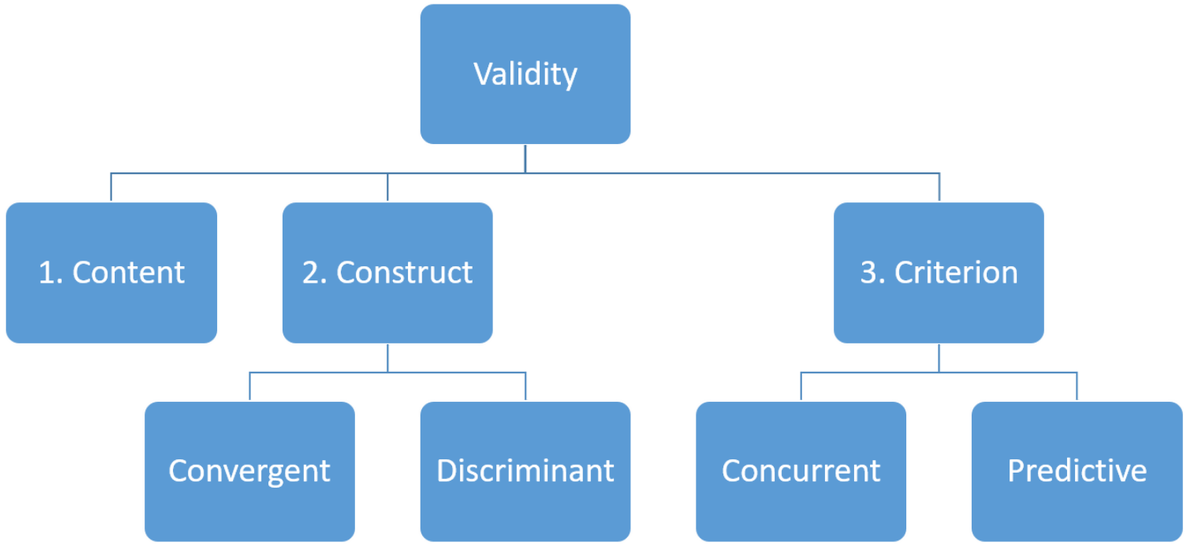
Concurrent validity & model diagrams

SURVEY AND MEASUREMENT DEVELOPMENT IN R

George Mount
Data analytics educator

datacamp

Concurrent validity



- Concurrent validity: Does the model correlate with another measure *now*?

Scaling the data

- Not all variables are on a scale of 1-5!
- Standardize variables

```
# Standardize our variables
b_loyal_age_scale <- scale(b_loyal_age)
```

Scaling the data

```
library(psych)

# Summary statistics - first five rows & columns
describe(b_loyal_age_scale)[1:5, 1:5]
```

```
     vars   n mean sd median
BL1     1 639    0  1   0.10
BL2     2 639    0  1   0.02
BL3     3 639    0  1   0.25
BL4     4 639    0  1  -0.13
BL5     5 639    0  1  -0.22
```

Building the model

1. "Latentize" the manifest variable: `=~`

```
b_loyal_age_model <- 'F1 =~ BL1 + BL2 + BL3
                      F2 =~ BL4 + BL5 + BL6
                      F3 =~ BL7 + BL8 + BL9 + BL10
                      age_fact =~ age'
```

Building the model

1. Correlate the manifest & latent statistics: `~~`

```
b_loyal_age_model <- 'F1 =~ BL1 + BL2 + BL3
                      F2 =~ BL4 + BL5 + BL6
                      F3 =~ BL7 + BL8 + BL9 + BL10
                      age_fact =~ age
                      age_fact ~~ F1 + F2 + F3'
```

Concurrent validity interpretation

1. Run & summarize model

```
b_loyal_age_cv <- sem(b_loyal_age_model, data = b_loyal_age_scale,
                      estimator = "MLR")

summary(b_loyal_age_cv, fit.measures = T, standardized = T)
```

Model interpretation

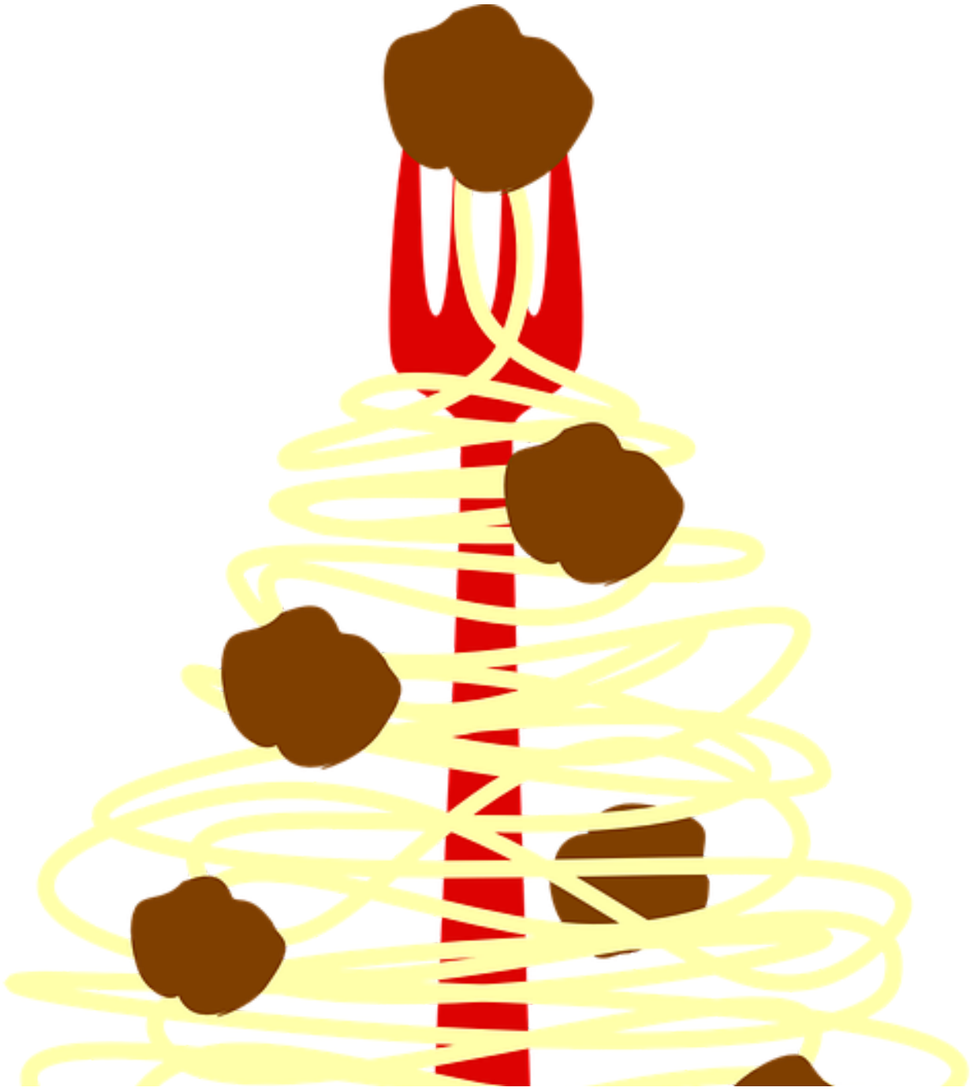
```
# Check the fit measures the same


  Robust Comparative Fit Index (CFI)    0.984
  Robust Tucker-Lewis Index (TLI)       0.978
```

Model interpretation

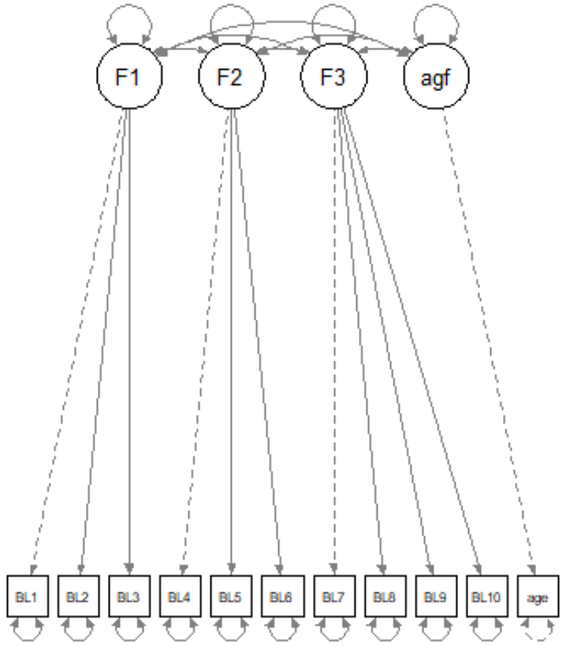
```
Covariances:
            Estimate  Std.Err  z-value  P(>|z|)
F1 ~~ age_fact  0.589    0.048   10.815    0.000
F2 ~~ age_fact  0.556    0.045   12.247    0.000
F3 ~~ age_fact  0.540    0.046   11.803    0.000
F1 ~~
    F2          0.330    0.035    9.401    0.000
    F3          0.214    0.029    7.461    0.000
F2 ~~
   F3           0.278    0.032    8.772    0.000
```

- Covariance of standardized items = correlation!

semPlot & "Spaghetti and meatballs modelling"

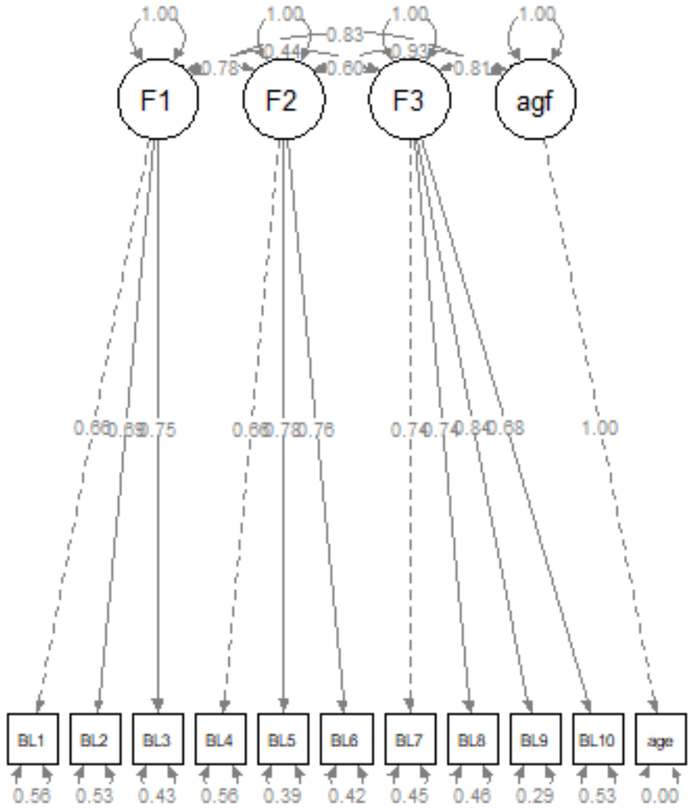


```
# Plot our model
library(semPlot)
semPaths(b_loyal_age_cv)
```



Modifying our diagram

```
# Plot our model with
#standardized estimates


semPaths(b_loyal_age_cv,
     whatLabels = "est.std",
     edge.label.cex = .8)
```



Let's practice!

SURVEY AND MEASUREMENT DEVELOPMENT IN R

Predictive validity & factor scores

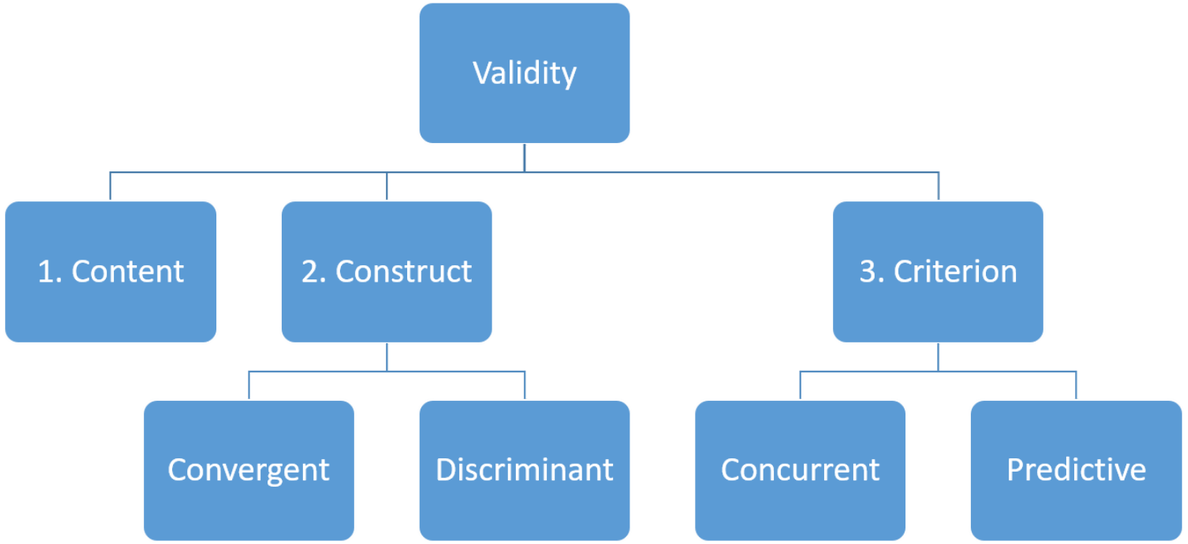
SURVEY AND MEASUREMENT DEVELOPMENT IN R

George Mount
Data analytics educator

datacamp

Predictive validity

- Can our model predict some *future* measure?
- Prediction & regression



Preparing our data for analysis

- Same as concurrent: bind and scale the data

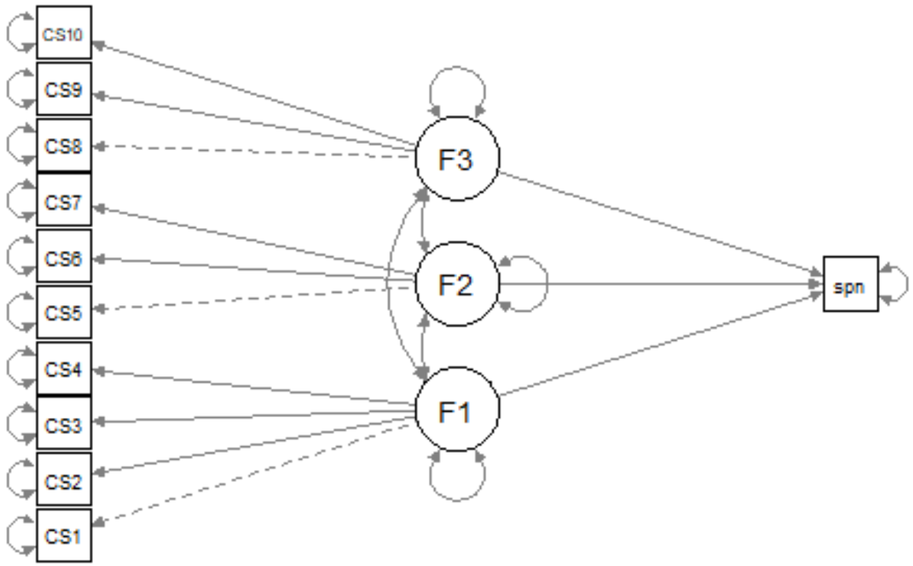
Regression in lavaan

- Same as in base `R` : use `~` !

```
# Regress total spending on our three
# dimensions of customer satisfaction

c_sat_model <- 'F1 =~ CS1 + CS2 + CS3 + CS4
                F2 =~ CS5 + CS6 + CS7
                F3 =~ CS8 + CS9 + CS10
                spend ~ F1 + F2 + F3'
```

Visualizing predictive validity with semPaths()

```
library(semPlot)

# plot regression model
# with IV's on left and
# DV on right

semPaths(c_sat_sem,
         rotation = 2)
```



Model interpretation

- p-values/standardized estimates:

```
# Get the standardized regression coefficients
# round the numeric output

library(dplyr)

standardizedSolution(c_sat_sem) %>%
    filter(op == "~") %>%
    mutate_if(is.numeric, round, digits = 3)
```

Model interpretation

```
   lhs op rhs est.std se    z   pvalue ci.lower ci.upper
1 spendf ~ F1 0.092 0.068 1.339 0.181 -0.042 0.226
2 spendf ~ F2 0.543 0.062 8.734 0.000 0.421 0.665
3 spendf ~ F3 0.395 0.048 8.148 0.000 0.300 0.490
```

Model interpretation

- R-squared

```
# Get the r-square
inspect(c_sat_sem, 'r2')
```

```
  CS1   CS2   CS3   CS4   CS5   CS6   CS7   CS8   CS9
0.536 0.410 0.397 0.543 0.429 0.413 0.448 0.617 0.467
 CS10   spend
0.539  0.736
```

Factor scores

- Factor score: relative standings on latent factor

```
# Compute factor scores based on CFA
csat_cfa <- cfa(model = csat_model, data = c_sat)

# Get factor scores as data frame
csat_scores <- as.data.frame(predict(csat_cfa))
```

Factor scores

```
# Factor scores for each respondent
nrow(csat_scores) == nrow(csat_cfa)
```

```
TRUE
```

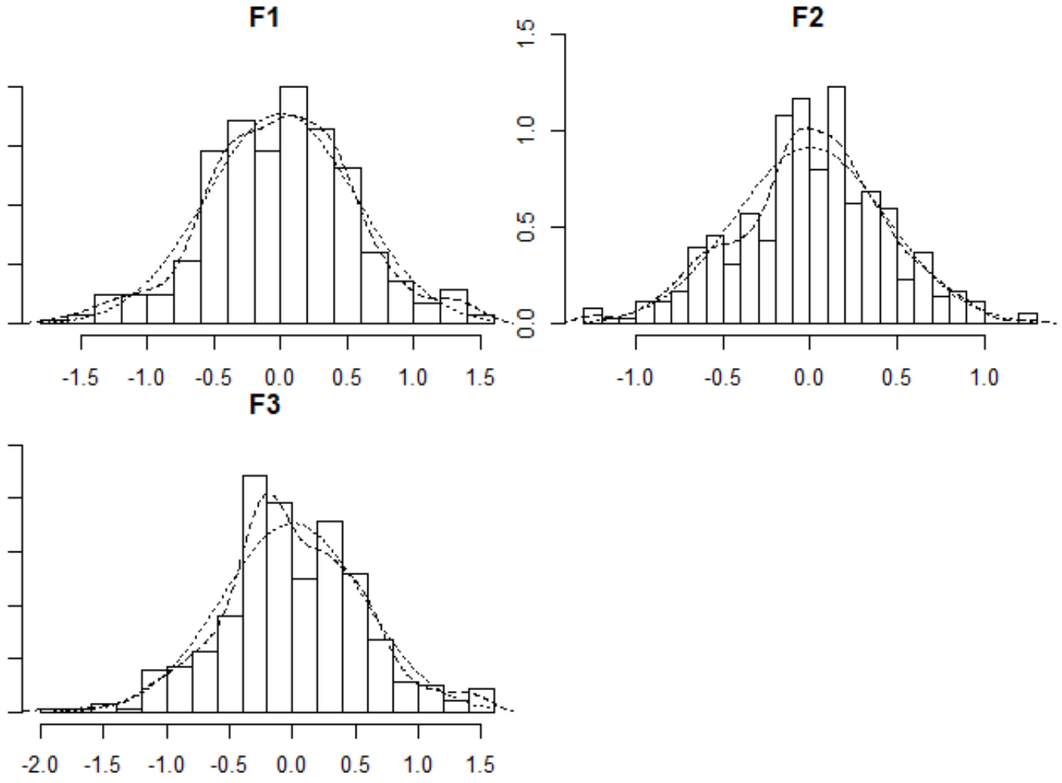
Describing factor scores

```
library(psych)
describe(csat_scores)
```

```
   vars   n mean   sd median trimmed  mad   min  max
F1    1 350    0 0.56   0.03    0.00 0.54 -1.71 1.49
F2    2 350    0 0.44   0.01    0.01 0.36 -1.28 1.26
F3    3 350    0 0.57  -0.06    0.00 0.49 -1.98 1.50

   range  skew kurtosis   se
F1  3.20 -0.06     0.31 0.03
F2  2.54 -0.18     0.16 0.02
F3  3.48 -0.01     0.53 0.03
```

Visualizing factor scores

```
# Plot histogram for each factor score
library(psych)
multi.hist(csat_scores)
```



Let's practice!

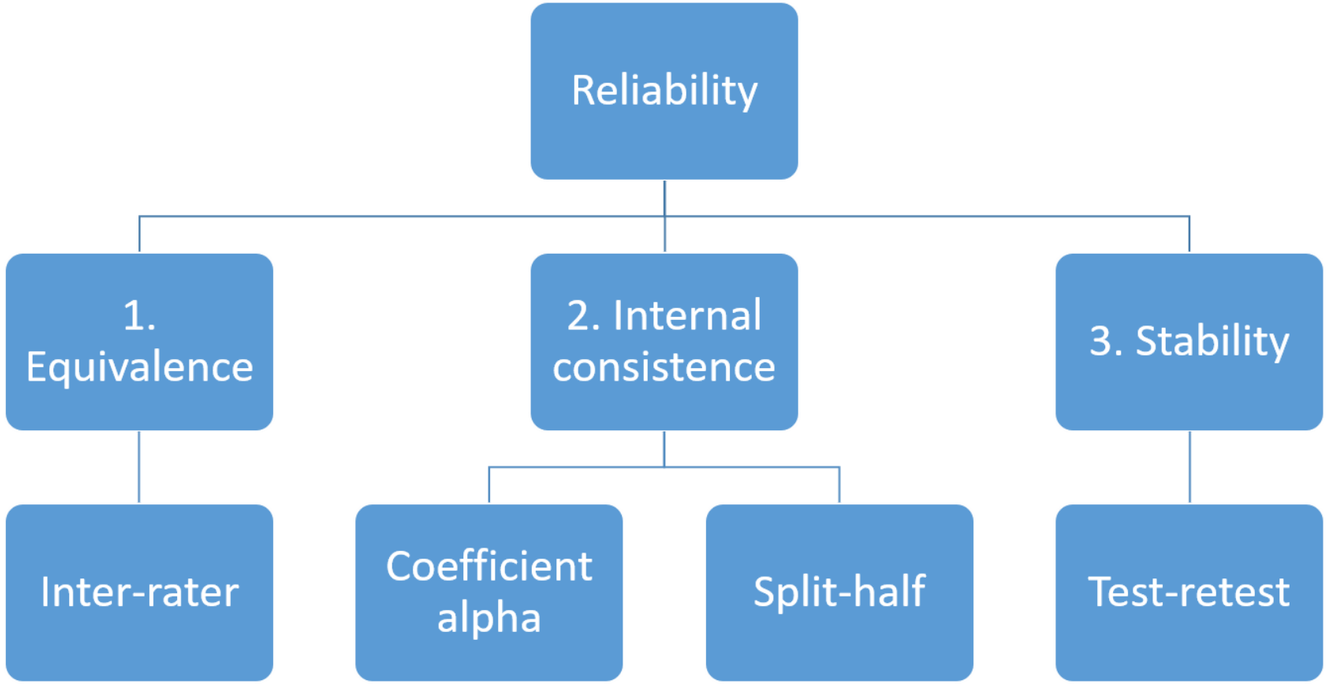
SURVEY AND MEASUREMENT DEVELOPMENT IN R

Repeated measures, replication & factor scores

SURVEY AND MEASUREMENT DEVELOPMENT IN R

George Mount
Data analytics educator

Reliability as stability over time



Test-retest reliability

- Are responses correlated from Time 1 to Time 2?
  - Repeated measures
- Correlations in scores of *the same respondents*...
- Answering at *closely spaced* points in time.

testRetest() in the psych package

```
# t1 and t2 are data frames of SAME respondents
    # at t1 and t2
survey_test_retest <- testRetest(t1 = survey_t_1,
                                 t2 = survey_t_2, id = "id")
```

testRetest() in the psych package

```
survey_test_retest$r12
```

```
0.9940203
```

- `r12` : The correlation of scaled scores across time 1 and 2

Test-retest interpretation

```
survey_test_retest$r12
```

```
0.9940203
```

Value	Interpretation
<.7	Unacceptable
.7 to .9	Good
.9	Very good

Scale validation & replication

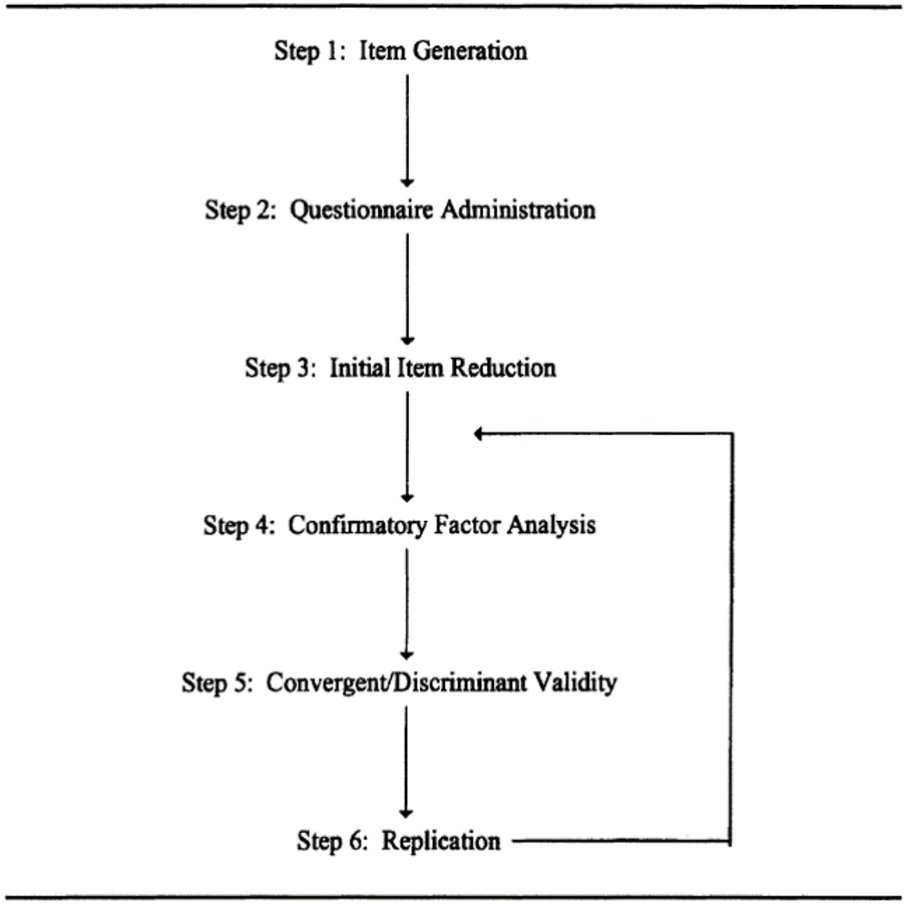


Figure 1: Scale Development Process

[1] Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. Organizational research methods, 1(1).

Splitting data

```
# Split the survey in half by rows
# Use one half for EFA and one for CFA

brand_rep_even <- brand_rep[c(TRUE,FALSE),]
brand_rep_odd <- brand_rep[c(FALSE,TRUE),]

dim(brand_rep_even)
dim(brand_rep_odd)
```

```
280   9
279   9
```

Let's practice!

SURVEY AND MEASUREMENT DEVELOPMENT IN R

Wrap-up: from generation to replication...

SURVEY AND MEASUREMENT DEVELOPMENT IN R

George Mount
Data analytics educator

Recap

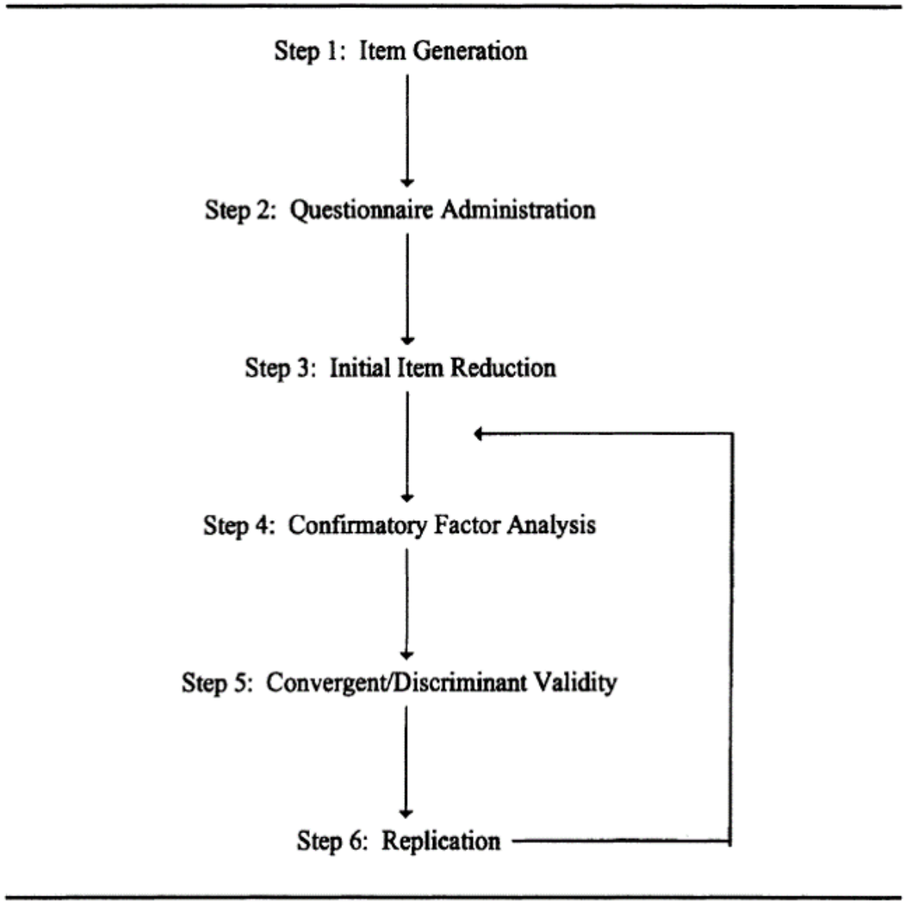
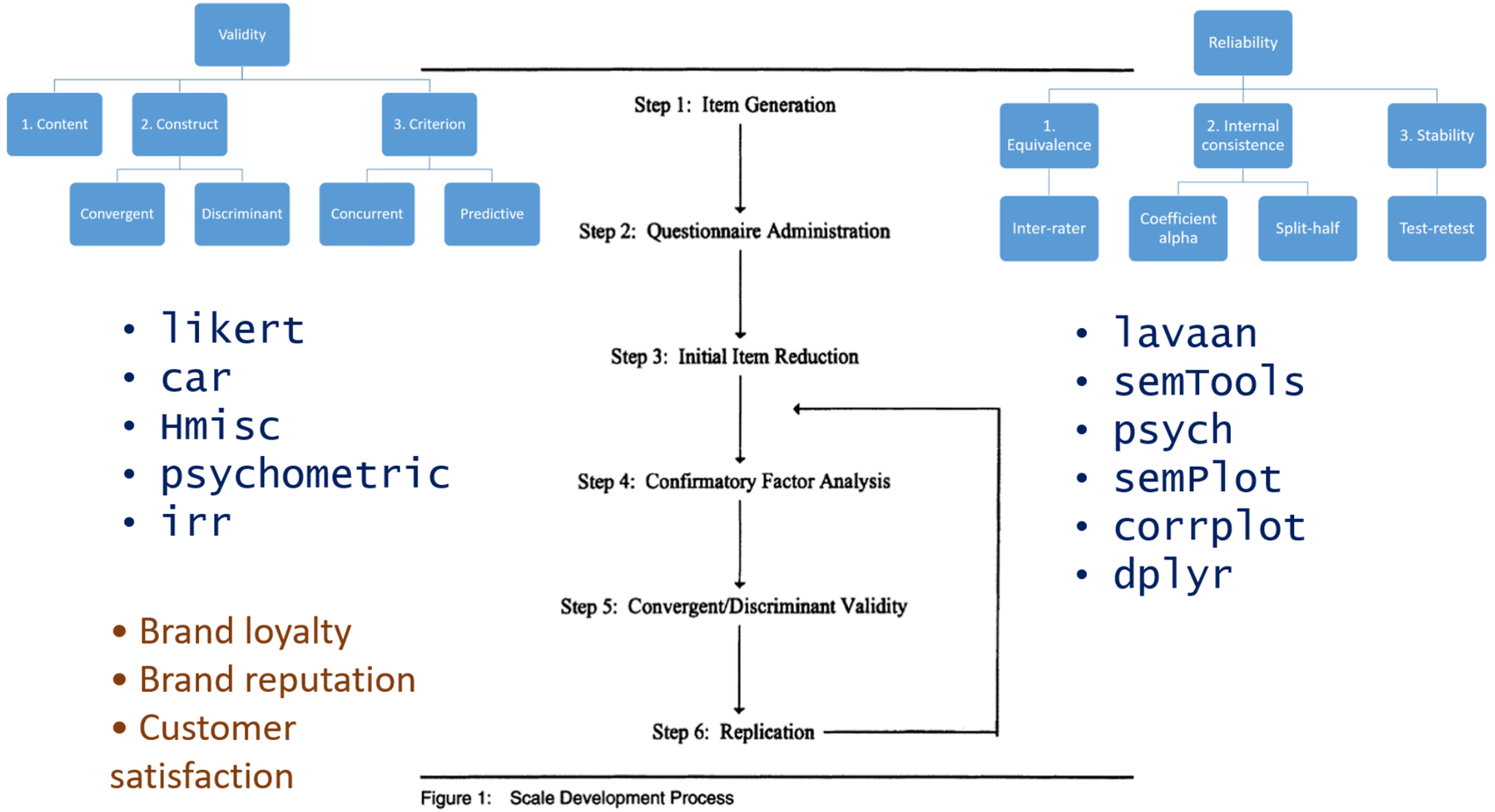


Figure 1: Scale Development Process

[1] Hinkin, Timothy R. "A brief tutorial on the development of measures for use in survey questionnaires." _Organizational research methods_ 1.1 (1998).

Recap



- `likert`
- `car`
- `Hmisc`
- `psychometric`
- `irr`

- `lavaan`
- `semTools`
- `psych`
- `semPlot`
- `corrplot`
- `dplyr`

- Brand loyalty
- Brand reputation
- Customer satisfaction

Recommendations

- Factor Analysis in R
- Structural Equation Modeling with lavaan in R
- Dimensionality Reduction in R
- Machine Learning for Marketing Analytics in R

Congratulations!

SURVEY AND MEASUREMENT DEVELOPMENT IN R