1/19/2010 36-402/608 ADA-II H. Seltman

Breakout #3 Results

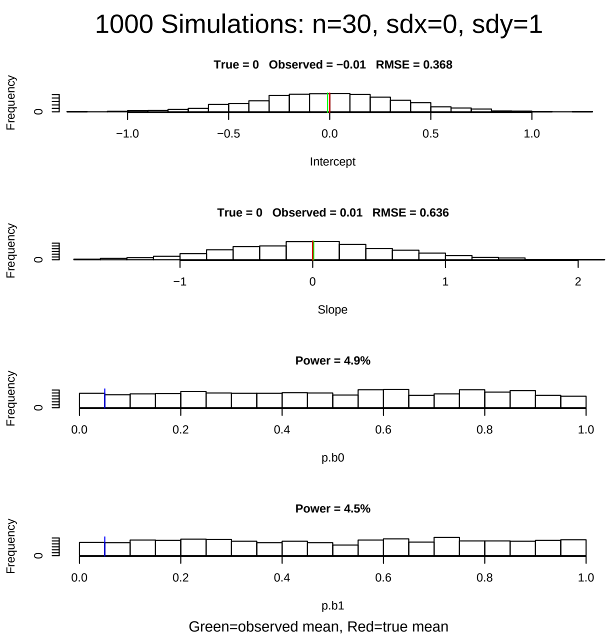
1. **Simple regression simulation**

   `eiv()` is a function that generates random data for a simple regression with "n" subjects, optionally adding measurement error to the x data (violating the fixed-x assumption and resulting in a so-called "errors in variables" model). It then calculates the standard regression results for the simulated data. The whole process is repeated "nsim" times, and the results are returned in a "simulation" object.

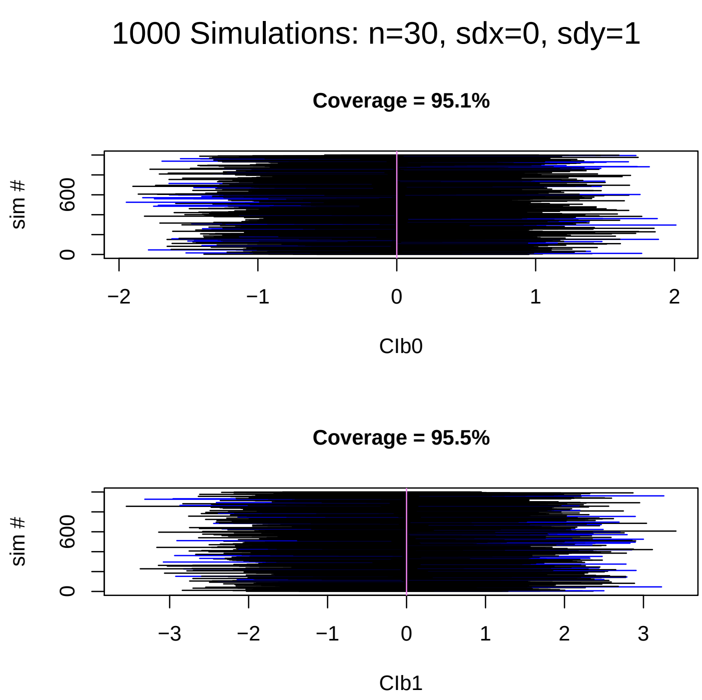
   The x data is generated using a uniform [0,1] distribution. You may set parameters for n, nsim, b (true intercept and slope), sdy (true error s.d.), and sdx (true x measurement error s.d.).

   Note: RMSE (root mean square error) is an estimate of σ.

```
> s1 = eiv(nsim=1000, n=30, b=c(0,0), sdy=1, sdx=0)
> print(s1)
Simulation: s1
1000 simulations with estimation of Intercept and Slope.
True values: 0, 0.
Tests: p.b0 and p.b1.
CIs: CIb0 and CIb1.
Options: nsim=1000, n=30, sdx=0, sdy=1.
> summary(s1)
$nsim
[1] 1000
$alpha
[1] 0.05
$estimates
           Statistic
Parameter     True Value         Mean         Bias     Variance         RMSE
  Intercept  0.000000000 -0.010462687 -0.010462687  0.135681577  0.368498364
  Slope      0.000000000  0.007420337  0.007420337  0.404349842  0.635928379
$power
 p.b0  p.b1
0.049 0.045
$coverage
     MeanWidth   MeanCenter True Coverage
CIb0  1.517241 -0.010462687    0    0.951
CIb1  2.639221  0.007420337    0    0.955
> plot(s1)
```




```
> plot(s1, doParam=FALSE, doP=FALSE, doCI=TRUE)
```



Question 1: What model was simulated, and how many times? What can you learn from the output, and how does it match your expectations?

If you remember something about the binomial distribution, see if you can explain how the following calculation relates to the uncertainty in the simulation results.

```
> p=0.05
> simVar = p*(1-p)/1000
> plusMinus = 2 *sqrt(simVar)
> plusMinus  # [1] 0.01378405
> 0.05 + c(-1,1)*plusMinus
[1] 0.03621595 0.06378405
> 0.95 + c(-1,1)*plusMinus
[1] 0.936216 0.963784
```

2. **Simulation when standard null hypotheses ($\beta_i = 0$) are false**

```
> s2 = eiv(nsim=1000, n=30, b=c(0.25,0.5), sdy=1, sdx=0)
> summary(s2)  # (abbreviated results:)
$estimates
           Statistic
Parameter    True Value        Mean        Bias    Variance        RMSE
  Intercept  0.25000000  0.26333971  0.01333971  0.14033487  0.37485039
  Slope      0.50000000  0.46705637 -0.03294363  0.42228688  0.65067055
$power
 p.b0  p.b1
0.098 0.113
$coverage
     MeanWidth MeanCenter True Coverage
CIb0  1.520409  0.2633397 0.25    0.948
CIb1  2.642626  0.4670564 0.50    0.953
```

Question 2: What model was simulated? Are H_0's true, false, or undetermined? What can you learn from the output, and how does it match your expectations?

3. **Errors in variables with b=c(0,0)**

```
> s3 = eiv(nsim=1000, n=30, b=c(0,0), sdy=1, sdx=1)
> summary(s3)
$estimates
           Statistic
Parameter    True Value        Mean        Bias    Variance        RMSE
  Intercept 0.000000000 0.003937849 0.003937849 0.043071478 0.207574047
  Slope     0.000000000 0.005046648 0.005046648 0.036418210 0.190902274
$power
```

```
 p.b0  p.b1
0.053 0.054
$coverage
     MeanWidth  MeanCenter True Coverage
CIb0 0.8400203 0.003937849    0    0.947
CIb1 0.7428122 0.005046648    0    0.946
```

Question 3: What model was simulated? What can you learn from the output, and how does it match your expectations?

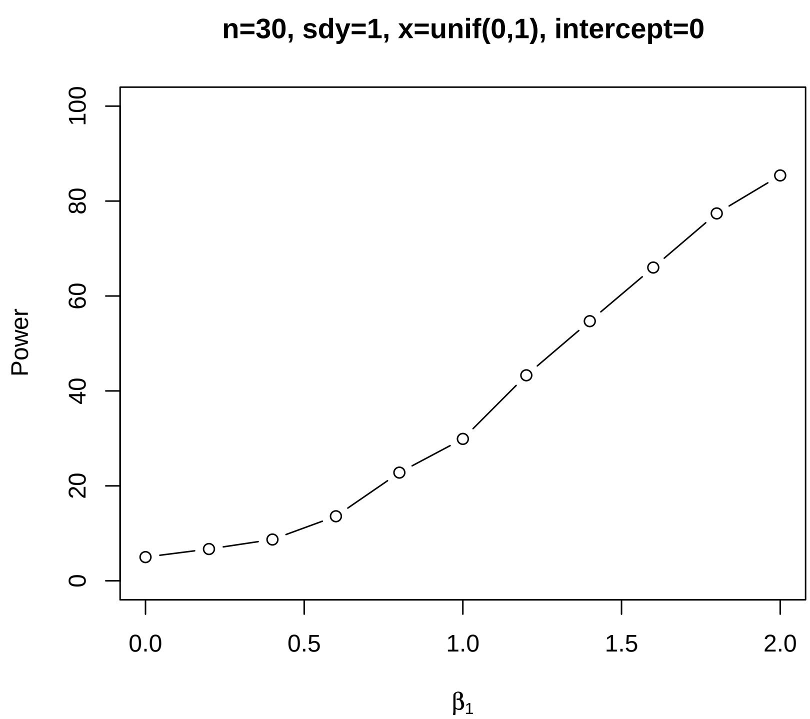
4. **Errors in variables with b=c(0.25,0.5)**

```
> s4 = eiv(nsim=1000, n=30, b=c(0.25,0.5), sdy=1, sdx=1)
> summary(s4)
$estimates
           Statistic
Parameter    True Value        Mean        Bias    Variance        RMSE
  Intercept  0.25000000  0.46990074  0.21990074  0.04658917  0.30813228
  Slope      0.50000000  0.04452477 -0.45547523  0.03760090  0.49503392
$power
 p.b0  p.b1
0.572 0.052
$coverage
     MeanWidth MeanCenter True Coverage
CIb0 0.8502235 0.46990074 0.25    0.810
CIb1 0.7596810 0.04452477 0.50    0.342
```

Question 4: What model was simulated? What can you learn from the output, and how does it match your expectations? Make educated guesses about what would happen if "sdx" were increased, and if β_1 were set to +1.0 or -0.5.

5. **Breakout #2: Power curve**

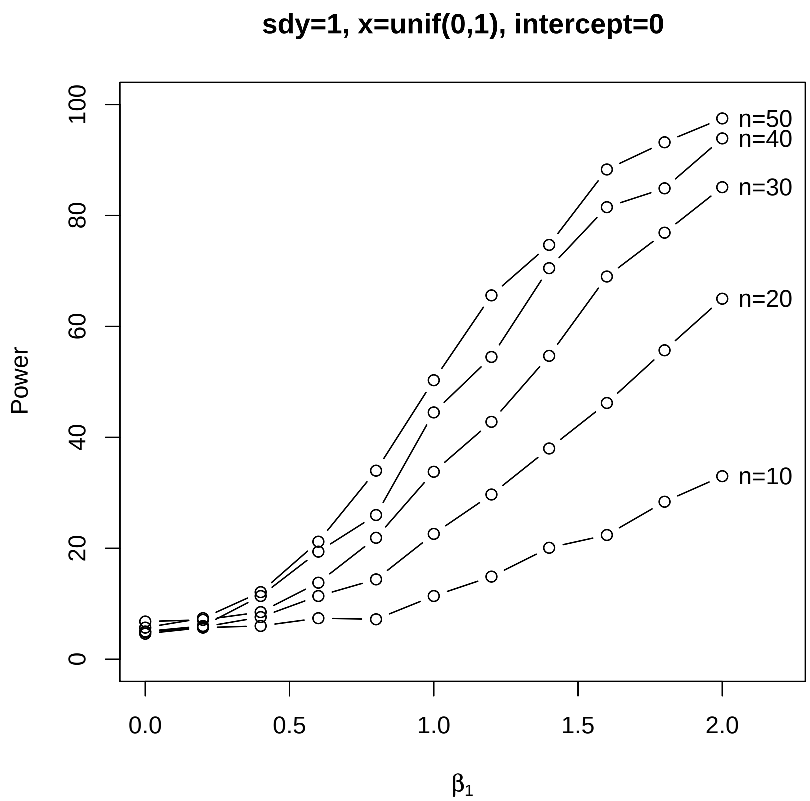
```
> summary(eiv(nsim=1000, n=30, b=c(0,0), sdy=1, sdx=0))$power[2]
p.b1
0.058
> b1s = seq(0,2,0.2)
> nb1 = length(b1s)
> pwr = rep(NA, nb1)
> for (i in (1:nb1))
+ pwr[i]=summary(eiv(nsim=1000, n=30, b=c(0,b1s[i]), sdy=1, sdx=0))$power[2]*100
> plot(b1s, pwr, type="b", xlab=expression(beta[1]), ylab="Power", ylim=c(0,100),
+ main=paste("n=30, sdy=1, x=unif(0,1), intercept=0"))
```



Question 5: How does the code work? What is the explanation for the shape of the curve? What are the asymptote values?

6. **Set of power curves**

```
# Set of power curves using an alternate approach:
> fpwr = function(beta1, n, nsim=1000)
+ return(summary(eiv(nsim=nsim, n=n, b=c(0,beta1), sdy=1, sdx=0))$power[2]*100)
> plot(c(0, 2.2), c(0,100), type="n", xlab=expression(beta[1]), ylab="Power",
+ main="sdy=1, x=unif(0,1), intercept=0")
> for (n in seq(from=10,to=50,by=10)) {
+   pwr = sapply(b1s, fpwr, n=n)
+   lines(b1s, pwr, type="b")
+   text(b1s[nb1], pwr[nb1], paste("n=",n,sep=""), adj=-0.3)
+ }
```



Question 6: How many simulated datasets are used in drawing this plot? How can we use this plot for designing an experiment?