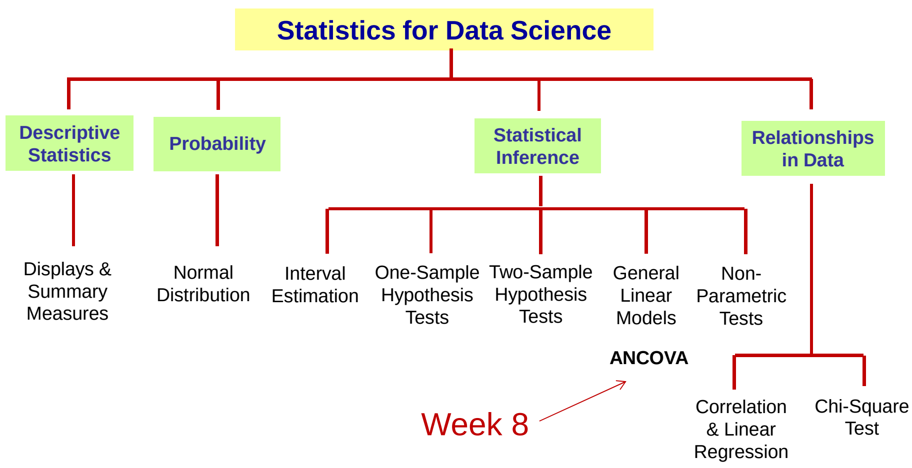
MATH 4044
Statistics for Data Science

Comparing Several Means
ANCOVA

Course outline



Field, A & Miles, J, *Discovering Statistics Using SAS*, Chapter 11.

Topics to be covered

- Analysis of Covariance:
  - When to use ANCOVA.
  - Theory and rationale.
  - Checking assumptions.
  - ANCOVA using SAS.

Example: Is there a difference?

- Effectiveness of three teaching approaches (traditional, fully online, blended) is to be compared.
- Students from the same course are randomly assigned to three instructors:
  - Each follows a different approach with their group of students (1 = traditional, 2 = fully online, 3 = blended).
- At the end of the course, students' scores on a common final exam are recorded.
  - Each student's GPA is also recorded, to adjust for the fact that students might have differing academic abilities.

Recall – theory of ANOVA

SS_T
Total variance in the data

SS_M
Variance explained by the model

SS_R
Unexplained variance

- If the experiment is successful, then the model will explain more variance than it can't:
  - SS_M will be greater than SS_R.

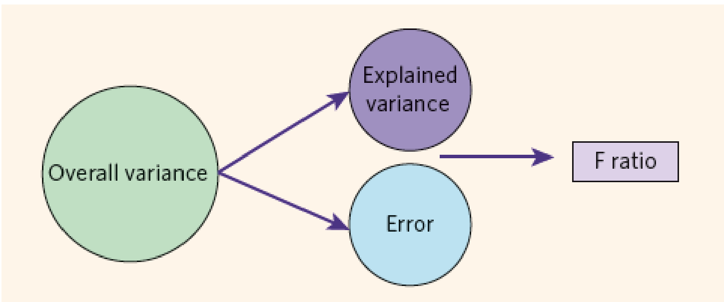
The ANalysis of COVAriance (ANCOVA)

- A single dependent variable (outcome) is assessed across one or more independent variables (factors), controlling for one or more covariates.
- Covariates are additional variables that are not part of the main analysis.
- We are aiming to explain as much variance as possible, while controlling for as many factors as possible.
  - The unexplained variance may be due to random factors.
  - Or it may be down to factors we 'know about' but do not want to measure.

ANCOVA – possible outcomes

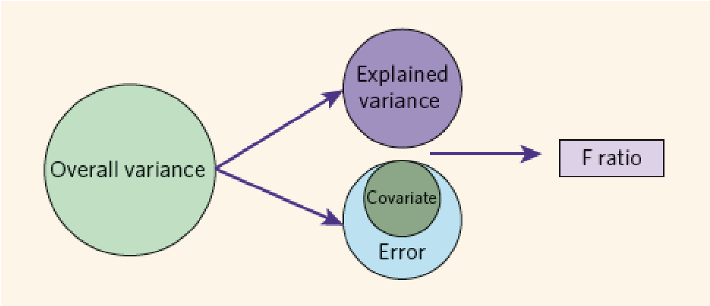
- If the covariates are not related to the explained variance, we can use ANCOVA to reduce the error variance and provide a clearer picture of the original analysis.
- If the covariates are even partially related to the explained variance, the covariates may be confounding the original outcome.
  - We can't use this covariate to reduce variance, but we could explore the extent to which this covariate is 'interfering' with the original outcome.

Main effect, prior to covariate



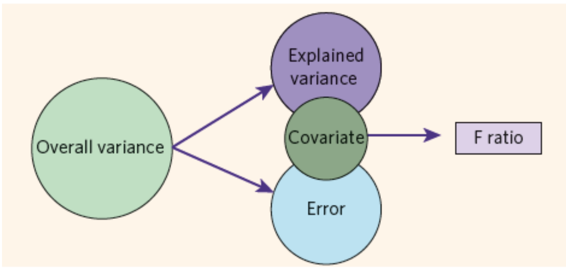
- To reduce error variance further, we can explore the effect of potential covariates using ANCOVA.
  - If we have measured additional variables, we can investigate whether they contribute to the variance in the outcome.

Independent covariate



- The covariate shares no variance with the explained variance.
- Adding the covariate to the model will reduce the error variance and will increase the F-ratio, producing a stronger effect.

Dependent covariate



- The covariate shares some variance with the explained variance; there is potential confounding effect on the outcome.
  - The covariate will reduce some of the error variance, but it will also reduce some of explained error variance.
  - This may reduce the F-ratio, producing a weaker effect.

ANCOVA – possible outcomes

If the covariate is not independent of an experimental effect, the following may occur:

- A previously significant effect may no longer be significant.
  - The covariate was entirely confounding the original outcome.
- A previously significant outcome is still significant, but the effect is reduced.
  - The covariate was partially confounding the outcome.
- A previously non-significant outcome is now significant.
  - The original outcome was being masked by the covariate.

Conditions for applying ANCOVA (to reduce error variance)

- Same assumptions as ANOVA.
- Two additional considerations:
  - Independence of the covariate and the treatment effect.
    - When treatment groups differ on the covariate, putting the covariate into the analysis will not ‘control for’ the differences.
  - Homogeneity of regression slopes.
    - The relationship (e.g. correlation) between the outcome and the covariate does not differ significantly across groups.

Example: Checking assumptions (Normality)

Instructor 1

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.94631	Pr < W	0.6251
Kolmogorov-Smirnov	D	0.195265	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.064305	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.33929	Pr > A-Sq	>0.2500

Instructor 2

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.915882	Pr < W	0.3239
Kolmogorov-Smirnov	D	0.177388	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.055293	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.376649	Pr > A-Sq	>0.2500

Instructor 3

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.973499	Pr < W	0.9213
Kolmogorov-Smirnov	D	0.126333	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.024558	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.165436	Pr > A-Sq	>0.2500

P-values for all Normality tests are greater than 0.05, which suggests that exam scores for all groups can be assumed to have come from Normal populations.

Example: Checking assumptions (equal variances)

Levene's Test for Homogeneity of Score Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Instructor	2	16009.6	8004.8	0.90	0.4177
Error	27	239650	8875.9		

Since the P-value = 0.4177 > 0.05, the assumption of equal variances cannot be rejected.

We have now verified all requirements for ANOVA.

Example: Results from ANOVA

Dependent variable: Exam score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	77.266667	38.633333	0.45	0.6428
Error	27	2322.100000	86.003704		
Corrected Total	29	2399.366667			

R-Square	Coeff Var	Root MSE	Score Mean
0.032203	11.25008	9.273818	82.43333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Instructor	2	77.26666667	38.63333333	0.45	0.6428

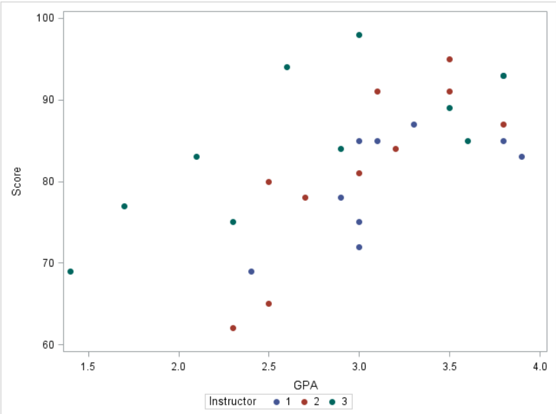
Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	84.70000000	B	2.93263881	28.88	<.0001
Instructor 1	-3.50000000	B	4.14737757	-0.84	0.4061
Instructor 2	-3.30000000	B	4.14737757	-0.80	0.4332
Instructor 3	0.00000000	B	.	.	.

The P-value is 0.6428 > 0.05 so the model is not statistically significant.

There is no significant difference in mean exam scores achieved by the three groups.

Could other variables be affecting this outcome?

Example: Include GPA as a covariate?



Pearson Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0	Score	GPA
Score	1.00000	0.61773 0.0003
GPA	0.61773 0.0003	1.00000

The correlation between GPA and final exam scores is r = 0.62.
Since the P-value = 0.003 < 0.05, this correlation is statistically significant.

The relationship appears to be linear.

Therefore, it makes sense to consider GPA as a covariate.

Example: Checking independence

Dependent variable: GPA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1.42466667	0.71233333	1.86	0.1755
Error	27	10.35400000	0.38348148		
Corrected Total	29	11.77866667			

R-Square	Coeff Var	Root MSE	GPA Mean
0.120953	20.82709	0.619259	2.973333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Instructor	2	1.42466667	0.71233333	1.86	0.1755

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	2.690000000	B	0.19582683	13.74	<.0001
Instructor 1	0.530000000	B	0.27694096	1.91	0.0663
Instructor 2	0.320000000	B	0.27694096	1.16	0.2580
Instructor 3	0.000000000	B	.	.	.

The difference in mean GPA between groups is not significant, so we can assume that GPA is independent of the teaching approach.

[All requirements for ANOVA are satisfied.]

Example: Results from ANCOVA

Dependent variable: Exam score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1325.616686	441.872229	10.70	<.0001
Error	26	1073.749981	41.298076		
Corrected Total	29	2399.366667			

R-Square	Coeff Var	Root MSE	Score Mean
0.552486	7.795824	6.426358	82.43333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Instructor	2	410.052091	205.026045	4.96	0.0149
GPA	1	1248.350019	1248.350019	30.23	<.0001

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	55.16299981	B	5.74384838	9.60	<.0001
Instructor 1	-9.31955766	B	3.06268118	-3.04	0.0053
Instructor 2	-6.81369519	B	2.94415505	-2.31	0.0288
Instructor 3	0.00000000	B	.	.	.
GPA	10.98029747		1.99715067	5.50	<.0001

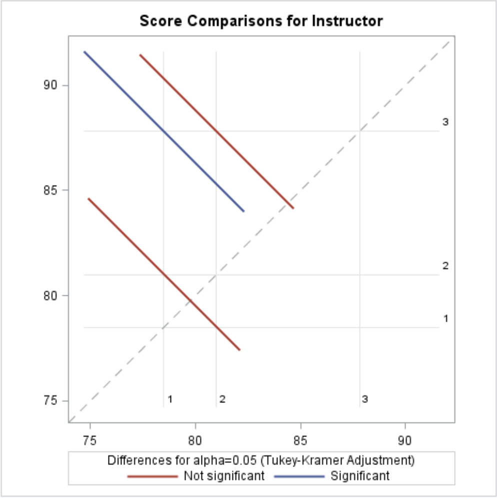
The model overall is statistically significant (P-value < 0.0001).

Use partial sums of squares, Type III SS, to assess significance of the factor and the covariate.

In this case, both instructor and GPA are statistically significant at 5% level.

All else equal, the exam score increases by 10.98 for one unit increase in GPA.

Example: Results from ANCOVA

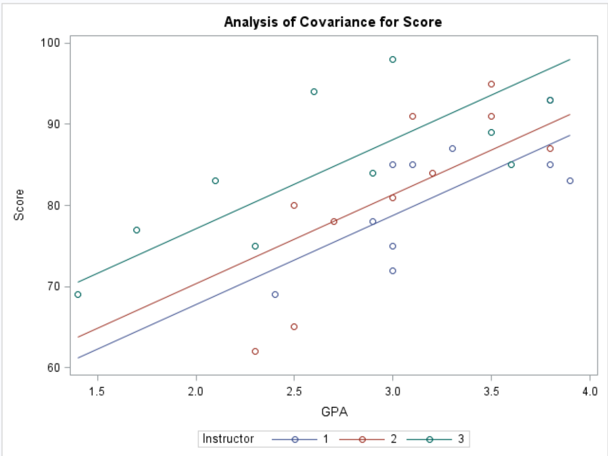


Instructor	Score LSMEAN	LSMEAN Number
1	78.4915266	1
2	80.9973891	2
3	87.8110843	3

There is no significant difference between the GPA-adjusted mean scores of the students in groups one and two (traditional and fully online), and two and three (fully online and blended).

GPA-adjusted mean score in group three (blended) is significantly higher than in group one (traditional).

Example: Regression lines without interaction



This assumes that the relationship between GPA and exam score is the same for each instructor (and their teaching approach).

This is a stringent assumption and one we should check.

Example: Testing homogeneity of slopes

Dependent variable: Exam score

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1490.345785	298.069157	7.87	0.0002
Error	24	909.020882	37.875870		
Corrected Total	29	2399.366667			

R-Square	Coeff Var	Root MSE	Score Mean
0.621141	7.465836	6.154338	82.43333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Instructor	2	255.692292	127.846146	3.38	0.0511
GPA	1	1305.710857	1305.710857	34.47	<.0001
GPA*Instructor	2	164.729099	82.364550	2.17	0.1355

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	62.86753204	B	7.02837832	8.94	<.0001
Instructor 1	-18.67581720	B	15.56770379	-1.20	0.2420
Instructor 2	-35.93672551	B	14.29768353	-2.51	0.0191
Instructor 3	0.00000000	B	.	.	.
GPA	8.11615909	B	2.51061551	3.23	0.0035
GPA*Instructor 1	3.37709717	B	4.95457657	0.68	0.5020
GPA*Instructor 2	9.97991847	B	4.79540918	2.08	0.0483
GPA*Instructor 3	0.00000000	B	.	.	.

The model overall is statistically significant (P-value = 0.0002).

Use partial sums of squares, Type III SS, to assess significance of the factor, the covariate and the interaction term.

In this case, the interaction term is not significant, so the assumption of homogeneity of slopes made previously was justified.

Example: Some conclusions

- How do the three teaching methods compare in terms of their effectiveness?
- Our results indicate that:
  - When only the teaching method is considered, there is no statistically significant difference in students' mean exam scores.
  - However, when GPA is added to the model, we find that mean exam scores adjusted for GPA become significantly different.
    - In particular, mean exam scores are significantly higher with blended teaching compared to the traditional approach.

Example: SAS code

```
proc glm data=mydata.instructors;
    class Instructor;
    model Score=Instructors GPA / solution ss3;
    lsmeans Instructor / adjust=Tukey diff;
    run;
quit;
```

Basic ANCOVA

```
proc glm data=mydata.instructors;
    class Instructor;
    model Score=Instructors GPA Instructors*GPA / solution ss3;
    lsmeans Instructor / adjust=Tukey diff;
    run;
quit;
```

ANCOVA with an interaction term

Example: Exercise and sleep quality

- Does exercise have an effect on perceived sleep quality?
- The following variables were measured:
  - Perceived sleep quality (scores based on sleep questionnaires);
  - Levels of exercise (frequent, infrequent, none);
  - Age of participant.
- Does age play a role?

[All assumption checking is left as an exercise.]

Example: Results from ANOVA

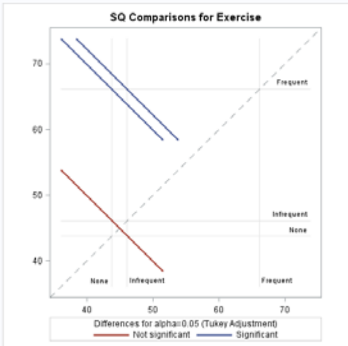
Dependent variable: Sleep Quality

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	3008.600000	1504.300000	7.74	0.0022
Error	27	5245.400000	194.274074		
Corrected Total	29	8254.000000			

R-Square	Coeff Var	Root MSE	SQ Mean
0.364502	26.80428	13.93822	52.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Exercise	2	3008.600000	1504.300000	7.74	0.0022

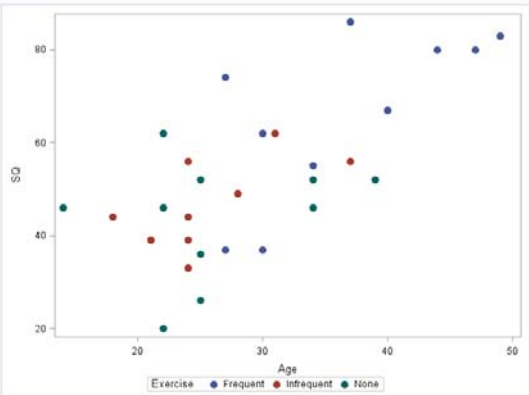
The P-value is 0.0022 < 0.05 so the model is significant. There are significant differences in sleep quality scores by exercise frequency.



The mean sleep quality score is significantly higher with frequent exercise, compared to no or infrequent exercise.

There is no significant difference in mean sleep quality scores between no and infrequent exercise.

Example: Age as a covariate?



Pearson Correlation Coefficients, N = 30 Prob > |r| under H0: Rho=0		
	Age	SQ
Age	1.00000	0.69531 <.0001
SQ	0.69531 <.0001	1.00000

The correlation between age and perceived sleep quality is r = 0.6953. Since the P-value = 0.0001 < 0.05, this correlation is statistically significant.

The relationship appears to be linear.

Therefore, it makes sense to consider age as a covariate.

Example: Checking independence

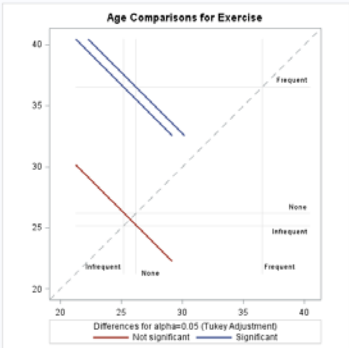
Dependent variable: Age

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	782.600000	391.300000	7.70	0.0023
Error	27	1371.700000	50.803704		
Corrected Total	29	2154.300000			

R-Square	Coeff Var	Root MSE	Age Mean
0.363273	24.32652	7.127672	29.30000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Exercise	2	782.6000000	391.3000000	7.70	0.0023

The P-value is 0.0023 < 0.05 so the model is significant. There are significant differences in age among the three groups.



The mean age is significantly higher for the frequent exercise group, compared to the other two groups.

There is no significant difference in mean age between no and infrequent exercise groups.

Example: ANCOVA with age as covariate

Dependent variable: Sleep quality

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4480.050383	1493.350128	10.29	0.0001
Error	26	3773.949617	145.151908		
Corrected Total	29	8254.000000			

R-Square	Coeff Var	Root MSE	SQ Mean
0.542773	23.16904	12.04790	52.00000

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Exercise	2	489.601513	244.800757	1.69	0.2048
Age	1	1471.450383	1471.450383	10.14	0.0037

Parameter	Estimate		Standard Error	t Value	Pr > |t|
Intercept	16.66408107	B	9.33560808	1.79	0.0859
Exercise Frequent	11.63206240	B	6.34481822	1.83	0.0782
Exercise Infrequent	3.33572210	B	5.39779591	0.62	0.5420
Exercise None	0.00000000	B	.	.	.
Age	1.03572210		0.32529831	3.18	0.0037

The model overall is statistically significant (P-value = 0.0001 < 0.05).

Use partial sums of squares, Type III SS, to assess significance of the factor and the covariate.

When age is included as a covariate, exercise frequency is no longer statistically significant.

Example: Exercise and sleep quality

- Does exercise have an effect on perceived sleep quality?
- Our results indicate that:
  - The explained variance in perceived sleep quality is shared with variance in age.
  - Age may have had a confounding effect on the results.
  - Another study might need to be conducted to tease out the effects, if any, of exercise frequency on perceived sleep quality.