Psychology 405: Psychometric Theory
More on Correlations

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2023

Outline

Predicting scores: Question 1

1. If a person has a GRE verbal score of 680, then what would you expect his/her GRE quantitative score to be?
2. Mean GRE V = 600 SD = 80 r = .72
3. Mean GRE Q = 650 SD = 100
4. Observe GRE V = 680
5. Predicted GRE Q = ?

Predicting scores: Answer 1

1. If a person has a GRE verbal score of 680, then what would you expect his/her GRE quantitative score to be?
2. Mean GRE V = 600 SD = 80 r = .72
3. Mean GRE Q = 650 SD = 100
4. Observe GRE V = 680
5. z GRE V = (680- 600)/80 = 1.0
6. predicted z GRE Q = rxy zx = .72 * (1) =.72
7. predicted GRE Q = .72 * 100 + 650 = 722

Predicting scores: Question 2

1. If a person has a GRE Quant score of 722, then what would you expect his/her GRE Verbal score to be?
2. Mean GRE V = 600 SD = 80 r = .72
3. Mean GRE Q = 650 SD = 100
4. Observe GRE Q = 722
5. Predicted GRE V = ?

Predicting scores: Answer 2

1. If a person has a GRE Quant score of 722, then what would you expect his/her GRE Verbal score to be?
2. Mean GRE V = 600 SD = 80 r = .72
3. Mean GRE Q = 650 SD = 100
4. Observed GRE Q = 722
5. Predicted GRE V = ?
6. $z_{GRE_Q} = (722 - 650)/100 = .72$
7. predicted $z_{GRE_V} = r_{xy} z_x = .72 * (.72) = .52$
8. predicted GRE Q = .52 * 80 + 600 = 642
9. Note that although 680 predicts 722, 722 predicts 642.

Predicting Scores: Question 3

1. For a person with an anxiety score of 16, what is the expected GPA?
2. Anxiety Mean = 12 sd = 4 r = -.39
3. GPA Mean = 3.0 sd = .5

Predicting Scores: Answer 3

1. For a person with an anxiety score of 16, what is the expected GPA?
2. Anxiety Mean = 12 sd = 4 r = -.39
3. GPA Mean = 3.0 sd = .5
4. z anx = (16- 12)/4 = 1.0
5. predicted z gpa = $r_{xy} z_x = -.39 * (1) = -.39$
6. predicted gpa = -.39 * .5 + 3 = 2.805

Partial Correlations: Question

1. Assuming the classical model of partial correlations, what is the correlation between GRE Quantitative and GPA with GRE Verbal held constant?
2. $r_{GREQ,GPA} = .34$
3. $r_{GREQ,V} = .72$ $r_{GREV,GPA} = .38$
4. We want to find the covariance of Q and GPA without V.
5. All correlations are $= \frac{C_{xy}}{\sqrt{V_x V_y}}$. So we just need to find the Covariances and Variances.

Partial Correlations: Answer

1. Assuming the classical model of partial correlations, what is the correlation between GRE Quantitative and GPA with GRE Verbal held constant?
2. $r_{GREQ,GPA} = .34$
3. $r_{GREQ,V} = .72$ $r_{GREV,GPA} = .38$
4. All correlations are $= \frac{C_{xy}}{\sqrt{V_x V_y}}$. So we just need to find the Covariances and Variances.
5. partial $r_{xy.z} = \frac{r_{xy} - r_{xz} * r_{yz}}{\sqrt{(1 - r_{xz}^2) * (1 - r_{yz}^2)}}$
6. $r_{qgpa.v} = \frac{(.34 - .72 * .38)}{\sqrt{(.482 * .856)}} = .103$ partial
7. part r = $\frac{r_{xy} - r_{xz} * r_{yz}}{\sqrt{1 - r_{xz}^2}} = .096$ part

Multiple Correlation: Question

1. What is the multiple correlation of GRE V and GRE Q with rated quality of the MA?
2. $r_{GREV,MA} = .32 \quad r_{GREQ,MA} = .29 r_{GREV,Q} = .72$
3. $\beta_{y.x} = \frac{r_{xy} - r_{xz} * r_{yz}}{1 - r_{xz}^2}$
4. $R^2 = \Sigma \beta_i r_{x_i y}$

Multiple Correlation: Answer

1. What is the multiple correlation of GRE V and GRE Q with rated quality of the MA?
2. r GRE V, MA = .32 r GRE Q, MA = .29 r GRE V,Q =.72
3. $beta_{y.x} = \frac{r_{xy} - r_{xz} * r_{yz}}{1 - r_{xz}^2}$
4. $\beta_{GREV,MA} = (.32 - .72 * .29)/(1 - .72^2) = .231$
5. $\beta_{GREQ,MA} = (.29 - .72 * .32)/(1 - .72^2) = .124$
6. $R^2 = \beta_{y.x} * r_{xy} + \beta_{y.z} * r_{yz} \ldots$
7. R2 = beta GRE Q, MA * r GRE Q, MA + beta GRE V, MA* r GRE V, MA =
8. R2 = .124 * .29 + .231 * .32 = .108
9. R = .329

Unit Weighted Multiple R

1. What is the unit weighted correlation of GREV and GRE Q with MA?
2.

Variable	GREV	GREQ	MA
GREV	1.00	0.72	0.32
GREQ	0.72	1.00	0.29
MA	0.32	0.29	1.00

3. All correlations are $= \frac{C_{xy}}{\sqrt{V_x V_y}}$. So we just need to find the Covariances and Variances.

Unit Weighted Multiple R

1. Weight the two predictors equally
2.

Variable	GREV	GREQ	MA
GREV	1.00	0.72	0.32
GREQ	0.72	1.00	0.29
MA	0.32	0.29	1.00

3. All correlations are $= \frac{C_{xy}}{\sqrt{V_x V_y}}$. So we just need to find the Covariances and Variances.
4. $C_{v+q,MA} = .32 + .29 = .61$
5. $V_{v+q} = 1.00 + .72 + .72 + 1.00 = 3.44$
6. $r_{v+q,MA} = \frac{.61}{\sqrt{3.44 * 1}} = .329$

Correlating two composites – unit weights

Table: Ability and Performance

Hypothetical relationships

Variable	GREV	GREQ	GREA	GPA	Pre	MA
GREV	1.00	0.72	0.54	0.38	0.32	0.27
GREQ	0.72	1.00	0.48	0.34	0.29	0.24
GREA	0.54	0.48	1.00	0.55	0.47	0.39
GPA	0.38	0.34	0.55	1.00	0.42	0.35
Pre	0.32	0.29	0.47	0.42	1.00	0.30
MA	0.27	0.24	0.39	0.35	0.30	1.00

1. What is the unit weighted correlation between the ability measures and the performance measures?
2. All correlations are $= \frac{C_{xy}}{\sqrt{V_x V_y}}$. So we just need to find the Covariances and Variances.

Correlating two composites – unit weights

Table: Ability and Performance

Hypothetical relationships

Variable	GREV	GREQ	GREA	GPA	Pre	MA
GREV	1.00	0.72	0.54	0.38	0.32	0.27
GREQ	0.72	1.00	0.48	0.34	0.29	0.24
GREA	0.54	0.48	1.00	0.55	0.47	0.39
GPA	0.38	0.34	0.55	1.00	0.42	0.35
Pre	0.32	0.29	0.47	0.42	1.00	0.30
MA	0.27	0.24	0.39	0.35	0.30	1.00

set

1. $V_{ability} = 1.0 + .72 + .54 + .72 + 1.0 + .48 + .54 + .48 + 1 = 3 * 1 + 2 * (.72 + .54 + .48) = 6.48$
2. $V_{performance} = 3 * 1 + 2 * (.42 + .35 + .30) = 5.14$
3. $C_{ability,performance} = .38 + .34 + .55 + .32 + .29 + .47 + .27 + .24 + .39 = 3.25$
4. $R_{ability,performance} = \frac{3.25}{\sqrt{6.48 * 5.14}} = .563$

R code

```
R <- structure(list(GREV = c(1, 0.72, 0.54, 0.38, 0.32, 0.27),
GREQ = c(0.72, 1, 0.48, 0.34, 0.29, 0.24),
GREA = c(0.54, 0.48, 1, 0.55, 0.47, 0.39),
GPA = c(0.38, 0.34, 0.55, 1, 0.42, 0.35),
 Pre = c(0.32, 0.29, 0.47, 0.42, 1, 0.3),
 MA = c(0.27, 0.24, 0.39, 0.35, 0.3,
1)), row.names = c("GREV", "GREQ", "GREA", "GPA", "Pre", "MA"
), class = "data.frame")
Rx <- R[1:3,1:3]
Ry <- R[4:6,4:6]
Rxy <- R[1:3,4:6]
sum(Rx); sum(Ry); sum(Rxy)
 sum(Rxy)/sqrt(sum(Rx)* sum(Ry))
```

```
R
     GREV GREQ GREA  GPA  Pre   MA
GREV 1.00 0.72 0.54 0.38 0.32 0.27
GREQ 0.72 1.00 0.48 0.34 0.29 0.24
GREA 0.54 0.48 1.00 0.55 0.47 0.39
GPA  0.38 0.34 0.55 1.00 0.42 0.35
Pre  0.32 0.29 0.47 0.42 1.00 0.30
MA   0.27 0.24 0.39 0.35 0.30 1.00
sum(Rx)
[1] 6.48
> sum(Ry)
[1] 5.14
> sum(Rxy)
[1] 3.25
 sum(Rxy)/sqrt(sum(Rx)* sum(Ry))   =  0.56
```

Instability of beta weights

An example from Niels Waller

R code

```
        Y    X1     X2    X3     C
Y   1.000 0.561  0.568 0.520  0.35
X1  0.561 1.000  0.720 0.526  0.40
X2  0.568 0.720  1.000 0.554 -0.25
X3  0.520 0.526  0.554 1.000  0.30
C   0.350 0.400 -0.250 0.300  1.00

 lmCor(Y ~ X1 + X2 + X3, data=niels)
```

```
lmCor(Y ~ X1 + X2 + X3, data=niels)
Call: lmCor(y = Y ~ X1 + X2 + X3, data = niels)

Multiple Regression from matrix input
 DV =  Y
   slope  VIF Vy.x
X1  0.25 2.18 0.14
X2  0.25 2.28 0.14
X3  0.25 1.52 0.13

 Multiple Regression
     R   R2  Ruw R2uw
Y 0.64 0.41 0.64 0.41
```

```
lmCor(Y ~ X1 + X2 + X3+C, data=niels)
Call: lmCor(y = Y ~ X1 + X2 + X3 + C, data = nie

Multiple Regression from matrix input
 DV =  Y
   slope   VIF  Vy.x
X1 -1.99 12.47 -1.12
X2  2.87 16.34  1.63
X3 -0.64  3.14 -0.33
C   2.06  8.66  0.72

 Multiple Regression
     R  R2  Ruw R2uw
Y 0.95 0.9 0.69 0.47
```

Cohen's Set correlation versus unit weighted correlations

1. What is the relationship between two sets of variables? Two alternative answers.
2. Set correlation: 1 - the ratio of determinants
   - $R = 1 - \frac{||\mathbf{R}_{xy}||}{||\mathbf{R}_x||||\mathbf{R}_y||}$
3. Unit weighted correlation
   - $R_{uw} = \frac{\mathbf{1R}_{yx}\mathbf{1}'}{(\mathbf{1R}_{yy}\mathbf{1}')^{.5}(\mathbf{1R}_{xx}\mathbf{1}')^{.5}}$
4. Both are found using the `lmCor` function

lmCor to find multiple correlations from correlation matrices

```
R
mod1 <- lmtCor(4:6,1:3,data=R)
 summary(mod1)


     GREV GREQ GREA  GPA  Pre   MA
GREV 1.00 0.72 0.54 0.38 0.32 0.27
GREQ 0.72 1.00 0.48 0.34 0.29 0.24
GREA 0.54 0.48 1.00 0.55 0.47 0.39
GPA  0.38 0.34 0.55 1.00 0.42 0.35
Pre  0.32 0.29 0.47 0.42 1.00 0.30
MA   0.27 0.24 0.39 0.35 0.30 1.00
Multiple Regression from matrix input
setCor(y = 4:6, x = 1:3, data = R)

Multiple Regression from matrix input

Beta weights
       GPA   Pre    MA
GREV 0.087 0.064 0.064
GREQ 0.047 0.045 0.030
GREA 0.481 0.414 0.341

Multiple R
 GPA  Pre   MA
0.56 0.48 0.40

Multiple R2
 GPA  Pre   MA
0.31 0.23 0.16

Cohen's set correlation R2
[1] 0.41
```

```
Unweighted multiple R
 GPA  Pre   MA
0.50 0.42 0.35
 Unweighted multiple R2
 GPA  Pre   MA
0.25 0.18 0.13

Various estimates of between set correlations
Squared Canonical Correlations
[1] 4.1e-01 6.0e-05 1.6e-06

 Average squared canonical correlation =  0.14
 Cohen's Set Correlation R2 =  0.41
Unweighted correlation between the two sets =  0
```

Just use Verbal and Quant

```
R
mod2 <- lmCor(4:6,1:2,R)
```

```
     GREV GREQ GREA  GPA  Pre   MA
GREV 1.00 0.72 0.54 0.38 0.32 0.27
GREQ 0.72 1.00 0.48 0.34 0.29 0.24
GREA 0.54 0.48 1.00 0.55 0.47 0.39
GPA  0.38 0.34 0.55 1.00 0.42 0.35
Pre  0.32 0.29 0.47 0.42 1.00 0.30
MA   0.27 0.24 0.39 0.35 0.30 1.00
```

```
 summary(mod2)

Multiple Regression from matrix input
lmCor(y = 4:6, x = 1:2, data = R)

Multiple Regression from matrix input

Beta weights
      GPA  Pre    MA
GREV 0.28 0.23 0.202
GREQ 0.14 0.12 0.095

Multiple R
 GPA  Pre   MA
0.39 0.33 0.28

Multiple R2
  GPA   Pre    MA
0.154 0.110 0.077
```

```
 Unweighted multiple R
 GPA  Pre   MA
0.39 0.33 0.27
 Unweighted multiple R2
 GPA  Pre   MA
0.15 0.11 0.08

Various estimates of between set correlations
Squared Canonical Correlations
[1] 2.0e-01 4.2e-05

 Average squared canonical correlation =  0.1
 Cohen's Set Correlation R2 =  0.2
Unweighted correlation between the two sets =  0
```

Comorbidity

1. Symptoms are said to be comorbid if one has both symptoms.
    - This is really just one cell in a 2 x 2 table
    - We need base rates as well
2. Consider Anxiety and Depression
    - 50 % of anxiety patients are also depressed
    - 67% of depressed patients are also anxious
    - base rates are 20% for anxiety, 15% for depression

R code

```
comorbidity(.2,.15,.1,c("Anxiety","Depression"))
```

```
Call: comorbidity(d1 = 0.2, d2 = 0.15, com = 0.1, labels = c("Anxiety",
    "Depression"))
Comorbidity table
            Anxiety -Anxiety
Depression      0.1     0.05
-Depression     0.1     0.75

implies phi =  0.49  with Yule =  0.87  and tetrachoric correlation of  0.75
and normal thresholds of  1.04 0.84
```

Can also treat this as an Area Under the Curve – signal detection problem

R code

```
com <- comorbidity(.2,.15,.1,c("Anxiety","Depression"))  #find the co
AUC(com$twobytwo)   #convert to a Signal Detection Approach
```

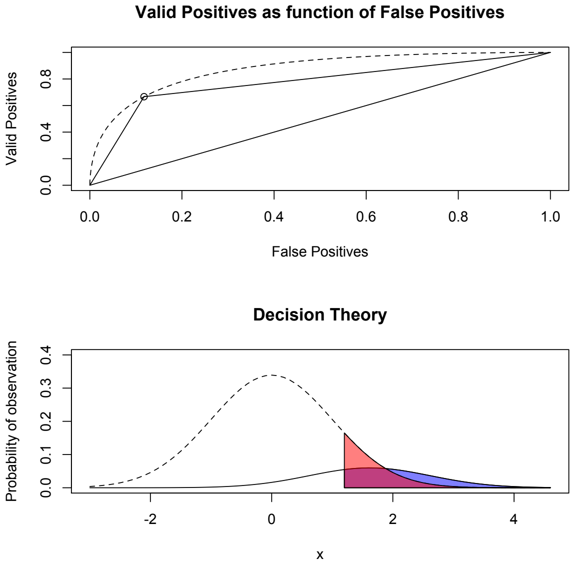
```
AUC(com$twobytwo)   #convert to a Signal Detection Approach
Decision Theory and Area under the Curve

The original data implied the following 2 x 2 table
         Predicted.Pos Predicted.Neg
True.Pos           0.1          0.05
True.Neg           0.1          0.75

Conditional probabilities of
         Predicted.Pos Predicted.Neg
True.Pos          0.67          0.33
True.Neg          0.12          0.88

Accuracy =  0.85  Sensitivity =  0.67   Specificity =  0.88
with Area Under the Curve =  0.87
d.prime =  1.62  Criterion =  1.19  Beta =  0.13
Observed Phi correlation =  0.49
Inferred latent (tetrachoric) correlation  =  0.75
```

Show the Signal Detection graphic



Three ways to compare categorical means

R code

```
dim(spi)
t.test(wellness ~ sex, data=spi)
t.test(education ~ sex, data=spi)
```

```
t.test(wellness ~ sex, data=spi)

        Welch Two Sample t-test

data:  wellness by sex
t = -6.021, df = 2872.4, p-value = 1.954e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.14075230 -0.07159818
sample estimates:
mean in group 1 mean in group 2
       1.479228        1.585404

> t.test(education ~ sex, data=spi)

        Welch Two Sample t-test

data:  education by sex
t = 4.5873, df = 2908.4, p-value = 4.68e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2040981 0.5088266
sample estimates:
mean in group 1 mean in group 2
       4.314540        3.958078
```

Try regression

R code

```
lm (wellness ~ sex, data=spi)
lm (education ~ sex, data=spi)
```

```
lm (wellness ~ sex, data=spi)

Call:
lm(formula = wellness ~ sex, data = spi)

Coefficients:
(Intercept)          sex
     1.3731       0.1062

> lm (education ~ sex, data=spi)

Call:
lm(formula = education ~ sex, data = spi)

Coefficients:
(Intercept)          sex
     4.6710      -0.3565
```

Regression with significance tests

R code

```
summary(lm (wellness ~ sex, data=spi))
```

```
 summary(lm (wellness ~ sex, data=spi))

Call:
lm(formula = wellness ~ sex, data = spi)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.5854 -0.4792  0.4146  0.4146  0.5208 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.37305    0.02926  46.925  < 2e-16 ***
sex          0.10618    0.01759   6.036 1.76e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4957 on 3278 degrees of freedom
  (720 observations deleted due to missingness)
Multiple R-squared:  0.01099,	Adjusted R-squared:  0.01069 
F-statistic: 36.43 on 1 and 3278 DF,  p-value: 1.756e-09
```

raw units

R code

```
spi.complete <- complete.cases(spi[c("wellness","sex")])
lmCor(wellness ~ sex, data=spi[spi.complete,],std=FALSE)
Call: lmCor(y = wellness ~ sex, data = spi[spi.complete, ], std = FALS
```

```
spi.complete <- complete.cases(spi[c("wellness","sex")])
Call: lmCor(y = wellness ~ sex, data = spi[spi.complete, ], std = FALSE)

Multiple Regression from raw data

 DV =  wellness
            slope   se     t       p lower.ci upper.ci VIF Vy.x
(Intercept)  1.37 0.03 46.93 0.0e+00     1.32     1.43   1 0.00
sex          0.11 0.02  6.04 1.8e-09     0.07     0.14   1 0.01

Residual Standard Error =  0.5  with  3278  degrees of freedom

 Multiple Regression
           R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1  df2        p
wellness 0.1 0.01 0.07 0.01        0.01        0     36.43   1 3278 1.76e-09
```

Do the lmCor operation with default values – standardized units

R code

```
lmCor(wellness ~ sex, data=spi[spi.complete,])
```

```
 setCor(wellness ~ sex, data=spi[spi.complete,])
Call: setCor(y = wellness ~ sex, data = spi[spi.complete, ])

Multiple Regression from raw data

 DV =  wellness
            slope   se    t       p lower.ci upper.ci VIF
(Intercept)   0.0 0.02 0.00 1.0e+00    -0.03     0.03   1
sex           0.1 0.02 6.04 1.8e-09     0.07     0.14   1

Residual Standard Error =  0.99  with  3278  degrees of freedom

 Multiple Regression
           R   R2 Ruw R2uw Shrunken R2 SE of R2 overall F df1  df2        p
wellness 0.1 0.01 0.1 0.01        0.01        0     36.43   1 3278 1.76e-09
```

Or, just show the correlations

R code

```
lowerCor(spi[cs(sex,wellness,education)])
corr.test(spi[cs(sex,wellness,education)])
```

```
lowerCor(spi[cs(sex,wellness,education)])
          sex   wllns edctn
sex        1.00
wellness   0.10  1.00
education -0.08  0.00  1.00

 corr.test(spi[cs(sex,wellness,education)])
Call:corr.test(x = spi[cs(sex, wellness, education)])
Correlation matrix
            sex wellness education
sex        1.00      0.1     -0.08
wellness   0.10      1.0      0.00
education -0.08      0.0      1.00
Sample Size
           sex wellness education
sex       3946     3280      3304
wellness  3280     3311      3000
education 3304     3000      3330
Probability values (Entries above the diagonal are adjusted for multiple tests.)
          sex wellness education
sex         0     0.00      0.00
wellness    0     0.00      0.92
education   0     0.92      0.00

 To see confidence intervals of the correlations, print with the short=FALSE option
```