

Chapter 2 Analysis of Variance (ANOVA)

2.1 Two-Sample t Tests in the TTEST Procedure

Objectives

- Use the TTEST procedure to analyze the differences between two population means.
- Verify the assumptions of a two-sample t test.


3

Test Score Data Set, Revisited

Gender	SATScore	IDNumber
Male	1170	61469897
Female	1090	33081197
Male	1240	68137597
Female	1000	37070397
Male	1210	64608797
Female	970	60714297
Male	1020	16907997
Female	1490	9589297
Male	1200	93891897
Female	1260	5859397
...

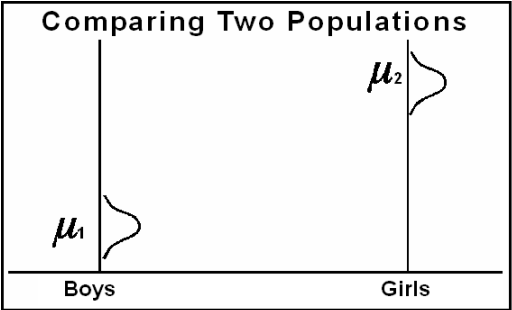
4

Recall the study in the previous chapter by students in Ms. Chao's statistics class. The Board of Education set a goal of having their graduating class scoring, on average, 1200 on the SAT. The students then investigated whether the school district met its goal by drawing a sample of 80 students at random. The conclusion was that it was reasonable to assume that the mean of all magnet students was, in fact, 1200. However, when they planned the project, an argument arose between the boys and the girls about whether boys or girls scored higher. Therefore, they also collected information about gender to test for differences.



Assumptions

Comparing Two Populations



Boys Girls

- independent observations
- normally distributed data for each group
- equal variances for each group

5

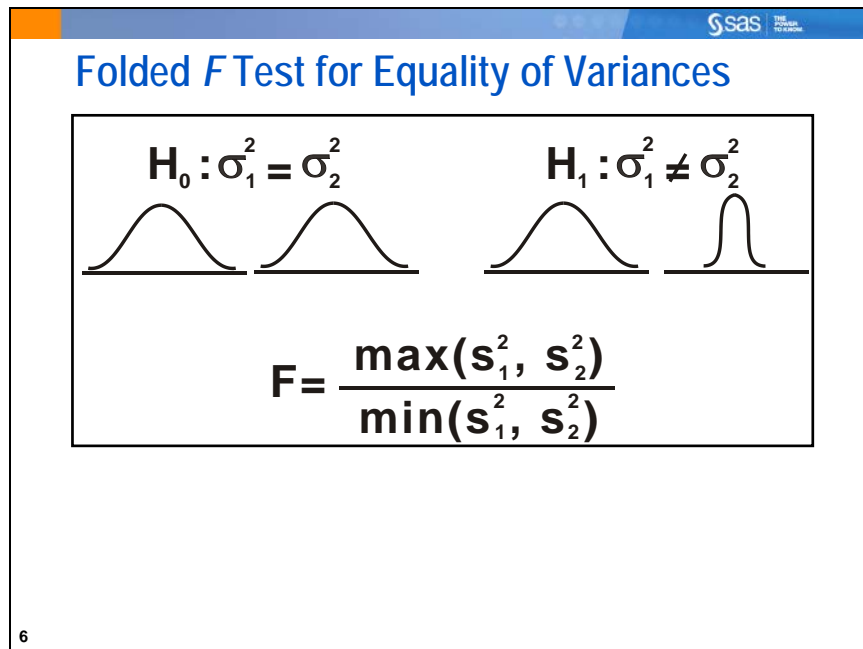
Before you start the analysis, examine the data to verify that the statistical assumptions are valid.

The assumption of independent observations means that no observations provide any information about any other observation that you collect. For example, measurements are not repeated on the same subject. This assumption can be verified during the design stage.

The assumption of normality can be relaxed if the data are approximately normally distributed or if enough data are collected. This assumption can be verified by examining plots of the data.

There are several tests for equal variances. If this assumption is not valid, an approximate t test can be performed.

If these assumptions are **not** valid and no adjustments are made, the probability of drawing incorrect conclusions from the analysis could increase.



To evaluate the assumption of equal variances in each group, you can use the Folded F test for equality of variances. The null hypothesis for this test is that the variances are equal. The F value is calculated as a ratio of the greater of the two variances divided by the lesser of the two. Thus, if the null hypothesis is true, F tends to be close to 1.0 and the p -value for F is statistically nonsignificant ($p > 0.05$).

This test is valid *only* for independent samples from normal distributions. Normality is required even for large sample sizes.

If your data are not normally distributed, you can look at plots to determine whether the variances are approximately equal.

If you reject the null hypothesis, it is recommended that you use the unequal variance t test in the PROC TTEST output for testing the equality of group means.

The TTEST Procedure

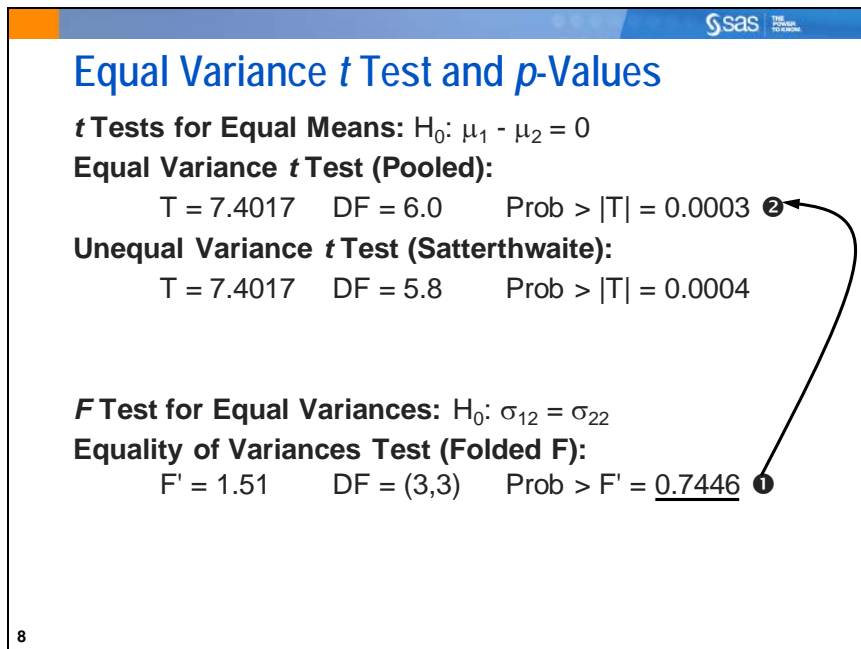
General form of the TTEST procedure:

```
PROC TTEST DATA=SAS-data-set;  
  CLASS variable;  
  VAR variables;  
  PAIRED variable1*variable2;  
RUN;
```

7

Selected TTEST procedure statements:

- | | |
|--------|---|
| CLASS | specifies the two-level variable for the analysis. Only one variable is allowed in the CLASS statement. |
| VAR | specifies numeric response variables for the analysis. If the VAR statement is not specified, PROC TTEST analyzes all numeric variables in the input data set that are not listed in a CLASS (or BY) statement. |
| PAIRED | specifies pairs of numeric response variables from which difference scores (variable1-variable2) are calculated. A one-sample t test is then performed on the difference scores. |
- If the CLASS statement and PAIRED statement are omitted, PROC TTEST performs a one-sample t test.
 - When the CLASS statement is present, a two-sample test is performed.
 - When a PAIRED statement is present instead, a paired t test is performed.

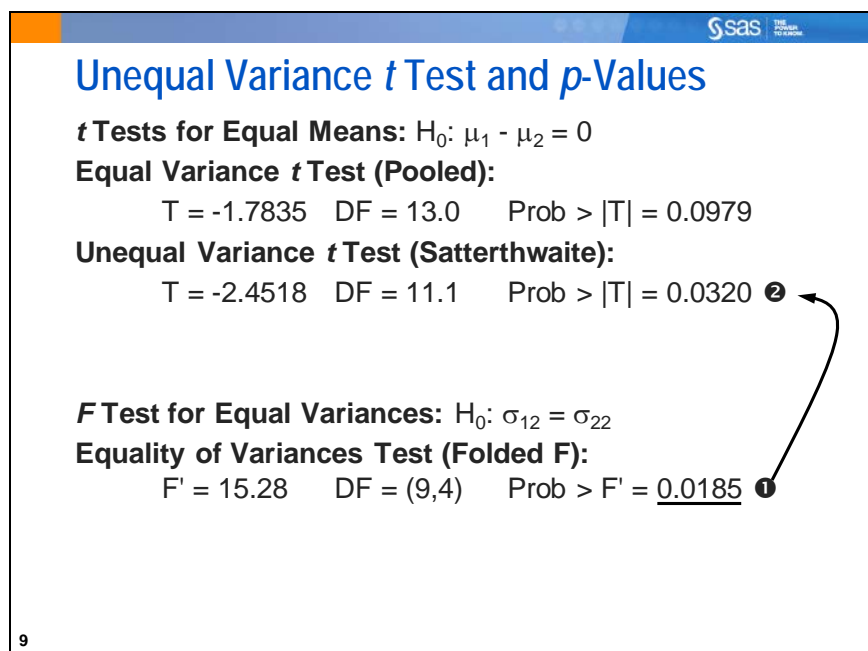


- ① Check the assumption of equal variances and then use the appropriate test for equal means. Because the *p*-value of the test *F* statistic is 0.7446, there is not enough evidence to reject the null hypothesis of equal variances.
- ② Therefore, use the equal variance *t*-test line in the output to test whether the means of the two populations are equal.

The null hypothesis that the group means are equal is rejected at the 0.05 level. You conclude that there is a difference between the means of the groups.




The equality of variances *F* test is found at the bottom of the PROC TTEST output.



- ❶ Again, check the assumption of equal variances and then use the appropriate test for equal means. Because the p -value of the test F statistic is less than $\alpha=0.05$, there is enough evidence to reject the null hypothesis of equal variances.
- ❷ Therefore, use the unequal variance t -test line in the output to test whether the means of the two populations are equal.

The null hypothesis that the group means are equal is rejected at the 0.05 level.

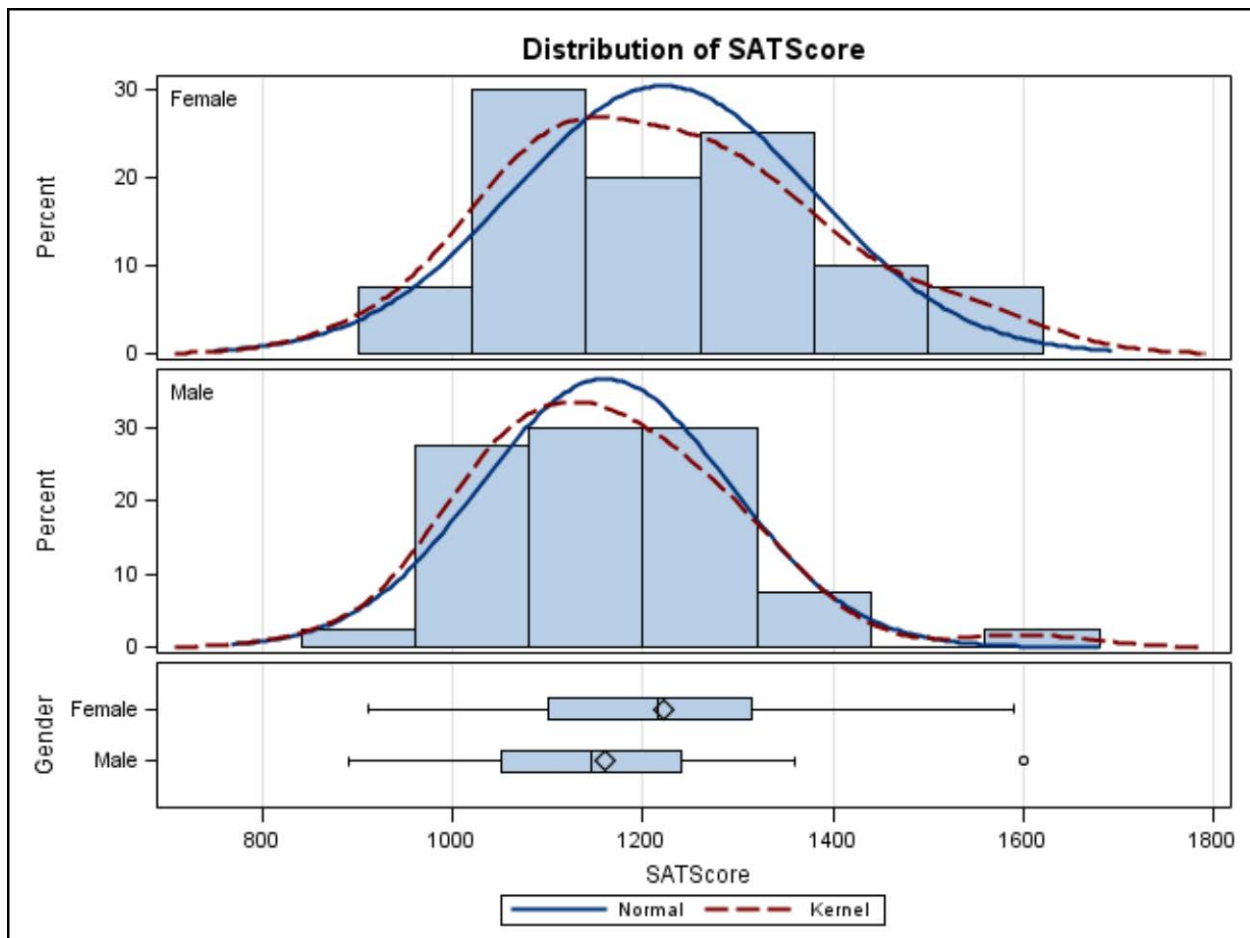
-  If you choose the equal variance t test, you would **not** reject the null hypothesis at the 0.05 level. This shows the importance of choosing the appropriate t test.

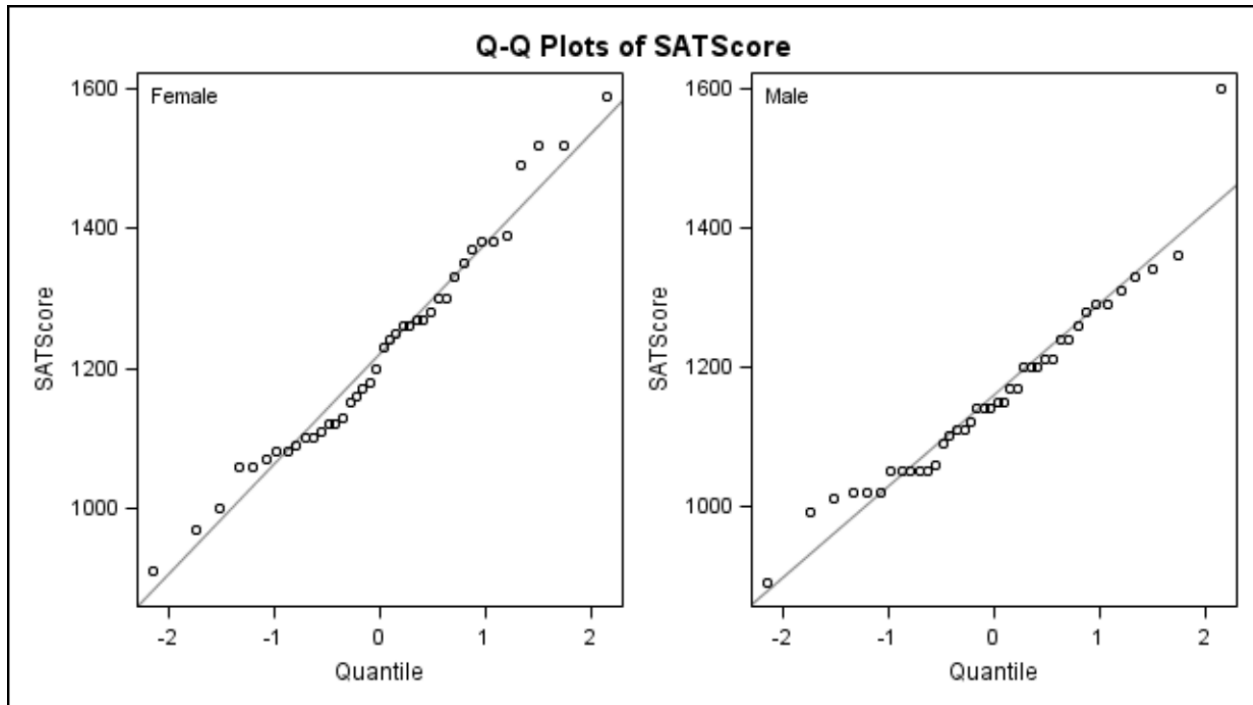


Two-Sample t Test

```
/*st102d01.sas*/
proc ttest data=sasuser.TestScores plots(shownull)=interval;
  class Gender;
  var SATScore;
  title "Two-Sample t-test Comparing Girls to Boys";
run;
```

First, it is advisable to verify the assumptions of t tests. There is an assumption of normality of the distribution of each group. This assumption can be verified with a quick check of the Summary panel and Q-Q plot.





The Q-Q plot (quantile-quantile plot) is similar to the Normal Probability plot that you saw earlier. The X-axis for this plot is scaled as quantiles, rather than probabilities. For each group it seems that the data approximate a normal distribution. There seems to be one potential outlier, a male scoring a perfect 1600 on the SAT, when no other male scored greater than 1400.



If assumptions are not met, you can do an equivalent nonparametric test, which does not make distributional assumptions. PROC NPAR1WAY is one procedure for performing this type of test. It is described in an appendix.

The statistical tables for the TTEST procedure are displayed below.

❶	Gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
	Female	40	1221.0	157.4	24.8864	910.0	1590.0
	Male	40	1160.3	130.9	20.7008	890.0	1600.0
	Diff (1-2)		60.7500	144.8	32.3706		

	Gender	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
	Female		1221.0	1170.7 1271.3	157.4	128.9 202.1
	Male		1160.3	1118.4 1202.1	130.9	107.2 168.1
❷	Diff (1-2)	Pooled	60.7500	-3.6950 125.2	144.8	125.2 171.7
	Diff (1-2)	Satterthwaite	60.7500	-3.7286 125.2		

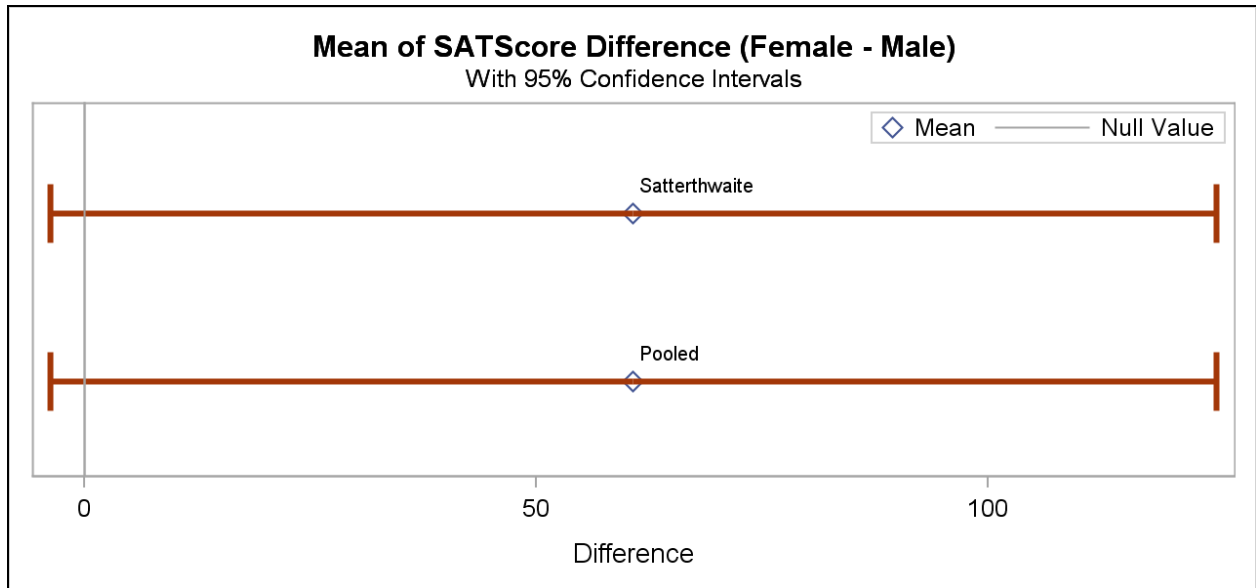
	Method	Variances	DF	t Value	Pr > t
❸	Pooled	Equal	78	1.88	0.0643
	Satterthwaite	Unequal	75.497	1.88	0.0644

Equality of Variances					
	Method	Num DF	Den DF	F Value	Pr > F
❹	Folded F	39	39	1.45	0.2545

- ❶ In the Statistics table, examine the descriptive statistics for each group and their differences.
- ❷ Look at the Equality of Variances table that appears at the bottom of the output. The F test for equal variances has a p -value of 0.2545. Because this value is greater than the alpha level of 0.05, do not reject the null hypothesis of equal variances (This is equivalent to saying that there is insufficient evidence to indicate that the variances are not equal.)
- ❸ Based on the F test for equal variances, you then look in the T-Tests table at the t test for the hypothesis of equal means. Using the equal variance (Pooled) t test, you do not reject the null hypothesis that the group means are equal. The mean difference between boys and girls is 60.75. However, because the p -value is greater than 0.05 ($\text{Pr}>|t|=0.0643$), you conclude that there is no significant difference in the average SAT score between boys and girls.



The confidence interval for the mean difference (-3.6950, 125.2) includes 0. This implies that you cannot say with 95% confidence that the difference between boys and girls is *not* zero. Therefore, it also implies that the p -value is greater than 0.05.



Confidence intervals are shown in the output object titled Difference Interval Plot. Because the variances here are so similar between males and females, the Pooled and Satterthwaite intervals (and p -values) are very similar. The lower bound of the Pooled interval extends past zero.



A good argument can be made that the point estimate for the difference between males and females is big from a practical standpoint. If the sample were a bit larger, that same difference might be significant because the pooled standard error would be smaller. However, such a sample needs to be drawn to confirm this hypothesis.



Exercises

1. Using PROC TTEST for Comparing Groups

Elli Sagerman, a Masters of Education candidate in German Education at the University of North Carolina at Chapel Hill in 2000, collected data for a study. She looked at the effectiveness of a new type of foreign language teaching technique on grammar skills. She selected 30 students to receive tutoring; 15 received the new type of training during the tutorials and 15 received standard tutoring. Two students moved away from the district before completing the study. Scores on a standardized German grammar test were recorded immediately before the 12-week tutorials and then again 12 weeks later at the end of the trial. Sagerman wanted to see the effect of the new technique on grammar skills. The data are in the **SASUSER.GERMAN** data set.

Change Change in grammar test scores

Group The assigned treatment, coded **Treatment** and **Control**

Analyze the data using PROC TTEST. Assess whether the treatment group improved more than the control group.

- a. Do the two groups appear to be approximately normally distributed?
- b. Do the two groups have approximately equal variances?
- c. Does the new teaching technique seem to result in significantly different change scores compared with the standard technique?

2.01 Multiple Answer Poll

How do you tell PROC TTEST that you want to do a two-sample *t* test?

- a. SAMPLE=2 option
- b. CLASS statement
- c. GROUPS=2 option
- d. PAIRED statement

2.2 One-Way ANOVA

Objectives

- Use the GLM procedure to analyze the differences between population means.
- Verify the assumptions of analysis of variance.

17

Overview of Statistical Models

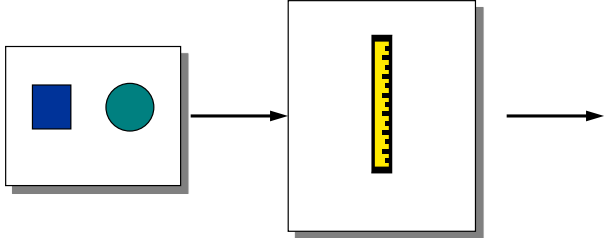
Type of Response \ Type of Predictors	Type of Predictors		
	Categorical	Continuous	Continuous and Categorical
Continuous	Analysis of Variance (ANOVA)	Ordinary Least Squares (OLS) Regression	Analysis of Covariance (ANCOVA)
Categorical	Contingency Table Analysis or Logistic Regression	Logistic Regression	Logistic Regression

18

Overview

Are there any differences among the population means?

Response




19

Analysis of variance (ANOVA) is a statistical technique used to compare the means of two or more groups of observations or treatments. For this type of problem, you have the following:

- a continuous dependent variable, or *response* variable
- a discrete independent variable, also called a *predictor* or *explanatory* variable.

Research Questions for One-Way ANOVA

Do accountants, on average, earn more than teachers? *



*** Is this a case for a *t* test?**


20

If you analyze the difference between two means using ANOVA, you reach the same conclusions as you reach using a pooled, two-group *t* test. Performing a two-group mean comparison in PROC GLM gives you access to different graphical and assessment tools than performing the same comparison in PROC TTEST.


sas THE POWER OF DATA

Research Questions for One-Way ANOVA


Do people treated with one of two new drugs have higher average T-cell counts than people in the control group?



Placebo



Treatment 1



Treatment 2

21

When there are three or more levels for the grouping variable, a simple approach is to run a series of t tests between all the pairs of levels. For example, you might be interested in T-cell counts in patients taking three medications (including one placebo). You could simply run a t test for each pair of medications. A more powerful approach is to analyze all the data simultaneously. The mathematical model is called a *one-way analysis of variance* (ANOVA), and the test statistic is the F ratio, rather than the Student's t value.


sas THE POWER OF DATA

Research Questions for One-Way ANOVA

Do people spend different amounts depending on which type of credit card they have?

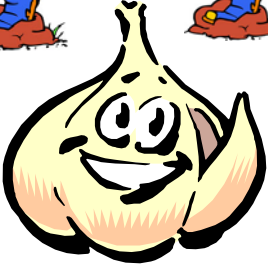




22



Research Questions for One-Way ANOVA

Does the type of fertilizer used affect the average weight of garlic grown at the Montana Gourmet Garlic Ranch?



23

Garlic Example



24

Example: Montana Gourmet Garlic is a company that grows garlic using organic methods. It specializes in hardneck varieties. Knowing a little about experimental methods, the owners design an experiment to test whether growth of the garlic is affected by the type of fertilizer used. They limit the experimentation to a Rocambole variety named Spanish Roja, and test three organic fertilizers and one chemical fertilizer (as a control). They blind themselves to the fertilizer by using containers with numbers 1 through 4. (In other words, they design the experiment in such a way that they do not know which fertilizer is in which container.) One acre of farmland is set aside for the experiment. It is divided into 32 beds. They randomly assign fertilizers to beds. At harvest, they calculate the average weight of garlic bulbs in each of the beds. The data are in the **sasuser.MGGarlic** data set.

These are the variables in the data set:

Fertilizer	The type of fertilizer used (1 through 4)
BulbWt	The average garlic bulb weight (in pounds) in the bed
Cloves	The average number of cloves on each bulb
BedID	A randomly assigned bed identification number



Descriptive Statistics across Groups

Example: Print the data in the **sasuser.MGGarlic** data set and create descriptive statistics.

```
/*st102d02.sas*/ /*Part A*/
proc print data=sasuser.MGGarlic (obs=10);
  title 'Partial Listing of Garlic Data';
run;
```

Part of the data is shown below.

Obs	Fertilizer	BulbWt	Cloves	BedID
1	4	0.20901	11.5062	30402
2	3	0.25792	12.2550	23423
3	2	0.21588	12.0982	20696
4	4	0.24754	12.9199	25412
5	1	0.24402	12.5793	10575
6	3	0.20150	10.6891	21466
7	1	0.20891	11.5416	14749
8	4	0.15173	14.0173	25342
9	2	0.24114	9.9072	20383
10	3	0.23350	11.2130	23306

Next, look at the fertilizer groups separately.

```
/*st102d02.sas*/ /*Part B*/
proc means data=sasuser.MGGarlic printalltypes maxdec=3;
  var BulbWt;
  class Fertilizer;
  title 'Descriptive Statistics of Garlic Weight';
run;

/*st102d02.sas*/ /*Part C*/
proc sgplot data=sasuser.MGGarlic;
  vbox BulbWt / category=Fertilizer datalabel=BedID;
  format BedID 5.;
  title "Box and Whisker Plots of Garlic Weight";
run;
```

Selected PROC MEANS statement option:

PRINTALLTYPES displays all requested combinations of class variables (all **_TYPE_** values) in the printed or displayed output.

Selected PROC MEANS statement:

CLASS variable(s) specifies the variables whose values define the subgroup combinations for the analysis. Class variables are numeric or character and can have continuous values, but they typically have a few discrete values that define levels of the variable. You do not have to sort the data by class variables.

Selected SGPLOT VBOX statement option:

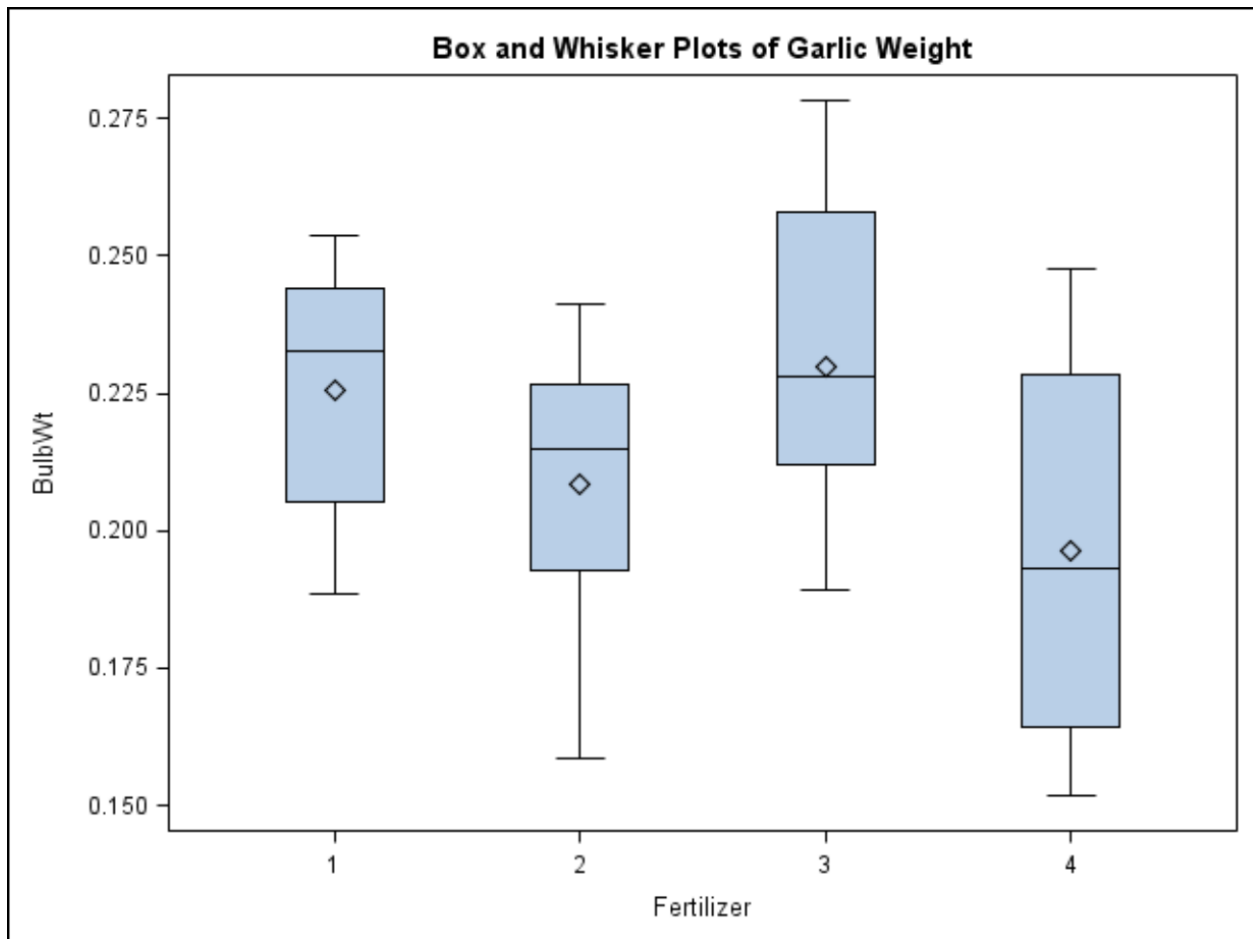
CATEGORY= produces separate box plots for each level of the variable listed.

Analysis Variable : BulbWt					
N					
Obs	N	Mean	Std Dev	Minimum	Maximum
32	32	0.219	0.029	0.152	0.278

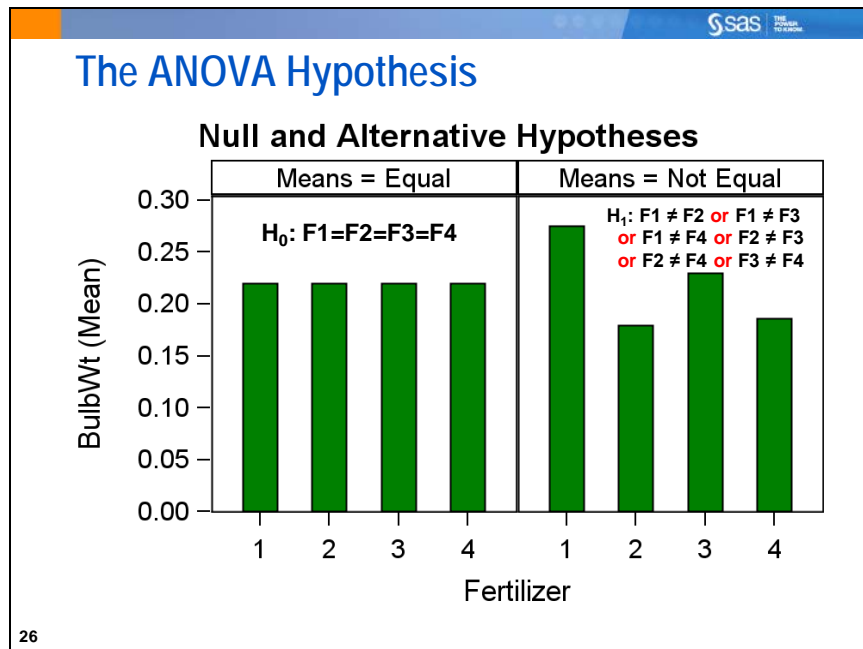
Analysis Variable : BulbWt						
Fertilizer	N					
	Obs	N	Mean	Std Dev	Minimum	Maximum
1	9	9	0.225	0.025	0.188	0.254
2	8	8	0.209	0.026	0.159	0.241
3	11	11	0.230	0.026	0.189	0.278
4	4	4	0.196	0.041	0.152	0.248



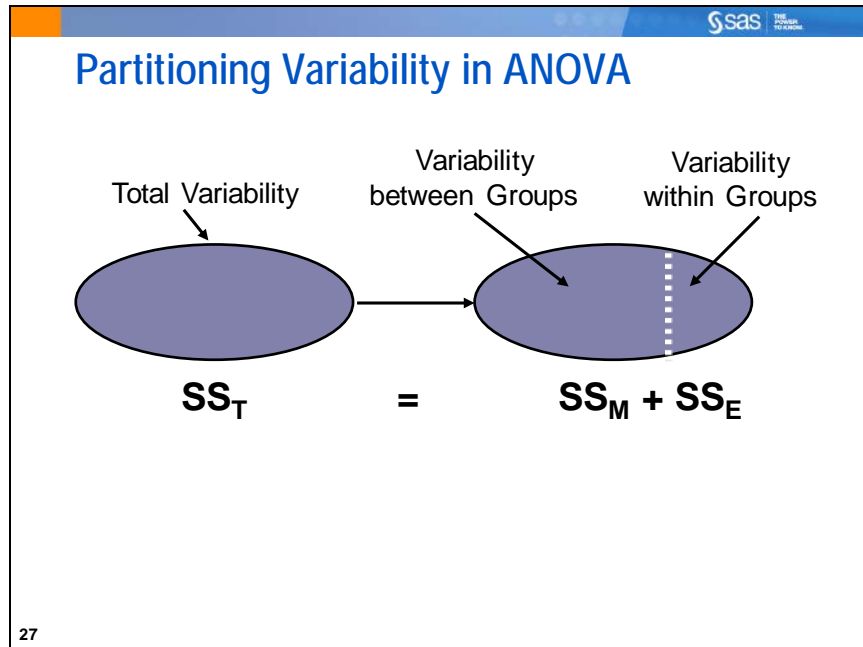
The design is not *balanced*. In other words, the groups are not equally sized.



Caution should be exercised when viewing the box plots since there are few observations per group, increasing variability.



Small differences between sample means are usually present. The objective is to determine whether these differences are statistically significant. In other words, is the difference more than what might be expected to occur by chance?



In ANOVA, the Total Variation (as measured by the corrected total sum of squares) is partitioned into two components, the Between Group Variation (displayed in the ANOVA table as the Model Sum of Squares) and the Within Group Variation (displayed as the Error Sum of Squares). As its name implies, ANalysis Of VAriance analyzes, or breaks apart, the variance of the dependent variable to determine whether the between-group variation is a significant portion of the total variation. ANOVA compares the portion of variation in the response variable attributable to the grouping variable to the portion of variability that is unexplained. The test statistic, the *F* Ratio, is only a ratio of the model variance to the error variance. The calculations are shown below.

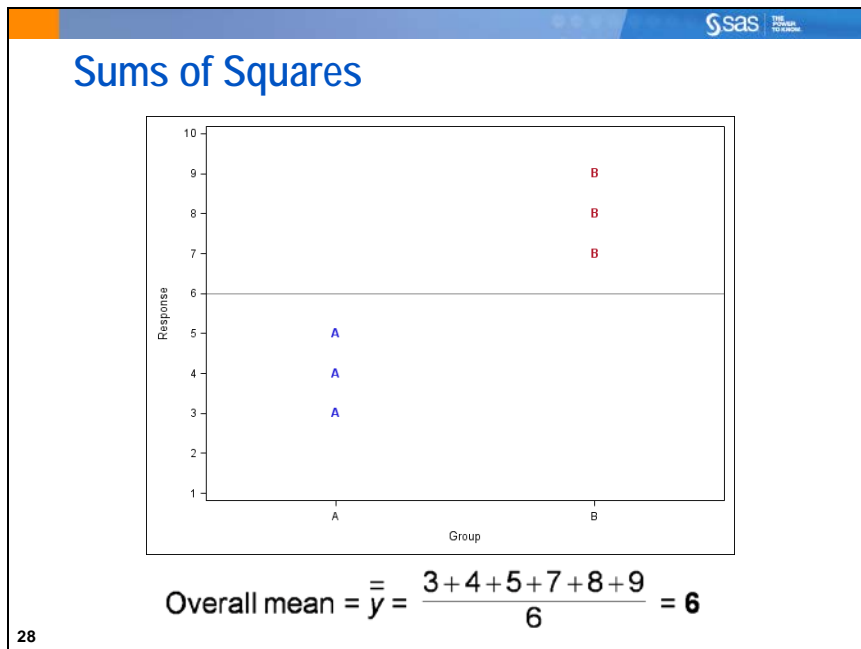
Total Variation the *overall* variability in the response variable. It is calculated as the sum of the squared differences between each observed value and the overall mean, $\sum \sum (Y_{ij} - \bar{\bar{Y}})^2$. This measure is also referred to as the *Total Sum of Squares* (SS_T).

Between Group Variation the variability explained by the independent variable and therefore represented by the between treatment sums of squares. It is calculated as the weighted (by group size) sum of the squared differences between the mean for each group and the overall mean, $\sum n_i (\bar{Y}_i - \bar{\bar{Y}})^2$. This measure is also referred to as the *Model Sum of Squares* (SS_M).

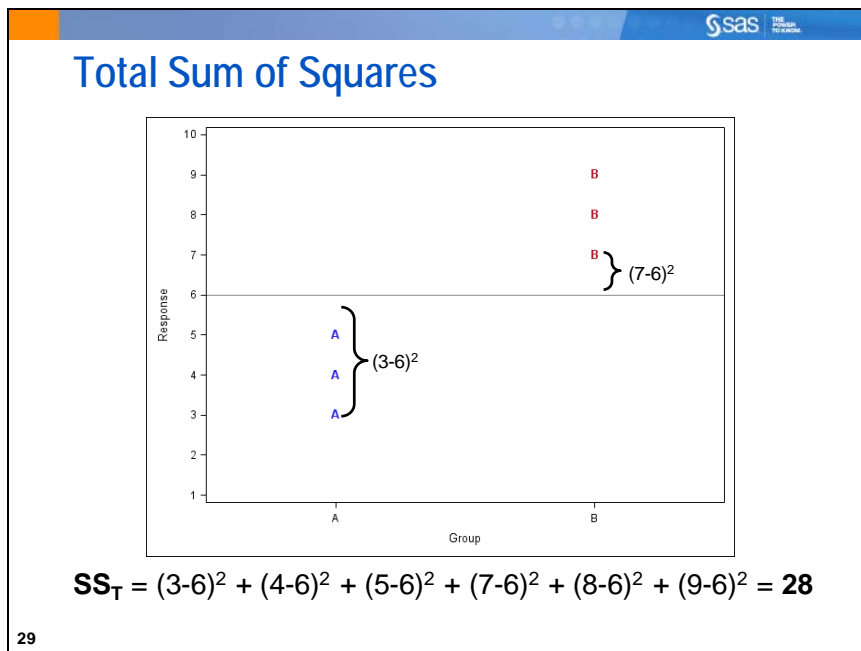
Within Group Variation the variability not explained by the model. It is also referred to as *within treatment variability* or *residual sum of squares*. It is calculated as the sum of the squared differences between each observed value and the mean for its group, $\sum \sum (Y_{ij} - \bar{Y}_i)^2$. This measure is also referred to as the *Error Sum of Squares* (SS_E).



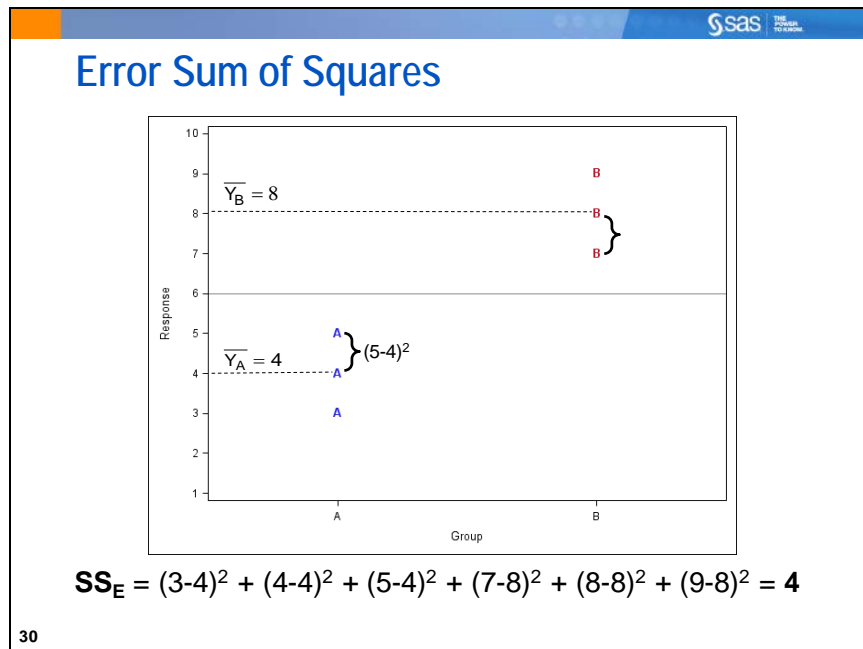
$SS_T = SS_M + SS_E$, meaning that the model sum of squares and the error sum of squares sums to the total sum of squares.



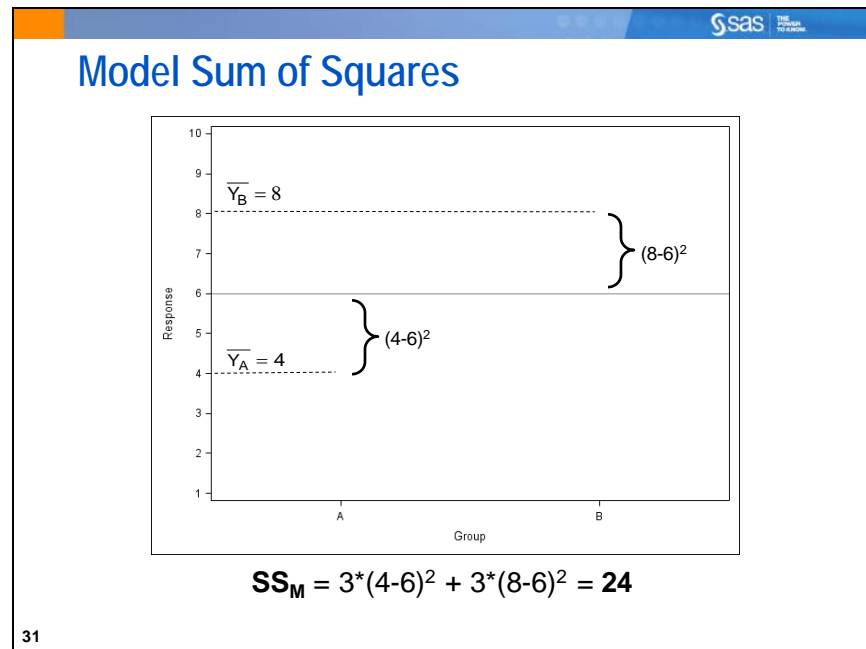
A simple example of the various sums of squares is shown in this set of slides. First, the overall mean of all data values is calculated.



The total sum of squares, SS_T , is a measure of the total variability in a response variable. It is calculated by summing the squared distances from each point to the overall mean. Because it is correcting for the mean, this sum is sometimes called the *corrected total sum of squares*.



The error sum of squares, SS_E , measures the random variability *within* groups; it is the sum of the squared deviations between observations in each group and that group's mean. This is often referred to as the *unexplained variation* or *within-group variation*.

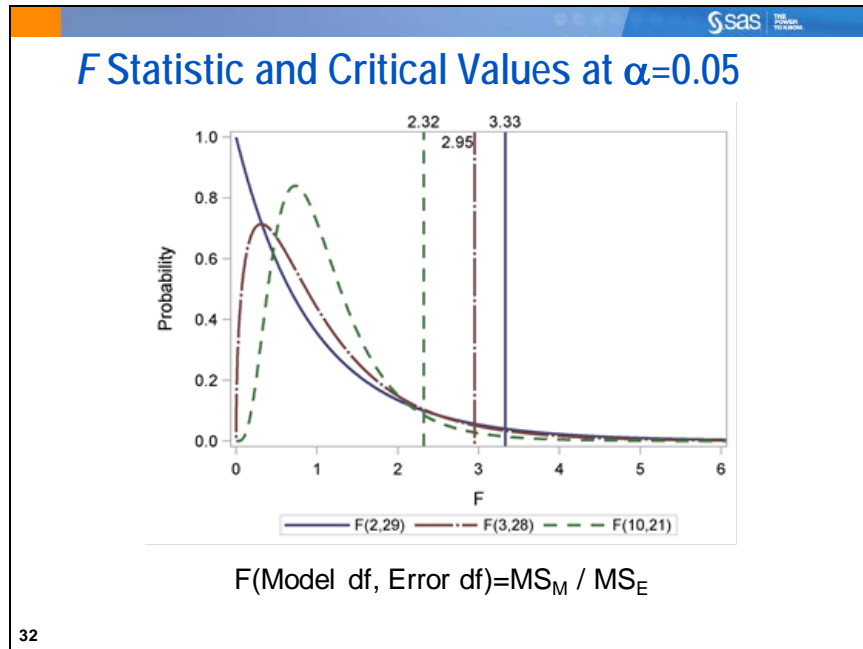


The model sum of squares, SS_M , measures the variability *between* groups; it is the sum of the squared deviations between each group mean and the overall mean, weighted by the number of observations in each group. This is often referred to as the *explained variation*. The model sum of squares can also be calculated by subtracting the error sum of squares from the total sum of squares: $SS_M = SS_T - SS_E$.

In this example, the model explains approximately 85.7%, $((SS_M / SS_T) * 100)\%$, of the variability in the response. The other 14.3% represents unexplained variability, or process variation. In other words, the variability due to differences between the groups (the explained variability) makes up a larger proportion of the total variability than the random error within the groups (the unexplained variability).

The total sum of squares (SS_T) refers to the *overall* variability in the response variable. The SS_T is computed under the null hypothesis (that the group means are all the same). The error sum of squares (SS_E) refers to the variability *within* the treatments not explained by the independent variable. The SS_E is computed under the alternative hypothesis (that the model includes nonzero effects). The model sum of squares (SS_M) refers to the variability *between* the treatments explained by the independent variable.

The basic measures of variation under the two hypotheses are transformed into a ratio of the model and the error variances that has a known distribution (a sample statistic, the F ratio) under the null hypothesis that all group means are equal. The F ratio can be used to compute a p -value.



The null hypothesis for analysis of variance is tested using an F statistic. The F statistic is calculated as the ratio of the Between Group Variance to the Within Group Variance. In the output of PROC GLM, these values are shown as the Model Mean Square and the Error Mean Square. The mean square values are calculated as the sum of square value divided by the degrees of freedom.

In general, *degrees of freedom* (DF) can be thought of as the number of independent pieces of information.


- Model DF is the number of treatments minus 1.
- Corrected total DF is the sample size minus 1.
- Error DF is the sample size minus the number of treatments (or the difference between the corrected total DF and the Model DF).

Mean squares are calculated by taking sums of squares and dividing by the corresponding degrees of freedom. They can be thought of as variances.

- Mean square error (MSE) is an estimate of σ^2 , the constant variance assumed for all treatments.
- If $\mu_i = \mu_j$, for all $i \neq j$, then the mean square for the model (MSM) is also an estimate of σ^2 .
- If $\mu_i \neq \mu_j$, for any $i \neq j$, then MSM estimates σ^2 plus a positive constant.
- $$F = \frac{MSM}{MSE} = \frac{SS_M / df_M}{SS_E / df_E}$$
- The p -value for the test is then calculated from the F distribution with appropriate degrees of freedom.



Variance is the traditional measure of precision. *Mean square error (MSE)* is the traditional measure of accuracy used by statisticians. MSE is equal to variance plus bias-squared. Because the sample mean (\bar{x}) is an unbiased estimate of the population mean (μ), bias=0 and MSE measures the variance.



Coefficient of Determination

$$R^2 = SS_M / SS_T$$

“Proportion of variance accounted for by the model”

33

The *coefficient of determination*, R^2 , is a measure of the proportion of variability explained by the independent variables in the analysis. This statistic is calculated as $R^2 = \frac{SS_M}{SS_T}$

The value of R^2 is between 0 and 1. The value is

- close to 0 if the independent variables do not explain much variability in the data
- close to 1 if the independent variables explain a relatively large proportion of variability in the data.

Although values of R^2 closer to 1 are preferred, judging the magnitude of R^2 depends on the context of the problem.

The *coefficient of variation* (denoted Coeff Var) expresses the root MSE (the estimate of the standard deviation for all treatments) as a percent of the mean. It is a unitless measure that is useful in comparing the variability of two sets of data with different units of measure.

SAS THE POWER OF DATA

The ANOVA Model

$$\text{BulbWt} = \text{Base Level} + \text{Fertilizer} + \text{Unaccounted for Variation}$$

$$Y_{ik} = \mu + \tau_i + \varepsilon_{ik}$$

34

The model, $Y_{ik} = \mu + \tau_i + \varepsilon_{ik}$, is one way of representing the relationship between the dependent and independent variables in ANOVA.

Y_{ik} the k^{th} value of the response variable for the i^{th} treatment.

μ the overall population mean of the response, for example, garlic bulb weight.

τ_i the difference between the population mean of the i^{th} treatment and the overall mean, μ . This is referred to as the *effect* of treatment i .

ε_{ik} the difference between the observed value of the k^{th} observation in the i^{th} group and the mean of the i^{th} group. This is called the *error term*.



PROC GLM uses a parameterization of categorical variables in its CLASS statement that will not directly estimate the values of the parameters in the model shown. The correct parameter estimates can be obtained by adding the SOLUTION option in the MODEL statement in PROC GLM and then using simple algebra. Parameter estimates and standard errors can also be obtained using ESTIMATE statements. These issues are discussed in depth in the Statistics 2: ANOVA and Regression course and in the SAS documentation.



The researchers are interested only in these four specific fertilizers. In some applications this would be considered a *fixed effect*. If the fertilizers used were a sample of many that can be used, the sampling variability of fertilizers would need to be taken into account in the model. In that case, the fertilizer variable would be treated as a *random effect*. (Random effects are not discussed in this course.)

The GLM Procedure

General form of the GLM procedure:

```

PROC GLM DATA=SAS-data-set PLOTS=options;
  CLASS variables;
  MODEL dependents=independents </ options>;
  MEANS effects </ options>;
  LSMEANS effects </ options>;
  OUTPUT OUT=SAS-data-set keyword=variable...;
RUN;
QUIT;

```

35

Selected GLM procedure statements:

CLASS	specifies classification variables for the analysis.
MODEL	specifies dependent and independent variables for the analysis.
MEANS	computes unadjusted means of the dependent variable for each value of the specified effect.
LSMEANS	produces adjusted means for the outcome variable, broken out by the variable specified and adjusting for any other explanatory variables included in the MODEL statement.
OUTPUT	specifies an output data set that contains all variables from the input data set and variables that represent statistics from the analysis.



PROC GLM supports RUN-group processing, which means the procedure stays active until a PROC, DATA, or QUIT statement is encountered. This enables you to submit additional statements followed by another RUN statement without resubmitting the PROC statement.

Assumptions for ANOVA

- Observations are independent.
- Errors are normally distributed.
- All groups have equal error variances.

36

The validity of the p -values depends on the data meeting the assumptions for ANOVA. Therefore, it is good practice to verify those assumptions in the process of performing the analysis of group differences.

- Independence implies that the ε_{ij} occurrences in the theoretical model are uncorrelated.
- The errors are assumed to be normally distributed for every group or treatment.
- Approximately equal error variances are assumed across treatments.

Assessing ANOVA Assumptions

- Good data collection designs help ensure the independence assumption.
- Diagnostic plots from PROC GLM can be used to verify the assumption that the error is approximately normally distributed.
- PROC GLM produces a test of equal variances with the HOVTEST option in the MEANS statement. H_0 for this hypothesis test is that the variances are equal for all populations.

37

Predicted and Residual Values

The predicted value in ANOVA is the *group mean*.

A *residual* is the difference between the observed value of the response and the predicted value of the response variable.

Observation	Fertilizer	Observed	Predicted	Residual
1	4	0.20901000	0.19635250	0.01265750
2	3	0.25792000	0.22982091	0.02809909
3	2	0.21588000	0.20856500	0.00731500
4	4	0.24754000	0.19635250	0.05118750
5	1	0.24402000	0.22540667	0.01861333

38

The residuals from the ANOVA are calculated as the actual values minus the predicted values (the group means in ANOVA). Diagnostic plots (including normal quantile-quantile plots of the residuals) can be used to assess the normality assumption. With a reasonably sized sample and approximately equal groups (balanced design), only severe departures from normality are considered a problem. Residual values sum to 0 in ANOVA.

In ANOVA with more than one predictor variable, the HOVTEST option is unavailable. In those circumstances, you can plot the residuals against their predicted values to visually assess whether the variability is constant across groups.

2.02 Multiple Choice Poll

If you have 20 observations in your ANOVA and you calculate the residuals, to which of the following would they sum?

- 20
- 0
- 20
- 400
- Unable to tell from the information given

40

2.03 Multiple Choice Poll

If you have 20 observations in your ANOVA and you calculate the squared residuals, to which of the following would they sum?

- a. -20
- b. 0
- c. 20
- d. 400
- e. Unable to tell from the information given



The GLM Procedure

```
/*st102d03.sas*/  /*Part A*/
proc glm data=sasuser.MGGarlic;
  class Fertilizer;
  model BulbWt=Fertilizer;
  title 'Testing for Equality of Means with PROC GLM';
run;
quit;
```

Turn your attention to the first two tables of the output. The first table specifies the number of levels and the values of the class variable.

Class Level Information				
Class	Levels	Values		
Fertilizer	4	1	2	3 4

The second table shows both the number of observations read and the number of observations used. These values are the same because there are no missing values in for any variable in the model. If any row has *missing data* for a predictor or response variable, that row is *dropped* from the analysis.

Number of Observations Read	32
Number of Observations Used	32

The second part of the output contains all of the information that is needed to test the equality of the treatment means. It is divided into three parts:

- the analysis of variance table
- descriptive information
- information about the effect of the independent variable in the model

Look at each of these parts separately.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.00457996	0.00152665	1.96	0.1432
Error	28	0.02183054	0.00077966		
Corrected Total	31	0.02641050			

The F statistic and corresponding p -value are reported in the Analysis of Variance table. Because the reported p -value (0.1432) is greater than 0.05, you do not reject the null hypothesis of no difference between the means.

R-Square	Coeff Var	Root MSE	BulbWt Mean
0.173414	12.74520	0.027922	0.219082

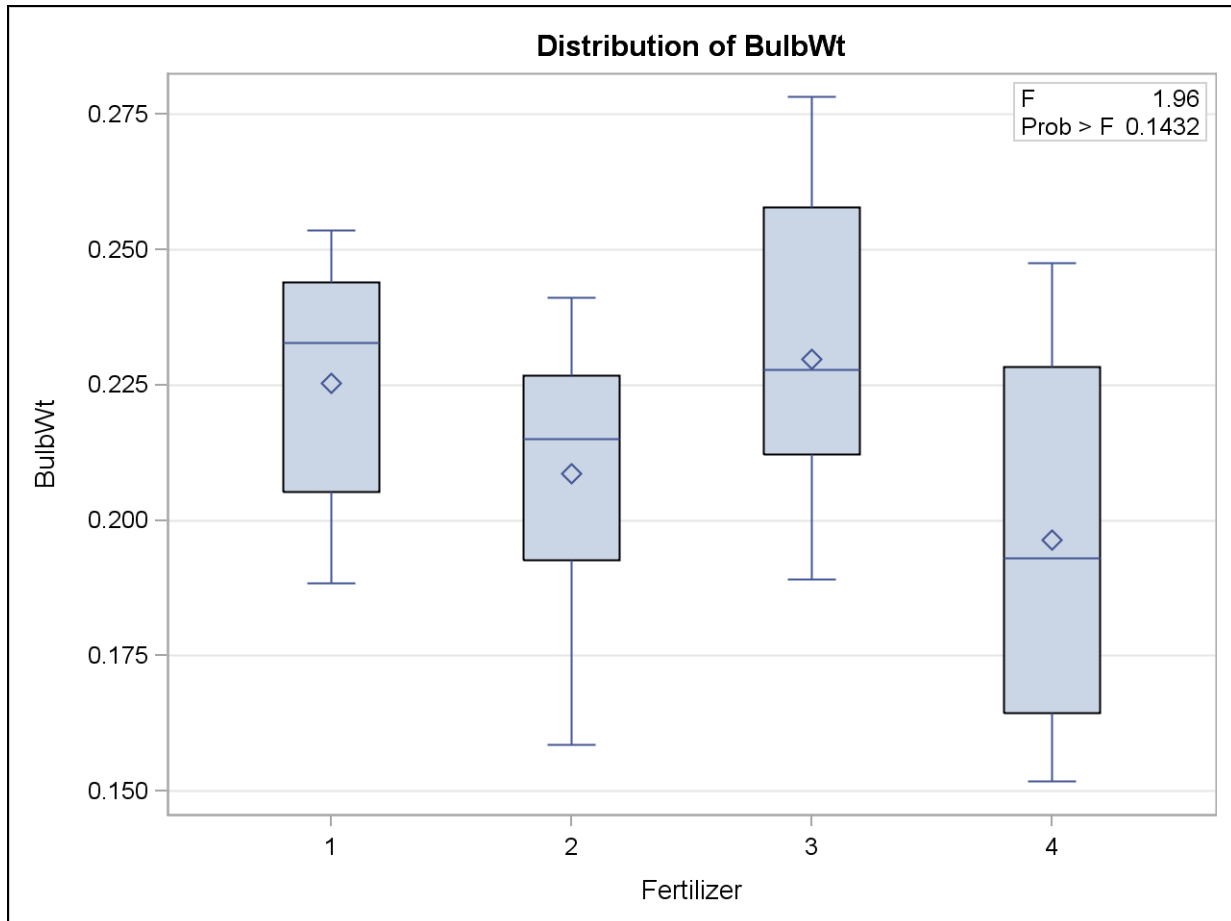
The **BulbWt** Mean is the mean of all of the data values in the variable **BulbWt** without regard to **Fertilizer**.

As discussed previously, the R^2 value is often interpreted as the “proportion of variance accounted for by the model.” Therefore, you might say that in this model, **Fertilizer** explains about 17% of the variability of **BulbWt**.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Fertilizer	3	0.00457996	0.00152665	1.96	0.1432

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Fertilizer	3	0.00457996	0.00152665	1.96	0.1432

For a one-way analysis of variance (only one classification variable), the information about the independent variable in the model is an exact duplicate of the model line of the analysis of variance table.



The default plot created with this code is a box plot.

It is good practice to check the validity of your ANOVA assumptions. The next part of the program is dedicated to verifying those statistical assumptions for inference tests.

```
/*st102d03.sas*/ /*Part B*/
proc glm data=sasuser.MGGarlic plots(only)=diagnostics;
  class Fertilizer;
  model BulbWt=Fertilizer;
  means Fertilizer / hovtest;
  title 'Testing for Equality of Means with PROC GLM';
run;
quit;
```

Selected MEANS statement option:

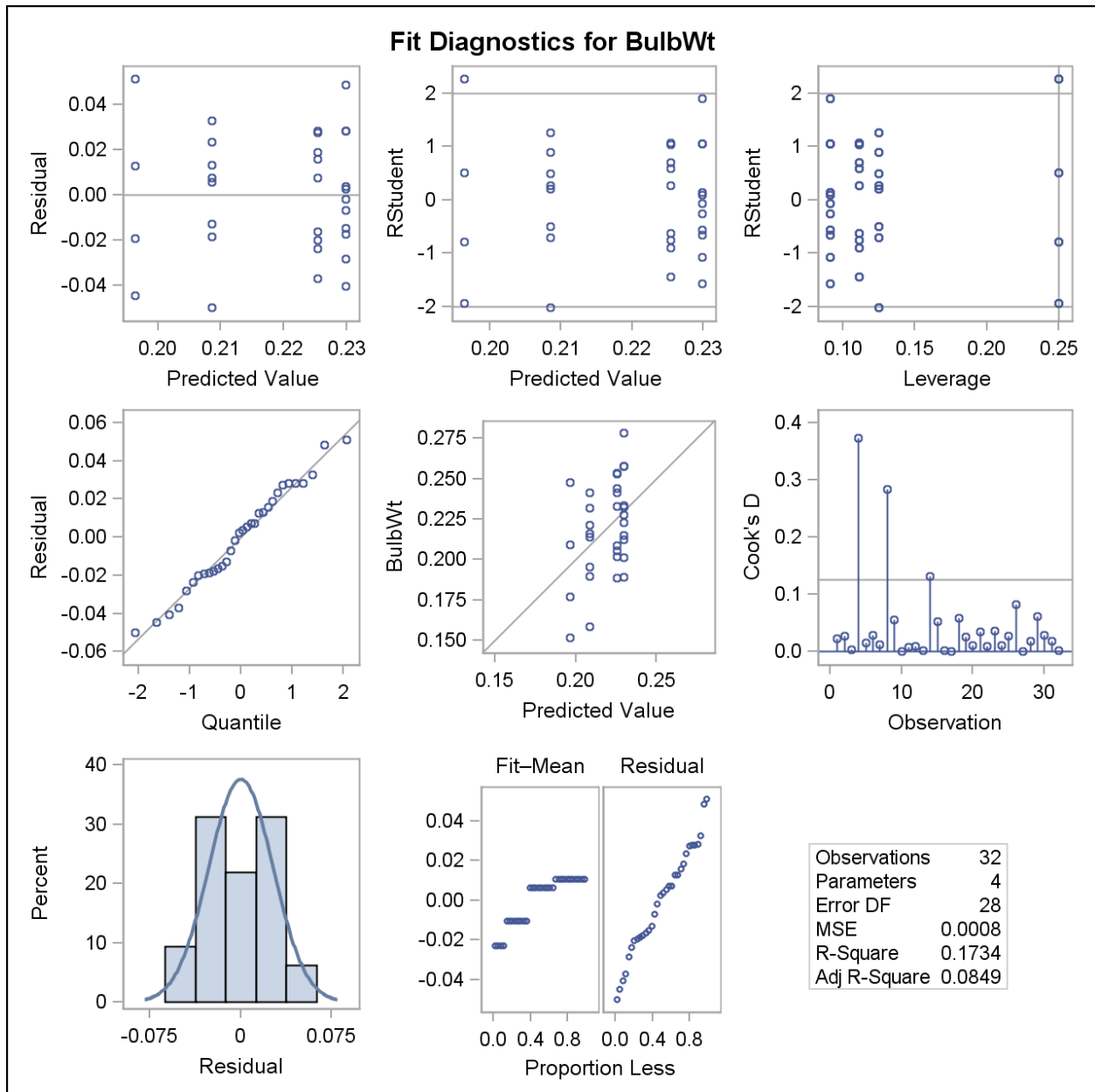
HOVTEST performs Levene's test for homogeneity (equality) of variances. The null hypothesis for this test is that the variances are equal. Levene's test is the default.

Selected PLOTS option:

DIAGNOSTICS produces a panel display of diagnostic plots for linear models.



The UNPACK option can be used in order to separate the individual plots in the panel display.



The panel in the upper left corner shows a plot of the residuals versus the fitted values from the ANOVA model. Essentially, you are looking for a random scatter within each group. Any patterns or trends in this plot can indicate model misspecification.

To check the normality assumption, open the residual histogram and Q-Q plot, which are at the bottom left and middle left, respectively.

The histogram has no unique peak and it has short tails. However, it is approximately symmetric.

The data values in the quantile-quantile plot stay close to the diagonal reference line and give strong support to the assumption of normally distributed errors.

Near the end of the tabular output, you can check the assumption of equal variances.

Levene's Test for Homogeneity of BulbWt Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Fertilizer	3	1.716E-6	5.719E-7	0.98	0.4173
Error	28	0.000016	5.849E-7		

The output above is the result of the HOVTEST option in the MEANS statement. Levene's test for homogeneity of variances is the default. The null hypothesis is that the variances are equal over all **Fertilizer** groups. The p -value of 0.4173 is not smaller than your alpha level of 0.05 and therefore you do not reject the null hypothesis. One of your assumptions is met.



At this point, if you determined that the variances were not equal, you could add the WELCH option to the MEANS statement. This requests Welch's (1951) variance-weighted one-way ANOVA. This alternative to the usual ANOVA is robust to the assumption of equal variances. This is similar to the unequal variance t test for two populations. See the appendix for more information.

Analysis Plan for ANOVA – Summary

Null Hypothesis: All means are equal.

Alternative Hypothesis: At least one mean is different.

1. Produce descriptive statistics.
2. Verify assumptions.
 - Independence
 - Errors are normally distributed.
 - Error variances are equal for all groups.
3. Examine the p -value in the ANOVA table. If the p -value is less than alpha, reject the null hypothesis.



Exercises

2. Analyzing Data in a Completely Randomized Design

Consider an experiment to study four types of advertising: local newspaper ads, local radio ads, in-store salespeople, and in-store displays. The country is divided into 144 locations, and 36 locations are randomly assigned to each type of advertising. The level of sales is measured for each region in thousands of dollars. You want to see whether the average sales are significantly different for various types of advertising. The **sasuser.ads** data set contains data for these variables:

Ad type of advertising

Sales level of sales in thousands of dollars

- a. Examine the data. Use the MEANS and SGLOT procedures. What information can you obtain from looking at the data?
- b. Test the hypothesis that the means are equal. Be sure to check that the assumptions of the analysis method that you choose are met. What conclusions can you reach at this point in your analysis?

2.3 ANOVA with Data from a Randomized Block Design

Objectives

- Recognize the difference between a completely randomized design and a randomized block design.
- Differentiate between observed data and designed experiments.
- Use the GLM procedure to analyze data from a randomized block design.

Observational or Retrospective Studies

- Groups can be naturally occurring.
 - for example, gender and ethnicity
- Random assignment might be unethical or untenable.
 - for example, smoking or credit risk groups
- Often you look at what already happened (retrospective) instead of following through to the future (prospective).
- You have little control over other factors contributing to the outcome measure.

50

In the original study, the Montana Gourmet Garlic growers randomly assigned their treatments (fertilizer) to plants in each of their Spanish Roja beds. They did this as an afterthought before they realized that they were going to do a statistical analysis. In fact, this could reasonably be thought of as a retrospective study. When you analyze the differences between naturally occurring groups, you are not actually manipulating a treatment. There is no true independent variable.

Many public health and business analyses are retrospective studies. The data values are observed as they occur, not affected by an experimental design. Often this is the best you can do. For example, you cannot ethically randomly assign people to smoking and nonsmoking groups.

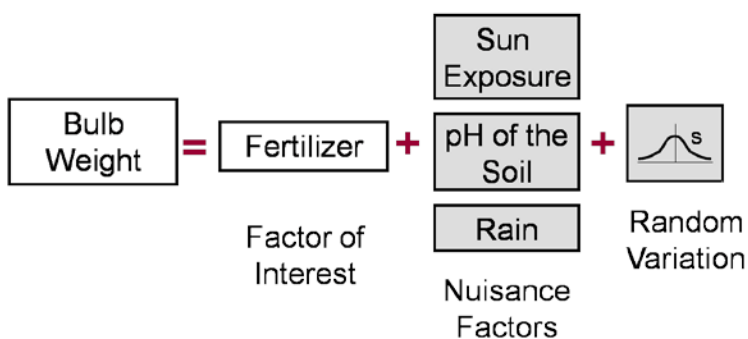
Controlled Experiments

- Random assignment might be desirable to eliminate selection bias.
- You often want to look at the outcome measure prospectively.
- You can manipulate the factors of interest and can more reasonably claim causation.
- You can design your experiment to control for other factors contributing to the outcome measure.

51

Given the negative results of the fertilizer study from 2006, the garlic growers planned a prospective study in 2007. They decided they needed to try more rigorously to control the influences on the growth of garlic.

Nuisance Factors



52

Factors that can affect the outcome but are not of interest in the experiment are called *nuisance factors*. The variation due to nuisance factors becomes part of the random variation.

2.04 Multiple Choice Poll

Which part of the ANOVA tables contains the variation due to nuisance factors?

- a. Sum of Squares Model
- b. Sum of Squares Error
- c. Degrees of Freedom

Assigning Treatments within Blocks

<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">1</div>	<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">4</div>
<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">3</div>	<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">2</div>

<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">2</div>	<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">1</div>
<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">3</div>	<div style="background-color: yellow; border: 1px solid black; padding: 5px; margin-bottom: 5px;">4</div>

56

A discussion with a statistician helped the farmers identify other determinants of garlic bulb weight. The statistician suggested that, although they could not actually apply those factors randomly (they could not change the weather or the soil pH or composition or sun exposure), they could control for those factors by blocking. He suggested that whatever the effects of those external influences are, the magnitudes of those nuisance factors should be approximately the same within sectors of the farm land. Therefore, instead of randomizing the **Fertilizer** treatment across all 32 beds, he suggested they only randomize the application of the four **Fertilizer** treatments within each of the eight sectors.

An experimental design such as this is often referred to as a *randomized block design*. In this case, **Sector** is the block. The blocking variable **Sector** is included in the model, but you are not interested in its effect, only in controlling the nuisance factor effects explained by it. By including **Sector** in the model, you could potentially account for many nuisance factors.

Blocking is a logical grouping of experimental units. In this study, applying all four fertilizers to the same sector makes sense from a practical point of view. There might be a great disparity in the presence of nuisance factors across sectors, but you can be reasonably confident that the nuisance factor influence is fairly even within sectors.

Blocking is a restriction on randomization and therefore must be taken into account in data analysis.

sas THE POWER OF DATA

Including a Blocking Variable in the Model

Bulb Weight = Base Level + Sector + Fertilizer Type + Unaccounted for Variation

$$Y_{ijk} = \mu + \alpha_i + \tau_j + \varepsilon_{ijk}$$

57

sas THE POWER OF DATA

Nuisance Factors

The diagram illustrates the components of an ANOVA model. On the left, a box labeled 'Bulb Weight' is followed by an equals sign. To the right of the equals sign is a box labeled 'Fertilizer', which is identified as the 'Factor of Interest'. This is followed by a plus sign and a group of three boxes: 'Sun Exposure', 'pH of the Soil', and 'Rain'. These three boxes are collectively labeled as 'Nuisance Factors'. Another plus sign follows, leading to a box containing a normal distribution curve, which is labeled 'Random Variation'.

58

Because **Sector** is included in the ANOVA model, any effect caused by the nuisance factors that are common within a sector are accounted for in the Model Sum of Squares and not the Error Sum of Squares, as was the case in the previous study. Removing significant effects from the Error Sum of Squares tends to give more power to the test of the effect of interest (in this case, **Fertilizer**). That is because the MSE, the denominator of the F statistic, tends to be reduced, increasing the F value and thereby decreasing the p -value.

2.05 Multiple Choice Poll

In a block design, which part of the ANOVA table contains the variation due to the nuisance factor?

- a. Sum of Squares Model
- b. Sum of Squares Error
- c. Degrees of Freedom

Including a Blocking Variable in the Model

Additional assumptions are as follows:

- Treatments are randomly assigned within each block.
- The effects of the treatment factor are constant across the levels of the blocking variable.

In the garlic example, the design is balanced, which means that there is the same number of garlic samples for every **Fertilizer/Sector** combination.

62

Typically, when the effects of the treatment factor are not constant across the levels of the other variable, then this condition is called *interaction*. However, when a randomized block design is used, it is assumed that the effects are the same within each block. In other words, it is assumed that there are no interactions with the block variable.



In most randomized block designs, the blocking variable is treated as a *random effect*. Treating an effect as random changes how standard errors are calculated and can give different answers from treating it as a fixed effect (as in the example).

In this example, you have the same number of garlic samples for every **Fertilizer/Sector** combination. This is a balanced design. When treatment groups are compared to each other (in other words, not to 0 or some other specified value), the results from treating the block as a fixed or random effect are exactly the same.

A model that includes both random and fixed effects is called a *mixed model* and can be analyzed with the MIXED procedure. The Mixed Models Analyses Using SAS® class focuses on analyzing mixed models. The Statistics 2: ANOVA and Regression class has more detail about how to analyze unbalanced designs and data that do not meet ANOVA assumptions.

For more information about mixed models in SAS, you can also consult the SAS online documentation or the SAS Books by Users book *SAS® System for Mixed Models*, which also goes into detail about the statistical assumptions for mixed models.



ANOVA with Blocking

```
/*st102d04.sas*/  
proc glm data=sasuser.MGGarlic_Block plots(only)=diagnostics;  
  class Fertilizer Sector;  
  model BulbWt=Fertilizer Sector;  
  title 'ANOVA for Randomized Block Design';  
run;  
quit;
```

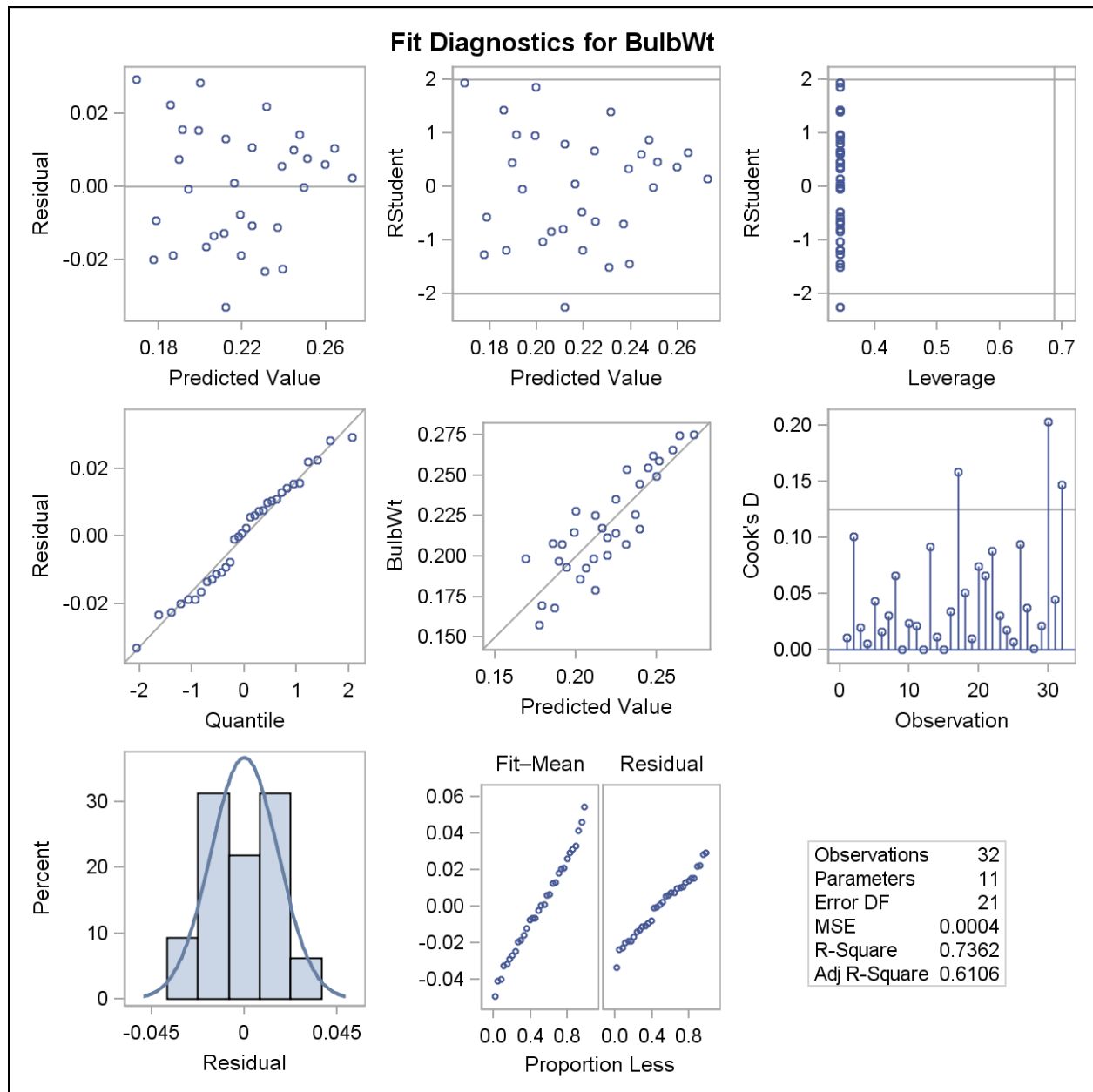
Selected PLOTS() option:

(ONLY) requests that only the requested plots be produced and no default plots.



The blocking variable must be in the model and it must be listed in the CLASS statement.

A check of the normality assumption using the Q-Q plot follows.



No severe departure from normality of the error terms seem to exist.



Validation of the equal variances assumption for models with more than one independent variable is beyond the scope of this text. This topic is discussed in the Statistics 2: ANOVA and Regression class.

Class Level Information	
Class	Levels Values
Fertilizer	4 1 2 3 4
Sector	8 1 2 3 4 5 6 7 8

The Class Level Information table reflects the addition of the eight-level **Sector** variable.

Number of Observations Read	32
Number of Observations Used	32

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	0.02307263	0.00230726	5.86	0.0003
Error	21	0.00826745	0.00039369		
Corrected Total	31	0.03134008			

R-Square	Coeff Var	Root MSE	BulbWt Mean
0.736202	9.085064	0.019842	0.218398

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Fertilizer	3	0.00508630	0.00169543	4.31	0.0162
Sector	7	0.01798632	0.00256947	6.53	0.0004

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Fertilizer	3	0.00508630	0.00169543	4.31	0.0162
Sector	7	0.01798632	0.00256947	6.53	0.0004

The overall F test ($F(10,21)=5.86$, $p=0.0003$) indicates that there are significant differences between the means of the garlic bulb weights across fertilizers or blocks (sectors). However, because there is more than one term in the model, you cannot tell whether the differences are due to differences among the fertilizers or differences across sectors. In order to make that determination, you must look at the subsequent tests for each factor.

What have you gained by including **Sector** in the model? If you compare the estimate of the experimental error variance (MSE), you note this is smaller compared to the data and model that included **Fertilizer** only (0.00039369 versus 0.00077966). Depending on the magnitude of the difference, this could affect the comparisons between the treatment means by finding more significant differences than the **Fertilizer**-only model, given the same sample sizes.

Also notice that the R square for this model is much greater than that in the previous model (0.736 versus 0.173). To some degree, this is a function of having more model degrees of freedom, but it is unlikely this is the only reason for this magnitude of difference.

Most important to the Montana Gourmet Garlic farmers is that the effect of **Fertilizer** in this model is now significant ($F=4.31$, $p=0.0162$). The Type III SS test is at the bottom of the output tests for differences due to each variable, controlling for (or “adjusting for”) the other variable.



The Type I SS test is *sequential*. In other words, the test for each variable only adjusts for the variables above it. In this case, because the design is completely balanced, the Type I and Type III tests would be exactly the same. In general that would not be true.

In determining the usefulness of having a blocking variable (**Sector**) included in the model, you can consider the F value for the blocking variable. Some statisticians suggest that if this ratio is greater than 1, then the blocking factor is useful. If the ratio is less than 1, then adding the variable is detrimental to the analysis. If you find that including the blocking factor is detrimental to the analysis, then you can exclude it from future studies, but it *must* be included in all ANOVA models calculated with the sample that you already collected. This is because blocking places a restriction on the random assignment of units to treatments, and modeling the data without the blocking variable treats the data as if that restriction did not exist.



Exercises

3. Analyzing Data in a Randomized Block Design

When you design the advertising experiment in the first question, you are concerned that there is variability caused by the area of the country. You are not particularly interested in what differences are caused by **Area**, but you are interested in isolating the variability due to this factor.

The **sasuser.ads1** data set contains data for the following variables:

Ad type of advertising

Area area of the country

Sales level of sales in thousands of dollars

Test the hypothesis that the means are equal. Include all of the variables in your MODEL statement.

- a. What can you conclude from your analysis?
- b. Was adding the blocking variable **Area** into the design and analysis detrimental to the test of **Ad**?

2.06 Multiple Answer Poll

If the blocking variable **Area** had a very small F value, what would be a valid next step? Select all that apply.

- a. Remove it from the MODEL statement and rerun the analysis.
- b. Test an interaction term.
- c. Report the F value and plan a new study.

67

My Groups Are Different. What Next?

- The p -value for **Fertilizer** indicates you should reject the H_0 that all groups are the same.
- From which pairs of fertilizers, are garlic bulb weights different from one another?
- Should you go back and do several t tests?

70

The garlic researchers know that not all fertilizers are created equal, but which one is the best?

2.4 ANOVA Post Hoc Tests

Objectives

- Perform pairwise comparisons among groups after finding a significant effect of an independent variable in ANOVA.
- Demonstrate graphical features in PROC GLM for performing post hoc tests.
- Interpret a diffogram.
- Interpret a control plot.

72

2.07 Multiple Choice Poll

With a fair coin, your probability of getting heads on one flip is 0.5. If you flip a coin once and got heads, what is the probability of getting heads on the second try?

- 0.50
- 0.25
- 0.00
- 1.00
- 0.75

74

2.08 Multiple Choice Poll

With a fair coin, your probability of getting heads on one flip is 0.5. If you flip a coin twice, what is the probability of getting **at least** one head out of two?

- a. 0.50
- b. 0.25
- c. 0.00
- d. 1.00
- e. 0.75

Multiple Comparison Methods		
Comparisonwise Error Rate ($\alpha=0.05$)	Number of Comparisons	Experimentwise Error Rate ($\alpha=0.05$)
.05	1	.05
.05	3	.14
.05	6	.26
.05	10	.40

$EER \leq 1 - (1 - \alpha)^{nc}$ where nc =number of comparisons

78

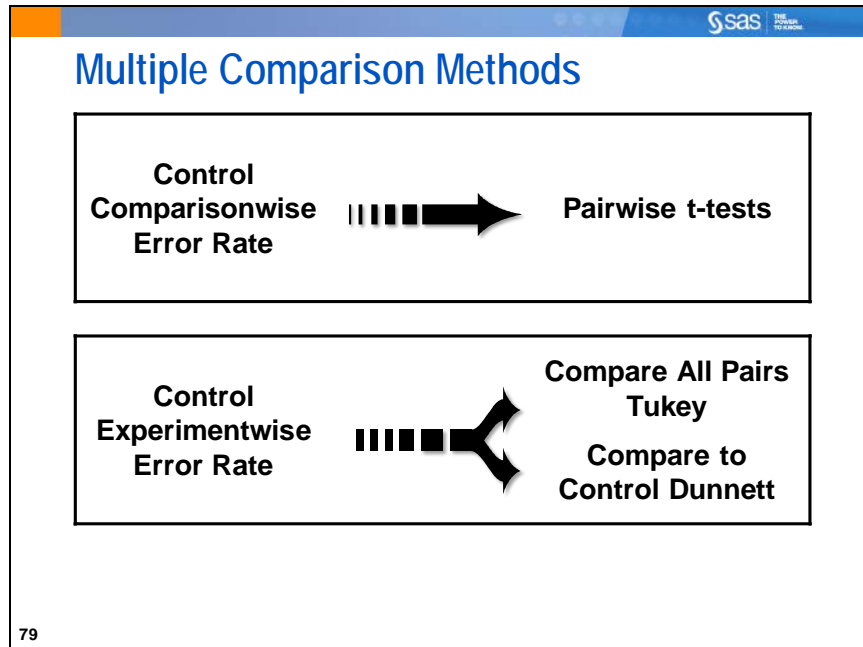
When you control the comparisonwise error rate (CER), you fix the level of alpha for a single comparison, without taking into consideration all the pairwise comparisons that you are making.

The experimentwise error rate (EER) uses an alpha that takes into consideration all the pairwise comparisons that you are making. Presuming no differences exist, the chance that you falsely conclude that **at least one** difference exists is much higher when you consider all possible comparisons.

If you want to make sure that the error rate is 0.05 for the entire set of comparisons, use a method that controls the experimentwise error rate at 0.05.



There is some disagreement among statisticians about the need to control the experimentwise error rate.



All of these multiple comparison methods are requested with options in the LSMEANS statement of PROC GLM.

In order to call for the statistical hypothesis tests for group differences and ODS Statistical Graphics to support them, turn on ODS Graphics and then:


- For Comparisonwise Control LSMEANS / PDIFF=ALL ADJUST=T
- For Experimentwise Control LSMEANS / PDIFF=ALL ADJUST=TUKEY or
PDIFF=CONTROL('control level') ADJUST=DUNNETT



Many other available options control the experimentwise error rate. For information about these options, see the SAS documentation.



One-tailed tests against a control level can be requested using the CONTROLL (lower tail) or CONTROLU (upper tail) options in the LSMEANS statement.



Tukey's Multiple Comparison Method

This method is appropriate when you consider pairwise comparisons only.

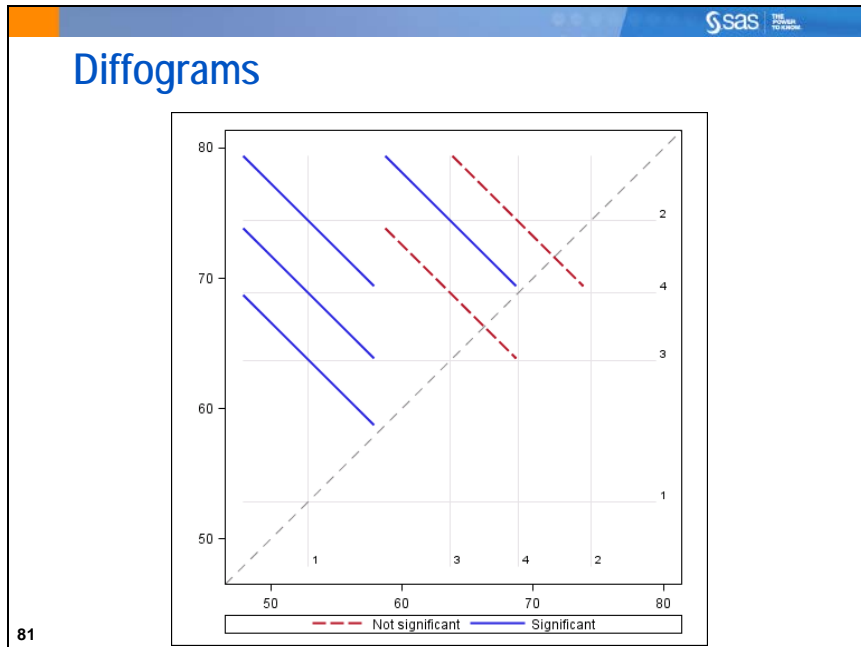
The experimentwise error rate is

- equal to alpha when **all** pairwise comparisons are considered
- less than alpha when **fewer** than all pairwise comparisons are considered.

80

A pairwise comparison examines the difference between two treatment means. “All pairwise comparisons” means all possible combinations of two treatment means.

Tukey's multiple comparison adjustment is based on conducting all pairwise comparisons and guarantees that the Type I experimentwise error rate is equal to alpha for this situation. If you choose to do fewer than all pairwise comparisons, then this method is more conservative.



A *diffogram* can be used to quickly tell whether two group means are statistically significant. The point estimates for the differences between pairs of group means can be found at the intersections of the vertical and horizontal lines drawn at group mean values. The downward-sloping diagonal lines show the confidence intervals for the differences. The upward-sloping line is a reference line showing where the group means would be equal. Intersection of the downward-sloping diagonal line for a pair with the upward-sloping, broken gray diagonal line implies that the confidence interval includes zero and that the mean difference between the two groups is not statistically significant. In that case, the diagonal line for the pair will be broken. If the confidence interval does not include zero, then the diagonal line for the pair will be solid. With ODS statistical graphics, these plots are automatically generated when you use the PDIFF=ALL option in the LSMEANS statement.

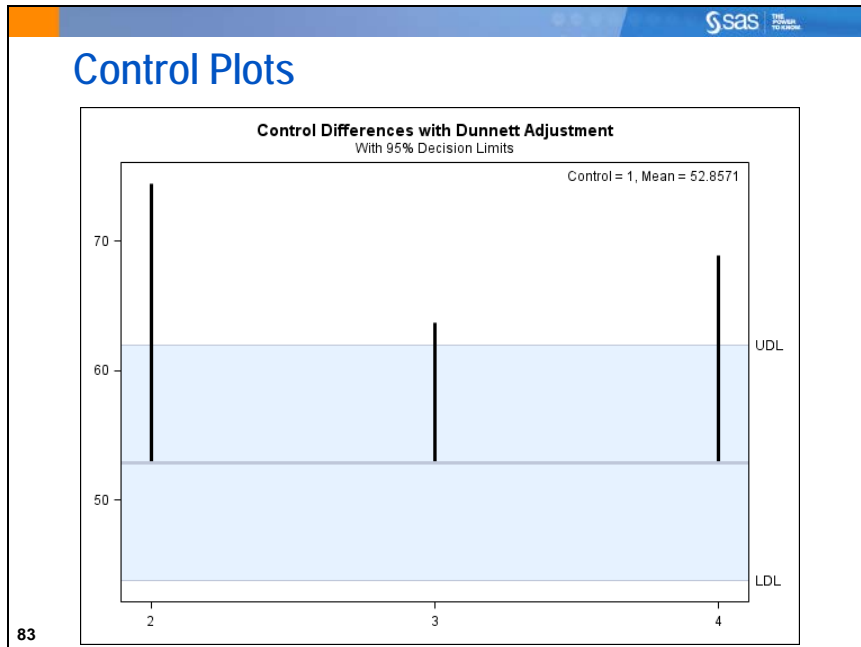
Special Case of Comparing to a Control

Comparing to a control is appropriate when there is a natural reference group, such as a placebo group in a drug trial.

- Experimentwise error rate is no greater than the stated alpha.
- Comparing to a control takes into account the correlations among tests.
- One-sided hypothesis tests against a control group can be performed.
- Control comparison computes and tests $k-1$ groupwise differences, where k is the number of levels of the CLASS variable.
- An example is the Dunnett method.

82

Dunnett's method is recommended when there is a true control group. When appropriate (when a natural control category exists, against which all other categories are compared) it is more powerful than methods that control for all possible comparisons. In order to do a one-sided test, use the option `PDIF=CONTROLL` (for lower-tail tests when the alternative hypothesis states that a group's mean is less than the control group's mean) or `PDIF=CONTROLU` (for upper-tail tests when the alternative hypothesis states that a group's mean is greater than the control group's mean).



LS-mean control plots are produced only when you specify `PDIFF=CONTROL` or `ADJUST=DUNNETT` in the `LSMEANS` statement, and in this case they are produced by default. The value of the control is shown as a horizontal line. The shaded area is bounded by the UDL and LDL (Upper Decision Limit and Lower Decision Limit). If the vertical line extends past the shaded area, that means that the group represented by that line is significantly different from the control group.



Post Hoc Pairwise Comparisons

Example: Use the LSMEANS statement in PROC GLM to produce comparison information about the means of the treatments.

```
/*st102d05.sas*/
proc glm data=sasuser.MGGarlic_Block
    plots(only)=(controlplot diffplot(center));
    class Fertilizer Sector;
    model BulbWt=Fertilizer Sector;
    lsmeans Fertilizer / pdiff=all adjust=tukey;
    lsmeans Fertilizer / pdiff=control('4') adjust=dunnett;
    lsmeans Fertilizer / pdiff=all adjust=t;
    title 'Garlic Data: Multiple Comparisons';
run;
quit;
```

Multiple LSMEANS statements are permitted, although typically only one type of multiple comparison method would be used for each LSMEANS effect. Three different methods are shown for illustration here. For this analysis, the garlic growers were unblinded to the fertilizers and number 4 is the chemical fertilizer. They might conceivably use Dunnett comparisons if they were only interested in knowing whether any of the organic fertilizers created differently sized bulbs compared with the chemical fertilizer.

Selected PLOTS= options:

CONTROLPLOT requests a display in which least squares means are compared against a reference level. LS-mean control plots are produced only when you specify **PDIF=CONTROL** or **ADJUST=DUNNETT** in the LSMEANS statement, and in this case they are produced by default.

DIFFPLOT modifies the diffogram produced by an LSMEANS statement with the **PDIF=ALL** option (or only **PDIF**, because **ALL** is the default argument). The **CENTER** option marks the center point for each comparison. This point corresponds to the intersection of two least squares means.

Selected LSMEANS statement options:

PDIF= requests *p*-values for the differences, which is the probability of seeing a difference between two means that is as large as the observed means or larger if the two population means are actually the same. You can request to compare all means using **PDIF=ALL**. You can also specify which means to compare. For details, see the documentation for LSMEANS under the GLM procedure.

ADJUST= specifies the adjustment method for multiple comparisons. If no adjustment method is specified, the Tukey method is used by default. The **T** option asks that no adjustment be made for multiple comparisons. The **TUKEY** option uses Tukey's adjustment method. The **DUNNETT** option uses Dunnett's method. For a list of available methods, check the documentation for LSMEANS under the GLM procedure.



The MEANS statement can be used for multiple comparisons. However, the results can be misleading if the groups that are specified have different numbers of observations.

The following output is for the Tukey LSMEANS comparisons.

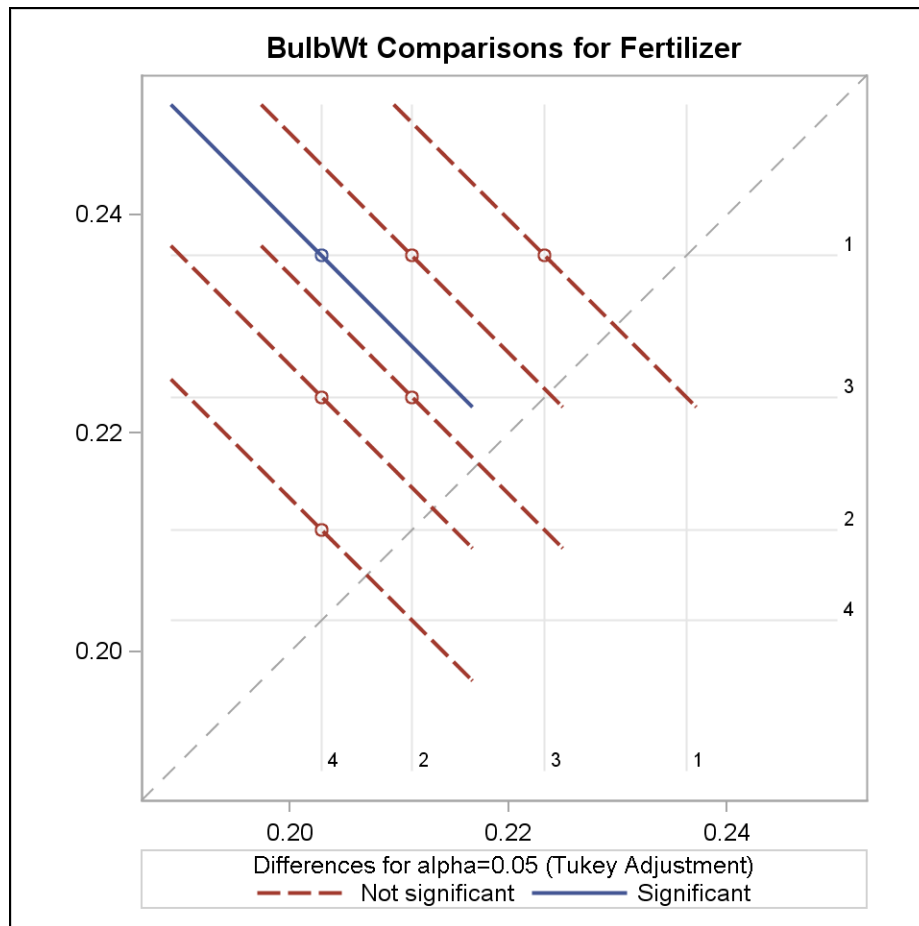
Fertilizer	BulbWt LSMEAN	LSMEAN Number
1	0.23625000	1
2	0.21115125	2
3	0.22330125	3
4	0.20288875	4

Least Squares Means for effect Fertilizer Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: BulbWt				
i/j	1	2	3	4
1		0.0840	0.5699	0.0144
2	0.0840		0.6186	0.8383
3	0.5699	0.6186		0.1995
4	0.0144	0.8383	0.1995	

The first part of the output shows the means for each group. The second part of the output shows p -values from pairwise comparisons of all possible combinations of means. Notice that row 2/column 4 has the same p -value as row 4/column 2 because the same two means are compared in each case. Both are displayed as a convenience to the user. Notice also that row 1/column 1, row 2/column 2, and so on, are blank, because it does not make any sense to compare a mean to itself.

The only significant pairwise difference is between fertilizer 1 and fertilizer 4 (p -value=0.0144).

The Least Square Means are shown graphically in the mean plot. The Tukey-adjusted differences among the LSMEANS are shown in the diffogram.



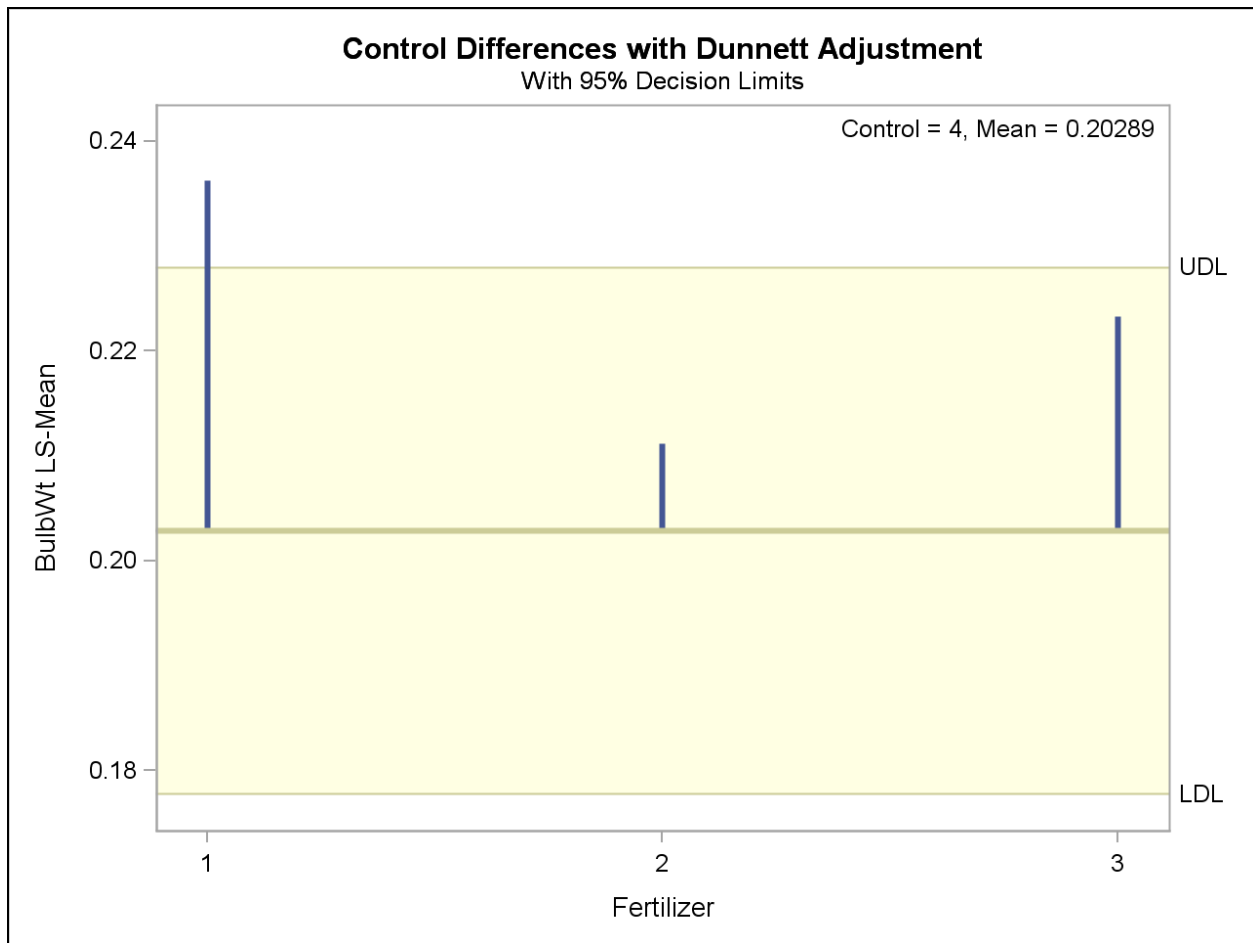
The solid line shows the significant difference between fertilizers 1 and 4. (The confidence limit for the difference does not cross the diagonal equivalence line.)

The following output is for the Dunnett LSMEANS comparisons:

Fertilizer	BulbWt LSMEAN	H0:LSMean=Control
		Pr > t
1	0.23625000	0.0080
2	0.21115125	0.7435
3	0.22330125	0.1274
4	0.20288875	

In this case, the first three fertilizers are compared to fertilizer 4, the chemical fertilizer. Even though the mean weights of garlic bulbs using any of the three organic methods are all greater than the mean weight of garlic bulbs grown using the chemical fertilizer, only fertilizer 1 can be said to be statistically significantly better.

The Control plot is below:



This plot corresponds to the tables that were summarized. The horizontal line is drawn at the least squared mean for group 4, which is 0.20289. The three other means are represented by the tops of the vertical lines extending from the horizontal control line. The only line that extended beyond the shaded area of nonsignificance is the line for fertilizer 1. That shows graphically that the mean bulb weight for fertilizer 1 is significantly different from the mean bulb weight for fertilizer 4.

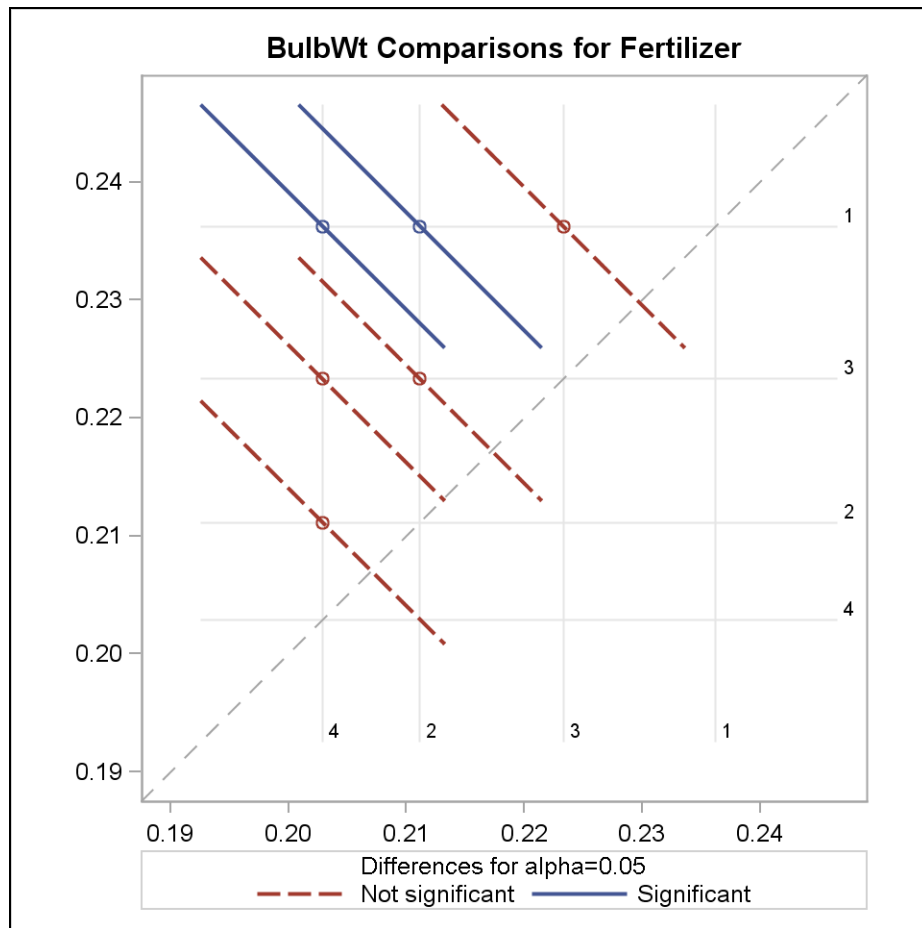
Finally, for comparison, the t tests are shown, which do not adjust for multiple comparisons and are therefore more liberal than tests that do control for experimentwise error:

Fertilizer	BulbWt LSMEAN	LSMEAN Number
1	0.23625000	1
2	0.21115125	2
3	0.22330125	3
4	0.20288875	4

Least Squares Means for effect Fertilizer Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: BulbWt				
i/j	1	2	3	4
1		0.0195	0.2059	0.0029
2	0.0195		0.2342	0.4143
3	0.2059	0.2342		0.0523
4	0.0029	0.4143	0.0523	

The p -values in this table are all smaller than those in the Tukey table. In fact, using this method shows one additional significant pairwise difference. Fertilizer 1 is significantly different from fertilizer 2 ($p=0.0195$). The comparison between 3 and 4 is nearly significant at $\alpha=0.05$ ($p=0.0523$).

The diffogram shows the additional significant difference:





Exercises

4. Post Hoc Pairwise Comparisons

Consider again the analysis of the **sasuser.Ads1** data set. There was a statistically significant difference among means for sales for the different types of advertising. Perform a post hoc test to look at the individual differences among means for the advertising campaigns.

- a. Conduct pairwise comparisons with an experimentwise error rate of $\alpha=0.05$. (Use the Tukey adjustment.) Which types of advertising are significantly different?
- b. Use **display** (case sensitive) as the control group and do a Dunnett comparison of all other advertising methods to see whether those methods resulted in significantly different amounts of sales compared with display ads in stores.

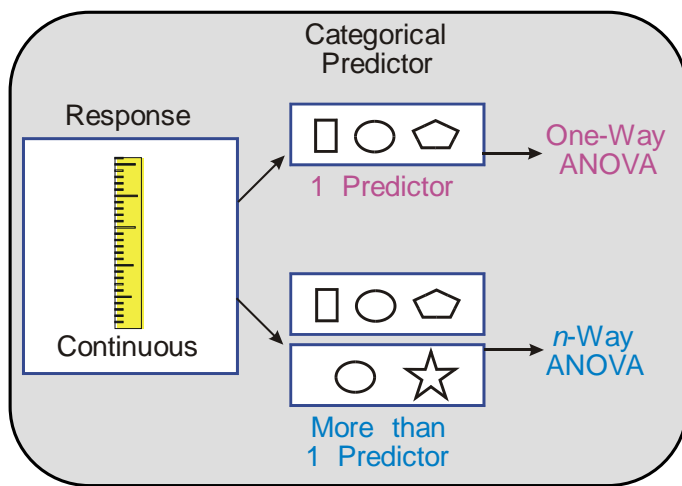
2.5 Two-Way ANOVA with Interactions

Objectives

- Fit a two-way ANOVA model.
- Detect interactions between factors.
- Analyze the treatments when there is a significant interaction.

88


n-Way ANOVA



89


In the previous section, you considered the case where you had one categorical predictor and a blocking variable. In this section, consider a case with two categorical predictors. In general, any time you have more than one categorical predictor variable and a continuous response variable, it is called *n*-way ANOVA. The *n* can be replaced with the number of categorical predictor variables.

The analysis for a randomized block design is actually a special type of *n*-way ANOVA.



Drug Example

The purpose of the study is to look at the effect of a new prescription drug on blood pressure.



90

Example: Data were collected in an effort to determine whether different dose levels of a given drug have an effect on blood pressure for people with one of three types of heart disease. The data are in the **sasuser.Drug** data set.

The data set contains the following variables:

DrugDose dosage level of drug (1, 2, 3, 4), corresponding to (Placebo, 50 mg, 100 mg, 200 mg)

Disease heart disease category

BloodP change in diastolic blood pressure after 2 weeks treatment

sas THE POWER OF DATA

The Model

BloodP = Base Level + Disease + Drug Dose + DrugDose and Disease + Unaccounted for Variation

↓

↓

↓

↓

↓

↓

$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

91

Y_{ijk} the observed **BloodP** for each subject

μ the overall base level of the response, **BloodP**

α_i the effect of the i^{th} **Disease**

β_j the effect of the j^{th} **DrugDose**

$(\alpha\beta)_{ij}$ the effect of the interaction between the i^{th} **Disease** and the j^{th} **DrugDose**

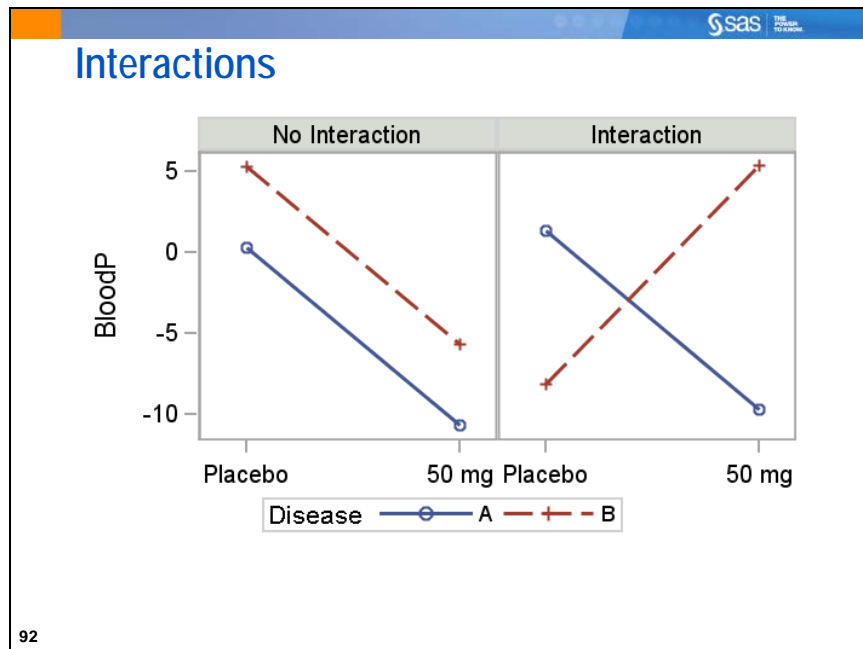
ϵ_{ijk} error term, or residual

In the model, the following is assumed:

- Observations are independent.
- Error terms are normally distributed for each treatment.
- Variances are equal across treatments.



Verifying ANOVA assumptions with more than two variables is discussed in the Statistics 2: ANOVA and Regression class.



An interaction occurs when the differences between group means on one variable change at different levels of another variable.

The average blood pressure change over different doses was plotted in mean plots and then connected for disease A and B.

In the left plot above, different types of disease show the same change across different levels of dose.

In the right plot, however, as the dose increases, average blood pressure *decreases* for those with disease A, but *increases* for those with disease B. This indicates an interaction between the variables **DrugDose** and **Disease**.

When you analyze an n -way ANOVA with interactions, you should first look at any tests for interaction among factors.

If there is no interaction between the factors, the tests for the individual factor effects can be interpreted as true effects of that factor.

If an interaction exists between any factors, the tests for the individual factor effects might be misleading, due to masking of the effects by the interaction. This is especially true for unbalanced data.

In the previous section, you used a blocking variable and a categorical predictor as effects in the model. It is generally assumed that blocks do not interact with other factors. In this section, neither independent variable is a blocking variable. An interaction between the two can be hypothesized and tested.

Nonsignificant Interaction

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

93

When the interaction is not statistically significant, the main effects can be analyzed with the model as originally written. This is generally the method used when analyzing designed experiments.

However, even when analyzing designed experiments, some statisticians suggest that if the interaction is nonsignificant, the interaction effect can be deleted from the model and then the main effects are analyzed. This increases the power of the main effects tests.

Neter, Kutner, Wasserman, and Nachtsheim (1996) suggest both guidelines for when to delete the interaction from the model:

- There are fewer than five degrees of freedom for the error.
- The F value for the interaction term is < 2 .



When you analyze data from an observational study, it is more common to delete the non-significant interaction from the model and then analyze the main effects.



Two-Way ANOVA with Interactions

Before conducting an analysis of variance, you should explore the data.

Presume that the initial data exploration was completed (output not shown here) and that no particular concerns were noted about unusual data values or the distribution of the data. During this exploration, you determine that the sample sizes for all treatments are not equal. The researchers recruited 240 patients (80 per heart disease category), but only 170 were randomized into the trial.

```
/*st102d06.sas*/ /*Part A*/
proc print data=sasuser.drug(obs=10);
  title 'Partial Listing of Drug Data Set';
run;
```

PROC PRINT Output

Obs	PatientID	DrugDose	Disease	BloodP
1	69	2 B		13
2	162	4 A		-47
3	181	1 B		12
4	209	4 A		-4
5	308	2 A		4
6	331	4 C		37
7	340	4 C		-19
8	350	1 B		-9
9	360	2 B		-17
10	363	4 A		-41

Negative values for **BloodP** mean that diastolic blood pressure was reduced, on average, by that amount. Positive values mean that blood pressure was raised, on average.

```
/*st102d06.sas*/ /*Part B*/
proc format;
  value dosefmt 1='Placebo'
                2='50 mg'
                3='100 mg'
                4='200 mg';
run;

proc means data=sasuser.drug
  mean var std nway;
  class Disease DrugDose;
  var BloodP;
  format DrugDose dosefmt.;
  output out=means mean=BloodP_Mean;
  title 'Selected Descriptive Statistics for Drug Data Set';
run;
```

Selected PROC MEANS statement:

OUTPUT This statement creates an output data set that contains values and statistics requested in the statement.

Selected PROC MEANS statement option:

NWAY When you include CLASS variables, NWAY specifies that the output data set contains only statistics for the observations with the highest _TYPE_ and _WAY_ values. NWAY corresponds to the combination of all class variables.

PROC MEANS Output

Analysis Variable : BloodP					
Disease	DrugDose	N Obs	Mean	Variance	Std Dev
A	Placebo	12	1.3333333	183.1515152	13.5333483
	50 mg	16	-9.6875000	356.7625000	18.8881577
	100 mg	13	-26.2307692	329.0256410	18.1390640
	200 mg	18	-22.5555556	445.0849673	21.0970369
B	Placebo	15	-8.1333333	285.9809524	16.9109714
	50 mg	15	5.4000000	479.1142857	21.8886794
	100 mg	14	24.7857143	563.7197802	23.7427838
	200 mg	13	23.2307692	556.3589744	23.5872630
C	Placebo	14	0.4285714	411.8021978	20.2929100
	50 mg	13	-4.8461538	577.6410256	24.0341637
	100 mg	14	-5.1428571	195.5164835	13.9827209
	200 mg	13	1.3076923	828.5641026	28.7847894

The mean blood pressure reduction seemed to change at different levels of **DrugDose**. These changes, however, do not seem to follow a consistent pattern across **Disease** categories.

To further explore the numerous treatments, examine the PROC MEANS output graphically.

```
/*st102d06.sas*/ /*Part C*/
proc sgplot data=means;
  series x=DrugDose y=BloodP_Mean / group=Disease markers;
  xaxis integer;
  title 'Plot of Stratified Means in Drug Data Set';
  format DrugDose dosefmt.;
run;
```

The SERIES statement creates a line plot.

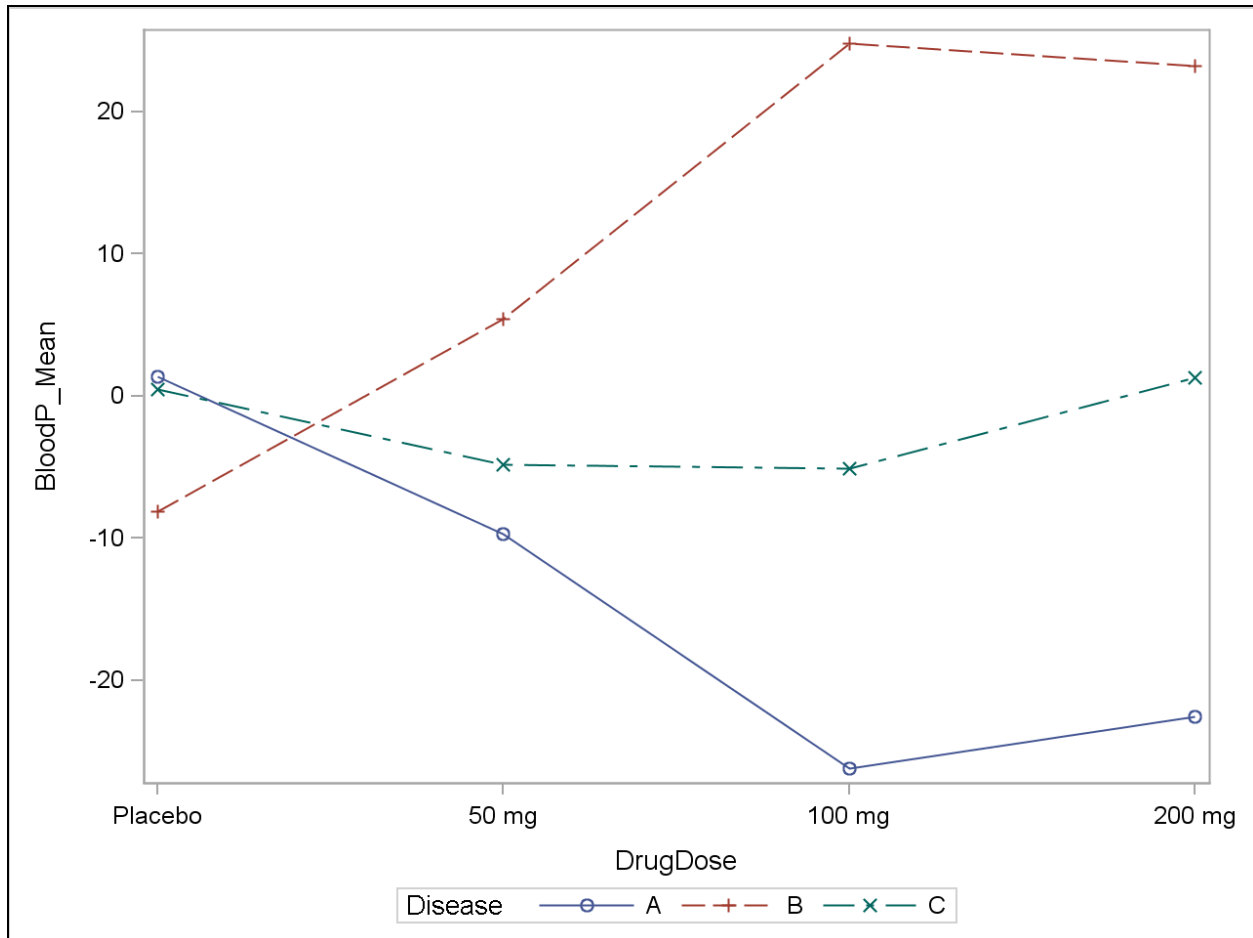
Selected SERIES statement option:

MARKERS adds data point markers to the series plot data points.

Selected AXIS statement:

XAXIS INTEGER forces the X-axis to have tick marks only at integer values.

PROC SGPLOT Output



From the graph, the relationship is clearer. For disease type A, blood pressure falls to a greater degree as the drug level increases through 100 mg, and then the change levels off. For disease type B, blood pressure climbs to a greater degree as the drug level increases through 100 mg, and then the change levels off. For disease type C, blood pressure change is relatively unchanged for different drug levels. In fact, the average change score stays close to 0, implying that the drug had neither positive nor negative effects on these patients, regardless of dose. This plot is exploratory, and helps you plan your analysis. Later you see similar plots output directly from PROC GLM.

You can use the GLM procedure to discover whether these differences and their interactions are statistically significant.

```
/*st102d06.sas*/ /*Part D*/
proc glm data=sasuser.drug order=internal;
  class DrugDose Disease;
  model Bloodp=DrugDose Disease DrugDose*Disease;
  title 'Analyze the Effects of DrugDose and Disease';
  title2 'Including Interaction';
  format DrugDose dosefmt.;
run;
quit;
```

Selected PROC GLM option:

ORDER=DATA | FORMATTED | FREQ | INTERNAL

specifies the sorting order for the levels of all classification variables. The ordering is important for the plot in this case.

As seen in the MODEL statement, the interaction term can be added to the model by using a * to separate the two main effects. It does *not* need to be created in a DATA step.

PROC GLM Output

Class Level Information		
Class	Levels	Values
DrugDose	4	Placebo 50 mg 100 mg 200 mg
Disease	3	A B C

Number of Observations Read	170
Number of Observations Used	170

The next part of the output shows the source table with the *F* test for the overall model. This tests the null hypothesis that none of the effects in the model is statistically different. In other words, that there are no differences among the 12 group means (one for each **DrugDose*Disease** combination).

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	36476.8353	3316.0759	7.66	<.0001
Error	158	68366.4589	432.6991		
Corrected Total	169	104843.2941			

R-Square	Coeff Var	Root MSE	BloodP Mean
0.347918	-906.7286	20.80142	-2.294118

The BloodP Mean value indicates that the average blood pressure change over all observations is -2.294118. (the same value as would be obtained using PROC MEANS.) The R square for this model is 0.347918.

The *p*-value is <.0001. Presuming an alpha equal to 0.05, you reject the null hypothesis and conclude that at least one treatment mean is different from one other treatment mean. Which factor(s) explain this difference?

The next part of the output shows tests of the main effects and the interaction.

Source	DF	Type I SS	Mean Square	F Value	Pr > F
DrugDose	3	54.03137	18.01046	0.04	0.9886
Disease	2	19276.48690	9638.24345	22.27	<.0001
DrugDose*Disease	6	17146.31698	2857.71950	6.60	<.0001

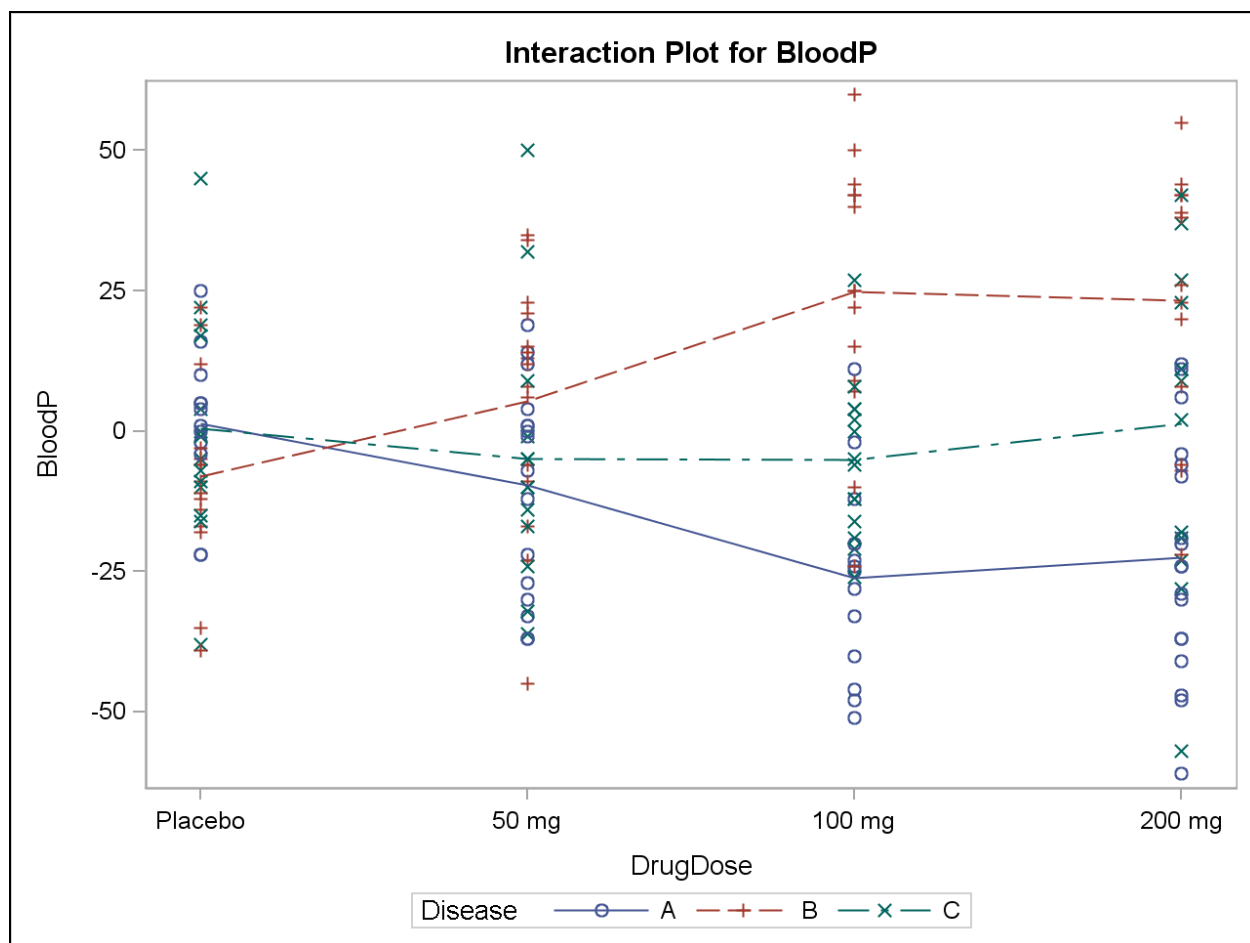
Source	DF	Type III SS	Mean Square	F Value	Pr > F
DrugDose	3	335.73526	111.91175	0.26	0.8551
Disease	2	18742.62386	9371.31193	21.66	<.0001
DrugDose*Disease	6	17146.31698	2857.71950	6.60	<.0001

The Type I SS are *model-order dependent*. Each effect is adjusted only for the preceding effects in the model. They are also known as *sequential sums of squares*. They are useful in cases where the marginal (additional) effect for adding terms in a specific order is important. An example is a test of polynomials, where X , X^2 , and X^3 are in the MODEL statement. Each term is only tested controlling for a lower order term. The TYPE I SS values are additive. They sum to the Model Sum of Squares for the overall model.

The Type III sums of squares are commonly called *partial sums of squares*. The Type III sum of squares for a particular variable is the increase in the model sum of squares due to adding the variable to a model that already contains all the other variables in the model. Type III sums of squares, therefore, do not depend on the order in which the explanatory variables are specified in the model. The Type III SS values are not generally additive (except in a completely balanced design). The values do not necessarily sum to the Model SS.

You will generally interpret and report results based on the Type III SS.

You should consider the test for the interaction first, because if there is an interaction, then by definition this means that the effect of each main effect is different at each level of the other main effect. The p -value for **DrugDose*Disease** is <.0001. Presuming an alpha of 0.05, you reject the null hypothesis. You have sufficient evidence to conclude that there is an interaction between the two factors. As shown in the graph, the effect of the level of drug changes for different disease types. With ODS Graphics, you get an interaction plot, as well.



This plot is similar to the mean plot from PROC SGPLOT.

Because of the interaction, you do not know the effect of **DrugDose** at any particular level of **Disease**. The LSMEANS statement can be used to test the effect of **DrugDose** at each level of **Disease**.

```
/*st102d06.sas*/ /*Part E*/
ods graphics off;
ods select LSMeans SlicedANOVA;
proc glm data=sasuser.drug order=internal;
  class DrugDose Disease;
  model Bloodp=DrugDose Disease DrugDose*Disease;
  lsmeans DrugDose*Disease / slice=Disease;
  title 'Analyze the Effects of DrugDose';
  title2 'at Each Level of Disease';
  format DrugDose dosefmt.;
run;
quit;
ods graphics on;
```

Selected ODS statement:

ODS SELECT specifies output objects for ODS destinations. To specify an output object, you need to know which output objects your SAS program produces. The ODS TRACE statement writes to the SAS log a trace record that includes the path, the label, and other information about each output object that your SAS program produces. The SAS documentation for each PROC also lists output object names.

Selected LSMEANS statement option:

SLICE= specifies effects within which to test for differences between interaction LS-mean effects. This can produce what are known as *tests of simple effects* (Winer 1971). For example, suppose that A*B is significant and you want to test for the effect of A within each level of B. The appropriate LSMEANS statement is as follows:

```
lsmeans A*B / slice=B;
```

PROC GLM Output

The first table is a report of the least squared mean for each unique **DrugDose*Disease** combination.

DrugDose	Disease	BloodP LSMEAN
Placebo	A	1.3333333
Placebo	B	-8.1333333
Placebo	C	0.4285714
50 mg	A	-9.6875000
50 mg	B	5.4000000
50 mg	C	-4.8461538
100 mg	A	-26.2307692
100 mg	B	24.7857143
100 mg	C	-5.1428571
200 mg	A	-22.5555556
200 mg	B	23.2307692
200 mg	C	1.3076923

The second table displays a test of the effect of **DrugDose** at each level of **Disease**.

DrugDose*Disease Effect Sliced by Disease for BloodP					
Disease	DF	Sum of Squares	Mean Square	F Value	Pr > F
A	3	6320.126747	2106.708916	4.87	0.0029
B	3	10561	3520.222833	8.14	<.0001
C	3	468.099308	156.033103	0.36	0.7815

The **DrugDose** effect is significant when used in patients with either disease A or disease B, but not in patients with disease C.

Given all of this information, it seems that you would want to aggressively treat blood pressure in people with disease A with high doses of the drug. For those with disease B (perhaps caused by a traumatic event), treating with the drug at all would be a mistake. For those with disease C, there seems to be no effect on blood pressure.



Exercises

5. Performing Two-Way ANOVA

Consider an experiment to test three different brands of concrete and see whether an additive makes the cement in the concrete stronger. Thirty test plots are poured and the following features are recorded in the **sasuser.concrete** data set:

Strength the measured strength of a concrete test plot

Additive whether an additive was used in the test plot

Brand the brand of concrete being tested

- a. Use the MEANS procedure to examine the data. Output the means to a data set and then plot them using the SGPLOT procedure, and put **Strength** on the Y axis, **Additive** on the X axis, and then stratify by **Brand**. What information can you obtain from looking at the data?
- b. Test the hypothesis that the means are equal, making sure to include an interaction term if the results from PROC SGPLOT indicate that would be advisable. What conclusions can you reach at this point in your analysis?
- c. Do the appropriate multiple comparisons test for statistically significant effects?

2.09 Multiple Answer Poll

A study is conducted to compare the average monthly credit card spending for males versus females. Which statistical method might be used?

- a. One-sample t test
- b. Two-sample t test
- c. One-way ANOVA
- d. Two-way ANOVA

2.6 Solutions

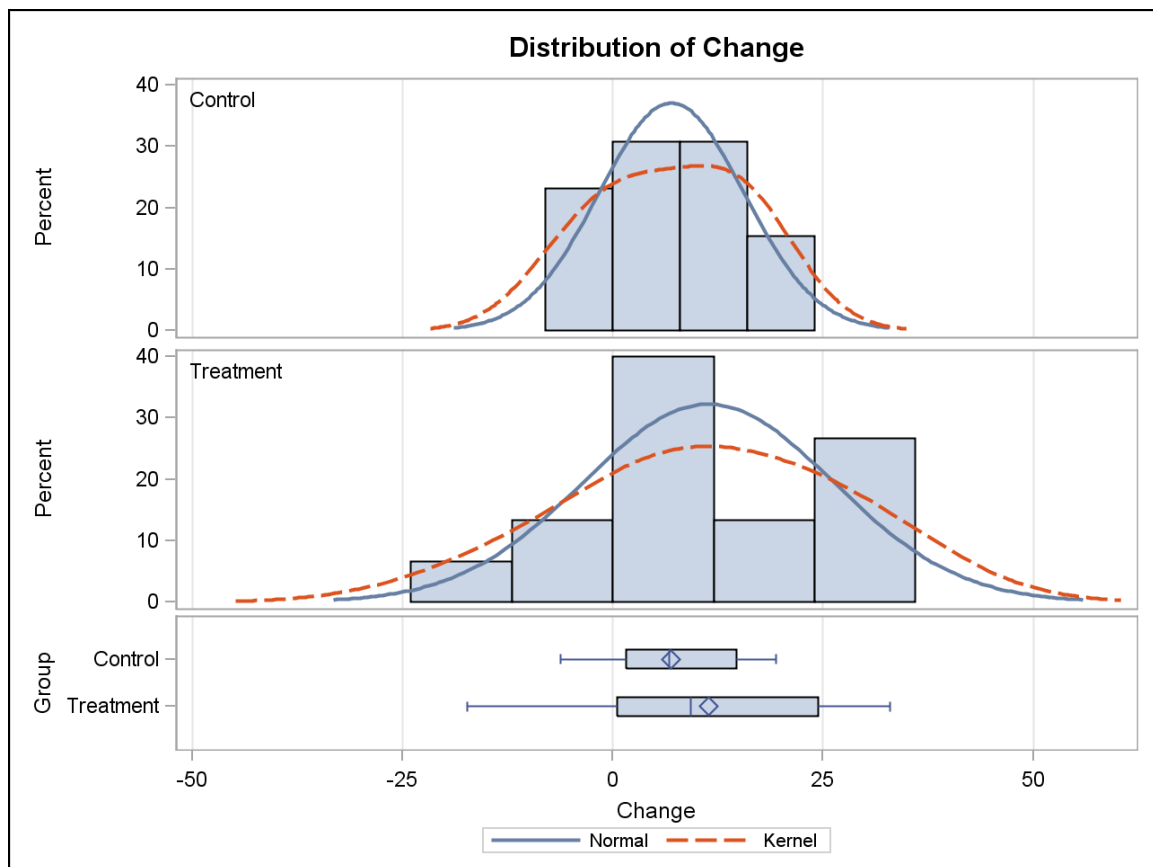
Solutions to Exercises

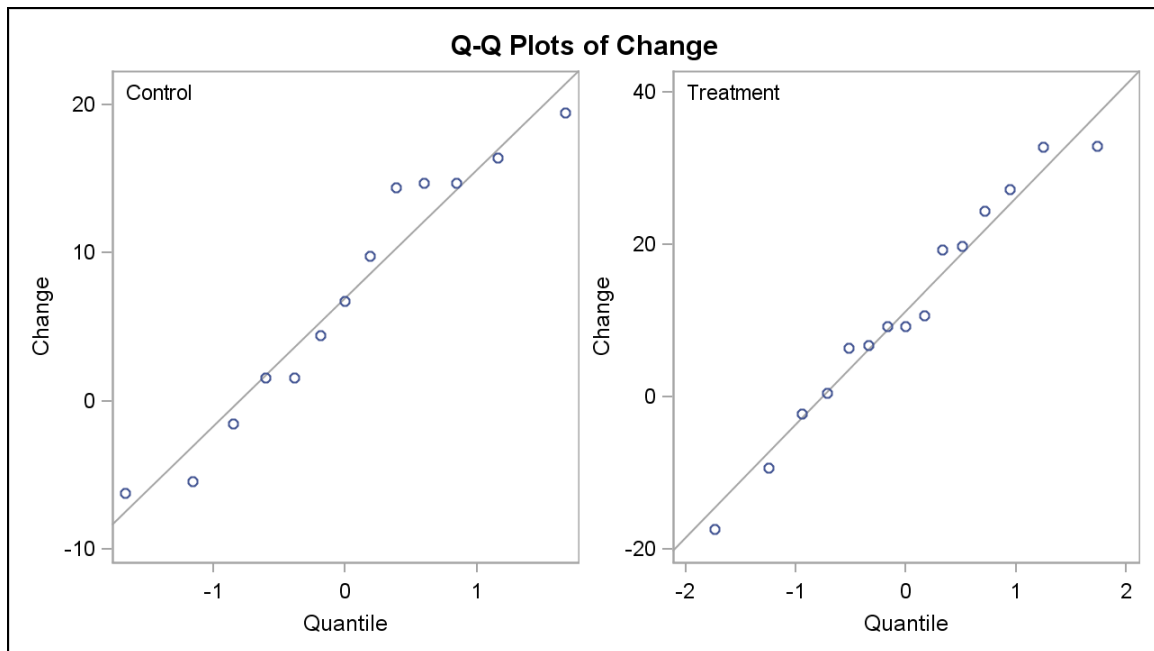
1. Using PROC TTEST for Comparing Groups

Assess whether the treatment group changed the same amount as the control group. Use a two-sided t test.

```
/*st102s01.sas*/
proc ttest data=sasuser.German plots(shownull)=interval;
  class Group;
  var Change;
  title "German Grammar Training, Comparing Treatment to Control";
run;
```

- a. Do the two groups appear to be approximately normally distributed?





The plots show evidence supporting approximate normality in both groups.

- b. Do the two groups have approximately equal variances?

From the bottom of the PROC TTEST output:

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	14	12	2.97	0.0660

Because the *p*-value for the Equality of Variances test is greater than the alpha level of 0.05, you would not reject the null hypothesis. This conclusion supports the assumption of equal variance (the null hypothesis being tested here).

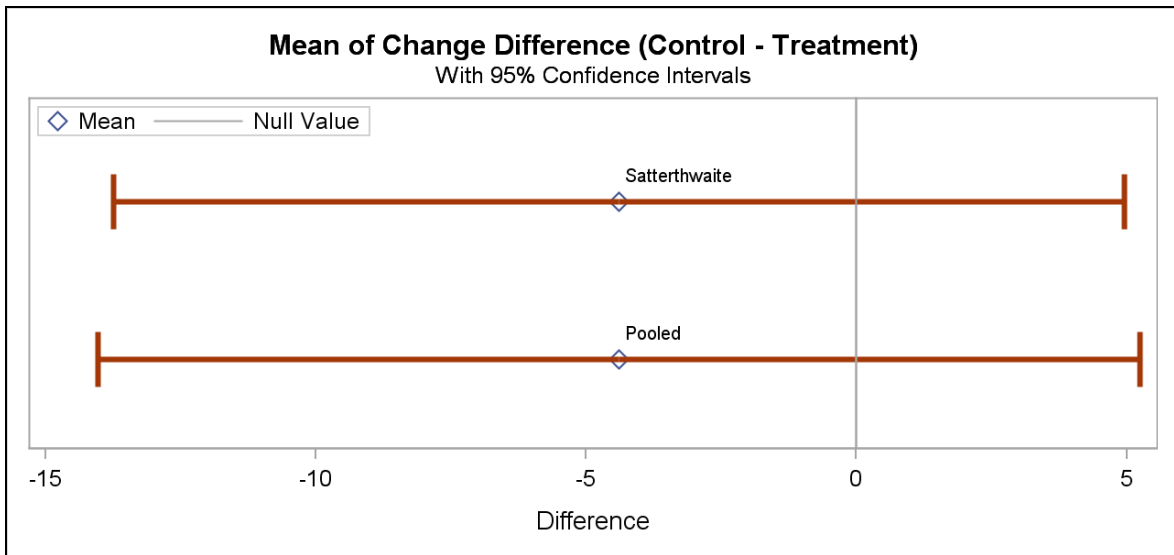
- c. Does the new teaching technique seem to result in significantly different change scores compared with the standard technique?

Group	N	Mean	Std Dev	Std Err	Minimum	Maximum
Control	13	6.9677	8.6166	2.3898	-6.2400	19.4100
Treatment	15	11.3587	14.8535	3.8352	-17.3300	32.9200
Diff (1-2)		-4.3910	12.3720	4.6882		

Group	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
Control		6.9677	1.7607 12.1747	8.6166	6.1789 14.2238
Treatment		11.3587	3.1331 19.5843	14.8535	10.8747 23.4255
Diff (1-2)	Pooled	-4.3910	-14.0276 5.2457	12.3720	9.7432 16.9550
Diff (1-2)	Satterthwaite	-4.3910	-13.7401 4.9581		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	26	-0.94	0.3576
Satterthwaite	Unequal	22.947	-0.97	0.3413

The p -value for the Pooled (Equal Variance) test for the difference between the two means shows that the two groups are not statistically significantly different. Therefore, there is not strong enough evidence to say conclusively that the new teaching technique is different from the old. The Difference Interval plot displays these conclusions graphically.



The confidence interval includes the value zero, indicating a lack of statistical significance of the mean difference.

2. Analyzing Data in a Completely Randomized Design

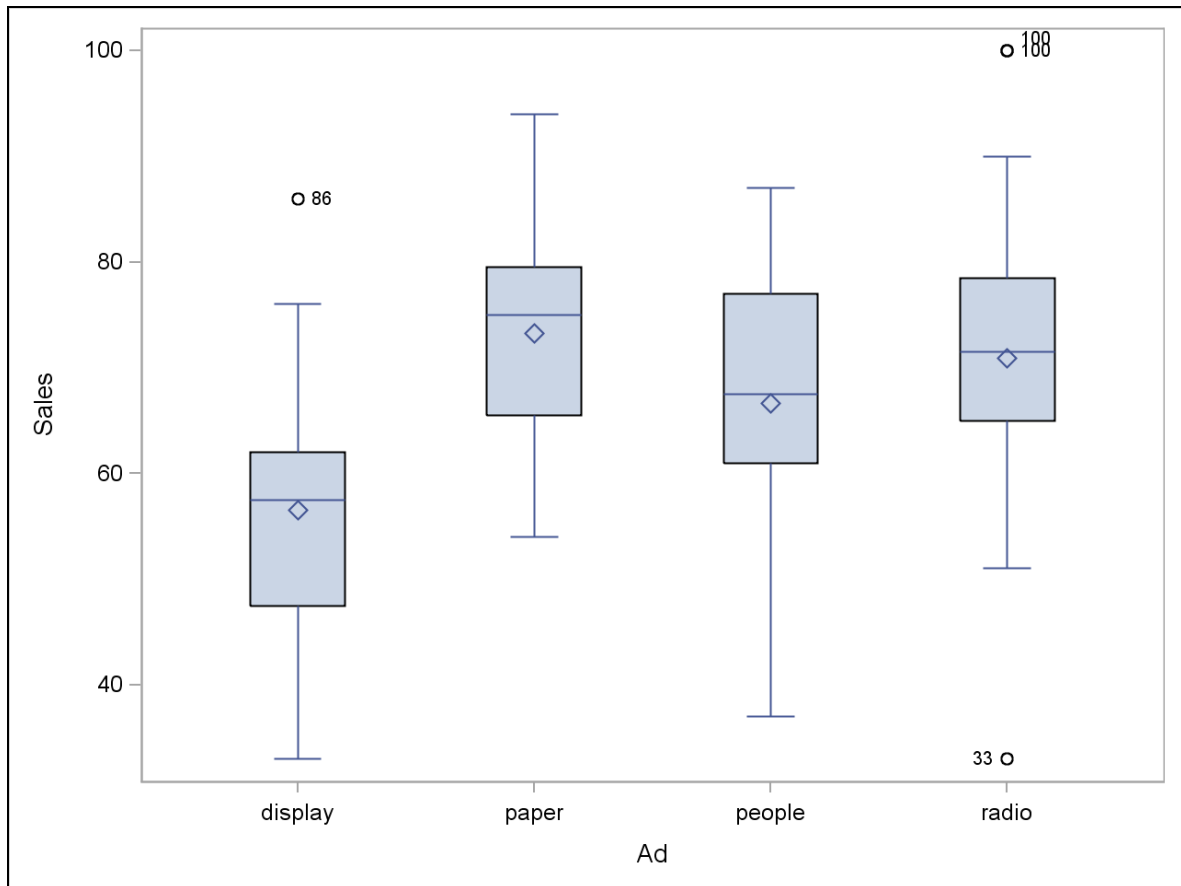
- a. Examine the data. Use the MEANS (using the SKEWNESS and KURTOSIS options in the PROC MEANS statement) and SGPLOT procedures. What information can you obtain from looking at the data?

```
/*st102s02.sas*/ /*Part A*/
proc means data=sasuser.Ads printalltypes n mean std skewness
  kurtosis;
  var Sales;
  class Ad;
  title 'Descriptive Statistics of Sales by Ad Type';
run;

proc sgplot data=sasuser.Ads;
  vbox Sales / category=Ad datalabel=Sales;
  title "Box and Whisker Plots of Sales by Ad Type";
run;
```

Analysis Variable : Sales						
N						
Obs	N	Mean	Std Dev	Skewness	Kurtosis	
144	144	66.8194444	13.5278282	-0.2547089	-0.1295813	

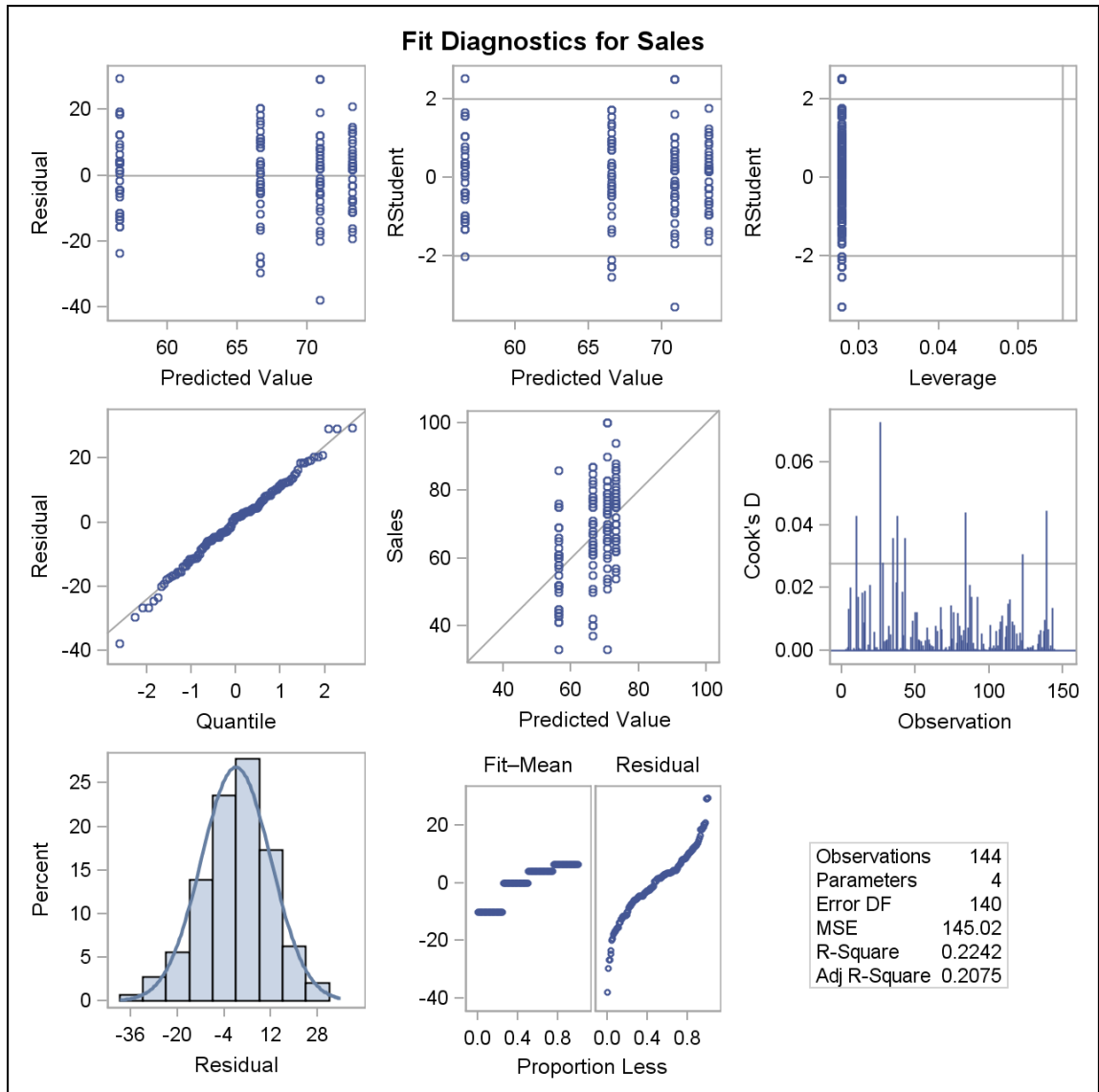
Analysis Variable : Sales						
Ad	Obs	N	Mean	Std Dev	Skewness	Kurtosis
display	36	36	56.5555556	11.6188134	0.3456470	0.0256814
paper	36	36	73.2222222	9.7339204	-0.0474705	-0.5475341
people	36	36	66.6111111	13.4976776	-0.5998808	-0.2130516
radio	36	36	70.8888889	12.9676031	-0.2172278	1.6565242



It appears that the in-store display mean is lower than the others. The value display has a positive outlier, and radio has outliers in both directions.

- b. Test the hypothesis that the means are equal. Be sure to check that the assumptions of the analysis method that you choose are met. What conclusions can you reach at this point in your analysis?

```
/*st102s02.sas*/ /*Part B*/
proc glm data=sasuser.Ads plots=diagnostics;
  class Ad;
  model Sales=Ad;
  means Ad / hovtest;
  title 'Testing for Equality of Ad Type on Sales';
run;
quit;
```



Both the histogram and Q-Q plot show that the residuals seem normally distributed (one assumption for ANOVA).

Levene's Test for Homogeneity of Sales Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Ad	3	154637	51545.6	1.10	0.3532
Error	140	6586668	47047.6		

The Levene's Test for Homogeneity of Variance shows a p -value greater than alpha. Therefore, do not reject the hypothesis of homogeneity of variances (equal variances across Ad types). This assumption for ANOVA is met.

Class Level Information	
Class	Levels Values
Ad	4 display paper people radio

Number of Observations Read	144
Number of Observations Used	144

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	5866.08333	1955.36111	13.48	<.0001
Error	140	20303.22222	145.02302		
Corrected Total	143	26169.30556			

R-Square	Coeff Var	Root MSE	Sales Mean
0.224159	18.02252	12.04255	66.81944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Ad	3	5866.08333	1955.36111	13.48	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Ad	3	5866.08333	1955.36111	13.48	<.0001

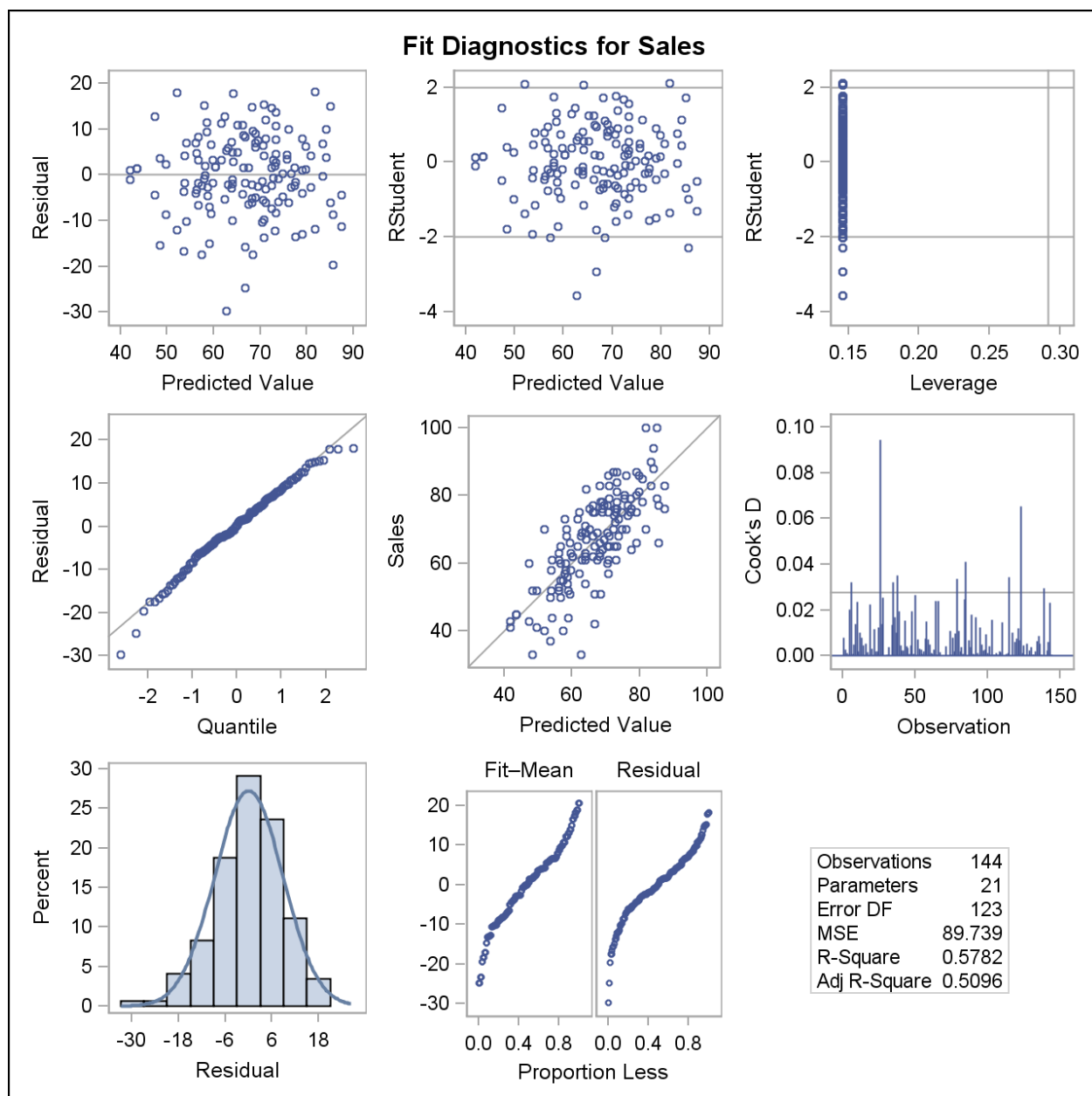
The overall F value from the analysis of variance table is associated with a p -value less than or equal to .0001. Presuming that all assumptions of the model are valid, you know that at least one treatment mean is different from one other treatment mean. At this point, you do not know which means are significantly different.

3. Analyzing Data in a Randomized Block Design

Test the hypothesis that the means are equal. Include all of the variables in your MODEL statement.

```
/*st102s03.sas*/
proc glm data=sasuser.Ads1 plots(only)=diagnostics;
  class Ad Area;
  model Sales=Ad Area;
  title 'ANOVA for Randomized Block Design';
run;
quit;
```

Partial PROC GLM Output



The Q-Q plot of residuals indicates that the normality assumption for ANOVA is met.

Class Level Information																			
Class	Levels	Values																	
Ad	4	display paper people radio																	
Area	18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18

Number of Observations Read	144
Number of Observations Used	144

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	15131.38889	756.56944	8.43	<.0001
Error	123	11037.91667	89.73916		
Corrected Total	143	26169.30556			

The ANOVA table shows that there is some difference in mean sales level across Ad types or across geographic areas (or both).

R-Square	Coeff Var	Root MSE	Sales Mean
0.578211	14.17712	9.473076	66.81944

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Ad	3	5866.083333	1955.361111	21.79	<.0001
Area	17	9265.305556	545.017974	6.07	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Ad	3	5866.083333	1955.361111	21.79	<.0001
Area	17	9265.305556	545.017974	6.07	<.0001

- a. What can you conclude from your analysis?

The p -value for Ad (<.0001) indicates that there was some difference in sales among the advertising campaign types, when controlling for Area.

- b. Was adding the blocking factor Area into the design and analysis detrimental to the test of Ad?

The large (statistically significant) F value for Area gives evidence that area of the country was a useful factor to block on. It was definitely not detrimental.

4. Post Hoc Pairwise Comparisons

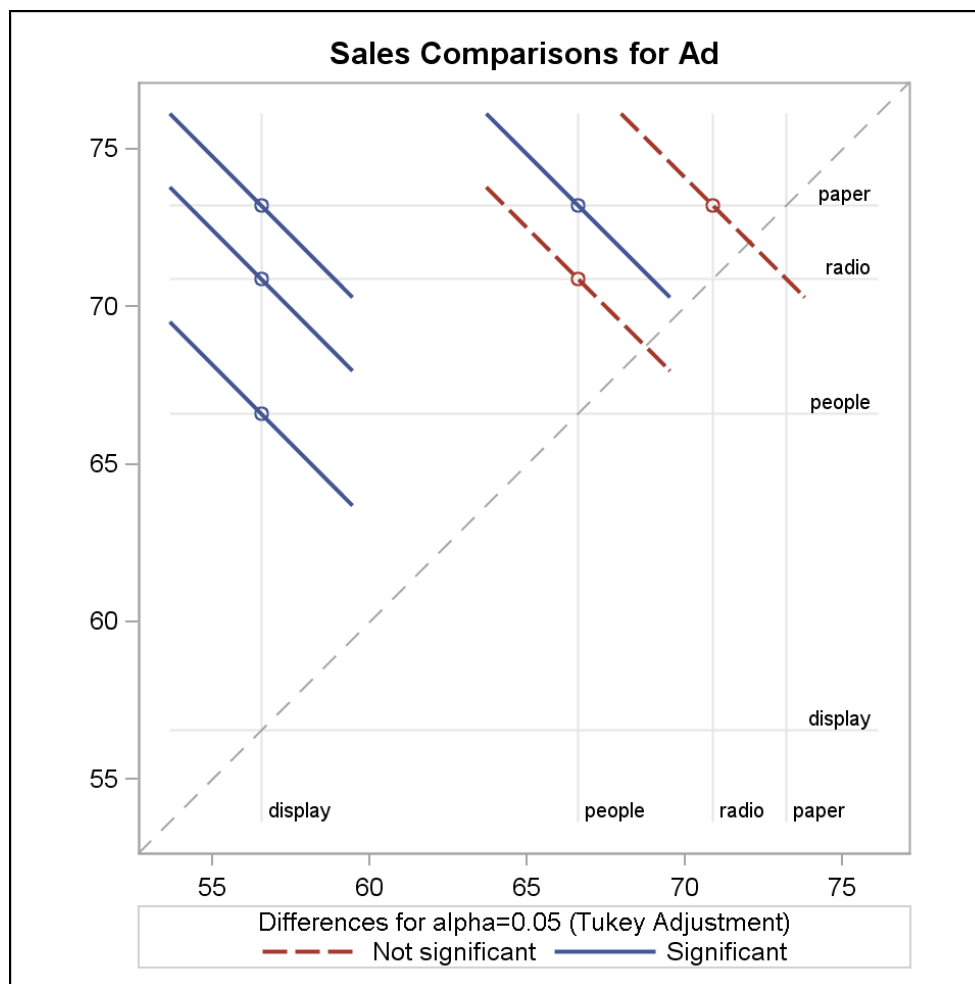
- a. Conduct pairwise comparisons with an experimentwise error rate of $\alpha=0.05$. (Use the Tukey adjustment.) Which types of advertising are significantly different?

```
/*st102s04.sas*/ /*Part A*/
proc glm data=sasuser.Ads1 plots(only)=diffplot(center);
  class Ad Area;
  model Sales=Ad Area;
  lsmeans Ad / pdiff=all adjust=tukey;
  title 'Tukey Pairwise Differences for Ad Types on Sales';
run;
quit;
```

Partial Output

Ad	Sales LSMEAN	LSMEAN Number
display	56.5555556	1
paper	73.2222222	2
people	66.6111111	3
radio	70.8888889	4

Least Squares Means for effect Ad Pr > t for H0: LSMean(i)=LSMean(j)				
Dependent Variable: Sales				
i/j	1	2	3	4
1		<.0001	<.0001	<.0001
2	<.0001		0.0190	0.7233
3	<.0001	0.0190		0.2268
4	<.0001	0.7233	0.2268	

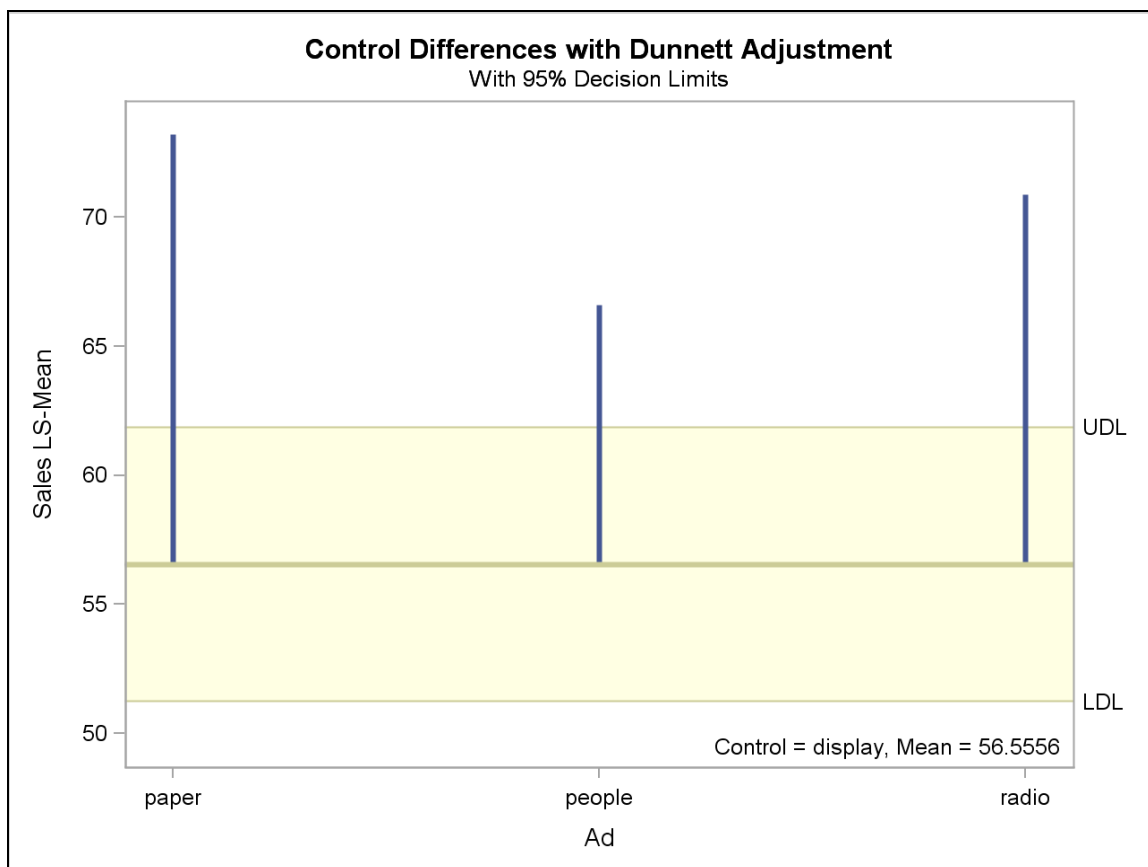


The Tukey comparisons show significant differences between display and all other types of advertising and between paper and people ($p=0.0190$).

- b. Use **display** (case sensitive) as the control group and do a Dunnett comparison of all other advertising methods to see whether those methods resulted in significantly different amounts of sales compared with display ads in stores.

```
/*st102s04.sas*/ /*Part B*/
proc glm data=sasuser.Ads1 plots(only)=controlplot;
  class Ad Area;
  model Sales=Ad Area;
  lsmeans Ad / pdiff=control('display') adjust=dunnett;
  title 'Dunnett Pairwise Differences for Ad Types on Sales';
run;
quit;
```

Ad	Sales LSMEAN	H0:LSMean=Control
		Pr > t
display	56.5555556	
paper	73.2222222	<.0001
people	66.6111111	<.0001
radio	70.8888889	<.0001



All other advertising campaigns resulted in significantly different (higher, in this case) average sales than display.

5. Performing Two-Way ANOVA

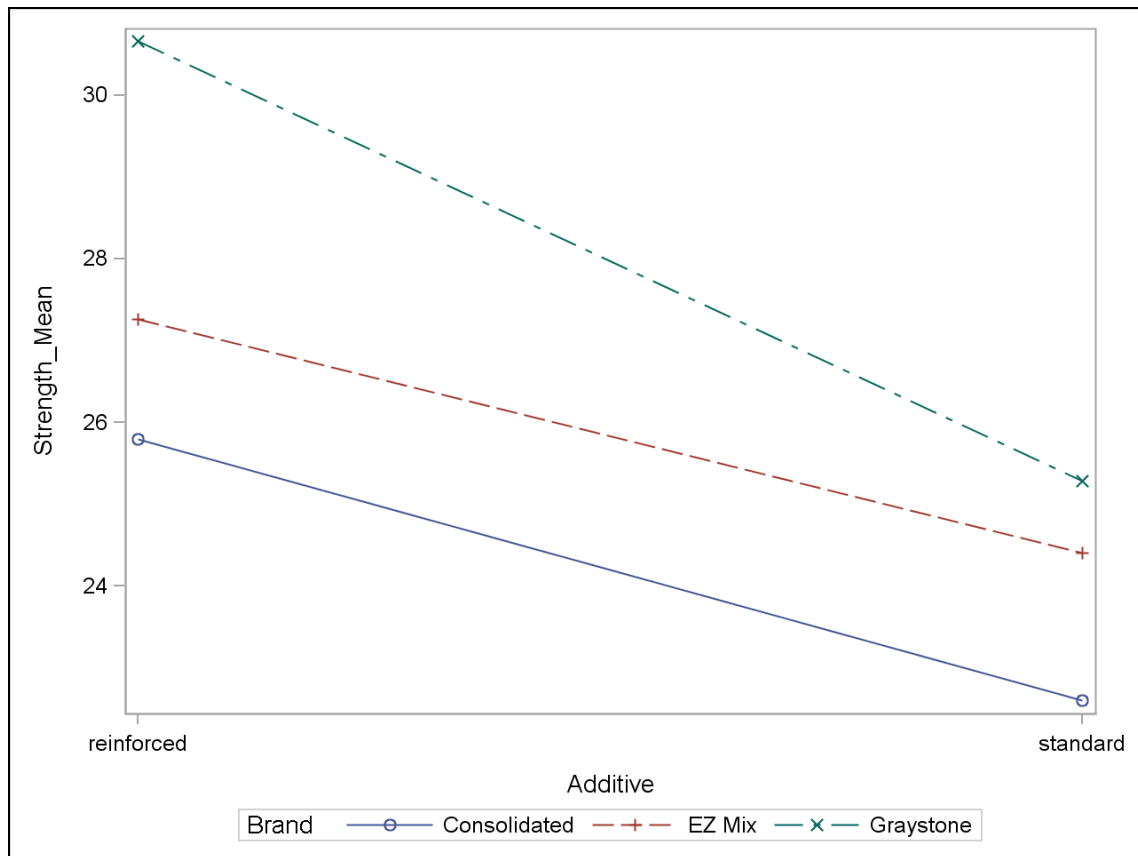
- a. Use the MEANS procedure to examine the data. Output the means to a data set and then plot them using the SGPLOT procedure, and put **Strength** on the Y axis, **Additive** on the X axis, and then stratify by **Brand**. What information can you obtain from looking at the data?

```
/*st102s05.sas*/ /*Part A*/
proc means data=sasuser.concrete
    mean var std nway;
    class Brand Additive;
    var Strength;
    output out=means mean=Strength_Mean;
    title 'Selected Descriptive Statistics for Concrete Data Set';
run;

proc sgplot data=means;
    series x=Additive y=Strength_Mean / group=Brand markers;
    xaxis integer;
    title 'Plot of Stratified Means in Concrete Data Set';
run;
```

Analysis Variable : Strength					
Brand	Additive	N Obs	Mean	Variance	Std Dev
Consolidated	reinforced	5	25.8000000	5.6300000	2.3727621
	standard	5	22.6000000	2.3450000	1.5313393
EZ Mix	reinforced	5	27.2600000	3.8430000	1.9603571
	standard	5	24.4000000	14.2350000	3.7729299
Graystone	reinforced	5	30.6600000	1.7930000	1.3390295
	standard	5	25.2800000	9.8920000	3.1451550

```
/*Alternative code to generate same plot but not using the MEANS procedure.*/
proc sgplot data=sasuser.concrete;
    vline Additive / group=Brand stat=mean response=Strength;
run;
```



The difference between means (reinforced minus standard) for Consolidated is 3.20; the mean difference for EZ Mix is 2.86; and for Graystone, the difference is approximately 5.38. It appears that the difference between concretes using standard and reinforced cements differs by brand. In other words, it appears that there is an interaction between Additive and Brand. That means that an interaction term in the ANOVA model would be appropriate to assess statistical significance of the interaction.

- b. Test the hypothesis that the means are equal, making sure to include an interaction term if the results from PROC SGPLOT indicate that would be advisable. What conclusions can you reach at this point in your analysis?

```
/*st102s05.sas*/ /*Part B*/
proc glm data=sasuser.concrete;
  class Additive Brand;
  model Strength=Additive Brand Additive*Brand;
  title 'Analyze the Effects of Additive and Brand';
  title2 'on Concrete Strength';
run;
quit;
```

Class Level Information		
Class	Levels	Values
Additive	2	reinforced standard
Brand	3	Consolidated EZ Mix Graystone

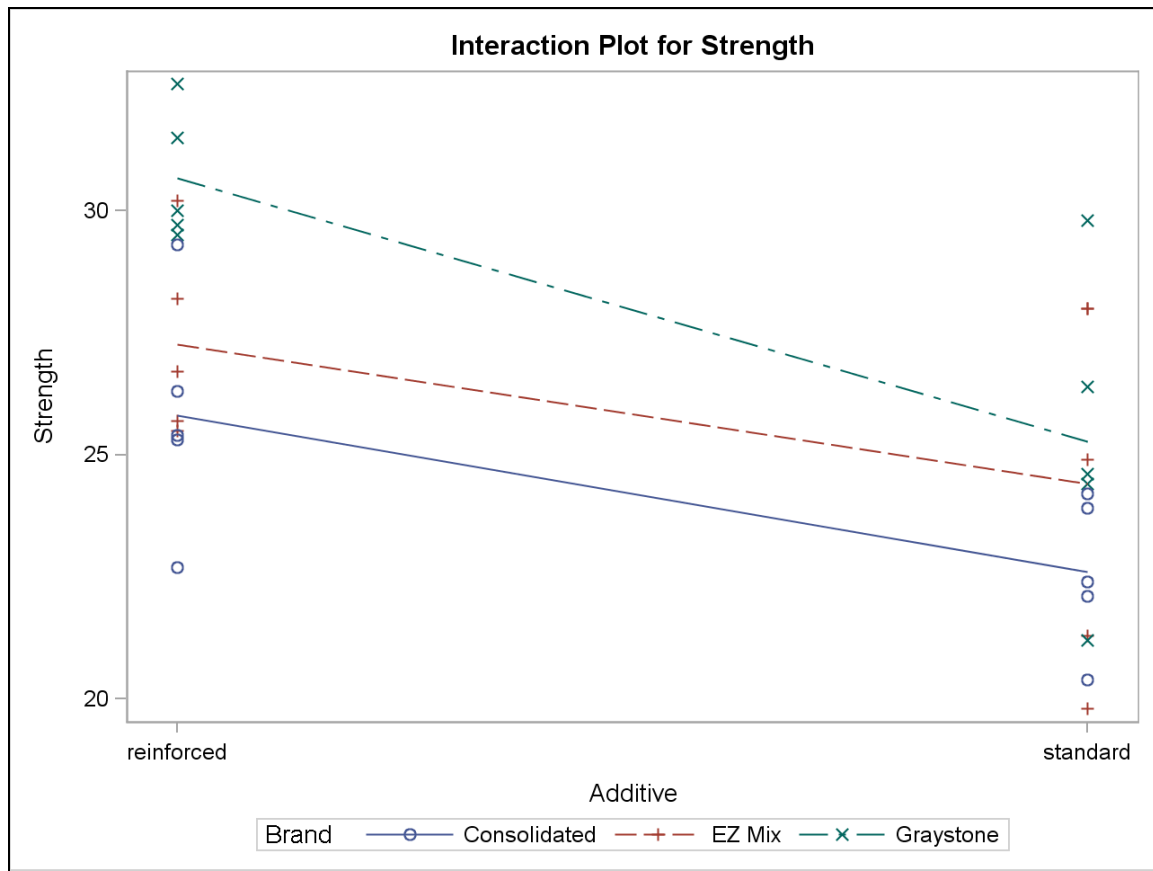
Number of Observations Read	30
Number of Observations Used	30

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	189.9080000	37.9816000	6.04	0.0009
Error	24	150.9520000	6.2896667		
Corrected Total	29	340.8600000			

R-Square	Coeff Var	Root MSE	Strength Mean
0.557144	9.645849	2.507921	26.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Additive	1	109.0613333	109.0613333	17.34	0.0003
Brand	2	71.4980000	35.7490000	5.68	0.0095
Additive*Brand	2	9.3486667	4.6743333	0.74	0.4862

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Additive	1	109.0613333	109.0613333	17.34	0.0003
Brand	2	71.4980000	35.7490000	5.68	0.0095
Additive*Brand	2	9.3486667	4.6743333	0.74	0.4862



There is no significant interaction between Additive and Brand, even though the plot shows slightly different slopes among the three brands of concrete. At this point, you can choose to remove the interaction term from the model and, if still significant, conclude that there is a difference in additive types.

- c. Do the appropriate multiple comparisons test for statistically significant effects?

Because the interaction was not significant, in many cases the interaction term can be removed from the model. A multiple comparisons option in the LSMEANS statement is unnecessary because there are only two levels of the Additive variable. Therefore, the p -value for Additive will also be the p -value for the comparison between group means for standard and reinforced.

```
/*st102s05.sas*/ /*Part C*/
ods graphics off;
proc glm data=sasuser.concrete;
  class Additive Brand;
  model Strength=Additive Brand;
  lsmeans Additive;
  title 'Analyze the Effects of Additive and Brand';
  title2 'on Concrete Strength without Interaction';
run;
quit;
ods graphics on;
```

Class Level Information		
Class	Levels	Values
Additive	2	reinforced standard
Brand	3	Consolidated EZ Mix Graystone

Number of Observations Read	30
Number of Observations Used	30

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	180.5593333	60.1864444	9.76	0.0002
Error	26	160.3006667	6.1654103		
Corrected Total	29	340.8600000			

R-Square	Coeff Var	Root MSE	Strength Mean
0.529717	9.550094	2.483024	26.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Additive	1	109.0613333	109.0613333	17.69	0.0003
Brand	2	71.4980000	35.7490000	5.80	0.0083

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Additive	1	109.0613333	109.0613333	17.69	0.0003
Brand	2	71.4980000	35.7490000	5.80	0.0083

The test for Additive is still significant. There is a difference between standard and reinforced. The estimate of the two least squared means is found in the results for LS Means.

Additive	Strength LSMEAN
reinforced	27.9066667
standard	24.0933333

Reinforced additive in cement in the concrete seems to add more strength than a standard additive does. The mean difference is about 3.8.

Solutions to Student Activities (Polls/Quizzes)

2.01 Multiple Answer Poll – Correct Answer

How do you tell PROC TTEST that you want to do a two-sample t test?

- a. SAMPLE=2 option
- ☒ b. CLASS statement
- c. GROUPS=2 option
- d. PAIRED statement

13

2.02 Multiple Choice Poll – Correct Answer

If you have 20 observations in your ANOVA and you calculate the residuals, to which of the following would they sum?

- a. -20
- ☒ b. 0
- c. 20
- d. 400
- e. Unable to tell from the information given

41

2.03 Multiple Choice Poll – Correct Answer

If you have 20 observations in your ANOVA and you calculate the squared residuals, to which of the following would they sum?

- a. -20
- b. 0
- c. 20
- d. 400
- ☒ e. Unable to tell from the information given

43

2.04 Multiple Choice Poll – Correct Answer

Which part of the ANOVA tables contains the variation due to nuisance factors?

- a. Sum of Squares Model
- ☒ b. Sum of Squares Error
- c. Degrees of Freedom

55

2.05 Multiple Choice Poll – Correct Answer

In a block design, which part of the ANOVA table contains the variation due to the nuisance factor?

- ☒ a. Sum of Squares Model
- ☐ b. Sum of Squares Error
- ☐ c. Degrees of Freedom

61

2.06 Multiple Answer Poll – Correct Answer

If the blocking variable **Area** had a very small F value, what would be a valid next step? Select all that apply.

- ☐ a. Remove it from the MODEL statement and rerun the analysis.
- ☐ b. Test an interaction term.
- ☒ c. Report the F value and plan a new study.

68

2.07 Multiple Choice Poll – Correct Answer

With a fair coin, your probability of getting heads on one flip is 0.5. If you flip a coin once and got heads, what is the probability of getting heads on the second try?

- ☒ a. 0.50
- b. 0.25
- c. 0.00
- d. 1.00
- e. 0.75

75

2.08 Multiple Choice Poll – Correct Answer

With a fair coin, your probability of getting heads on one flip is 0.5. If you flip a coin twice, what is the probability of getting **at least** one head out of two?

- a. 0.50
- b. 0.25
- c. 0.00
- d. 1.00
- ☒ e. 0.75

77

2.09 Multiple Answer Poll – Correct Answers

A study is conducted to compare the average monthly credit card spending for males versus females. Which statistical method might be used?

- a. One-sample t test
- ☒ b. Two-sample t test
- ☒ c. One-way ANOVA
- d. Two-way ANOVA