

Topic 13. Analysis of Covariance (ANCOVA, ST&D Chapter 17)

13. 1. Introduction

- The ANCOVA is useful for improving the precision of an experiment.
- Suppose that Y is linearly related to X . Furthermore, suppose that the experimenter cannot control X but can observe it along with Y . The variable X is called a **covariate** or **concomitant variable**.
- The analysis of covariance involves **adjusting the observed response** variable for the effect of the **covariate**. If such an adjustment is not performed, the covariate could inflate the error mean square and make true differences in the response due to treatments harder to detect.
- The concept is **similar to** the use of **blocks** to reduce the experimental error. However, when the blocking variable is a **continuous variable**, the delimitation of the blocks can be very subjective.
- The procedure is a combination of **analysis of variance** and **regression** analysis. We will review briefly some concepts of regression analysis to facilitate the discussion of the analysis of covariance.

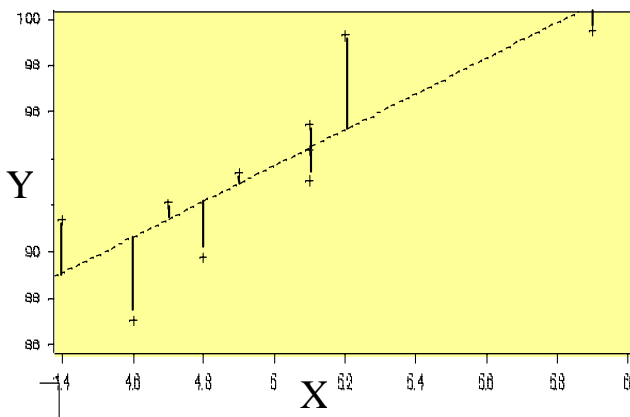
13. 2. Review of Regression concepts.

The equation of a straight line is $Y = a + bX$,

Y is the **dependent** variable and X is the **independent** variable.

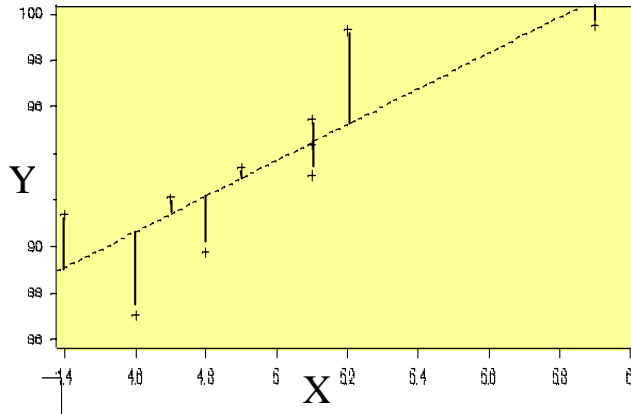
“ a ” is called the **intercept**. The coefficient b is the **slope**

Any point (X, Y) on this line has an X coordinate, or **abscissa**, and a Y coordinate, or **ordinate**.



Body weight, X	Food consumption, Y
4.6	87.1
5.1	93.1
4.8	89.8
4.4	91.4
5.9	99.5
4.7	92.1
5.1	95.5
5.2	99.3
4.9	93.4
5.1	94.4

$$Y_{estimated} = 55.26 + 7.69 * X$$



13. 2. 1. The principle of least squares: the line that best fits the data.

- For each point we find its **vertical distance** from the straight line, square this distance, and then add together all the squared distances.
- The **best-fitting line** is the one that **minimizes the sum of squared vertical deviations**.

13. 2. 2. Residuals: The distance from a point to the straight line is a **residual**

- The residuals represent the behavior of Y that the independent variables don't account for—the error in the model.

13. 2. 3. Formulas to calculate a and b : The equations for the intercept a and the slope b that minimize the SS of the residuals are:

$$b = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{S(XY)}{SS(X)} \text{ and } a = \bar{Y} - b\bar{X}$$

13. 2. 4. Covariance: $S(XY)$ is called the **corrected sum of cross products**.

- Dividing $S(XY)$ by $n-1$ produces a statistic called the **sample covariance**.
- If **high** values of Y are associated with **high** of X the **covariance** will be **positive**.
- If **high** values of Y are associated with **low** values of X the **covariance** will be **negative**.
- If there is no association the covariance will be close to zero.

13. 2. 5. Using SAS for regression analysis:

PROC REG and PROC GLM.

```

data STp254_r;
input x y @@;
cards;
4.6 87.1    5.1 93.1    4.8 89.8    4.4 91.4    5.9 99.5
4.7 92.1    5.1 95.5    5.2 99.3    4.9 93.4    5.1 94.4
;
proc glm;
model y= x;
run; quit;

```

Note that the independent variable is not included in a CLASS statement

Output:

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	90.84	90.84	16.23	0.0038
Error	8	44.77	5.60		
Corr. Total	9	135.60			

F: tests if the model as a whole accounts for a significant proportion of variation in Y .

R-Square	C.V.	Root MSE	Y Mean
0.670	2.528	2.366	93.560

R-Square: measures how much variation in Y the model can account for.

Parameter	Estimate	T for H0: Parameter=0	Pr > T	Std Error of Estimate
INTERCEPT (a)	55.26	5.80	0.0004	9.5349
X ($=b$)	7.69	4.03	0.0038	1.9087

Estimates: Calculates the INTERCEPT ($a= 55.26$) and the slope ($b= 7.69$) and test if they are significantly different from 0.

13. 2. 6. ANOVA of the adjusted Y's

- The MS_{error} (44.77) from the previous analysis represents the variation in Y (food consumption) that would have been obtained if all the animals used in the experiment had had the same initial (X) body weight.
- In the following table each Y value is adjusted using the regression equation to a common X.
- Any value of X can be used to adjust the Y's but the mean of the X ($\bar{X} = 4.98$) value is used as a representative value:

X	Y	Adjusted Y = $Y + b(\bar{X} - X)$
4.6	87.1	90.02
5.1	93.1	92.18
4.8	89.8	91.18
4.4	91.4	95.86
5.9	99.5	92.43
4.7	92.1	94.25
5.1	95.5	94.58
5.2	99.3	97.61
4.9	93.4	94.02
5.1	94.4	93.48
<hr/>		
$\bar{X} = 4.98$		
<hr/>		
SS Y	135.604	44.77
<hr/>		

- The first adjusted value. 90.02, is the expected food consumption expected for this animal if its initial body weight would have been 4.98.
- Note that the **SS of the Y's** is similar to the **Total SS** of the previous ANOVA and that the **SS of the adjusted Y's** is similar to the SS_{error} .
- The SS_{error} is the variation in food consumption that we would have found if all the animals used in the experiment had had the same weight (assuming that "b" was estimated without error).
- Note the large reduction in the variation of the Y's that is obtained when the variation due to the regression is eliminated.

13. 3. ANCOVA example

The goal of this experiment with **oysters** is to determine

- if exposure to water **heated** artificially affects growth
- if the **position** in the water column (surface or bottom) affects growth

Four bags (e.u.) with ten oysters in each bag are randomly placed at each of 5 locations in the cooling water canal of a power-generating plant.

TRT1: cool-bottom, **TRT2:** cool-surface, **TRT3:** hot-bottom,
TRT4: hot-surface, **TRT5:** control mid-depth and mid-temperature.

Each bag is **weighted at the beginning** and at the end of the experiment.

```
data oyster;  
input trt rep initial final;  
cards;
```

1	1	27.2	32.6
1	2	32.0	36.6
1	3	33.0	37.7
1	4	26.8	31.0
2	1	28.6	33.8
2	2	26.8	31.7
2	3	26.5	30.7
2	4	26.8	30.4
3	1	28.6	35.2
3	2	22.4	29.1
3	3	23.2	28.9
3	4	24.4	30.2
4	1	29.3	35.0
4	2	21.8	27.0
4	3	30.3	36.4
4	4	24.3	30.5
5	1	20.4	24.6
5	2	19.6	23.4
5	3	25.1	30.3
5	4	18.1	21.8

```
;
```

```
proc GLM;
```

```
Title 'One-way Anova';
```

```
class trt;
```

```
model final= trt;
```

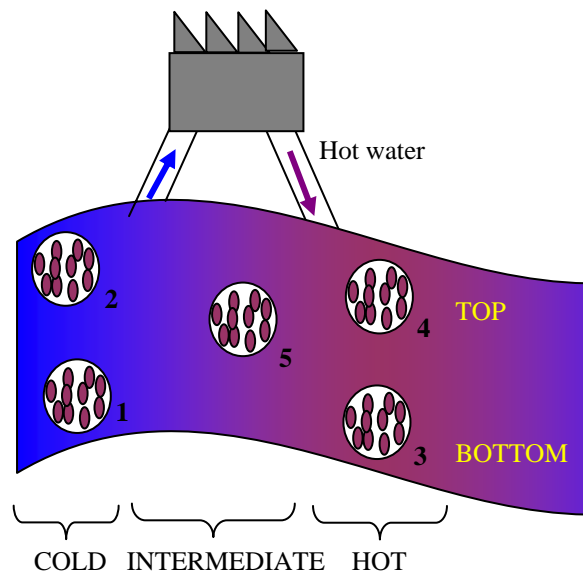
```
proc GLM;
```

```
Title 'ANCOVA';
```

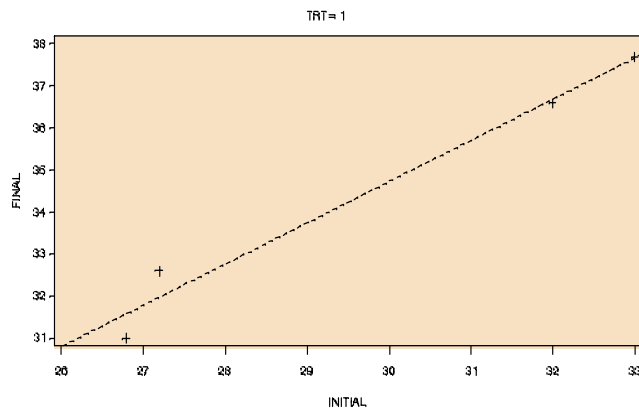
```
class trt;
```

```
model final= trt initial;
```

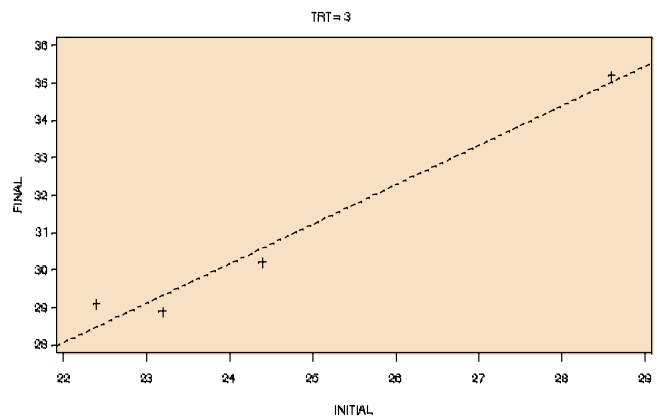
```
run; quit;
```



- The CLASS statement specifies that TRT is a classification variable.
- The variable “**initial**” is the **covariate**. Note that “initial” is a continuous variable and is **NOT included in the CLASS statement**.
- This is similar to the regression example in 13. 2. 4.
- A simple regression of Final versus Initial showed a large significant regression ($P < 0.01$, $R^2 = 0.95$).
- Note that the slopes of the regressions within the different treatments are similar. Examples of final vs. initial regressions for treatments 1 and 3 are shown below.



$$\text{TRT1: } Y = 5.24 + 0.98X$$



$$\text{TRT3: } Y = 4.82 + 1.06X$$

SAS Output:

One-way ANOVA

Dependent Variable: FINAL

Source	DF	SS	MS	F Value	Pr > F
TRT(= Model)	4	198.407	49.60	4.64	0.0122
Error	15	160.262	10.68		
Correc. Total	19	358.669			

R-Square	C.V.	Root MSE	FINAL Mean
0.553175	10.597	3.2687	30.845

The simple one-way ANOVA discovers treatment differences ($P = 0.0122$) in final weight even when the initial weights are not considered

ANCOVA

Dependent Variable: FINAL

Source	DF	SS	MS	F Value	Pr > F
Model	5	354.45	70.89	235.05	0.0001
Error	14	4.22	0.30		
Corrected Total	19	358.67			

R-Square	C.V.	Root MSE	FINAL Mean
0.988228	1.780438	0.5492	30.845

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRT	4	198.40700	49.60175	164.47	0.0001
INITIAL	1	156.04018	156.04018	517.38	0.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRT	4	12.08936	3.02234	10.02	0.0005
INITIAL	1	156.04018	156.04018	517.38	0.0001

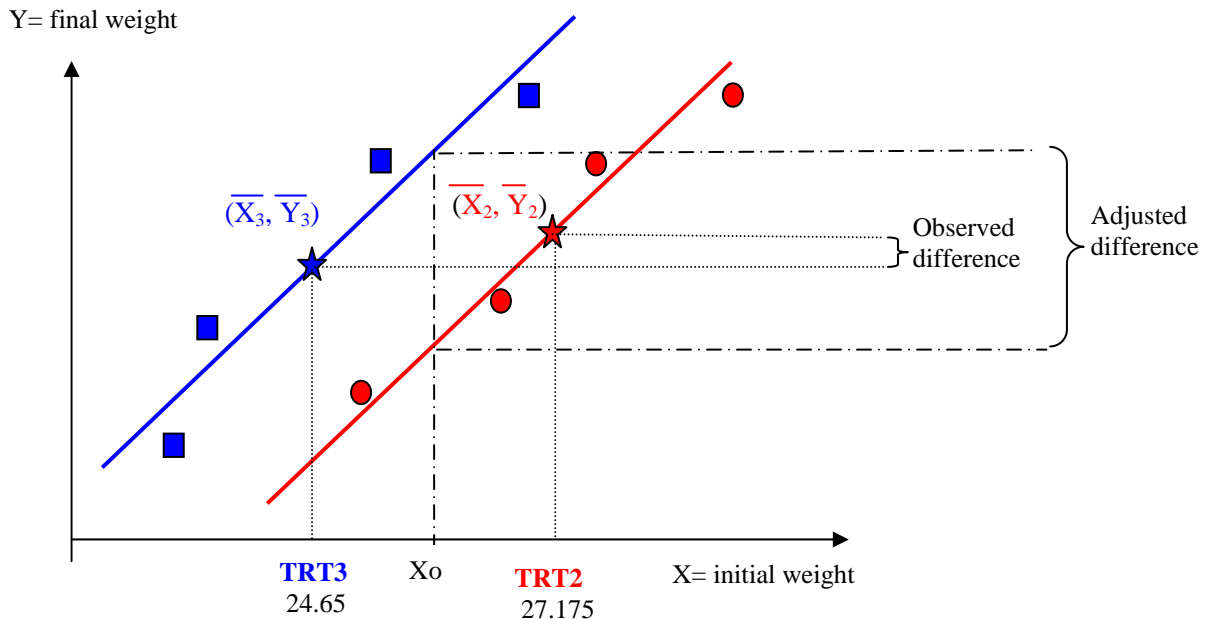
The Type I SS for TRT is the unadjusted treatment SS and is the same as the one found in the one-way ANOVA.

The **Type III** SS (12.089) is the **adjusted** treatment SS and enable us to test the treatment effects adjusted for all other factors included in the model.

Type III SS will produce the appropriate results for the ANCOVA.

- Though this SS is smaller than the unadjusted TRT SS (198.407) the reduction in the error term is even larger (from 10.684 to 0.30159).
- This allows an increase in the **F statistic** of the test from **4.642** in the simple one-way **ANOVA** to **10.02** in the **ANCOVA**.
- The power of the test for treatment differences increases when the covariate is included because most of the error in the simple ANOVA is due to variation in **INITIAL** values.
- Finally, the **INITIAL SS** is used to test the significance of the regression between **INITIAL** and FINAL once the SS of INITIAL was adjusted for TRT.

13. 3. 1. Graphic interpretation of the ANCOVA example



- The **red circles** represent the data from the four bags of **treatment 2**. The **blue squares** represent the data from **treatment 3**.
- The mean final weight of **treatment 3** (30.85) is slightly lower than the mean of **treatment 2** (31.65).
- For each treatment, **variation in X** is seen to **contribute** to **variation in Y**.
- The distance between the initial weight averages of the oysters assigned to each treatment can contribute greatly to the difference between the final average weights.
- Treatment 3 started with an average **initial weight** of 24.65 and treatment 2 with an average weight of 27.175.
- If the treatments means had been observed from some **common average X**, say X_0 , then they would be comparable. Thus the need for adjusting treatment means is apparent.
- This X_0 can be thought as a common mean obtained after moving the values of treatment 3 upwards along the regression line and the values of treatment 2 downward along the regression line.
- **The adjusted mean of treatment 3 is significantly larger than that of treatment 2.**

13. 3. 2. Least squares means or adjusted means

- **MEANS** statement: unadjusted treatment means of all continuous (non-CLASS) variables in the model. These means or their comparison are not strictly appropriate.
- To be comparable, treatment means should be adjusted to make them the best estimates of what they would have been if all treatment independent means had been the same.
- These adjusted means can be calculated in SAS by using the **LSMEANS** (least-squares means) statement.
- This statement is the same one used to obtain adjusted means for the unbalanced two-way classification. These adjusted means are obtained using the following formula:

$$LS \bar{Y}_i = \bar{Y}_i - \hat{\beta}(\bar{X}_i - \bar{X}_{..})$$

The statement: **lsmeans trt /stderr pdiff adjust=tukey;**

will print **lsmeans** followed by their **standard errors** and significance probabilities for all **pairwise tests** of treatment differences.

Note the large changes between unadjusted and adjusted treatments means:

TRT	Unadjusted Means	Adjusted LS Means	Calculation
1	34.475	30.153	34.475 - 1.08318 (29.75 - 25.76)
2	31.650	30.117	31.650 - 1.08318 (27.18 - 25.76)
3	30.850	32.052	30.850 - 1.08318 (24.65 - 25.76)
4	32.225	31.504	32.225 - 1.08318 (26.43 - 25.76)
5	25.025	30.398	25.025 - 1.08318 (20.80 - 25.76)

These changes are due to the large trt differences for the variable INITIAL. Some treatments, particularly TRT 5, received smaller oysters than other trt.

The coefficient $\hat{\beta}=1.08318 =$ weighted average of the regression coefficients of FINAL on INITIAL, for each of the 5 treatment groups.

To obtain this coefficient:

model final = trt initial/ solution;

13. 3. 3. Contrasts

The adjusted treatment means from the analysis of covariance can be analyzed further with four orthogonal contrasts implemented by the following **CONTRAST** statements:

```
proc GLM;
  class trt;
  model final= trt initial;
  contrast 'control vs. Treatment'      TRT -1 -1 -1 -1 4;
  contrast 'bottom vs. top'             TRT -1  1 -1  1 0;
  contrast 'cool vs, hot'                TRT -1 -1  1  1 0;
  contrast 'interactions depth*temp'    TRT  1 -1 -1  1 0;
```

Output:

Contrast	DF	Contrast SS	MS	F Value	Pr > F
control vs. Trt.	1	0.520	0.520	1.72	0.2103
bottom vs. top	1	0.339	0.339	1.12	0.3071
cool vs. hot	1	8.591	8.591	28.49	0.0001
interac. depth*T	1	0.229	0.229	0.76	0.3979
		9.679			

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	4	12.089	3.022	10.02	0.0005

The only significant difference is **cool vs. hot**.

These contrasts are **not orthogonal to the covariable**; hence, their sums of squares do not add to the adjusted treatment SS.

If the covariable is **not** included in the model, the same contrast statements produce completely different results!

Contrast	DF	Contrast SS	MS	F Value	Pr > F
control vs. Trt	1	169.36	169.36	15.85	0.0012
bottom vs. top	1	2.10	2.10	0.20	0.6637
cool vs. hot	1	9.30	9.30	0.87	0.3655
interac. depth*T	1	17.64	17.64	1.65	0.2183

The significance of the control vs. treatment contrast is due to the lower initial weight of the oysters placed in the control bags.

In this case the contrast SS will add to the TRT SS (198.407).

The **contrast** statement will **partition the TRT SS** resulting for the **specified model**.

ANCOVA of Y-X: Oyster Example using Final – Initial weight

TRT:5 Levels:1 2 3 4 5 Number of observations in data set = 20

ANOVA (no covariable)

Dependent Variable: **FINAL(weight)**

Source	DF	SS	MS	F Value	Pr > F
TRT	4	198.41	49.60	4.64	0.0122
Error	15	160.26	10.68		
Corrected Total	19	358.67			

Dependent Variable: **DIF (Final - Initial)**

Source	DF	SS	MS	F Value	Pr > F
TRT	4	11.98	3.00	8.74	0.0008
Error	15	5.14	0.34		
Corrected Total	19	17.13			

ANCOVA (using Initial as covariable)

Dependent Variable: **FINAL(weight)**

Source	DF	SS	MS	F Value	Pr > F
Model	5	354.45	70.89	235.05	0.0001
Error	14	4.22	0.30		
Corrected Total	19	358.67			

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	4	12.09	3.02	10.02	0.0005
INITIAL	1	156.04	156.04	517.38	0.0001

Dependent Variable: **DIF(Final - Initial)**

Source	DF	SS	MS	F Value	Pr > F
Model	5	12.90	2.58	8.56	0.0007
Error	14	4.22	0.30		
Corrected Total	19	17.13			

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	4	12.09	3.02	10.02	0.0005
INITIAL	1	0.92	0.92	3.05	0.1026

Comparison between ANCOVA and ANOVA of ratios

Effect of stress on the presence of enzyme A in the liver. The researcher measured the total activity of enzyme A from liver homogenates of 10 control and 10 shocked animals and the total amount of N to correct by the total protein present in the liver. A/N= activity of enzyme per unit of protein.

Control animals			Shocked animals		
N	A	A/N	N	A	A/N
84	76	90.4	122	108	88.5
28	38	133.9	98	158	161.2
166	72	43.4	115	58	50.0
98	64	65.3	86	65	75.5
105	53	50.0	69	40	58.0
84	28	32.8	86	65	75.5
72	31	43.0	102	82	80.3
80	28	34.3	112	94	84.1
84	28	32.7	98	65	66.3
105	49	46.1	74	76	102.7

ANOVA for the variable A/N. Dependent Variable: A/N

Source	DF	SS	MS	F Value	Pr > F
Model = TRT	1	3650.40	3650.40	3.66	0.0719
Error	18	17970.72	998.37		
Correc. Total	19	21621.12			

The ANOVA indicates that there are no significant differences between treatments.

ANCOVA for the variable A using N as covariable.

Dependent Variable: A

Source	DF	SS	MS	F Value	Pr > F
Model	2	8079.39	4039.70	6.31	0.0089
Error	17	10882.41	640.14		
Correc. Total	19	18961.80			

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	1	5108.81	5108.81	7.98	0.0117
N	1	2162.59	2162.59	3.38	0.0836

The ANCOVA shows significant differences between treatments.

- The use of ANOVA to analyze ratios $Z = Y/X$ is not that correct.
- Both X and Y have random variation
- Variation in the numerator (Y) affect Z in a lineal way but variation in the denominator affect Z in a hyperbolic way

The magnitude of the error of Z depends not only on the error of X but also on the absolute value of X .

13. 4 ANCOVA model

The ANOVA model for a CRD is $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

The regression model is: $Y_i = \mu + \beta(X_i - \bar{X}_{..}) + \varepsilon_i$

The ANCOVA model for a CRD is: $Y_{ij} = \mu + \tau_i + \beta(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij}$

The linear additive model for any given design is that for the ANOVA plus an additional term for the concomitant or independent variable.

Y generally denotes the dependent variable, whereas the variable used in the control of error and adjustment of means, is denoted by X .

The rearranged formula

$$Y_{ij} - \beta(X_{ij} - \bar{X}_{..}) = \mu + \tau_i + \varepsilon_{ij}$$

ANCOVA is a regular ANOVA of values that have been adjusted for regression on an independent variable.

13. 5 Assumptions of the ANCOVA

a. The X 's are fixed, measured without error, and independent of treatments.

This means that the measurement error is trivial relative to the observed variation, and that the treatments itself will not affect the X values.

b. The regression of Y on X after removal of the treatment differences is linear and independent of treatments.

This means that the regression is assumed to be approximately linear within the range of X values, and that the slopes of the regressions within the treatments are not significantly different.

c. The residuals are normally and independently distributed with zero mean and a common variance.

These are the normal assumptions for the validity of the F tests.

13. 5. 1. Independence of X values from the treatments

- If the covariable is measured **before** the experiment, like in the previous oyster example (13.3) the **independence** of the treatments and the concomitant variable is **always satisfied**.
- If the concomitant variable is measured **after** the experiment the independence of the covariable and the treatments **should be tested**.

An analysis of variance of the covariable using the treatments as CLASS variable is appropriate to test this hypothesis.

- The null hypothesis is that there are no significant differences among treatments for the covariable. We expect to find no significant differences in order to be able to perform a standard covariance analysis.
- The following statements are included only as an example, because the test is not required in the oyster example.

```
proc glm;  
  title 'Test for independence of treatments and  
covariable';  
  class trt;  
  model initial= trt;  
run; quit;
```

Dependent Variable: INITIAL

Source	DF	SS	MS	F Value	Pr > F
TRT (Model)	4	176.793	44.198	4.98	0.0093
Error	15	132.995	8.866		

R-Square	Coeff Var	Root MSE	Initial Mean
0.570690	11.55916	2.977639	25.76

- In this case the differences in initial weight are highly significant.
- The selected covariable will most likely have an effect in the final results.

Covariance can be used where the X values are affected by the treatments but it must be interpreted with caution.

Values in **red** will be used later for a correction in the **calculation of the efficiency** of the ANCOVA relative to the ANOVA

The **Mean of the covariable** will be used to obtain the corrected Y values

13. 5. 2. Test for heterogeneity of slopes.

- The adjustment of the Y values using a single β for all treatments is based on the assumption of a constant regression relationship among groups
- The test for **heterogeneity of slopes tests** the validity of this assumption
- The null hypothesis is $H_0: \beta_1 = \beta_2 = \dots = \beta_i$

Regression relationships that differ among treatments reflect an **interaction between the treatment groups and the covariates**.

The **GLM** procedure analyzes this phenomenon as an interaction.

The expression **$X*TRT$** produces the appropriate statistics for estimating different regressions of Y on X for the different classes specified by TRT.

For the previous example (Topic 13.3) the appropriate statements are

```
proc glm;
  title 'Test for heterogeneity of slopes';
  class trt;
  model final= trt initial trt*initial;
run; quit;
```

Output:

Dependent Variable: FINAL

Source	DF	SS	MS	F Value	Pr > F	
Model		9	355.84	39.54	139.51	0.0001
Error		10	2.83	0.28		
Corrected Total		19	358.67			

Source	DF	Type I SS	MS	F Value	Pr > F
TRT	4	198.407	49.60	175.02	0.0001
INITIAL	1	156.040	156.04	550.60	0.0001
INITIAL*TRT	4	1.388	0.35	1.22	0.3602

Source	DF	Type III SS	MS	F Value	Pr > F
TRT	4	1.697	0.424	1.50	0.2752
INITIAL	1	68.529	68.529	241.81	0.0001
INITIAL*TRT	4	1.388	0.347	1.22	0.3602

The last row **INITIAL*TRT** is an additional SS due to different regression coefficients for the groups specified by TRT.

If **INITIAL*TRT** $P > 0.05 \Rightarrow$ don't reject hypothesis of slope homogeneity.

13. 5. 3. Analysis of residuals.

Similar statements to those used in previous analysis can be included to check the normality and homogeneity of variance of the residuals:

```
proc GLM;
  class trt;
  model final= trt initial;
  output out=check p=predi r=resi;
proc univariate data=check normal;
  var resi;
proc plot;
  plot resi*predi=trt;
run; quit;
```

13. 6. Increase in precision due to covariance (ST&D 17.6)

- To test the **effectiveness of covariance** to control error we compare the variance of treatment mean with and without the covariance adjustment.
- The **ANOVA** on the unadjusted Y's has error MS 10.68417 with **15** df
- The **ANCOVA** on the adjusted Y's has error MS 0.30159 with **14** df
- This last value must be adjusted upward to allow for sampling error in the regression coefficient.
- The adjustment involves the TRT SS (176.793) and the error SS (132.995) from an ANOVA on X.

The **effective error MS** after adjustment for X is given by

$$MS_{\text{Error Adjusted Y}} \left[1 + \frac{TRT \ SS_X \ variable}{(t-1) \ Error \ SS_X \ variable} \right] = 0.30159 \left[1 + \frac{176.793}{4 * 132.995} \right] = 0.402$$

An estimate of the relative precision is :

$$MS_{\text{Error Unadjusted Y}} / MS_{\text{Error Adjusted Y}} = 10.68417 / 0.402 = \mathbf{26.6},$$

This indicates that each replication in the **ANCOVA** is as effective as **26.6** replications in the **ANOVA**.

Assumptions

Data Adjust;

Input Block Variety X Y @@;

Z= Y - 8.39163498 * (X - 4.98); * Z is adjusted Y;

Cards;

```
1 1 4.6 87.1 2 1 4.7 92.1
1 2 5.1 93.1 2 2 5.1 95.5
1 3 4.8 89.8 2 3 5.2 99.3
1 4 4.4 91.4 2 4 4.9 93.4
1 5 5.9 99.5 2 5 5.1 94.4;
```

Proc GLM Data = Adjust; * Model to test ANOVAS;

Class Block Variety;

Model X Y = Block Variety;

Proc GLM Data = Adjust;* The ANCOVA;

Class Block Variety;

Model Y = Block Variety X / solution;

lsmeans var;

Proc GLM Data = Adjust; * The ANOVA of Z;

Class Block Variety;

Model Z = Block Variety; * Produces similar P ANCOVA;

Output Out = ZZ p = PZ r = RZ;

means var; * Produces = result as LSMEANS in ANCOVA;

Proc GLM Data = ZZ;

Class Variety;

Model Z = Variety;

means Variety / hovtest= Levene; * Levene's test adjusted Z;

Proc GLM Data = ZZ;

Class Block Variety;

Model Z = Block Variety PZ*PZ; *Tukey test non-additivity Z;

Run; quit;

ANOVA Dependent Variable: X

Source	DF	SS	MS	F Value	Pr > F
Model	5	1.01	0.20	1.54	0.3492
Error	4	0.53	0.13		
Corrected Total	9	1.54			

R-Square	Coeff Var	Root MSE	X
Mean			
0.657552	7.281713	0.362629	4.98

solution

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	52.00817490 B	19.46736895	2.67	0.0756
Block 1	-2.42433460 B	1.61546874	-1.50	0.2304
Block 2	0.00000000 B	.	.	.
Variety 1	-0.21711027 B	3.92035053	-0.06	0.9593
Variety 2	0.70665399 B	2.90598043	0.24	0.8236
Variety 3	1.79581749 B	3.09073299	0.58	0.6020
Variety 4	2.58288973 B	3.92035053	0.66	0.5571
Variety 5	0.00000000 B	.	.	.
X	8.39163498	3.50857389	2.39	0.0966

13. 8. Uses of ANCOVA (ST&D p 429)

1. To **control error** and increase precision
2. To **adjust treatment means** of the dependent variable for differences in sets of values of corresponding independent variables
3. **To assist in the interpretation of data**, especially with regard to the nature of treatment effects
4. To estimate missing data

13. 8. 1. Error control.

Experimental design, a covariable, or both can be used to **reduce MS error**

Covariance can be used to reduce error when variation in the dependent variable Y is partly attributable to variation in the independent variable X .

Removes, by regression, certain recognized effects that cannot be or have not been controlled effectively by experimental design.

Example: cattle-feeding experiment to compare the effects of several rations on gain in weight, using initial weight as covariable.

13. 8. 3. To adjust treatment means

When variation in Y is partly attributable to variation in X , variation in treatment \bar{Y} must also be affected by variation among \bar{X} .

To be comparable, treatment \bar{Y} should be **adjusted** to make them the best estimates of what they would have been if all treatment \bar{X} had been =.

Example1: canning peas yield increases rapidly with increase in **maturity**. **Maturity can be used as a covariate**. A comparison of yields adjusted for maturity differences would be more meaningful than a comparison among unadjusted yields.

Example2: In animal feeding experiments, differences among unadjusted treatment means may be due to differences in the nutritive value of the rations, to differences in the amounts consumed, or to both.

If differences among mean gains in weight for the different rations are adjusted to a common food intake, the adjusted means indicate whether or not the rations differ in nutritive value.

13. 7. 3. Interpretation of data.

Covariance analysis often aids the experimenter in understanding the principles underlying the results of an investigation.

If the **independent variable** is **influenced by the treatments**, the **interpretation of the data is changed**.

Adjustment removes part of the treatment effects when means of the independent variable are affected by treatments. The adjusted treatment means estimate values expected when treatment means for X are =.

Example 1: In a **fertilizer trial** on sugar beets, the treatments may cause **differences in stand**. When stand, the **independent variable**, is influenced by treatments, the analysis of yield adjusted for stand differences removes part of the treatment effect.

An analysis of covariance can still supply useful information.

Total yield= average weight per beet x No. of beet (stand).

If stand is influenced by treatments, the **ANCOVA** of yield **adjusted for stand** differences would indicate treatments affects in weight per beet.

Example 2: An adjustment in proportion to the **number of plants** is sometimes practiced. The use of ratios is not recommended, because it usually results in an over-correction for the plots with smallest stand since yields are rarely proportional to the number of plants per plot. The analysis of covariance provides a more satisfactory and appropriate method of adjusting the experimental data.

In situations where real **differences among treatments for the independent variable** do occur but are **not the direct effect of the treatments**, **adjustment is warranted**. For example, consider a variety trial for which seed of the various varieties or strains has been produced in different areas. Such seed may differ widely in germination, not because of inherent differences, but as a result of the environment in which it was grown. Differences in stand may occur even if planting rate is controlled. In this situation, the use of covariance for both error control and yield adjustment is warranted.

13. 7. 4. Estimation of missing data: The use of covariance to estimate missing plots results in a minimum residual sum of squares and an unbiased treatment sum of squares. The covariance procedure is simple to carry out though more difficult to describe than previous procedures which required little more than a formula.

Interpretation of examples of ANCOVA

1) Effect of a fertilizer treatment on sugar beets.

Variable: yield

Covariable: number of plants.

	Treatment	Treatment	Treatment
ANOVA	significant	significant	NS
ANCOVA	significant	NS	significant

2) Effect of different cat foods on weight gain during the first year.

Variable: gained weight

Covariable: food intake.

	Treatment	Treatment	Treatment
ANOVA	significant	significant	NS
ANCOVA	significant	NS	significant

3) Effect of two storage protein genes on bread loaf volume (2x2 factorial)

Variables: loaf volume

Covariable: protein content

	Treatment	Treatment	Treatment
ANOVA	G1= S G2=S	G1=NS G2=NS	G1=S G2=S
ANCOVA	G1=NS G2=NS	G1=S G2=S	G1=S G2=S

4) Effect of hardness genes and environment in grain hardness

Variables: grain hardness

Covariable: grain weight

	Treatment	Treatment
ANOVA	G=S E=S GxE=S	G=NS E=S GxE=S
ANCOVA	G=S E=NS GxE=NS	G=S E=NS GxE=S