Lectures 33-34: Two-Way ANOVA

When the effects of two qualitative factors upon a quantitative response variable are investigated, the procedure is called two-way ANOVA. Although a model exists for two-way analysis of variance, similar to the multiple regression model, it will not be covered in this lecture. we will also postpone the details of the ANOVA Table. Nevertheless, there are some new considerations in two-way ANOVA stemming from the presence of the second factor in the model.

Example**:** The EPA (Environmental Protection Agency) tests public bodies of water for the presence of *coliform* bacteria. Aside from being potentially harmful to people in its own right, this bacteria tend to proliferate in polluted water, making the presence of *coliform* bacteria a surrogate for pollution. Water samples are collected off public beaches, and the number of *coliform* bacterial per cc is determined. (See the file *Coliform Bacteria*.

The EPA is interested in determining the factors that affect *coliform* bacterial formation in a particular county. The county has beaches adjacent to the ocean, a bay, and a sound. The EPA believes that the amount of “flushing” a beach gets may affect the ability of pollution to accumulate in the waters off the beach. The EPA also believes that the geographical location of the beach may be significant. (There could be several reasons for this: the climate may be different in different parts of the county, or land-use may vary across the county, etc.)

As luck would have it, there is at least one beach for each combination of type (ocean, bay, sound) and location (west, central, east) within the county. Because of this, the EPA decides to sample a beach at each of the 9 possible combinations of type and location and conduct a two-way analysis of variance for *coliform* bacterial count. Two independent samples are taken at each beach to allow for an estimation of the natural variation in *coliform* bacterial count (this “repetition” is needed for the computation of *MSE*, which estimates the sample-to-sample variance in bacterial counts).

Two-Way ANOVA Using Statgraphics

To perform a two-way analysis of variance in Statgraphics, follow *Compare* > *Analysis of Variance* > *Multifactor ANOVA* and enter the response and factors into the dependent variable and factor fields, respectively.

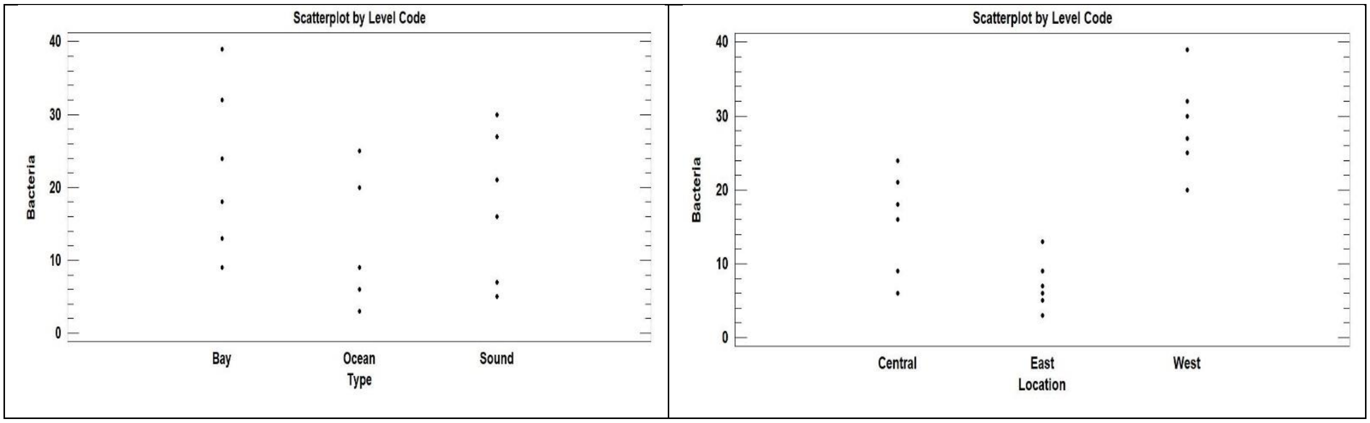
Example 2 (continued)**:** Since data from such a study often appears in the form of a two-way table, with one factor as the row variable, the second as the column variable, and the observations as values in the row-by-column cells, it is important to remember that each variable must have its own column in the spreadsheet as in the example below. (This may require that you re-format the original spreadsheet prior to beginning the analysis.)

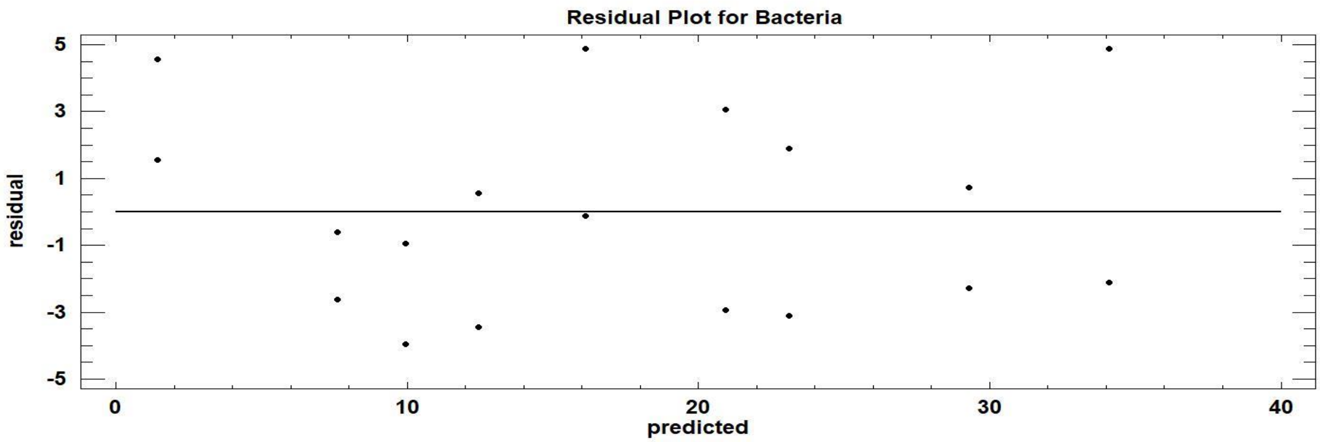
Bacteria	Type	Location
25	Ocean	West
20	Ocean	West
9	Ocean	Central
6	Ocean	Central
3	Ocean	East
6	Ocean	East
32	Bay	West
39	Bay	West

The default ANOVA Table below has separate rows for the factors Type (called factor A) and Location (called factor B). A test of the significance of each factor is performed and the corresponding *P*-value displayed. It appears that both the type of beach and its location affect *coliform* bacterial count.

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:Type	364.778	2	182.389	16.00	0.0003
B:Location	1430.11	2	715.056	62.71	0.0000
RESIDUAL	148.222	13	11.4017		
TOTAL (CORRECTED)	1943.11	17			

We'll make the case stronger by checking the reasonableness of the assumption that errors about the model (which we have yet to specify in detail) are normally distributed with equal variance. Below the scatterplots by type of beach and location. Variance looks fairly constant by level. Finally, The plot of residuals versus predicted coliform bacterial count shows no signs of heteroscedasticity.



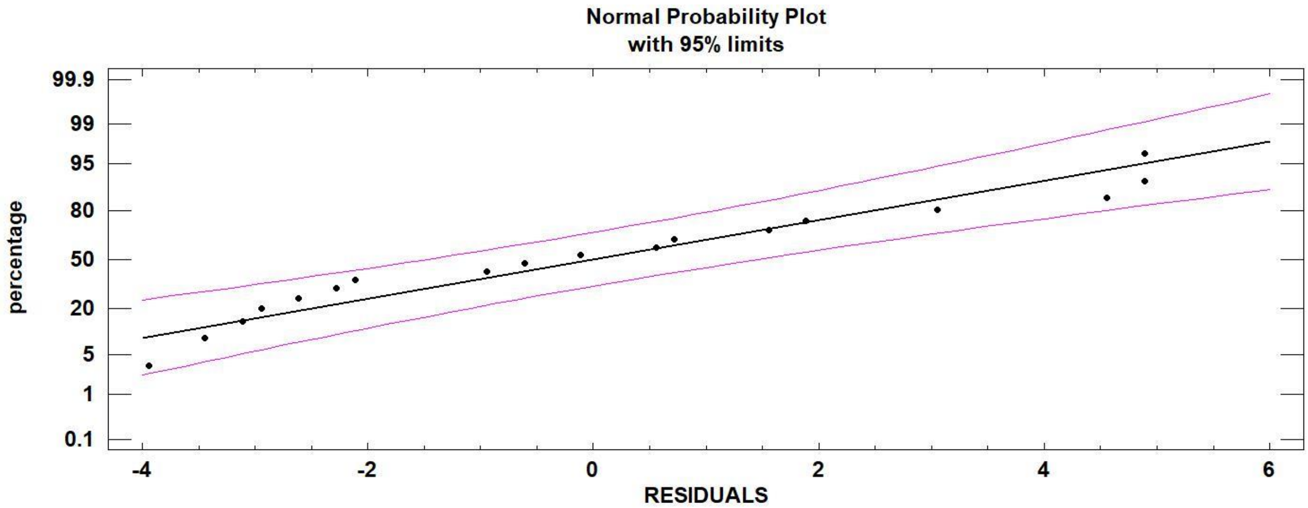


Plot of the Residuals vs Predicted Bacterial Count shows no evidence of Heteroscedasticity

The normal probability plot of residuals shown below, and the accompanying standardized skewness and kurtosis value, won't make us reject the normality assumption.

Stnd. skewness	0.756176
Stnd. kurtosis	-0.908106

The StatAdvisor: The standardized skewness and standardized kurtosis are within the range expected for data from a normal distribution.



But does the effect of the beach type on bacteria count depend upon its location within the county? If the particular pairings of factor levels are important, the factors are said to “interact.”

Analysis of Variance for Bacteria - Type III Sums of Squares

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:Type	364.778	2	182.389	16.00	0.0003
B:Location	1430.11	2	715.056	62.71	0.0000
RESIDUAL	148.222	13	11.4017		
TOTAL (CORRECTED)	1943.11	17			

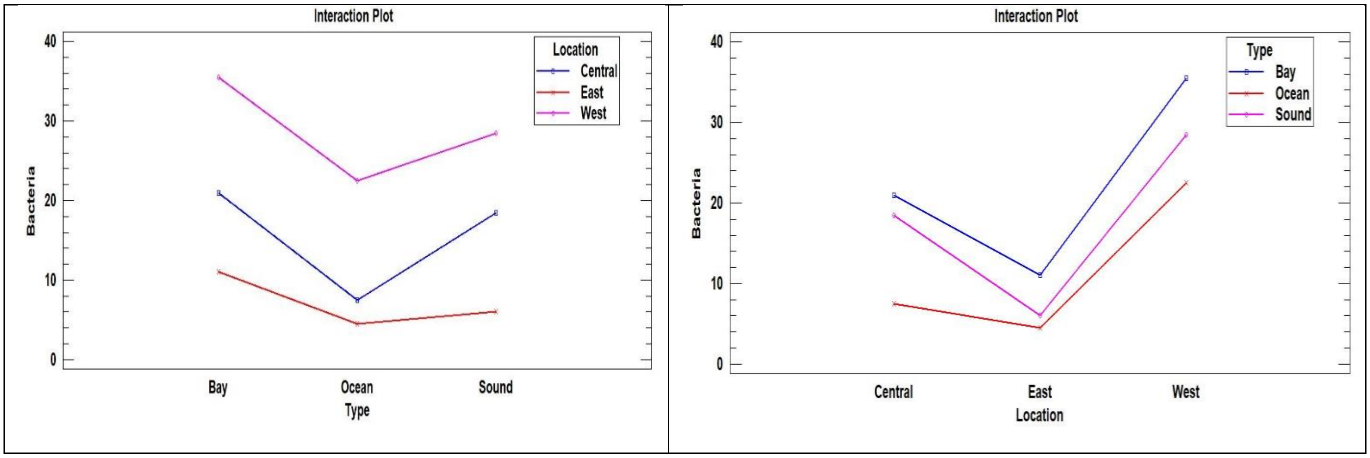
Before interpreting the results in the ANOVA table above, we should consider the role that interaction plays. If the effect of beach type on bacteria formation depends on the location of the beach, then it is better to investigate the *combinations* of the levels of the factors type and location for their effect on coliform bacteria. It should come as no surprise that there is a hypothesis test for interactions.

H_0: The factors Type and Location do *not* interact.
H_A: The factors Type and Location *do* interact

To check for interaction, use the right mouse button and *Analysis Options* and enter “2” for the *Maximum Order Interaction*. The resulting output for our example below shows a *P*-value of 0.3047 for the test for interactions. Thus, the evidence for interaction is not particularly strong. The practical effect of discounting interaction is that we are able to return to the previous output (the one without interactions) and interpret the *P*-values for the factors Type and Location separately. Since the *P*-values for both factors are significant, we conclude that factors affect coliform bacteria growth. Below is the Statgraphics output for an analysis of variance with interactions.

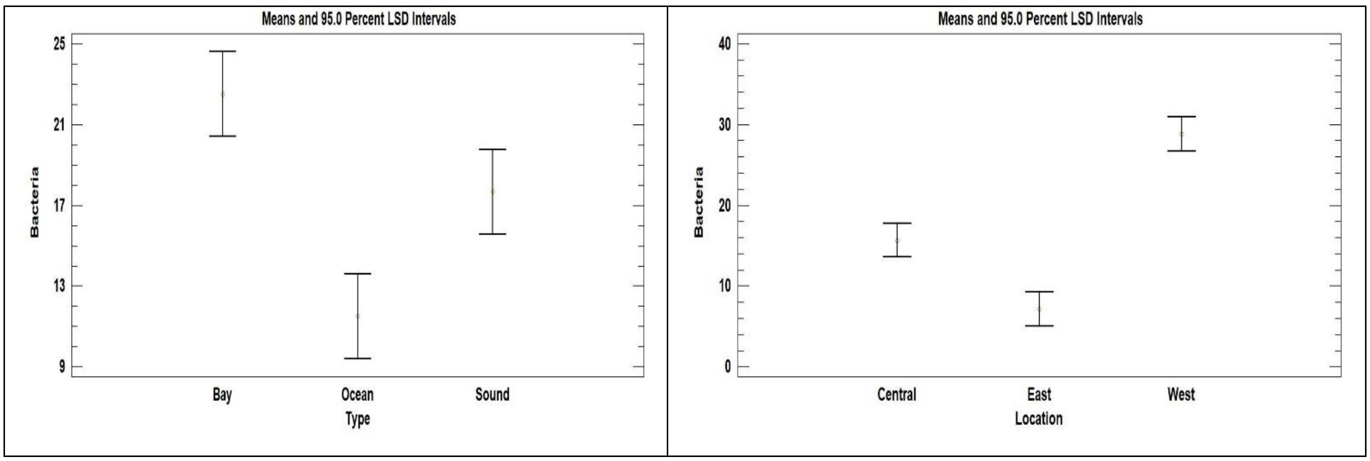
Analysis of Variance for Bacteria - Type III Sums of Squares

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:Type	364.778	2	182.389	18.04	0.0007
B:Location	1430.11	2	715.056	70.72	0.0000
INTERACTIONS					
AB	57.2222	4	14.3056	1.41	0.3047
RESIDUAL	91.0	9	10.1111		
TOTAL (CORRECTED)	1943.11	17			



The Interaction Plots above support the conclusion that interactions aren't significant

Having determined that the type of beach and the beach's location are both significant, we wish next investigate the nature of the relationship between these factors and bacteria count. We shouldOnce again, we turn to the means plots under *Graphical Options*. Statgraphics defaults to a means plot for the factor *Type* because this was the first factor entered in the *Input Dialog Box*. To get a means plot for the factor *Location,* use *Pane Options* to select it. The two means plots appear below.



Individually, these means plots are interpreted as in one-way ANOVA. There is evidence, at the 5% level of significance, that the mean bacteria count at ocean beaches is less than for other types, and that the mean count is highest at bay beaches. Similarly, the mean count is lowest in the east and greatest in the west, with all differences being statistically significant at the 5% level of significance. Furthermore, ***because interactions were judged not-significant***, we can add the main effects together and say that the least polluted beaches tend to be located in the east on the ocean, while the most polluted tend to be in the west on bays. We could not have added the separate (or main) effects in this way if there had been significant interaction, for in that case the effect upon bacteria count at a particular type of beach (ocean, for example) may be very different for different locations.

We finish by reproducing below the Statgraphics output in the *Multiple Range Tests* window.

Multiple Range Tests for Bacteria by Type

Method: 95.0 percent LSD

Type	*Count*	*LS Mean*	*LS Sigma*	*Homogeneous Groups*
Ocean	6	11.5	1.37851	X
Sound	6	17.6667	1.37851	X
Bay	6	22.5	1.37851	X

Contrast	*Sig.*	*Difference*	*+/- Limits*
Bay - Ocean	*	11.0	4.21166
Bay - Sound	*	4.83333	4.21166
Ocean - Sound	*	-6.16667	4.21166

* denotes a statistically significant difference.

Multiple Range Tests for Bacteria by Location

Method: 95.0 percent LSD

Location	*Count*	*LS Mean*	*LS Sigma*	*Homogeneous Groups*
East	6	7.16667	1.37851	X
Central	6	15.6667	1.37851	X
West	6	28.8333	1.37851	X

Contrast	*Sig.*	*Difference*	*+/- Limits*
Central - East	*	8.5	4.21166
Central - West	*	-13.1667	4.21166
East - West	*	-21.6667	4.21166

* denotes a statistically significant difference.