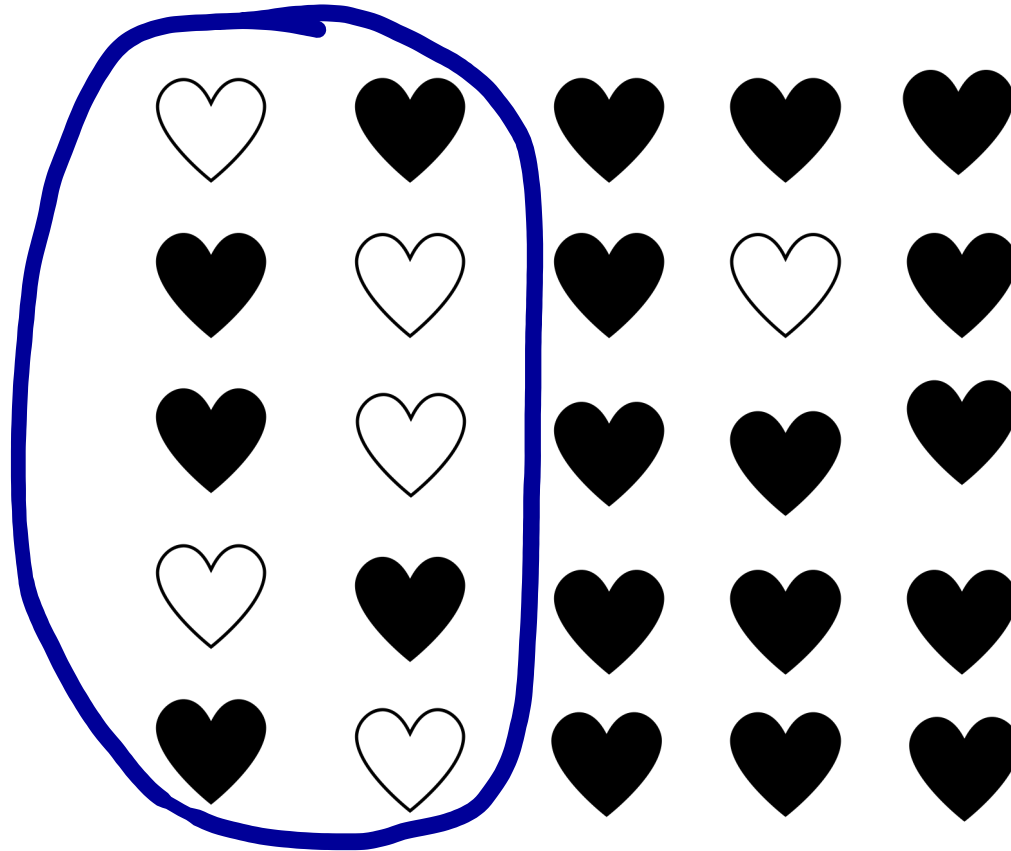
Weight

Height

Hair color

Diet

Physical activity

Eye color

Fat mass

+ Age

Weight

− Height

Hair color

Diet

+ Physical activity

Eye color

Fat mass

Pitt

Age
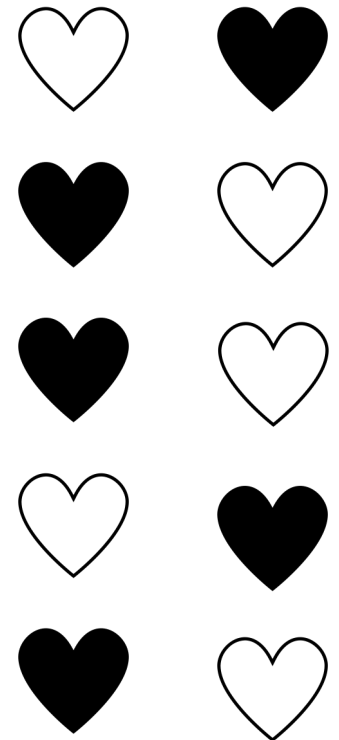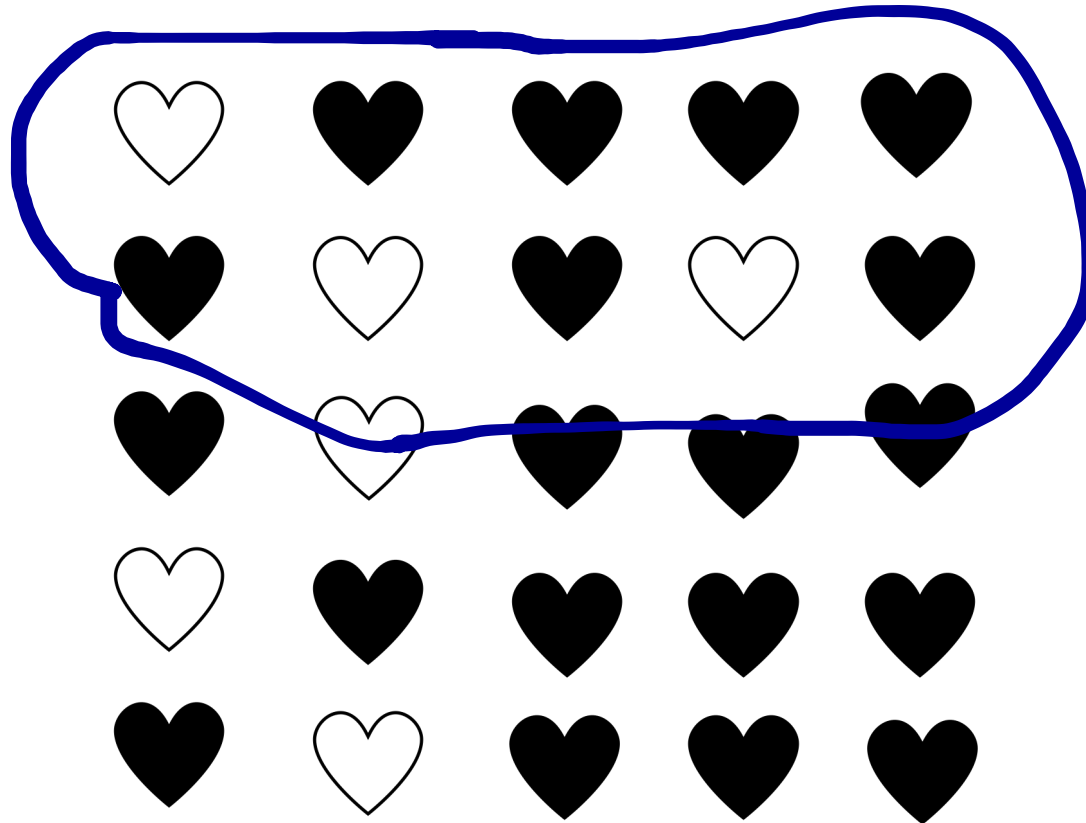
Weight

— Height

Hair color

Diet

Physical activity

— Eye color

Fat mass
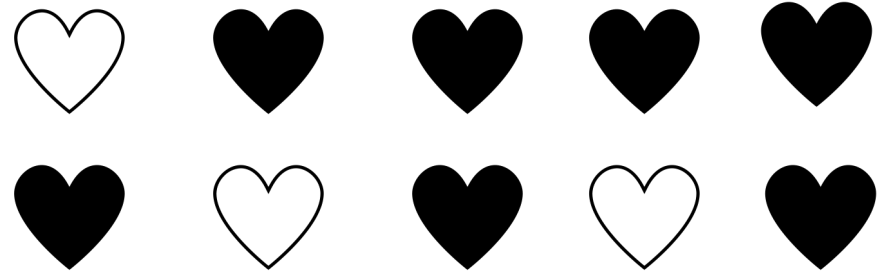
Samples are often not truly representative of the population

Heinze, Wallisch, and Dunkler (2017). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal.*
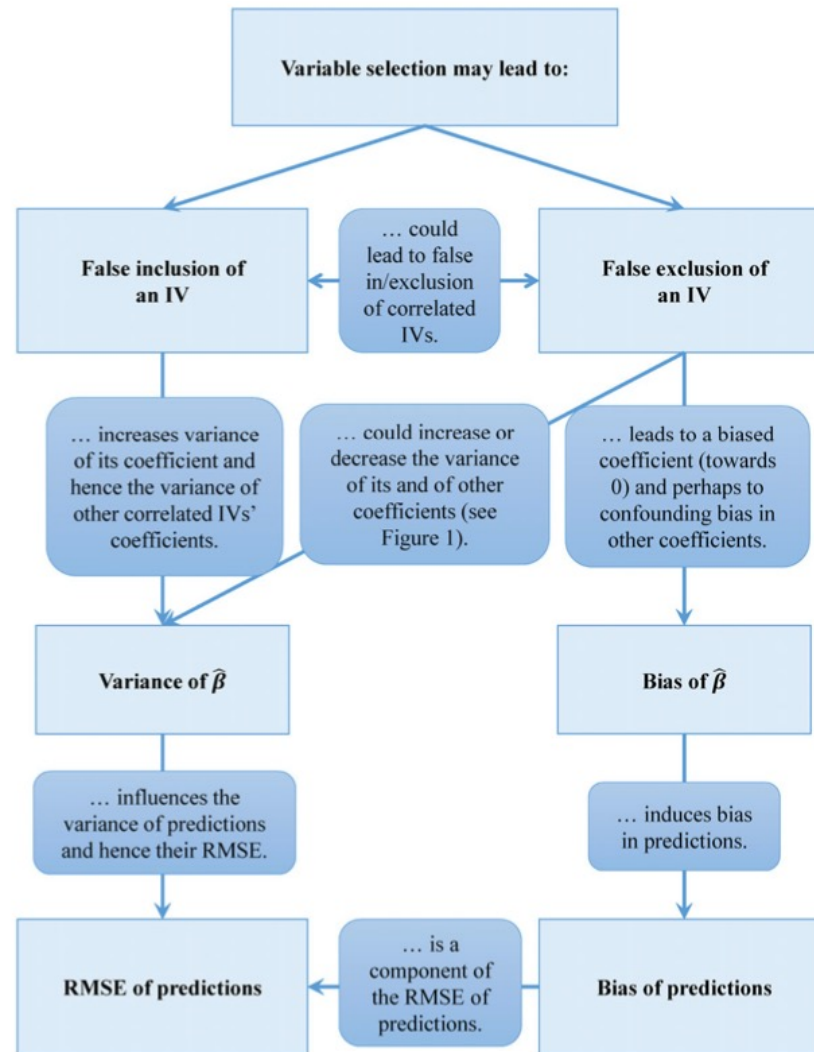
FIGURE 2 A schematic network of dependencies arising from variable selection. β, regression coefficient; IV, independent variable; RMSE, root mean squared error

Pitt

TABLE 1 Four potential models to estimate body fat in %

Model	Intercept		Weight in kg			Height in cm		Abdomen circumference		R^2_{adj}
	Estimate	SE	Estimate		SE	Estimate	SE	Estimate	SE	
1	−14.892	2.762	0.420	+	0.034					0.381
2	76.651	9.976	0.582	+	0.034	−0.586	0.062			0.543
3	−47.659	2.634	−0.292	−	0.047			0.979	0.056	0.722
4	−30.364	11.432	−0.215	−	0.068	−0.096	0.062	0.910	0.071	0.723

R^2_{adj}, adjusted R^2; SE, standard error

Heinze, Wallisch, and Dunkler (2017). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*.

Questions

How stable is variable selection?

Does variable selection lead to biased or inaccurate predictions?

Pitt

Road map

Introduction of the **problem**

Introduction to a **solution**

Practical **application**

Pitt

Problem: Using the same data for variable selection AND post-selection inference

Potential solutions

Discovery / replication

Cross validation

Bootstrapping stability investigations

Pitt

Potential solutions

Discovery / replication

Cross validation

Bootstrapping stability investigations

Pitt

Bootstrap to perform stability investigation

1. Draw B samples with replacement from original data set

2. Perform model selection on each sample (and store results)

3. Repeat

Etc.
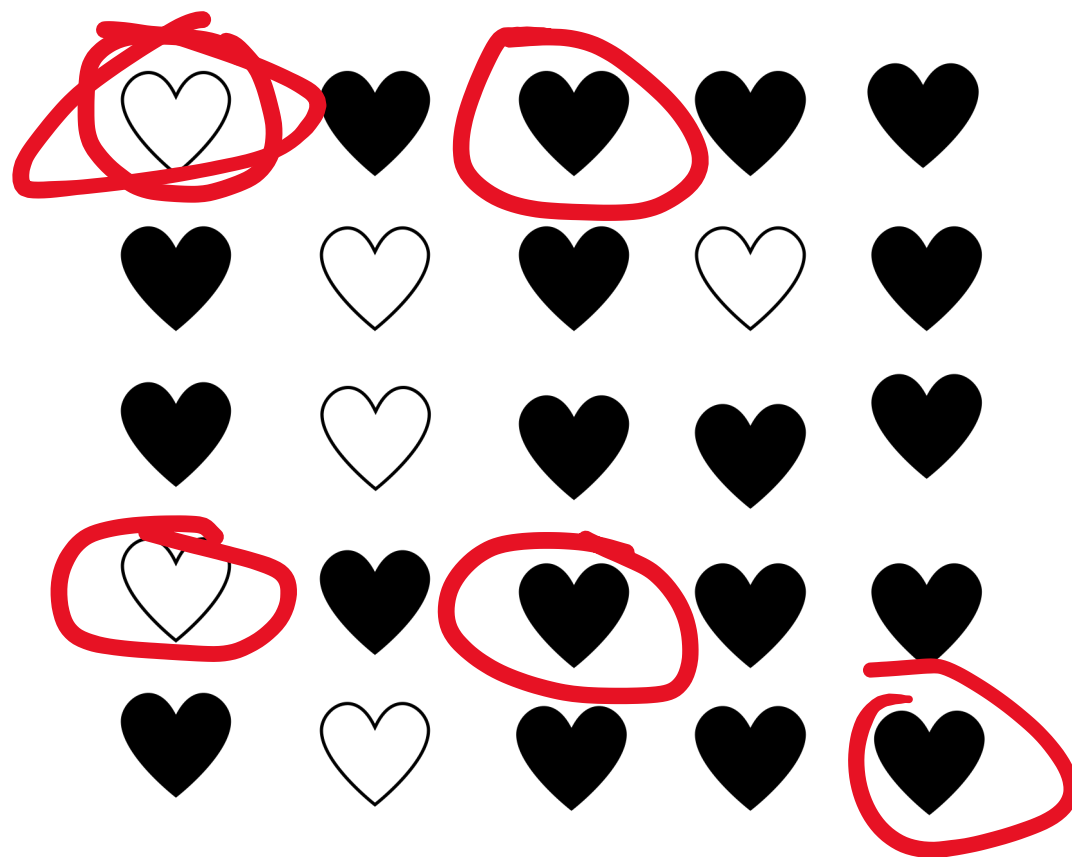
Bootstrap sample 1

Pitt
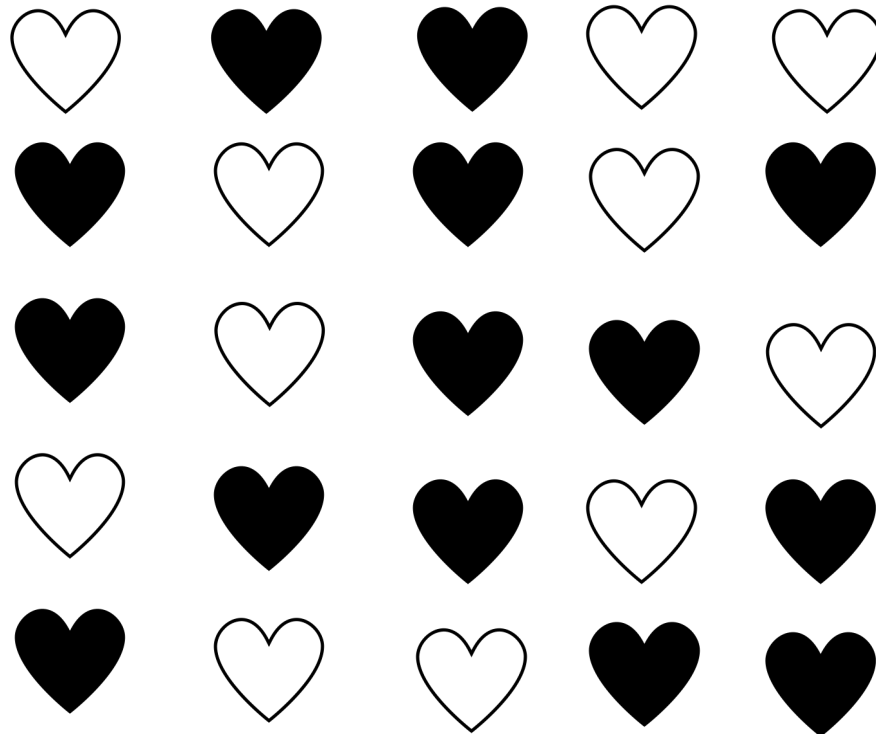
Bootstrap sample 2
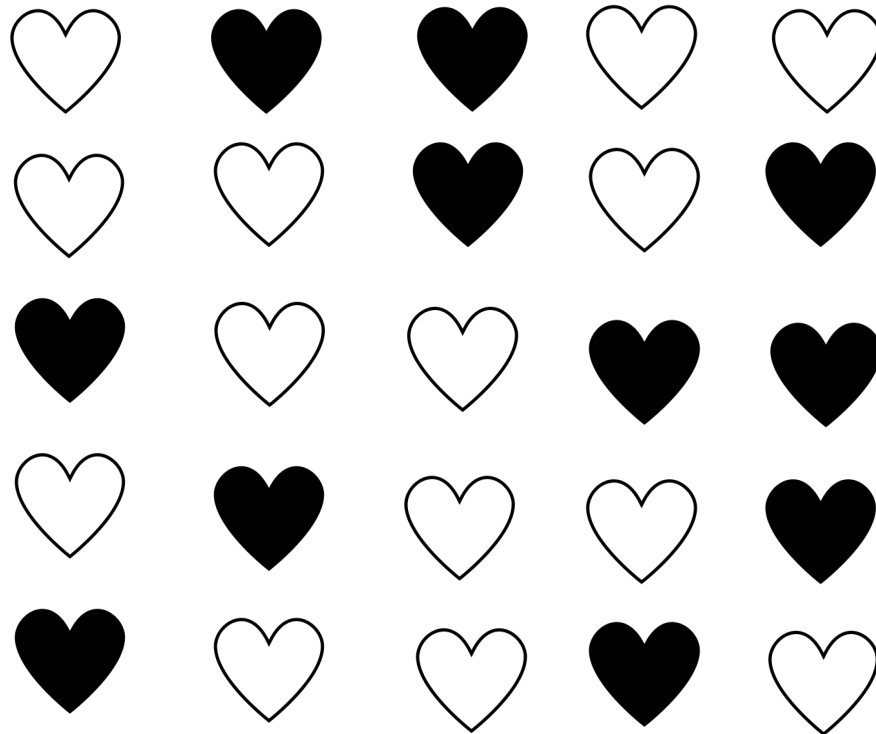
Bootstrap sample 3
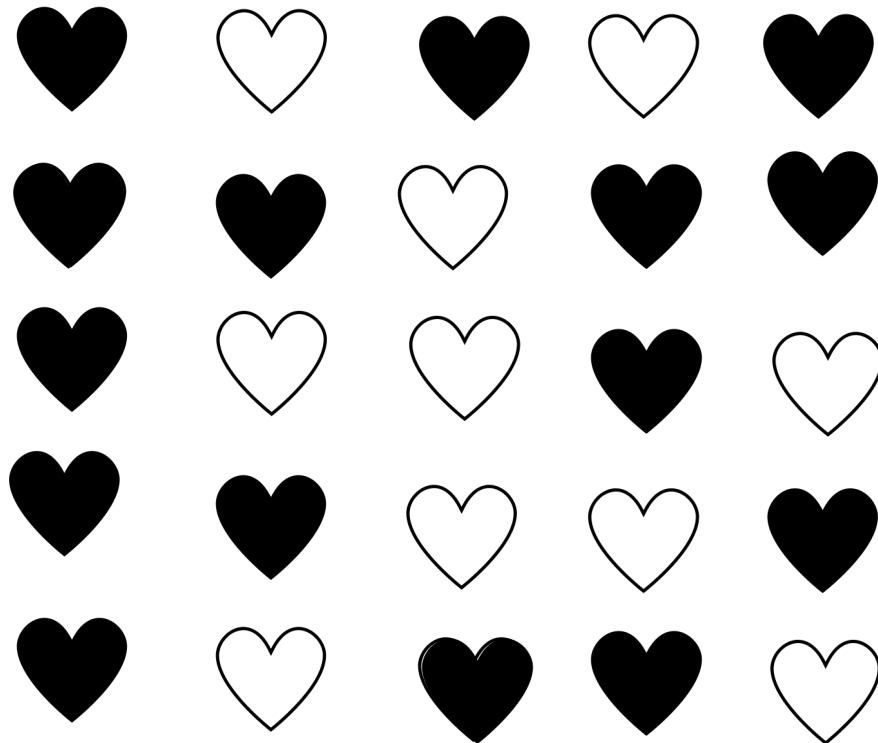
Bootstrap sample B
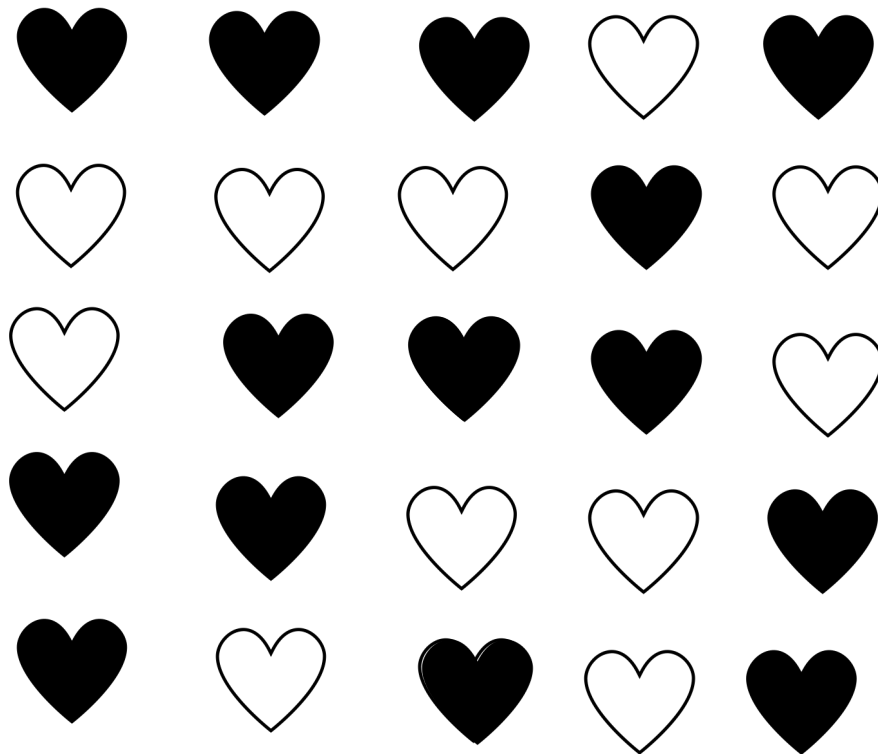
Bootstrap stability investigation

Bootstrap inclusion frequencies (BIF) for each variable

Distribution of coefficients

Model selection frequencies

Pairwise inclusion tables

Pitt

Road map

Introduction of the **problem**

Introduction to a **solution**

Practical **application**

Pitt

(Fictitious) **Case Study:**

Variable selection and **post-selection inference** for identifying drivers of **fat mass**

www.github.com/ lwheinsberg/NUR_VarSel

Adapted from
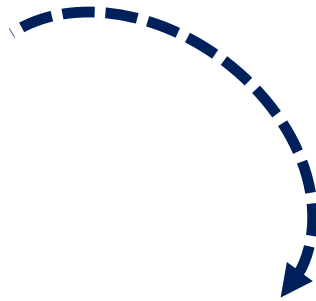
Heinze, Wallisch, and Dunkler (2017). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal.*

Pitt

Case study objective

This fictitious study aims to investigate the influence of **genetic variation** (obesity-related SNPs), **anthropometric measures**, **social drivers of health** (SDOH), and **behavioral factors** on a **continuous measure of body fat mass** in a sample of adults.

Pitt

Predictors

Genetic variation (rs1-rs6 + gxe)

Demographics (e.g., age)

Anthropometrics (e.g., weight, height, circumferences)

Family dynamics (e.g., mealtime frequency, no. of kids)

Environmental factors (e.g. walkability)

Psychosocial factors (e.g., stress)

Behavioral factors (e.g., diet, sleep)

Modeling approaches

A priori (model is predefined)

Data-driven approach to select variables

Pitt

A note about events per variable (EPV)

TABLE 3 Some recommendations on variable selection, shrinkage, and stability investigations based on events-per-variable ratios

Situation	Recommendation
For some IVs it is known from previous studies that their effects are strong, for example age in cardiovascular risk studies or tumor stage at diagnosis in cancer studies.	Do not perform variable selection on IVs with known strong effects.
$EPV_{global} > 25$	Variable selection (on IVs with unclear effect size) should be accompanied by stability investigation.
$10 < EPV_{global} \leq 25$	Variable selection on IVs with unclear effect size should be accompanied by postestimation shrinkage methods (e.g. Dunkler et al., 2016), or penalized estimation (LASSO selection) should be performed. In any case, a stability investigation is recommended.
$EPV_{global} \leq 10$	Variable selection not recommended. Estimate the global model with shrinkage factor, or penalized likelihood (ridge regression). Interpretation of effects may become difficult because of biased effect estimation.

Heinze, Wallisch, and Dunkler (2017). Variable selection – A review and recommendations for the practicing statistician. *Biometrical Journal*.

Data-driven variable-selection approaches

~~Bivariable p values~~

~~Best subset~~

+ Stability investigation

~~Forward selection~~

Backward elimination

Pitt

Statistical analysis

R statistical software

Multiple linear regression and backwards elimination + bootstrapping (stability investigation and post-selection inference)

Abdominal circumference and height will be "forced" into the model

Pitt

Predictors	Global model			Selected model				Final bootstrap model		
	Est.	SE	Bootstrap inclusion frequency	Est.	SE	RMSD ratio	Relative conditional bias	Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22		-15.01	14.15	1.22	-24.79	-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	100 (forced)	-0.04	0.05	1.12	157.35	-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**	**1.18**	**-1.26**	**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**100**	**-1.14**	**0.14**	**1.56**	**0.90**	**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**100**	**0.82**	**0.16**	**1.39**	**1.22**	**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**99.9**	**-1.92**	**0.42**	**1.02**	**0.32**	**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**99.2**	**-1.26**	**0.33**	**1.03**	**2.92**	**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	95.3	0.44	0.14	0.95	0.76	0.40	0	0.65
Stress	-0.29	0.11	87.3	-0.33	0.11	1.31	13.24	-0.30	-0.55	0
rs1	3.87	1.56	84.6	4.09	1.49	1.36	15.14	3.95	0	7.70
rs5	0.83	0.36	81.3	0.80	0.35	1.28	14.61	0.83	0	1.54
Neck circumference (cm)	-0.40	0.20	77.8	-0.36	0.19	1.25	25.16	-0.43	-0.81	0
Age (years)	0.05	0.03	76.0	0.05	0.02	1.24	20.37	0.05	0	0.10
Weight (kg)	-0.20	0.10	75.8	-0.19	0.09	1.27	15.23	-0.19	-0.39	0
rs3	0.71	0.40	67.5	0.60	0.39	1.22	27.47	0.72	0	1.44
Thigh circumference (cm)	0.19	0.11	62.6	0.15	0.09	1.21	19.66	0.16	0	0.34
Health eaty index	0.03	0.02	57.4	0.04	0.02	1.26	53.52	0.04	0	0.08
Bicep circumference (cm)	0.15	0.14	37.5			1.09	95.16	0	0	0.42
No. of children	0.15	0.15	36.9			1.19	87.26	0	-0.22	0.50
rs4	-0.32	0.33	33.7			1.06	103.26	0	-0.93	0
Ankle circumference (cm)	-0.12	0.22	30.1			1.05	145.05	0	-0.59	0.36
Physical activity	0.37	0.48	29.6			0.99	132.23	0	0	1.29
Knee circumference (cm)	-0.17	0.20	29.3			1.03	109.50	0	-0.56	0
Family mealtime frequency (meals/wk)	0.03	0.05	24.5			0.97	143.39	0	-0.08	0.14
rs6	0.08	0.33	21.4			0.87	259.41	0	-0.56	0.73
Physical environment score	-0.0004	0.02	18.1			0.78	-504.14	0	-0.04	0.04

Predictors
(Intercept)
Height (cm)
Abdominal circumference (cm)
Sleep duration (avg. hr/night)
GxE (Sleep*rs1)
Wrist circumference (cm)
rs2
Forearm circumference (cm)
Stress
rs1
rs5
Neck circumference (cm)
Age (years)
Weight (kg)
rs3
Thigh circumference (cm)
Health eaty index
Bicep circumference (cm)
No. of children
rs4
Ankle circumference (cm)
Physical activity
Knee circumference (cm)
Family mealtime frequency (meals/wk)
rs6
Physical environment score

Predictors	Global model	
	Est.	SE
(Intercept)	-17.44	15.22
Height (cm)	-0.01	0.06
Abdominal circumference (cm)	**0.96**	**0.07**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**
GxE (Sleep*rs1)	**0.86**	**0.17**
Wrist circumference (cm)	**-1.83**	**0.44**
rs2	**-1.27**	**0.33**
Forearm circumference (cm)	0.41	0.15
Stress	-0.29	0.11
rs1	3.87	1.56
rs5	0.83	0.36
Neck circumference (cm)	-0.40	0.20
Age (years)	0.05	0.03
Weight (kg)	-0.20	0.10
rs3	0.71	0.40
Thigh circumference (cm)	0.19	0.11
Health eaty index	0.03	0.02
Bicep circumference (cm)	0.15	0.14
No. of children	0.15	0.15
rs4	-0.32	0.33
Ankle circumference (cm)	-0.12	0.22
Physical activity	0.37	0.48
Knee circumference (cm)	-0.17	0.20
Family mealtime frequency (meals/wk)	0.03	0.05
rs6	0.08	0.33
Physical environment score	-0.0004	0.02

| | Global model | | | Selected model | |
Predictors	Est.	SE		Est.	SE
(Intercept)	-17.44	15.22		-15.01	14.15
Height (cm)	-0.01	0.06		-0.04	0.05
Abdominal circumference (cm)	**0.96**	**0.07**		**0.95**	**0.07**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**		**-1.14**	**0.14**
GxE (Sleep*rs1)	**0.86**	**0.17**		**0.82**	**0.16**
Wrist circumference (cm)	**-1.83**	**0.44**		**-1.92**	**0.42**
rs2	**-1.27**	**0.33**		**-1.26**	**0.33**
Forearm circumference (cm)	0.41	0.15		0.44	0.14
Stress	-0.29	0.11		-0.33	0.11
rs1	3.87	1.56		4.09	1.49
rs5	0.83	0.36		0.80	0.35
Neck circumference (cm)	-0.40	0.20		-0.36	0.19
Age (years)	0.05	0.03		0.05	0.02
Weight (kg)	-0.20	0.10		-0.19	0.09
rs3	0.71	0.40		0.60	0.39
Thigh circumference (cm)	0.19	0.11		0.15	0.09
Health eaty index	0.03	0.02		0.04	0.02
Bicep circumference (cm)	0.15	0.14			
No. of children	0.15	0.15			
rs4	-0.32	0.33			
Ankle circumference (cm)	-0.12	0.22			
Physical activity	0.37	0.48			
Knee circumference (cm)	-0.17	0.20			
Family mealtime frequency (meals/wk)	0.03	0.05			
rs6	0.08	0.33			
Physical environment score	-0.0004	0.02			

Predictors	Global model		Selected model		Final bootstrap model		
	Est.	SE	Est.	SE	Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22	-15.01	14.15	-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	-0.04	0.05	-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**0.95**	**0.07**	**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**-1.14**	**0.14**	**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**0.82**	**0.16**	**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**-1.92**	**0.42**	**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**-1.26**	**0.33**	**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	0.44	0.14	0.40	0	0.65
Stress	-0.29	0.11	-0.33	0.11	-0.30	-0.55	0
rs1	3.87	1.56	4.09	1.49	3.95	0	7.70
rs5	0.83	0.36	0.80	0.35	0.83	0	1.54
Neck circumference (cm)	-0.40	0.20	-0.36	0.19	-0.43	-0.81	0
Age (years)	0.05	0.03	0.05	0.02	0.05	0	0.10
Weight (kg)	-0.20	0.10	-0.19	0.09	-0.19	-0.39	0
rs3	0.71	0.40	0.60	0.39	0.72	0	1.44
Thigh circumference (cm)	0.19	0.11	0.15	0.09	0.16	0	0.34
Health eaty index	0.03	0.02	0.04	0.02	0.04	0	0.08
Bicep circumference (cm)	0.15	0.14			0	0	0.42
No. of children	0.15	0.15			0	-0.22	0.50
rs4	-0.32	0.33			0	-0.93	0
Ankle circumference (cm)	-0.12	0.22			0	-0.59	0.36
Physical activity	0.37	0.48			0	0	1.29
Knee circumference (cm)	-0.17	0.20			0	-0.56	0
Family mealtime frequency (meals/wk)	0.03	0.05			0	-0.08	0.14
rs6	0.08	0.33			0	-0.56	0.73
Physical environment score	-0.0004	0.02			0	-0.04	0.04

Predictors	Global model		Bootstrap inclusion frequency	Selected model		Final bootstrap model		
	Est.	SE		Est.	SE	Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22		-15.01	14.15	-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	100 (forced)	-0.04	0.05	-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**	**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**100**	**-1.14**	**0.14**	**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**100**	**0.82**	**0.16**	**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**99.9**	**-1.92**	**0.42**	**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**99.2**	**-1.26**	**0.33**	**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	95.3	0.44	0.14	0.40	0	0.65
Stress	-0.29	0.11	87.3	-0.33	0.11	-0.30	-0.55	0
rs1	3.87	1.56	84.6	4.09	1.49	3.95	0	7.70
rs5	0.83	0.36	81.3	0.80	0.35	0.83	0	1.54
Neck circumference (cm)	-0.40	0.20	77.8	-0.36	0.19	-0.43	-0.81	0
Age (years)	0.05	0.03	76.0	0.05	0.02	0.05	0	0.10
Weight (kg)	-0.20	0.10	75.8	-0.19	0.09	-0.19	-0.39	0
rs3	0.71	0.40	67.5	0.60	0.39	0.72	0	1.44
Thigh circumference (cm)	0.19	0.11	62.6	0.15	0.09	0.16	0	0.34
Health eaty index	0.03	0.02	57.4	0.04	0.02	0.04	0	0.08
Bicep circumference (cm)	0.15	0.14	37.5			0	0	0.42
No. of children	0.15	0.15	36.9			0	-0.22	0.50
rs4	-0.32	0.33	33.7			0	-0.93	0
Ankle circumference (cm)	-0.12	0.22	30.1			0	-0.59	0.36
Physical activity	0.37	0.48	29.6			0	0	1.29
Knee circumference (cm)	-0.17	0.20	29.3			0	-0.56	0
Family mealtime frequency (meals/wk)	0.03	0.05	24.5			0	-0.08	0.14
rs6	0.08	0.33	21.4			0	-0.56	0.73
Physical environment score	-0.0004	0.02	18.1			0	-0.04	0.04

Predictors	Est.	SE	Bootstrap inclusion frequency	Est.	SE	Bootstrap inclusion frequency	RI
			Global model				
(Intercept)	-17.44			-17.44	15.22	14.15	
Height (cm)	-0.01		100 (forced)	-0.01	0.06	0.05	100 (forced)
Abdominal circumference (cm)	0.96		100 (forced)	0.96	0.07	0.07	100 (forced)
Sleep duration (avg. hr/night)	-1.17			-1.17	0.14	0.14	100
GxE (Sleep*rs1)	0.86	0.17	100	0.86	0.17	0.16	100
Wrist circumference (cm)	1.83	0.44	99.9	1.83	0.44	0.42	99.9
rs2	-1.27	0.33	99.2	-1.27	0.33	0.33	99.2
Forearm circumference (cm)	0.41	0.15	95.3	0.41	0.15	0.14	95.3
Stress	-0.29	0.11	87.3	-0.29	0.11	0.11	87.3
rs1	3.87	1.56	84.6	3.87	1.56	1.49	84.6
rs5	0.83	0.36	81.3	0.83	0.36	0.35	81.3
Neck circumference (cm)	-0.40	0.20	77.8	-0.40	0.20	0.19	77.8
Age (years)	0.05	0.03	76.0	0.05	0.03	0.02	76.0

Predictors	Est.	SE	Global model			
			Bootstrap inclusion frequency	Est.	SE	Bootstrap inclusion frequency
(Intercept)	-17.44			-17.44	15.22	14.15
Height (cm)	-0.01		100 (forced)	-0.01	0.06	100 (forced)
Abdominal circumference (cm)	0.96		100 (forced)	0.96	0.07	100 (forced)
Sleep duration (avg. hr/night)	1.17		100	1.17	0.14	100
GxE (Sleep*rs1)	0.86		100	0.86	0.17	100
Wrist circumference (cm)	1.83		99.9	1.83	0.44	99.9
rs2	-1.27		99.2	-1.27	0.33	99.2
Forearm circumference (cm)	0.41		95.3	0.41	0.15	95.3
ss	-0.29			-0.29	0.11	87.3
rs1	3.87	1.56	84.6	3.87	1.49	84.6
rs5	0.83	0.36	81.3	0.83	0.35	81.3
Neck circumference (cm)	-0.40	0.20	77.8	-0.40	0.19	77.8
Age (years)	0.05	0.03	76.0	0.05	0.02	76.0
Weight (kg)	-0.20	0.10	75.8	-0.20	0.09	75.8
rs3	0.71	0.40	67.5	0.71	0.39	67.5
Thigh circumference (cm)	0.19	0.11	62.0	0.19	0.09	62.6

Predictors	Est.		Global model Bootstrap inclusion frequency	Est.	SE	Bootstrap inclusion frequency	SE
(Intercept)	-17.44			-17.44	15.22	14.15	
Height (cm)	-0.01		100 (forced)	-0.01	0.06	100 (forced)	0.05
Abdominal circumference (cm)	0.96		100 (forced)	0.96	0.07	100 (forced)	0.07
Sleep duration (avg. hr/night)	1.17		100	1.17	0.14	100	0.14
GxE (Sleep*rs1)	0.86		100	0.86	0.17	100	0.16
Wrist circumference (cm)	1.83		99.9	1.83	0.44	99.9	0.42
rs2	-1.27		99.2	-1.27	0.33	99.2	0.33
Forearm circumference (cm)	0.41		95.3	0.41	0.15	95.3	0.14
...ess	-0.29			-0.29	0.11	87.3	0.18
rs1	3.87	1.56	84.6	3.87	1.49	84.6	
rs5	0.83	0.36	81.3	0.83	0.35	81.3	
Neck circumference (cm)	0.40	0.20	77.8	0.40	0.19	77.8	
Age (years)	0.05	0.03	76.0	0.05	0.02	76.0	
Weight (kg)	-0.20	0.10	75.8	-0.20	0.09	75.8	
rs3	0.71	0.40	67.5	0.71	0.39	67.5	
Thigh circumference (cm)	0.19	0.11	62.6	0.19			

Variable	Estimate	SE	Estimate	SE	Score
rs1	3.87	1.56	3.84	1.56/4.09	84.49
rs5	0.83	0.36	0.83	0.36/0.80	80.35
Neck circumference (cm)	-0.40	0.20	-0.40	0.20/-0.36	77.89
Age (years)	0.05	0.03	0.05	0.03/0.05	76.02
Weight (kg)	-0.20	0.10	-0.20	0.10/-0.19	75.89
rs3	0.71	0.40	0.71	0.40/0.60	67.59
Thigh circumference (cm)	0.19	0.11	0.19	0.11/0.15	62.69
Health eaty index	0.03		0.03	0.02	57.42
Bicep circumference (cm)	0.15		0.15	0.14	37.5
No. of children	0.15		0.15	0.15	36.9
rs4	-0.32		-0.32	0.33	33.7
Ankle circumference (cm)	0.12		0.12	0.22	30.1
Physical activity	0.37		0.37	0.48	29.6
Knee circumference (cm)	-0.17		-0.17	0.20	29.3
Family mealtime frequency (meals/wk)	0.03		0.03	0.05	24.5
rs6	0.08		0.08	0.33	21.4
Physical environment score	-0.0004		-0.0004	0.02	18.1

Pitt

Predictors	Global model		Bootstrap inclusion frequency	Selected model		Final bootstrap model		
	Est.	SE		Est.	SE	Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22		-15.01	14.15	-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	100 (forced)	-0.04	0.05	-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**	**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**100**	**-1.14**	**0.14**	**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**100**	**0.82**	**0.16**	**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**99.9**	**-1.92**	**0.42**	**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**99.2**	**-1.26**	**0.33**	**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	95.3	0.44	0.14	0.40	0	0.65
Stress	-0.29	0.11	87.3	-0.33	0.11	-0.30	-0.55	0
rs1	3.87	1.56	84.6	4.09	1.49	3.95	0	7.70
rs5	0.83	0.36	81.3	0.80	0.35	0.83	0	1.54
Neck circumference (cm)	-0.40	0.20	77.8	-0.36	0.19	-0.43	-0.81	0
Age (years)	0.05	0.03	76.0	0.05	0.02	0.05	0	0.10
Weight (kg)	-0.20	0.10	75.8	-0.19	0.09	-0.19	-0.39	0
rs3	0.71	0.40	67.5	0.60	0.39	0.72	0	1.44
Thigh circumference (cm)	0.19	0.11	62.6	0.15	0.09	0.16	0	0.34
Health eaty index	0.03	0.02	57.4	0.04	0.02	0.04	0	0.08
Bicep circumference (cm)	0.15	0.14	37.5			0	0	0.42
No. of children	0.15	0.15	36.9			0	-0.22	0.50
rs4	-0.32	0.33	33.7			0	-0.93	0
Ankle circumference (cm)	-0.12	0.22	30.1			0	-0.59	0.36
Physical activity	0.37	0.48	29.6			0	0	1.29
Knee circumference (cm)	-0.17	0.20	29.3			0	-0.56	0
Family mealtime frequency (meals/wk)	0.03	0.05	24.5			0	-0.08	0.14
rs6	0.08	0.33	21.4			0	-0.56	0.73
Physical environment score	-0.0004	0.02	18.1			0	-0.04	0.04

Predictors	Global model		Bootstrap inclusion frequency	Selected model			Final bootstrap model		
	Est.	SE		Est.	SE		Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22		-15.01	14.15		-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	100 (forced)	-0.04	0.05		-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**		**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**100**	**-1.14**	**0.14**		**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**100**	**0.82**	**0.16**		**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**99.9**	**-1.92**	**0.42**		**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**99.2**	**-1.26**	**0.33**		**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	85.3	0.44	0.14		0.40	0	0.65

Predictors	Global model Est.	SE	Bootstrap inclusion frequency	Selected model Est.	SE	Final bootstrap model Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22		-15.01	14.15	-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	100 (forced)	-0.04	0.05	-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**	**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**100**	**-1.14**	**0.14**	**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**100**	**0.82**	**0.16**	**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**99.9**	**-1.92**	**0.42**	**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**99.2**	**-1.26**	**0.33**	**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	95.3	0.44	0.14	0.40	0	0.65
Stress	-0.29	0.11	87.3	-0.33	0.11	-0.30	-0.55	0
rs1	3.87	1.56	84.6	4.09	1.49	3.95	0	7.70
rs5	0.83	0.36	81.3	0.80	0.35	0.83	0	1.54
Neck circumference (cm)	-0.40	0.20	77.8	-0.36	0.19	-0.43	-0.81	0
Age (years)	0.05	0.03	76.0	0.05	0.02	0.05	0	0.10
Weight (kg)	-0.20	0.10	75.8	-0.19	0.09	-0.19	-0.39	0
rs3	0.71	0.40	67.5	0.60	0.39	0.72	0	1.44
Thigh circumference (cm)	0.19	0.11	62.6	0.15	0.09	0.16	0	0.34
Health oety index	0.03	0.02	57.4	0.04	0.02	0.04	0	0.08

Predictors	Global model Est.	SE	Bootstrap inclusion frequency	Selected model Est.	SE	RMSD ratio	Relative conditional bias	Final bootstrap model Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22		-15.01	14.15	1.22	-24.79	-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	100 (forced)	-0.04	0.05	1.12	157.35	-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**	**1.18**	**-1.26**	**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**100**	**-1.14**	**0.14**	**1.56**	**0.90**	**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**100**	**0.82**	**0.16**	**1.39**	**1.22**	**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**99.9**	**-1.92**	**0.42**	**1.02**	**0.32**	**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**99.2**	**-1.26**	**0.33**	**1.03**	**2.92**	**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	95.3	0.44	0.14	0.95	0.76	0.40	0	0.65
Stress	-0.29	0.11	87.3	-0.33	0.11	1.31	13.24	-0.30	-0.55	0
rs1	3.87	1.56	84.6	4.09	1.49	1.36	15.14	3.95	0	7.70
rs5	0.83	0.36	81.3	0.80	0.35	1.28	14.61	0.83	0	1.54
Neck circumference (cm)	-0.40	0.20	77.8	-0.36	0.19	1.25	25.16	-0.43	-0.81	0
Age (years)	0.05	0.03	76.0	0.05	0.02	1.24	20.37	0.05	0	0.10
Weight (kg)	-0.20	0.10	75.8	-0.19	0.09	1.27	15.23	-0.19	-0.39	0
rs3	0.71	0.40	67.5	0.60	0.39	1.22	27.47	0.72	0	1.44
Thigh circumference (cm)	0.19	0.11	62.6	0.15	0.09	1.21	19.66	0.16	0	0.34
Health eaty index	0.03	0.02	57.4	0.04	0.02	1.26	53.52	0.04	0	0.08
Bicep circumference (cm)	0.15	0.14	37.5			1.09	95.16	0	0	0.42
No. of children	0.15	0.15	36.9			1.19	87.26	0	-0.22	0.50
rs4	-0.32	0.33	33.7			1.06	103.26	0	-0.93	0
Ankle circumference (cm)	-0.12	0.22	30.1			1.05	145.05	0	-0.59	0.36
Physical activity	0.37	0.48	29.6			0.99	132.23	0	0	1.29
Knee circumference (cm)	-0.17	0.20	29.3			1.03	109.50	0	-0.56	0
Family mealtime frequency (meals/wk)	0.03	0.05	24.5			0.97	143.39	0	-0.08	0.14
rs6	0.08	0.33	21.4			0.87	259.41	0	-0.56	0.73
Physical environment score	-0.0004	0.02	18.1			0.78	-504.14	0	-0.04	0.04

Predictors	Global model			Selected model				Final bootstrap model		
	Est.	SE	Bootstrap inclusion frequency	Est.	SE	RMSD ratio	Relative conditional bias	Est.	2.5th %	97.5th %
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**	**1.18**	**-1.26**	**0.95**	**0.78**	**1.10**

	Global model			Selected model				Final bootstrap model		
Predictors	Est.	SE	Bootstrap inclusion frequency	Est.	SE	RMSD ratio	Relative conditional bias	Est.	2.5th %	97.5th %
(Intercept)	-17.44	15.22		-15.01	14.15	1.22	-24.79	-14.50	-46.37	20.22
Height (cm)	-0.01	0.06	100 (forced)	-0.04	0.05	1.12	157.35	-0.03	-0.15	0.08
Abdominal circumference (cm)	**0.96**	**0.07**	**100 (forced)**	**0.95**	**0.07**	**1.18**	**-1.26**	**0.95**	**0.78**	**1.10**
Sleep duration (avg. hr/night)	**-1.17**	**0.14**	**100**	**-1.14**	**0.14**	**1.56**	**0.90**	**-1.18**	**-1.64**	**-0.76**
GxE (Sleep*rs1)	**0.86**	**0.17**	**100**	**0.82**	**0.16**	**1.39**	**1.22**	**0.84**	**0.45**	**1.32**
Wrist circumference (cm)	**-1.83**	**0.44**	**99.9**	**-1.92**	**0.42**	**1.02**	**0.32**	**-1.83**	**-2.73**	**-0.92**
rs2	**-1.27**	**0.33**	**99.2**	**-1.26**	**0.33**	**1.03**	**2.92**	**-1.31**	**-1.93**	**-0.65**
Forearm circumference (cm)	0.41	0.15	95.3	0.44	0.14	0.95	0.76	0.40	0	0.65
Stress	-0.29	0.11	87.3	-0.33	0.11	1.31	13.24	-0.30	-0.55	0
rs1	3.87	1.56	84.6	4.09	1.49	1.36	15.14	3.95	0	7.70
rs5	0.83	0.36	81.3	0.80	0.35	1.28	14.61	0.83	0	1.54
Neck circumference (cm)	-0.40	0.20	77.8	-0.36	0.19	1.25	25.16	-0.43	-0.81	0
Age (years)	0.05	0.03	76.0	0.05	0.02	1.24	20.37	0.05	0	0.10
Weight (kg)	-0.20	0.10	75.8	-0.19	0.09	1.27	15.23	-0.19	-0.39	0
rs3	0.71	0.40	67.5	0.60	0.39	1.22	27.47	0.72	0	1.44
Thigh circumference (cm)	0.19	0.11	62.6	0.15	0.09	1.21	19.66	0.16	0	0.34
Health eaty index	0.03	0.02	57.4	0.04	0.02	1.26	53.52	0.04	0	0.08
Bicep circumference (cm)	0.15	0.14	37.5			1.09	95.16	0	0	0.42
No. of children	0.15	0.15	36.9			1.19	87.26	0	-0.22	0.50
rs4	-0.32	0.33	33.7			1.06	103.26	0	-0.93	0
Ankle circumference (cm)	-0.12	0.22	30.1			1.05	145.05	0	-0.59	0.36
Physical activity	0.37	0.48	29.6			0.99	132.23	0	0	1.29
Knee circumference (cm)	-0.17	0.20	29.3			1.03	109.50	0	-0.56	0
Family mealtime frequency (meals/wk)	0.03	0.05	24.5			0.97	143.39	0	-0.08	0.14
rs6	0.08	0.33	21.4			0.87	259.41	0	-0.56	0.73
Physical environment score	-0.0004	0.02	18.1			0.78	-504.14	0	-0.04	0.04

Other stability measures

Model selection frequency

Pairwise inclusion frequencies

Sensitivity analyses

Summary

Variable selection invalidates classical inference, can be unstable, and can lead to biased output

Pitt

Recommendations

Generate an initial working set of variables

If variable selection is appropriate, choose method

If using variable selection, ALWAYS perform stability investigations and sensitivity analyses

Pitt

Acknowledgements

Daniel E. Weeks, PhD
Heinze, Wallisch, and Dunkler (2018)

Pitt

Thank you!

Lacey W. Heinsberg, PhD, RN

Pronouns: she/her/hers

law145@pitt.edu

@lwheinsberg

University of Pittsburgh