Psychology 205: Research Methods in Psychology
Simulation results

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

May, 2021

Outline

Two Experimenters

1. Graduate students 1 and 2 are both interested in studying the same phenomenon. They are both equally talented, but differ in other ways
2. Graduate student 1 is very conscientious and runs all subjects at the same time of day, in the same room, and uses a block randomization schedule for counterbalancing. She always dresses and acts professionally. Any time there is an equipment failure, that subject is replaced with the next randomly assigned subject. She runs 100 subjects recruited from Psych 110 in a 2 x 2 design with 25 cases per cell.
3. Graduate student 2 is much more careless. He runs subjects at all times of day, sometimes is dressed in a coat and tie, sometimes in blue jeans. He recruits subjects from the Reader and the Daily. He randomly assigns subjects to condition and replaces any subject lost due to equipment failure with the next available participant. He runs 100 subjects in a 2 x 2 design with cell sizes ranging from 22 to 28.

Conscientious versus sloppy: comparing results

1. Which graduate student is more likely to find a significant result?
2. The results of which graduate student are more likely to generalize to other conditions and subjects?
3. What basic principles are shown in this example?

Control variability through design or randomization

1. Precision vs. generalizability
2. By controlling for more variables, we reduce the error (unexplained) variance.
3. But we limit the degree to which we can generalize.

Experimental design to control for sources of variance

1. By fixing other variables to constant values
2. By fixed Expected score of other variables to be the same across conditions (by randomizattion)

Simulation of theory

1. In real research, one never knows what is correct answer
2. Simulations are usually done to test model against data.
3. Simulations are are used to test how well a new method performs with artificial data.
4. Our simulation was true to basic model of performance with one powerful addition
5. This simulation was "truth", and the experiments were to try to find "truth"

A "meta analysis" of reults

1. Reviews used to be just written summaries of what the review thought was there (reflecting the reviewer's biases).
2. More recent procedures treat the studies as if they were the data.
3. When the literature consists of multiple studies examining similar hypotheses, we can do a "meta-analysis" of the results.
4. Each individual study reflects true effects + error.
5. The expectation is that if we average over studies, the errors, being random, will average out.
6. Naive summary is just a "box score" of results
7. Modern meta-analytic techniques average the effect sizes across studies.

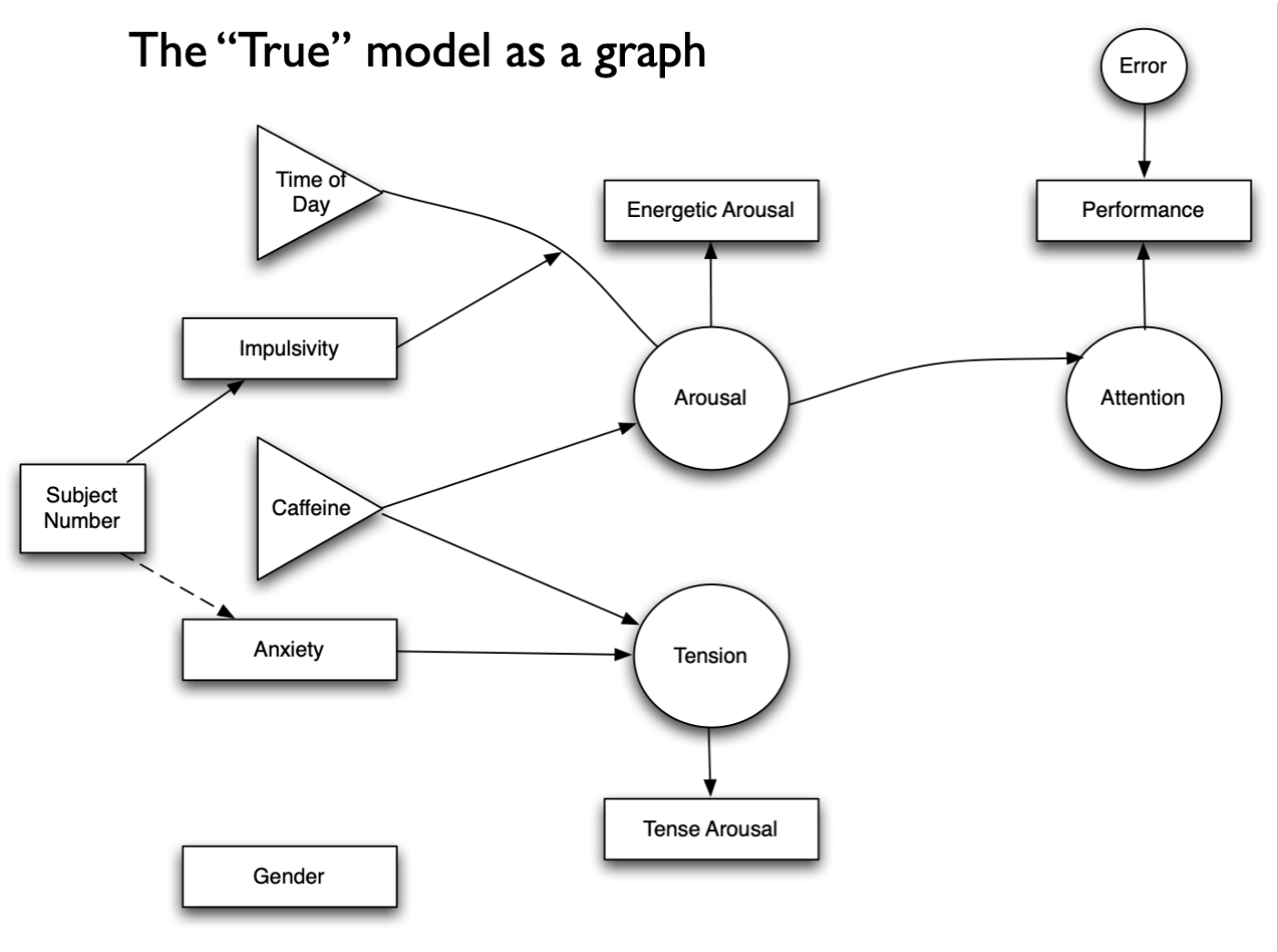
A box score from prior years

	Main effect	Interaction with			
		Drug	Impulsivity	Anxiety	Sex
Time	7/7	1/3	1/1	1/3	0/2
Drug	17/17	-	1/5	1/5	2/12
Impulsivity	6/6	-	-	0/0	1/5
Anxiety	1/4	-	-	-	1/1
Sex	2/13	-	-	-	-

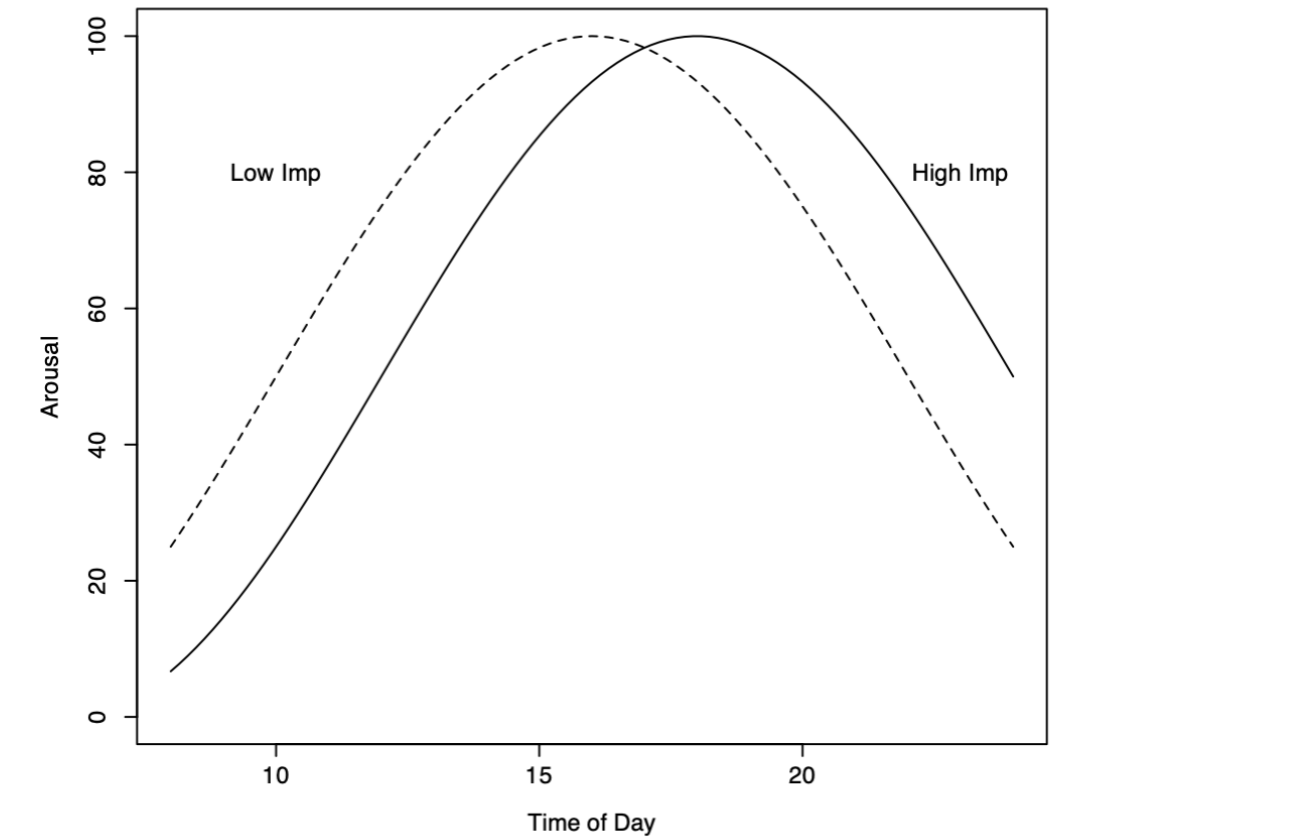
The True model

1. Arousal was cosine function of time of day + impulsivity + caffeine
2. Impulsivity tended to increase by subject number.
3. The assumption was that subjects were more bored as the term increases.
4. Tension was a function of anxiety and caffeine and was decreased by subject number.
5. Performance was a logistic function of arousal.
6. Arousal and Tension as latent variables affected energetic arousal and tense arousal (observed variables)

The true model

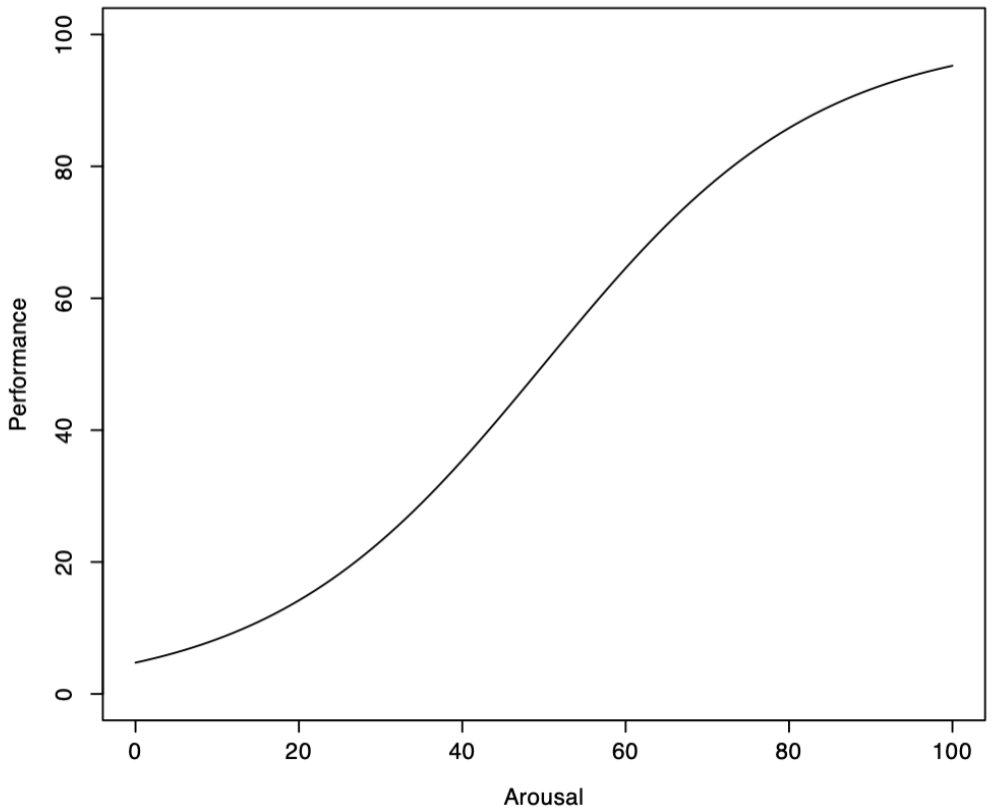


Impulsivity, Time of Day and Arousal

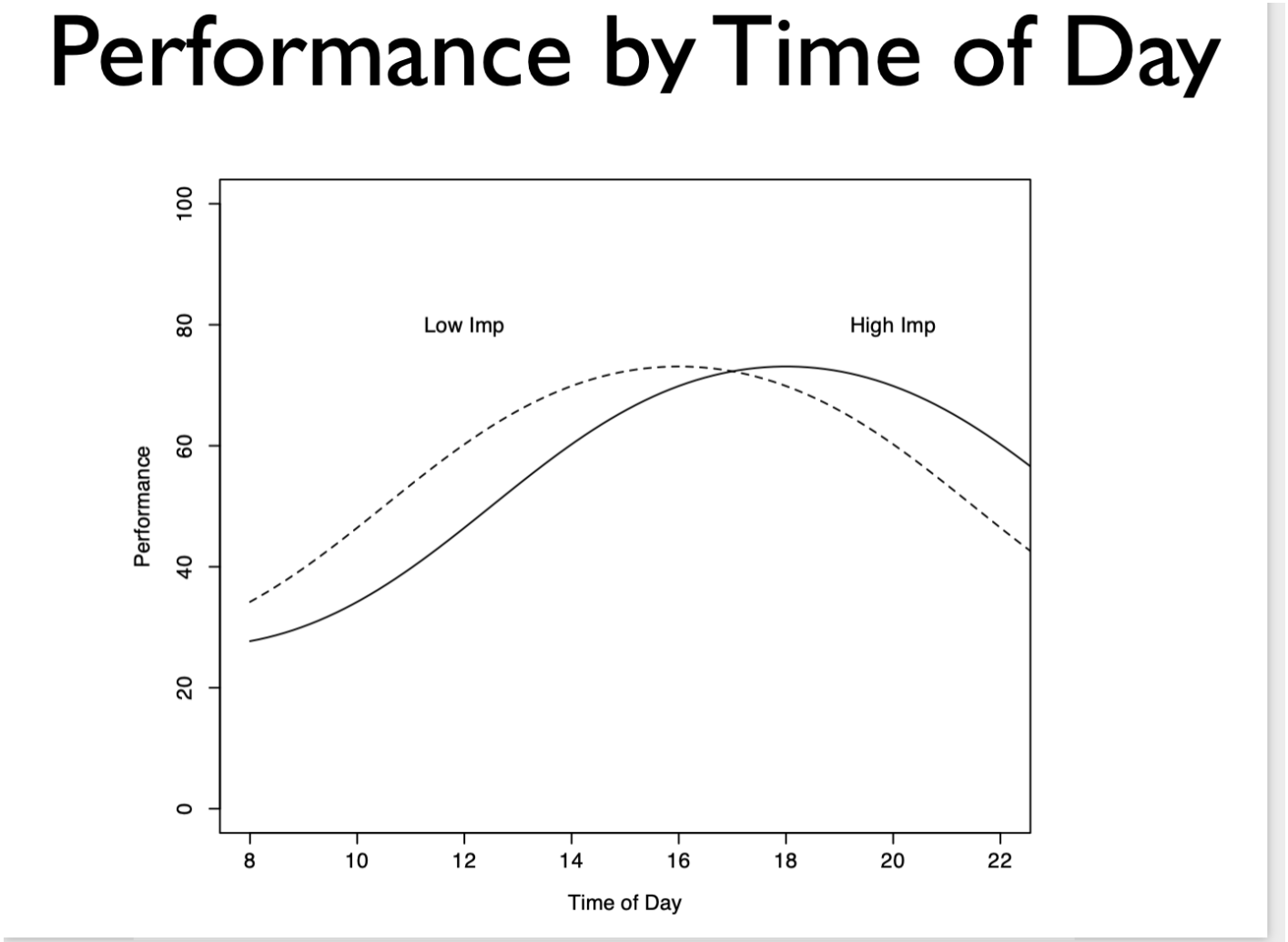


Arousal and performancel

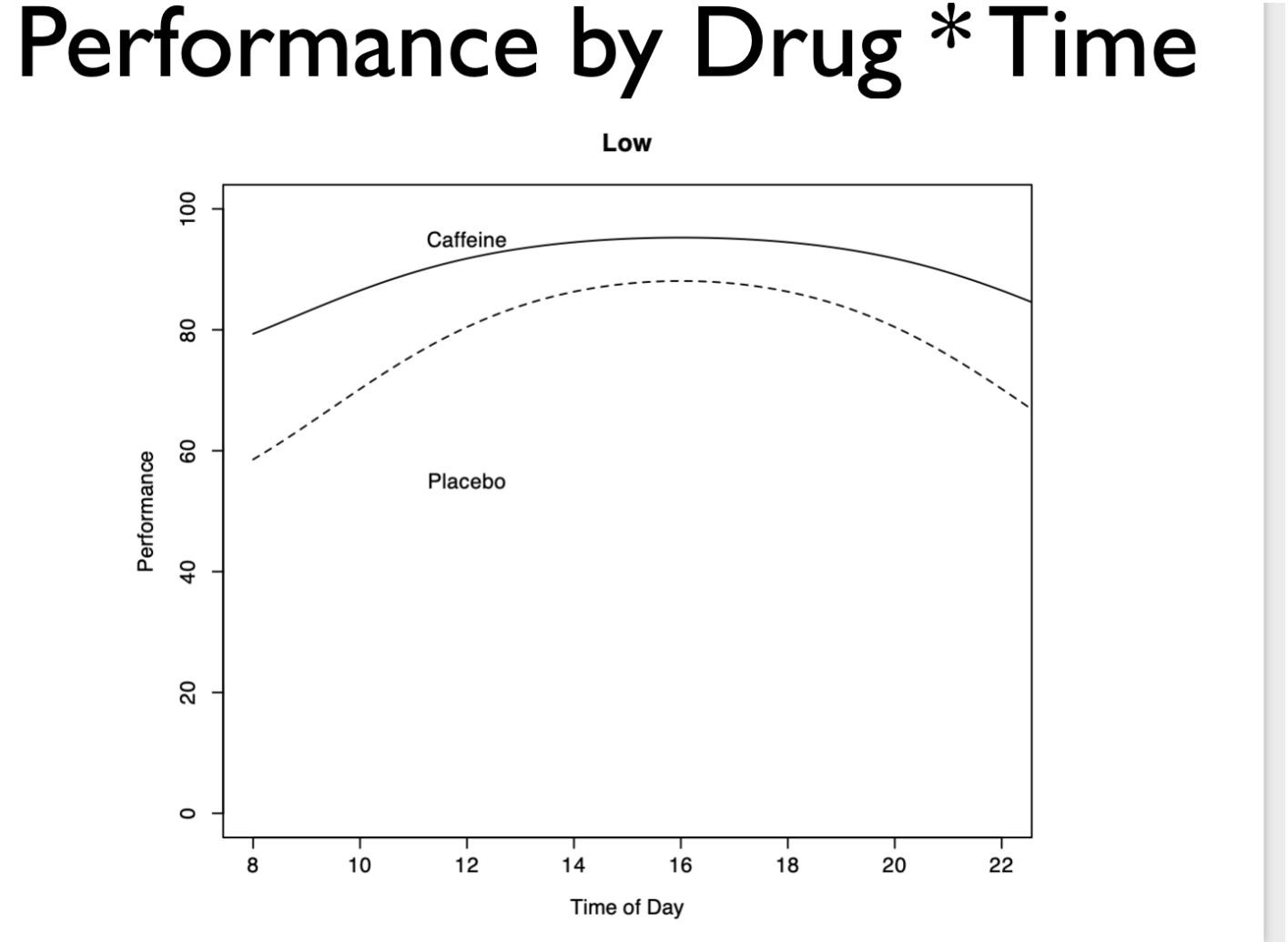
Performance=f(Arousal)



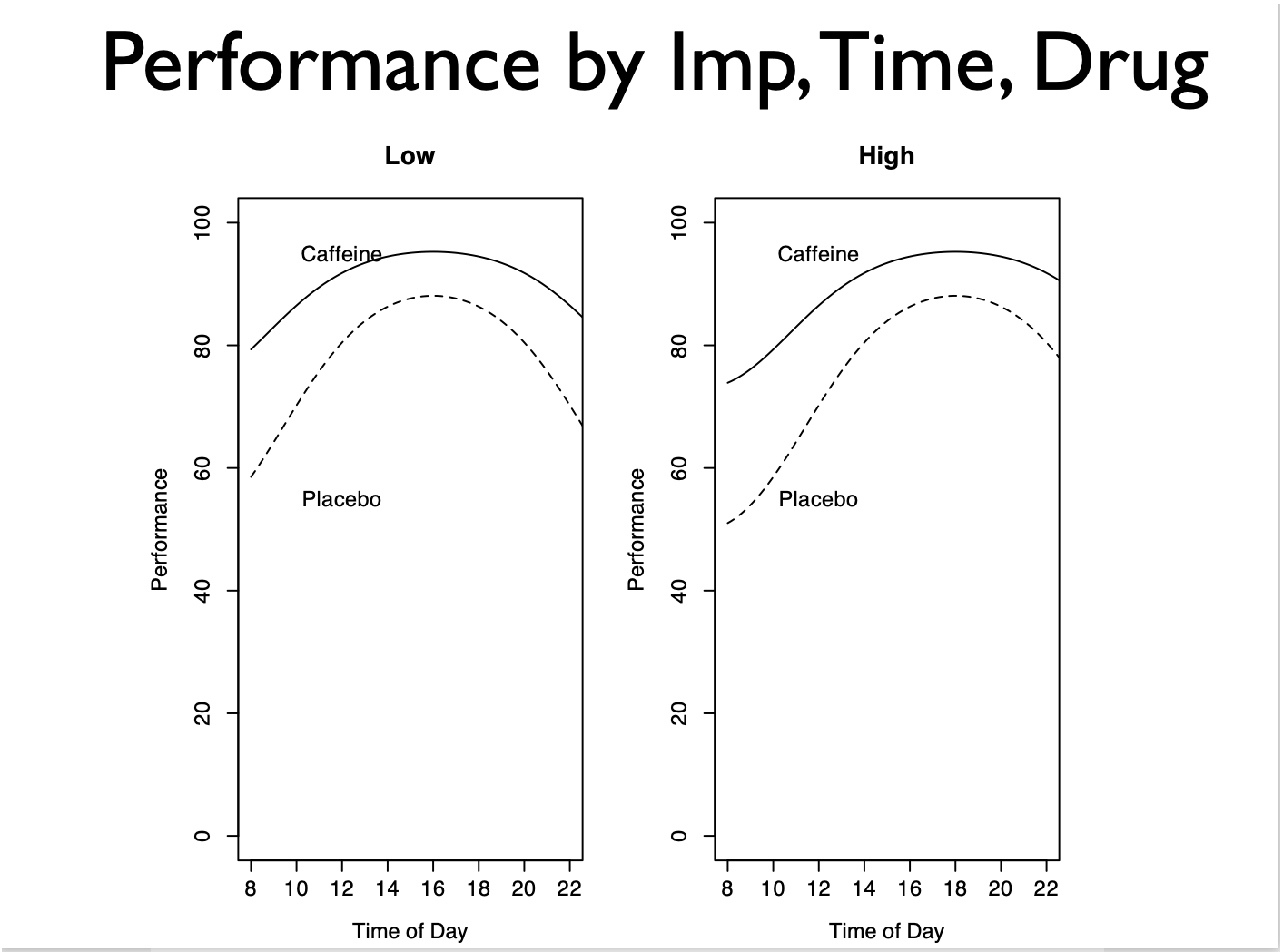
Performance by Impulsivity * time of day



Performance by Drug x Time



Performance by Impulsivity * Drug * Time



Simulation "Truth"

```
diurnal <- cos(2*pi*((time-impulsivity)-16)/24)
arousal <- 1+diurnal+caffeine -ID+z_normal(0, 1)
 tense <- anxiety+caffeine- ID +  z_normal(0, 1)
   performance <-  1/ (1 + exp(-arousal))
```

Lets try to get some clean results

Repeated the experiment 5 times with 200 subjects per study

R code

```
sim.data <- read.clipboard()
describe(sim.data)
```

```
describe(sim.data)
             vars    n   mean    sd median trimmed   mad min max range  skew kurtosis   se
snum            1 1000 100.50 57.76  100.5  100.50 74.13   1 200   199  0.00    -1.20 1.83
sex             2 1000   1.51  0.50    2.0    1.52  0.00   1   2     1 -0.06    -2.00 0.02
drug            3 1000   0.52  0.50    1.0    0.52  0.00   0   1     1 -0.06    -2.00 0.02
time            4 1000  15.05  4.36   15.0   15.06  5.93   8  22    14 -0.04    -1.20 0.14
anxiety         5 1000   4.05  2.02    4.0    4.05  1.48  -1  10    11  0.04    -0.39 0.06
impulsivity     6 1000   5.03  1.97    5.0    5.05  1.48   0  10    10 -0.09    -0.25 0.06
arousal         7 1000  61.88 26.19   64.0   63.55 25.20   0  99    99 -0.39    -0.58 0.83
tension         8 1000  24.12 13.87   22.0   22.92 14.83   2  70    68  0.68    -0.26 0.44
performance     9 1000  64.07  8.65   65.0   64.45  8.90  35  86    51 -0.45     0.01 0.27
```

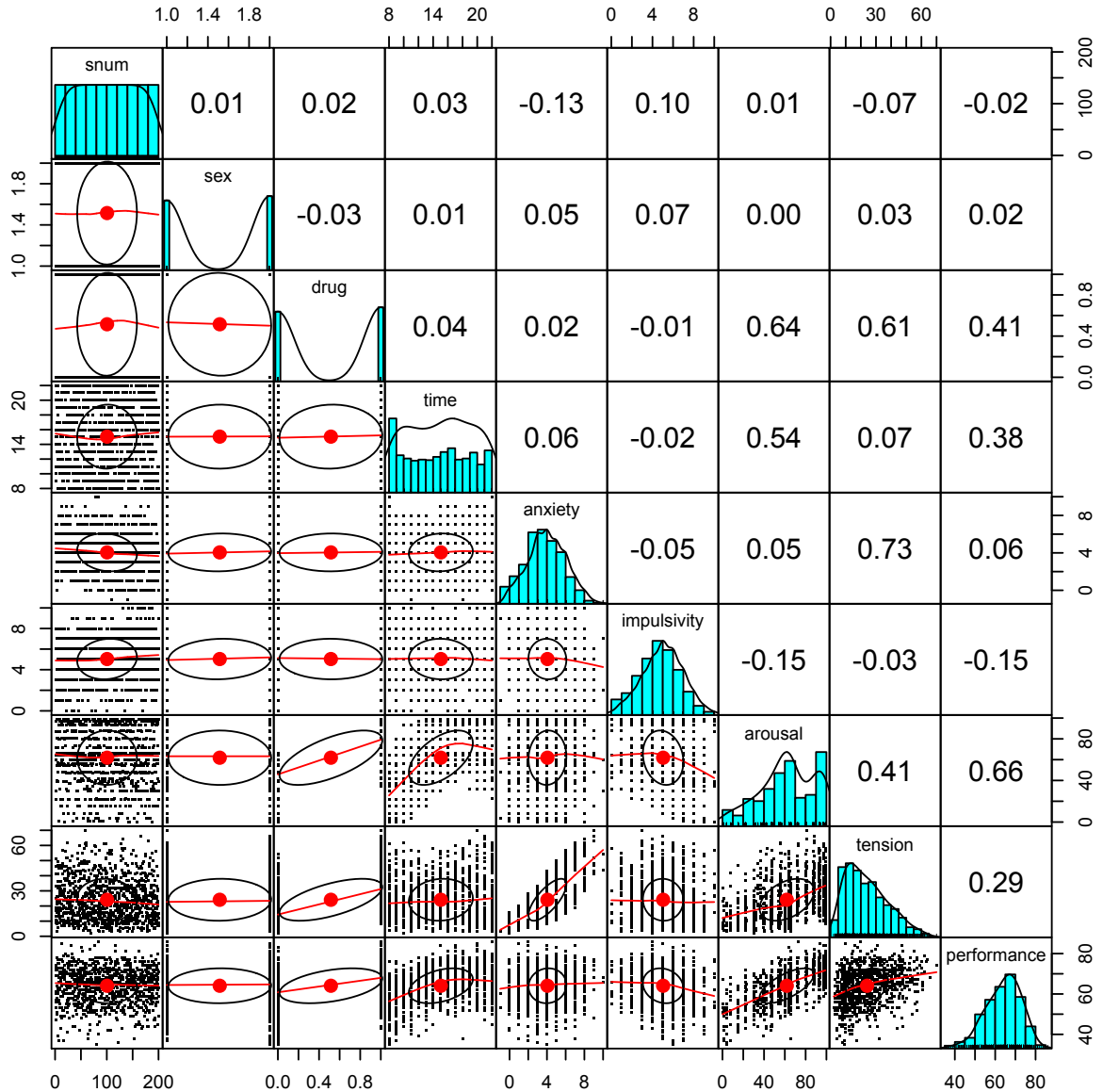
Correlations

R code

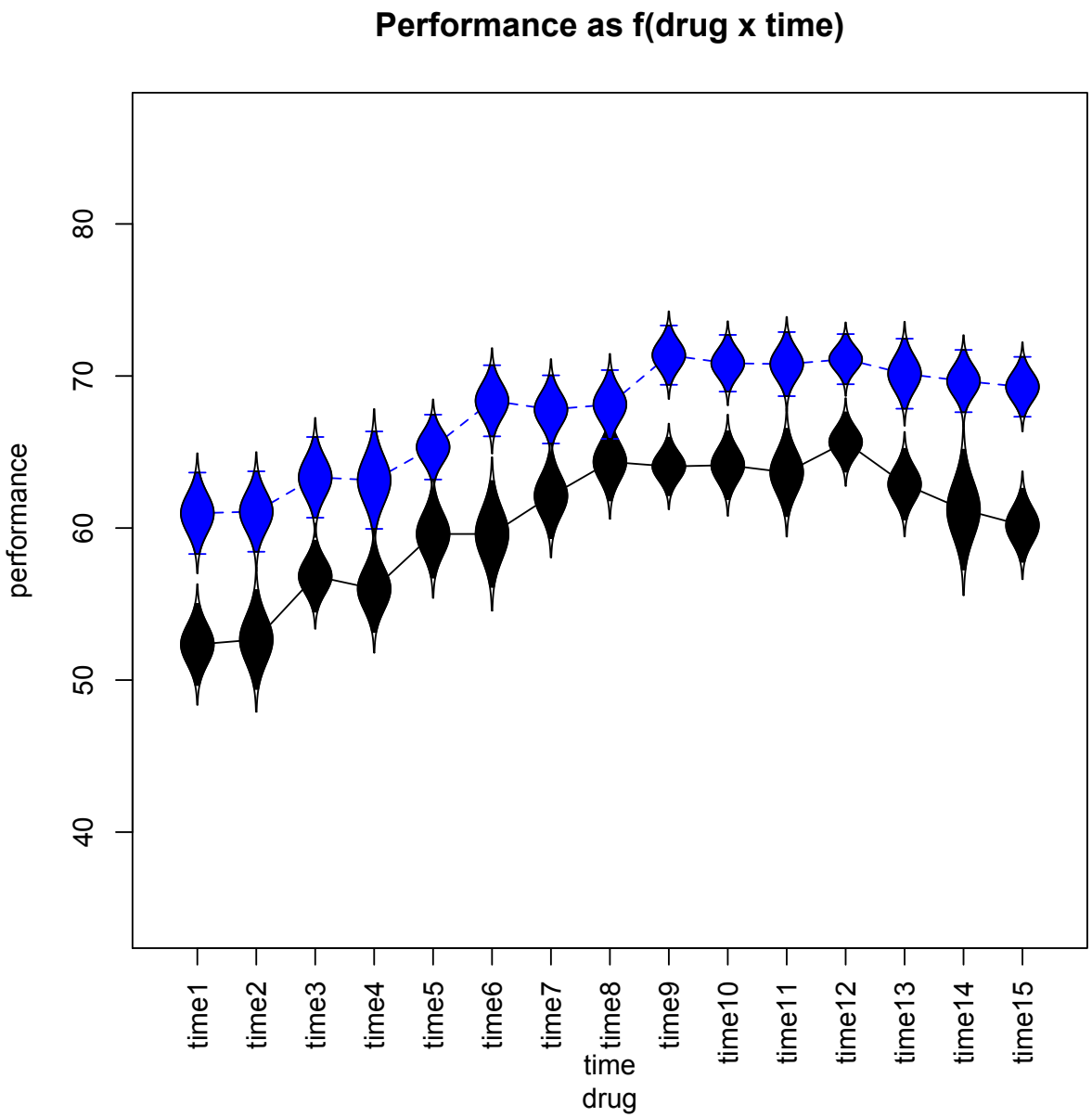
```
lowerCor(sim.data[-10])
```

```
            snum  sex   drug  time  anxty impls arosl tensn prfrm
snum         1.00
sex          0.01  1.00
drug         0.02 -0.03  1.00
time         0.03  0.01  0.04  1.00
anxiety     -0.13  0.05  0.02  0.06  1.00
impulsivity  0.10  0.07 -0.01 -0.02 -0.05  1.00
arousal      0.01  0.00  0.64  0.54  0.05 -0.15  1.00
tension     -0.07  0.03  0.61  0.07  0.73 -0.03  0.41  1.00
performance -0.02  0.02  0.41  0.38  0.06 -0.15  0.66  0.29  1.00
```

Pairs panels of the data



Performance by time and drug



The basic regression with 1000 subjects

R code

```
setCor(performance ~ drug * time * impulsivity, data=sim.data)
```

```
Call: setCor(y = performance ~ drug * time * impulsivity, data = sim.data)

Multiple Regression from raw data

 DV =  performance
                      slope   se     t       p lower.ci upper.ci  VIF
(Intercept)            0.00 0.03  0.00 1.0e+00    -0.05     0.05 1.00
drug                   0.40 0.03 15.87 1.0e-50     0.35     0.45 1.00
time                   0.36 0.03 14.32 2.1e-42     0.31     0.41 1.00
impulsivity           -0.15 0.03 -5.90 5.1e-09    -0.20    -0.10 1.02
drug*time             -0.02 0.03 -0.78 4.3e-01    -0.07     0.03 1.00
drug*impulsivity      -0.02 0.03 -0.60 5.5e-01    -0.07     0.03 1.01
time*impulsivity       0.22 0.03  8.74 9.6e-18     0.17     0.27 1.01
drug*time*impulsivity  0.00 0.03  0.17 8.6e-01    -0.05     0.05 1.01

Residual Standard Error =  0.8  with  992  degrees of freedom

 Multiple Regression
               R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2        p
performance 0.61 0.37 0.44 0.19        0.37     0.02     83.62   7 992 1.75e-95
```

Showing the impulsivity by time interaction

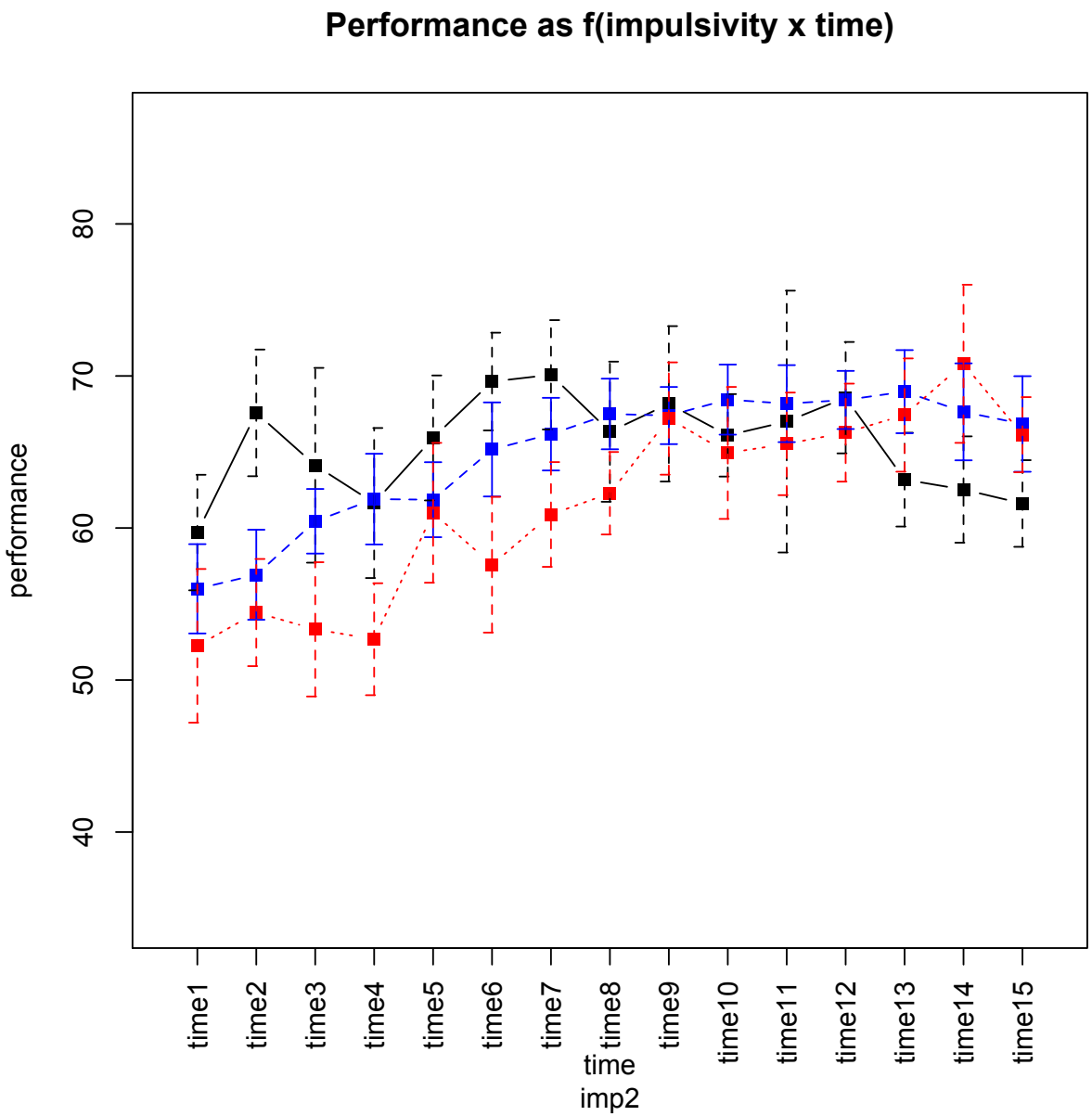
1. There are complicated ways to show the interaction
2. Or, we can just recode impulsivity into 3 levels and then draw those three levels

R code

```
imp2 <- round((sim.data$impulsivity-1)/4)
table(imp2)
sim.data2 <- data.frame(sim.data[,1:9],imp2=imp2)
error.bars.by(performance ~ time * imp2,data=sim.data2,las=3,main="Perfor
```

```
table(imp2)
imp2
  0   1   2
217 554 229
```

Performance by time and 3 levels of impusivity



Simulation: confounds

1. Although at the correlation level, there was a relationship between tension and performance, this was because both of them were affected by ID number.
2. At the correlation level, it seems as if Tension and Performance are correlated, but this is the confounding effect of caffeine.
3. We can examine this through partial correlations
4. What is the effect of tension controlling for arousal vs.
5. What is the effect of arousal controlling for tension?

Partial correlations

R code

```
setCor(performance ~ tension -arousal,data=sim.data)
setCor(performance ~ arousal -tension,data=sim.data)
```

```
setCor(performance ~ tension -arousal,data=sim.data)
Call: setCor(y = performance ~ tension - arousal, data = sim.data)

Multiple Regression from raw data
The following variables were partialed out: arousal
 and are included in the calculation of df1 and df2
 DV =  performance*
             slope   se    t    p lower.ci upper.ci VIF
(Intercept)*  0.00 0.02 0.00 1.00    -0.05     0.05   1
tension*      0.02 0.03 0.88 0.38    -0.03     0.07   1
Residual Standard Error =  0.75  with  997  degrees of freedom

Call: setCor(y = performance ~ arousal - tension, data = sim.data)

Multiple Regression from raw data
The following variables were partialed out: tension
 and are included in the calculation of df1 and df2
 DV =  performance*
             slope   se     t        p lower.ci upper.ci VIF
(Intercept)*  0.00 0.02  0.00  1.0e+00    -0.05     0.05   1
arousal*      0.66 0.03 25.29 2.2e-109     0.60     0.71   1

Residual Standard Error =  0.75  with  997  degrees of freedom

 Multiple Regression
                R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2         p
performance 0.63 0.39 0.66 0.44        0.39     0.02    639.39   1 997 2.15e-109
```

Or, just put them both in the model

R code

```
setCor(performance ~ tension +arousal,data=sim.data)
```

```
Call: setCor(y = performance ~ tension + arousal, data = sim.data)

Multiple Regression from raw data

 DV =  performance
            slope   se     t        p lower.ci upper.ci VIF
(Intercept)  0.00 0.02  0.00  1.0e+00    -0.05     0.05 1.0
tension      0.02 0.03  0.88  3.8e-01    -0.03     0.07 1.2
arousal      0.66 0.03 25.29 2.2e-109     0.60     0.71 1.2

Residual Standard Error =  0.75  with  997  degrees of freedom

 Multiple Regression
               R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2         p
performance 0.67 0.44 0.57 0.32        0.44     0.02    395.65   2 997 3.23e-127
```

The effects of drug and time and impulsivity are all due to arousal

R code

```
setCor(performance ~ drug * time * impulsivity + arousal, data=sim.data)
```

```
 setCor(performance ~ drug * time * impulsivity + arousal, data=sim.data)
Call: setCor(y = performance ~ drug * time * impulsivity + arousal,
    data = sim.data)

Multiple Regression from raw data

 DV =  performance
                       slope   se     t       p lower.ci upper.ci  VIF
(Intercept)             0.00 0.02  0.00 1.0e+00    -0.05     0.05 1.00
drug                    0.01 0.04  0.25 8.0e-01    -0.07     0.09 3.04
time                    0.04 0.04  1.04 3.0e-01    -0.03     0.11 2.40
impulsivity            -0.06 0.03 -2.22 2.7e-02    -0.11    -0.01 1.13
arousal                 0.63 0.05 11.53 6.2e-29     0.52     0.73 5.29
drug*time               0.00 0.02  0.01 1.0e+00    -0.05     0.05 1.01
drug*impulsivity       -0.01 0.02 -0.25 8.0e-01    -0.05     0.04 1.02
time*impulsivity        0.01 0.03  0.22 8.2e-01    -0.05     0.07 1.63
drug*time*impulsivity   0.00 0.02 -0.01 9.9e-01    -0.05     0.05 1.01

Residual Standard Error =  0.75  with  991  degrees of freedom

 Multiple Regression
               R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2         p
performance 0.67 0.45 0.54 0.29        0.44     0.02      99.5   8 991 2.44e-121
>
```

Types of errors

1. Type 2: Failing to detect a difference that is in fact there due to lack of power.
2. Small effects are harder to detect than large effect
3. Type 3: Failing to detect an effect because you did not study it! (a failure of theory, not of statistics)

Science and Error

1. Type 1 errors can happen to you (or me)!
2. Experiment wide error rate is a function of the number of tests run = 1- (1-alpha)n
3. Bonferoni correction sets experiment wide error rate by using a correction for the number of tests = alpha/n
4. This is somewhat conservative but better than pretending that type 1 errors don't happen. The Holm test is probably a better test.
5. Type 2 errors happen due to lack of power
6. If the study is too small, important effects will probably not be detected

Precision vs. generalizability

1. By controlling for more variables, we reduce the error (unexplained) variance
2. But we limit the degree to which one can generalize