Stat 145, Thu 15-Apr-2021 -- Thu 15-Apr-2021
Biostatistics
Spring 2021

Thursday, April 15th 2021

Due:: PS12 due at 11 pm

Other calendar items

Thursday, April 15th 2021

Wk 11, Th
Topic:: 1-way ANOVA
Read:: Lock5 8.1

Last time:
- considered bivariate data:
  non-binary categorical explanatory variable
  quantitative response variable

grp	val
A	15
A	18
A	17
B	14
B	11
B	13
C	16
C	17
C	19
C	17

- encountered SST, SSG, SSE
  each is similar/connected to a variance calculation: for instance

$$\text{SST} = \sum (x - \bar{x})^2 = (n-1)\underbrace{\frac{1}{n-1}\sum (x-\bar{x})^2}_{\text{Variance}} = (n-1)s^2$$

```
SST = SSG + SSE  breaks total sum-of-squares into group and residual parts
  not a true variance until tempered by size of data (divided by df)


- Why this breakup is useful
   picture from textbook
   app at  https://shiny.calvin.edu:3838/scofield/ssQuantDataMultipleGroups/

   idea: compare across-group variance (MSG) with within-group variance (MSE)
     F-statistic:   F = MSG / MSE



- randomization
   in StatKey, or
   app at  https://shiny.calvin.edu:3838/scofield/fstatRandomizationDist/

- use of F-distribution model
   when appropriate (rules of thumb)
     know your group sample means have normal distributions, either by
       underlying populations are normal, or
       sample sizes are at least 30
     2:1 rule for sample standard deviations

- commands in R
   pf( F-statistic, df1= ..., df2 = ...)
   anova( lm(respVar ~ explVar, data=dFrame) )
   anova( lm(respVar ~ explVar, data=dFrame) )["F value"][1,1]
```

ANOVA table

	df	SS	MS	F
Groups ("Source")	4	2234.4	2234.4/4	MSG/MSE
Error ("Residual")	35	798	798/35	
Total	39	3032.4		

$39 = n - 1 \implies 40$ total observations

$df_1 = 4 = I - 1 \implies$ So $I = 5$ groups/populations

$$F = \frac{558.6}{22.8} = \frac{\text{across-group variability}}{\text{within-group variability}}$$

larger this is, the stronger evidence against the population means being equal.

Start from scratch:

Have bivariate data, expl. variable identifies group/pop.

H_0: means for all groups/populations are equal

$$\mu_1 = \mu_2 = \cdots = \mu_I$$

H_a: At least one group mean is different from others

Statistic used : $F = \dfrac{MSG}{MSE}$

larger F corresponds to stronger evidence against H_0.

ANOVA = ANalysis Of VAriance