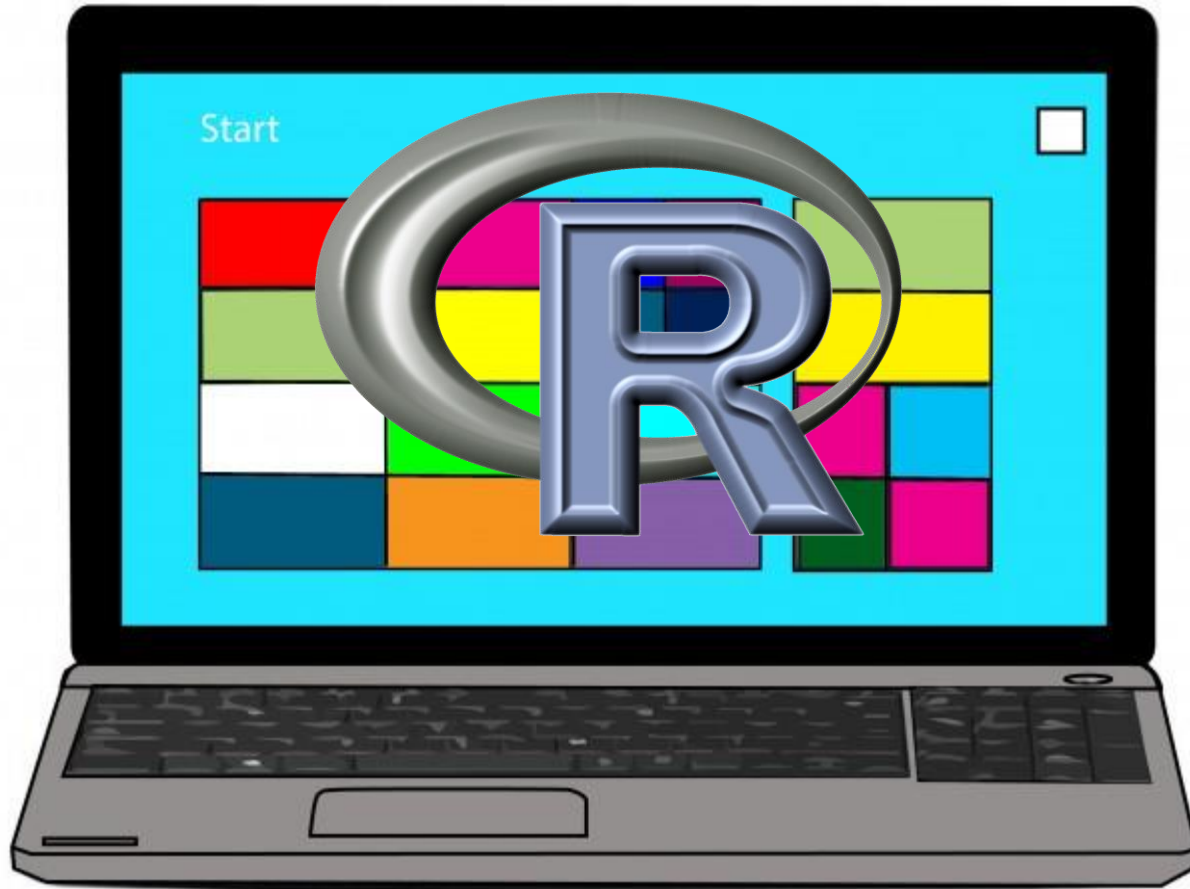




Advanced STATS and R Regression and interaction with categorical variables



Bring your laptop or use Network.



The University of Oklahoma

Using R

- # Multiple Linear Regression (MLR) can
- # be modeled in the following way
- $fit = lm(Y \sim X_1 + X_2 + X_3 + \dots + X_r, data = mydata)$
- `summary(fit)` # show results
- What if an X variable is categorical
 - Use dummy variables (treatment)
 - There will be a base line level to which all other levels are compared.

Regression – Factors

How to use Dummy Variables

- Dummy Variables:

	lang	IQ	method	meth1	meth2	meth3	
1	72	87	1	1	0	0	
2	75	119	1	1	0	0	
3	85	121	1	1	0	0	method 1
...	
8	68	109	1	1	0	0	
9	91	139	1	1	0	0	
10	75	105	1	1	0	0	
11	90	110	2	0	1	0	
12	98	128	2	0	1	0	
13	73	117	2	0	1	0	method 2
...	
18	81	80	2	0	1	0	
19	84	123	2	0	1	0	
20	79	96	2	0	1	0	
21	59	95	3	0	0	1	
22	65	120	3	0	0	1	
23	67	125	3	0	0	1	
...	
28	70	138	3	0	0	1	method 3
29	59	112	3	0	0	1	
30	48	90	3	0	0	1	



Look at the design matrix for a parallel lines model

```
> data(teach.df)
> head(teach.df)
  lang  IQ method
1   72  87      1
2   75 119      1
3   85 121      1
4   70 112      1
5   73 100      1
6   86 133      1
```

$$Xa = 0$$
$$X(1 \ 0 \ -1 \ -1 \ -1)' = 0$$
$$int - dum1 - dum2 - dum3 = 0$$
$$\therefore X'Xa = 0$$
$$\therefore (X'X) \text{ is singular}$$

```
> X=mydummy(teach.df[, -1])
> head(X)
      int  IQ dum1 dum2 dum3
[1,]   1  87   1   0   0
[2,]   1 119   1   0   0
[3,]   1 121   1   0   0
[4,]   1 112   1   0   0
[5,]   1 100   1   0   0
[6,]   1 133   1   0   0
```

Get the Data and Book

- Link below:
- statsandr.oucreate.com



Courses

- **Bayesian Stats MATH 4803/5803**
- Advanced Applied STATS MATH 4793/5793
- **Applied Statistical Methods MATH 4753**



Next time bring your laptop (if you want to).

