ANCOVA

Combining Quantitative and
Qualitative Predictors

ANCOVA

- In an ANCOVA we try to adjust for differences in the quantitative variable.
- For example, suppose that we were to compare men's average faculty income to women's average faculty income here at OU faculty.
  - Looking for a difference involves an ANOVA
  - Explaining the difference (if one is found) involves an ANCOVA

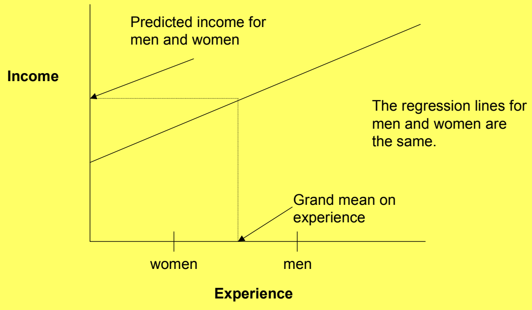
Explaining the Difference

- In trying to explain the difference between men and women, we may want to control for certain variables:
  - Experience
  - Rank
  - Performance record
  - Etc.
- That is, we would like to show that the difference is due to relevant performance criteria. If we can't show that this is the case, then we have a serious discrimination problem.

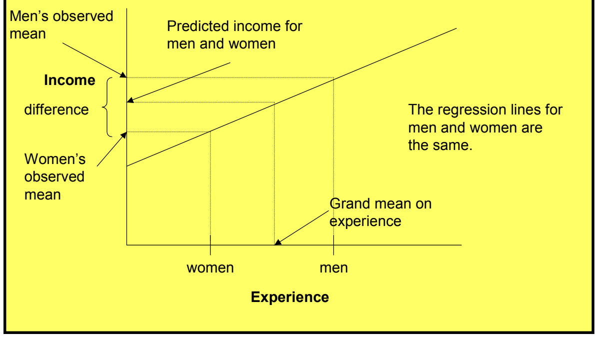
What if?

- The ANCOVA is many ways is a what if analysis– what if men and women had the same amount of experience? Would we still see the difference in income?
- Inherent in this analysis is the possibility that the what if question is relevant. For example, it would be silly to compare basketball teams adjusting for the heights of players. This would be a meaningless comparison.

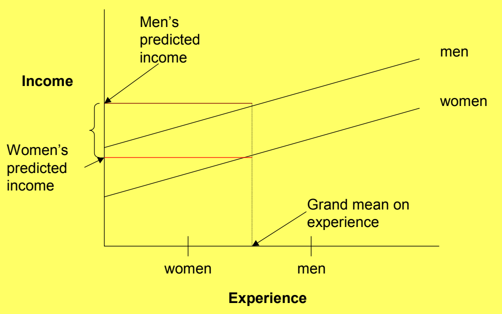
Difference Are Due to Experience



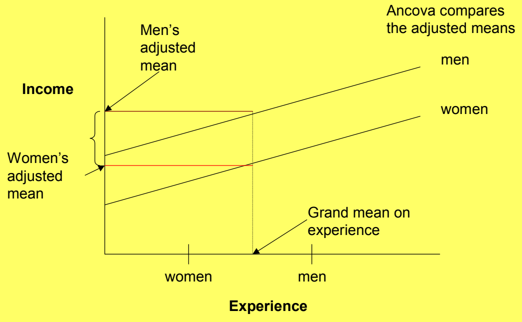
Difference Are Due to Experience



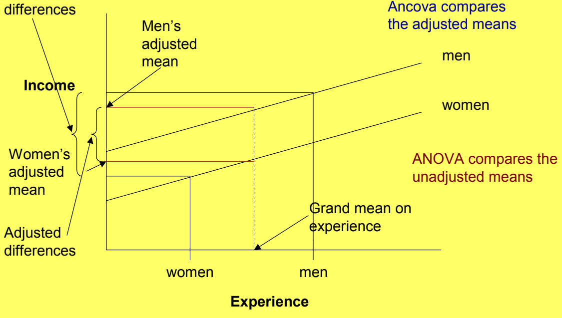
Difference Due to Something Besides Experience



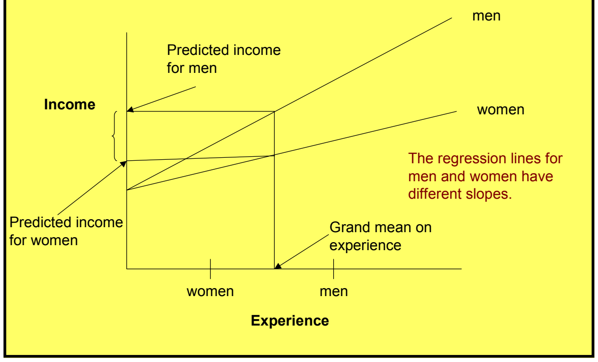
Difference Due to Something Besides Experience



Difference Due to Something Besides Experience



Difference Are Due to Interaction



Situational and Individual Differences

- In the social sciences researcher use ANCOVA to adjust the results for individual differences.
- Suppose that you are looking at ethical decisions under a variety of situations (personal gain, accountability, etc.)
  - You would also like to see if certain individual difference variables (introversion, conscientiousness, cognitive style, etc.) moderate the situational results, you can adjust for these individual difference variables using an ANCOVA design.

SAS ANCOVA Setup
Annual income for three groups

```
proc glm; class race;
model inc = educ race / solution;
means race / tukey;
lsmeans race / tdiff adj=tukey;
/* Note that contrast and estimate statements
 are based on the adjusted means */
contrast 'black vs white' race 1 0 -1;
estimate 'black vs white' race 1 0 -1; run;
```

ANCOVA

```
proc glm data=anc; class race;
model inc= race;
estimate 'black vs white' race 1 0 -1;
contrast 'black vs white' race 1 0 -1;run;
```

ANOVA

ANOVA: Unadjusted Means

Blacks Variable	N	Mean	Std Dev	Minimum	Maximum
Income	16	13.87	6.64	8.00	33.00
education	16	12.25	3.35	7.00	19.00

Hispanics Variable	N	Mean	Std Dev	Minimum	Maximum
Income	14	15.50	6.40	8.00	29.00
education	14	11.64	2.30	8.00	16.00

Whites Variable	N	Mean	Std Dev	Minimum	Maximum
Income	50	21.24	11.43	9.00	60.00
education	50	13.12	2.80	7.00	20.00

ANOVA Results
(Comparing the Unadjusted Means)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	838.117	419.05	4.24	0.0178
Error	77	7602.370	98.73		
Corrected Total	79	8440.487			

Tukey on the Unadjusted Means

Comparisons significant at the 0.05 level are indicated by ***.

race Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
3 - 2	5.740	-1.440	12.920	
3 - 1	7.365	0.544	14.186	***
2 - 3	-5.740	-12.920	1.440	
2 - 1	1.625	-7.065	10.315	
1 - 3	-7.365	-14.186	-0.544	***
1 - 2	-1.625	-10.315	7.065	

ANCOVA Results

Source	DF	Type I SS	Mean Square	F Value	Pr > F
race	2	838.117	419.058	7.01	0.0016
ed	1	3061.307	3061.307	51.23	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
race	2	365.145	182.572	3.06	0.0529
ed	1	3061.307	3061.307	51.23	<.0001

Notice that the Type I and Type III Sums of Squares are different

Adjusted Means for Income

race	inc LSMEAN	LSMEAN Number
1	14.8444275	1
2	17.8147554	2
3	20.2816517	3

We see that even after we adjust for education
there is still a difference between the averages.

Adjustment for Multiple Comparisons: Tukey-Kramer

Least Squares Means for Effect race
t for H0: LSMean(i)=LSMean(j) / Pr > |t|

Dependent Variable: inc

i/j	1	2	3
1		-1.04771 0.5493	-2.43112 0.0453
2	1.047706 0.5493		-1.03581 0.5567
3	2.43112 0.0453	1.035813 0.5567	

Using the Solution option in SAS

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	-7.831	B	4.2060	-1.86	0.0665
race 1 (Blacks)	-5.437	B	2.2365	-2.43	0.0174
race 2 (Hispanics)	-2.466	B	2.3816	-1.04	0.3036
race 3 (Whites)	0.000	B	.	.	.
ed	2.215		0.3095	7.16	<.0001

Blacks $y_1 = (-7.83 - 5.44) + 2.22x$

$y_1 = -13.27 + 2.22x$

Regression Equation for each Group

$$y_1 = -13.27 + 2.22x$$

$$y_2 = -10.3 + 2.22x$$

$$y_3 = -7.83 + 2.22x$$

Adjusted Means

To obtain the adjusted means we use the regression equation for each group and the **overall** x mean

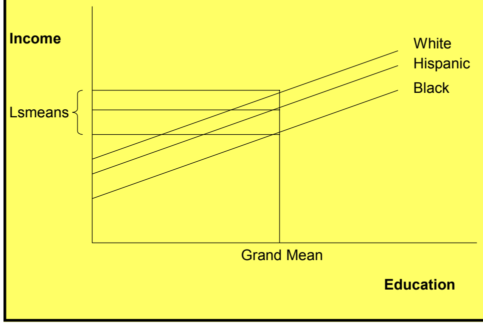
$$14.92 = -13.27 + 2.22(12.7)$$

$$17.89 = -10.3 + 2.22(12.7)$$

$$20.36 = -7.83 + 2.22(12.7)$$

lsmeans
14.84
17.81
20.28

The Adjusted Means



Estimate Statements

ANCOVA Parameter	Estimate	Standard Error	t Value	Pr > |t|
black vs white	-5.43722419	2.23651016	-2.43	0.0174

ANOVA Results Parameter	Estimate	Standard Error	t Value	Pr > |t|
black vs white	-7.36500000	2.85401409	-2.58	0.0118

SAS Type Sum of Squares for unequal n's

Source	**SS I**	**SS II**	**SS III**
A	SS(A| μ)	SS(A| μ,B)	SS(A| μ, B,AB)
B	SS(B| μ,A)	SS(B| μ ,A)	SS(B| μ ,A,AB)
A*B	SS(AB| μ ,A,B)	SS(AB| μ ,A,B)	SS(AB| μ ,A,B)

Unequal n's Designs and Ancova Models

- Under the MCAR (Missing data complete at random) assumption:
  - SAS Type III Sum of Squares provides a test of the partial effects, all submodels are compared to the overall model,

$$\begin{aligned} y_{ij} &= \mu \ + \tau_j + \beta \ x_i + e_{ij} \\ &= \beta_{oj} + \beta \ x_i + e_{ij} \end{aligned}$$

Sequential Sum of Squares SAS Type I

- SAS model statement: (testing the equality of slopes assumption in ancova)

  model y= trt cov trt*cov;

  SS(trt | μ)

  SS(cov | μ, trt)

  SS(trt*cov | μ, trt, cov)

For Type I SS, the sum of all effects add up to the model SS:

SS(trt)+SS(cov)+SS(trt*cov)+SS(error)=SS(total)

SS's are also independent

SAS Type II SS

- SAS model statement:

  model y= trt cov trt*cov;

  SS(trt | μ,cov)

  SS(cov | μ, trt)

  SS(trt*cov | μ, trt, cov)

Type II SS do not necessarily add up to the model SS.

The SS's are not independent.

SAS Type III SS Partial Sum of Squares

- SAS model statement:

  model y= trt cov trt*cov;

  SS(trt | μ, cov, trt*cov)

  SS(cov | μ, trt, trt*cov)

  SS(trt*cov | μ, trt, cov)

Type III SS do not necessarily add up to the model SS.

- The SS's are not independent

Partial F Test Type III SS

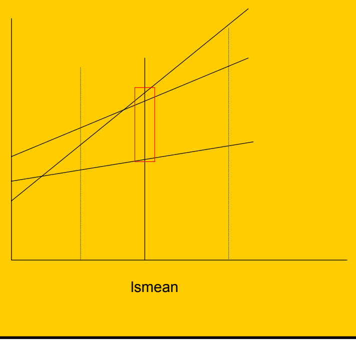
$$\begin{aligned} F &= \frac{(SS_r - SS_f)/1}{df_e} \\ &= \frac{(R_f^2 - R_r^2)}{(1-R_f^2)/df_e} \end{aligned}$$

The reduced model is the full model minus the element being tested.

SAS ANCOVA Setup Unequal Slopes Model

```
proc glm; class race;
model inc = educ race educ*race / solution;
means race / tukey;
lsmeans race / tdiff adj=tukey;
run;
```

Unequal Slopes
Comparing Means



Testing for Difference in Means for a Given Value

```
proc glm; class race;
model inc = educ race educ*race / solution;
means race / tukey;
/* Comparison at the mean */
lsmeans race / tdiff adj=tukey;
/* Comparison at 10 grade level */
lsmeans race / tdiff adj=tukey at educ=10;
estimate `one vs two` race 1 -1 0
educ*race 10 -10 0;
run;
```