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

Breakout #3 Comments

1. **Simple regression simulation**

   With sdx=0, this is the usual “fixed x” regression model. With b=c(0,0), the true values of β_1 and β_2 are set to zero which are the standard “null hypothesis” values. We simulate 1000 experiments with 30 subjects each. The distribution of the x values is (mostly) immaterial.

```
$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)
```

   For each simulated dataset, we calculate the linear regression results and store the two coefficient estimates, the two p-values (for $H_0 : \beta_0 = 0$ and for $H_0 : \beta_1 = 0$, and the two 95% confidence intervals.

   Under “estimates” we see that the mean and variance of the 1000 slope estimates are 0.00742 and 0.404 respectively, i.e., using either the normality assumption or the central limit theorem, the empiric (observed) distribution of b_1 is $\text{N}(0.00742, 0.404)$. We can construct a confidence interval for the population mean that these 1000 values are estimating, using a Z-test (with df=999, there is no practical difference between t and Z). SE(estimate) $= \sqrt{s^2/N} = 0.404/\sqrt{1000} = 0.0128$. The confidence interval is $0.00742 \pm 1.96\text{SE} = 0.00742 \pm 0.025$, so the 95% CI is [-0.0176, 0.0325] which includes the true value of 0. Therefore we conclude that these results are consistent with the supposition that the simple linear regression is unbiased.

   Note that a b_1 s.d. of 0.404 is consistent with the normal shaped histogram of all 1000 b_1 values shown in the plots.

   The MSE (mean squared error) is define as the square of the bias plus the variance, and equals the variance for unbiased estimators. RMSE is the square root of MSE, and can be thought of a measure of how far a single estimate might be from the true value: roughly 1/3 of values are more that "RMSE" away from the true value if the estimates are normally distributed.

Since this is a "null" model, the power values of 4.9% and 4.5% should be just estimates of α. Using the binomial variance of 0.05*0.95/1000, and a normal approximation this estimate should be within about 0.4% of the true value (between 4.6 and 5.4%) if the t-test works correctly, and it is.

The coverage tells how often, over the 1000 repeated "experiments", the calculated 95%CI includes the true value (which is known only because we simulated the data). We expect 95%, and observe values quite near that.

The plots of the 1000 p-values for a true null hypothesis are uniform because that is the only way for it to be true that the probability of incorrectly rejecting H_0 (finding $p \leq \alpha$) equals α for any given value of α.

The plots of CI's show about 5% of the 1000 missing the true value.

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

   Here we simulate data with the true parameter values $\beta_0 = 0.25$ and $\beta_1 = 0.5$, so the standard null hypothesis are both definitely false (and we'd like to see both p-value be ≤ 0.05.

```
> 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
```

   Again, we conclude that the tests are unbiased with appropriate coverage. Unfortunately the power is quite low, and only a small minority of the repeat "experiments" correctly rejected the null hypotheses.

   We would need to increase the number of subjects, or decrease the outcome variability (by reducing subject, environmental, measurement, or treatment application error) to improve the power into an acceptable (> 80%) range.

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

   Now the null hypotheses are true, but the fixed-x assumption is violated, and the explanatory variable is measured imprecisely (added measurement error is sdx=1).

```
> 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
```

   We still have unbiased estimation, appropriate coverage and appropriate near 5% type 1 error (power under H_0). Somewhat surprisingly the variability of the estimates is reduced.

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

   This model violates fixed-x and has non-null values of the betas.

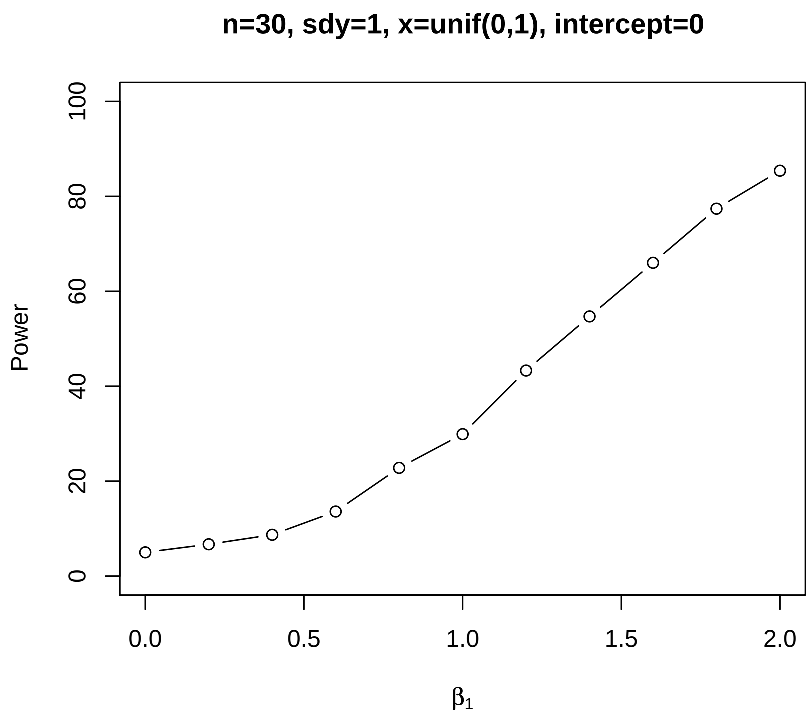
```
> s4 = eiv(nsim=1000, n=30, b=c(0.25,0.5), sdy=1, sdx=1)
> summary(s4)
$estimates
           Statistic
Parameter    True Value        Mean        Bias    Variance        RMSE
  Intercept  0.25000000  0.46990074  0.21990074  0.04658917  0.30813228
  Slope      0.50000000  0.04452477 -0.45547523  0.03760090  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
```

This assumption violation has destroyed the reliability of the regression analysis. The estimate of the intercept is biased away from zero, and the estimate of the slope (which is usually of primary interest) is biased towards zero. The coverage is horribly far from 95%, especially for the slope, so our usual statement for any one experiment that we are 95% confident that the true slope is in some interval will be mostly incorrect. Also, the power to detect any effect of x on y has been reduced from 10% to 5% which is really no power at all (power=α).

Violation of the fixed x assumption is a large problem, most commonly resulting in underestimation of the (absolute) magnitude of the slope.

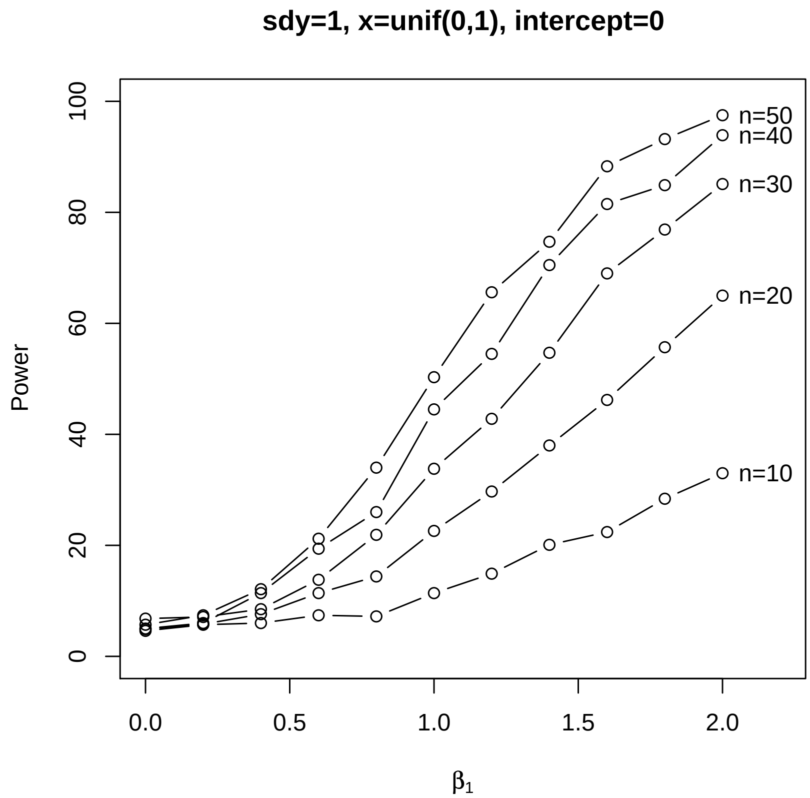
5. **Breakout #2: Power curve**

```
> 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"))
```



The code calculates power for each of several possible values of β_1. It is a fact that any experiment has more power to detect a large effect than it has to detect a small effect. Power does not drop below α because we set things up for any standard test to reject H_0 5% of the time even if there is no real treatment effect. Any of course power can't get above 100% (rejecting H_0 every time). So we expect a curve with asymptotes at 5% and 100% and therefore it must have a sigmoid shape.

6. **Set of power curves**



With 11 beta values and 5 n values, there are 55,000 simulated data sets (1,650,000 simulated subjects) used to make the plot. In practice, for any study we can and should judge a "smallest meaningful effect size". For example, for a t-test of the effects of a drug on cholesterol lowering performed by a drug company, they may decide that a lowering of less than 15 mg/dL on average is not commercially viable, so they would calculate power for that effect size, knowing that if the true effect size is larger the power would be larger, and if the true effect size is smaller, then the would have less power and a greater chance of a type 2 error (but they wouldn't be practically concerned by that).

We can use this plot to find an appropriate sample size for the experiment (before performing it, of course), for whatever particular value of β_1 is judged to be the "smallest meaningful effect size".