Lecture 16: Two Way ANOVA   Stat 102

- General Description (Read Chapter 9.2 & 9.3)
  - Additive Model (≈ "Randomized Block Design", Chapter 9.2)
  - Model with interactions (Chapter 9.3)
- Example; Agent service times
- General Description (cont.)
  - Theory for the Additive Model
  - Formulas for analysis of the Balanced Additive Model
- Application to the example
- More Theory and formulas; Model with interactions
- Example (cont.)
- Comments on the unbalanced case & other notes

General Description

- The 2-way model is an extension of the 1-way model
- The 1-way model has **1** "factor" with ***I*** categorical "levels"
  - Example in One Way ANOVA had "factor" = Fund Type, with ***I*** = **4** different fund types, *ie* **4** "levels"
- The 2-way model has **2** "factor" types that can be combined in various ways. (Typically, they can be combined in all possible ways.)
- In the Additive Model the effects of these 2 factor types add together
  - This can occur in a "Randomized Block Design" as described in Chapter 9.2; but it can also occur in other situations.
- In a more general (interaction) model the factor effects may also interact with each other.

Example: Agent Service Times

- In a telephone call-center, the Agent-Service-Time of a call is the amount of time the agent spends on a given call.
- This depends on Two types of Factors
  - The Agent (identified here by their name)
  - The Type of Call. There are 3 Call-Types in our data:

    Regular**, Stock** transaction**, New** Customer
- There are many other factors, but these others are ignored in the following discussion.
- It is desirable to see which agent(s) serve their customers faster, after controlling for call-type. (In general, faster is better.)
- *For example, the manager may wish to give a bonus to the fastest agent(s)*

Example: Data

- From an Israeli Bank Call Center
- Service times for a *random sample* of calls handled by **8** different agents in Nov 1999
- There are 3 major types of calls handled by center:

  Regular**, Stock** transaction**, New** Customer
- Random Sample, by design, has **25** calls of each of the **3** types from each of the **8** agents. [= "***Balanced***" sample]
- Traditionally, Service Times (of many sorts) have been treated as if they had an exponential distribution.
  – *Beginning with the research of Erlang in 1911+.*
- However, *Brown,…, Zhao (2005)* found that such times seem to have a lognormal distribution. [*ie*, Their logs are normally distributed.]
  – *We don't really understand WHY?*

  But this has now been verified in several call-center situations

Example: Use of 1-Way ANOVA

- A 1-way ANOVA (on "Server") – ignoring "Call-Type" – is a useful **preliminary analysis**

– Can help us understand the data, and the relation of Servers to call times

– Will help prepare for explanation of the difference between the 1-way and 2-way analyses

- The statistical MODEL in this 1-way analysis has

Y_{ij} = **LogServiceTime** of *j*-th call to service agent *I*,

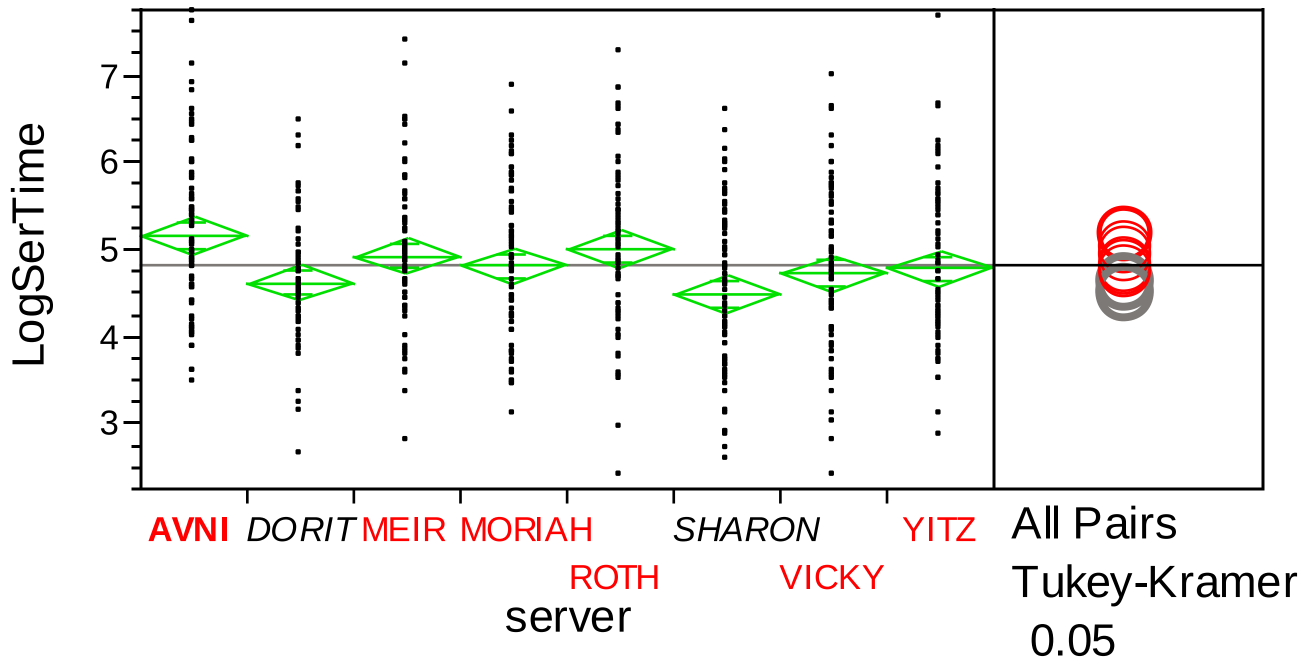
$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} : \varepsilon_{ij} \sim N\left(0, \sigma_e^2\right)$$

$i = 1,..,8;\ j = 1,..,75 =$ total # calls handled by agent i.

*(Note this **model** is homoscedastic.)*

- Here is output from the 1-way ANOVA

Side-by-Side Plot from 1-way ANOVA



- Note that The Tukey-Kramer multiple comparison test shows that **Avni** is worse than Dorit and Sharon; not controlling for Call-Type

2-Way ANOVA: General Model

- Observe

  Y_{ijk} with $i = 1,..,I$; $j = 1,.., J$; and $k = 1,.., K_{ij}$.

- The subscripts ***i*** and ***j*** index the two types of factors
- The subscript ***k*** indexes the repetitions of a type of observation
- A ***Balanced Model*** has $K_{ij} = K$ (a constant) for all i,j.

- Our Example has $I = 8$ (servers), $J = 3$ (Call-Types) and $K=25$ (repetitions for each server&Type combination). Our model is balanced.

Model (without interactions), cont

- The additive model decomposes the population means μ_{ij} = E(Y_{ijk}) as a sum of effects of the corresponding *i* and *j* factors. Thus,

$$\mu_{ij} = \mu + \alpha_i + \beta_j.$$

- We also assume normality, homoscedasticity, & independence – *ie*,

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \text{ with } \varepsilon_{ijk} \text{ independent } N\left(0, \sigma_e^2\right).$$

- The general principle for estimating the unknown values μ, α_i, β_j is – as before – **Least Squares**.
- Thus, these are chosen to min'ize

$$SSE = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K_{ij}} \left(Y_{ijk} - \hat{Y}_{ij}\right)^2$$

where $\hat{Y}_{ij} = \text{est of } \mu_{ij} \triangleq \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$.

Hypothesis Tests (in additive model)

- There is an **overall** null hypothesis to be tested: There are no factor effects, *ie*

  H_0: $\alpha_1 = \ldots = \alpha_I = 0$ **and** $\beta_1 = \ldots = \beta_J = 0$

  H_a: H_0 is not true.

- There are also two separate **sub-hypotheses** of additional interest:

No Factor A effects: H_{0A}: $\alpha_1 = \ldots = \alpha_I = 0$ vs H_{aA}: H_{0A} is not true, &

No Factor B effects: H_{0B}: $\beta_1 = \ldots = \beta_J = 0$ vs H_{aB}: H_{0B} is not true.

1. In ordinary language, these should be interpreted as testing whether there are factor A effects after controlling (additively) for the effects of factor B, and vice-versa.
2. Normally you would only be interested in these tests if the overall test of H_0 rejects.

- There are also tests and CIs for individual estimates of μ, α_i, β_j.
- And CIs and prediction CIs for the estimates of μ_{ij}.

Example: Results for Additive Model

- **ANOVA Table**: Gives test of the overall null hypothesis H_0.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	**I+J-2**=9	80.45	8.94	12.87
Error	**n-I-J+1**=590	409.71	0.69	Prob > F
C. Total	**n-1**=599	490.16		**<.0001**

- F = 12.87 with 9 & 590 DF has P<.0001
- Hence, we REJECT H_0.
- Here DFTotal = $n - 1$; DFModel = $(I\text{-}1) + (J\text{-}1) = I + J - 2$;
  DFError = $(n\text{-}1) - (I+J\text{-}2) = n - I - J + 1$.
- Mean Square = Sum of Squares/DF
- Also, SSTotal = $\sum_{i,j,k}\left(Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet}\right)^2$, as always & $SSE = \sum_{ijk}\left(Y_{ijk} - \hat{Y}_{ij}\right)^2$.

Results, cont

- **Effect Test Table**: Gives tests of H_{0A} and H_{0B}.

Effect Tests

Source	DF	Sum of Squares	F Ratio	Prob > F
CallType	**I-1=**2	56.19	40.46	<.0001
server	**J-1**=7	24.26	4.99	<.0001

- DF(CallType) = $I - 1$ = 2; DF(server) = $J - 1$ = 7.
- The F-ratios test H_{0A} and H_{0B} respectively, with DF^s being DF(effect) & DFError, and P-values as given in the table.
- Both null hypotheses are Rejected here.
- The Effect Sums of Squares have the usual interpretation: They are decrease in SSE when going from the 1-way model without the Effect type to the complete model with both types of effects.
- *When the model is balanced (as here), the Sums of Squares will add to be the SSModel; otherwise usually not.*

Results, cont

- **<u>Parameter Estimates</u>**: We recommend using the "**Expanded Estimates**" table. This has **all** the μ, α_i, β_j parameter estimates.
- Here are some of the entries:

Expanded Estimates

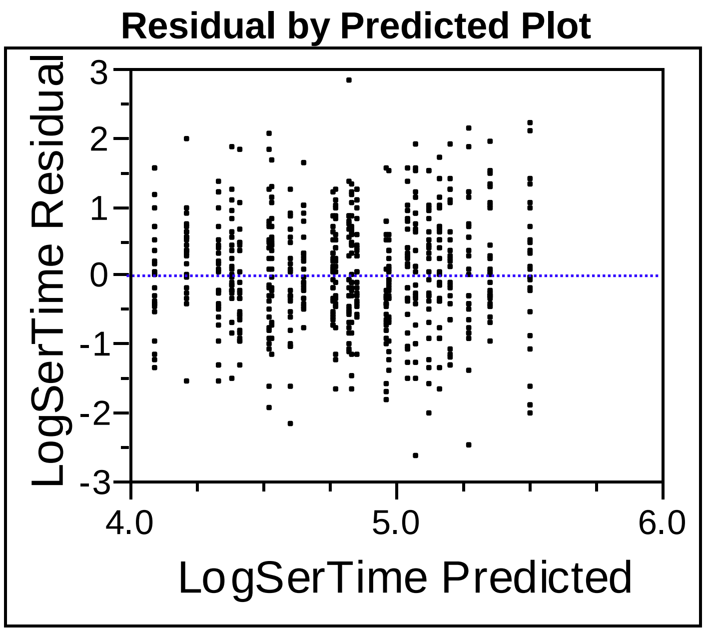
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	**4.823**	0.0340	141.78	0.0000
CallType[New]	-0.395	0.0481	-8.21	<.0001
CallType[Reg]	0.0438	0.0481	0.91	0.3628
CallType[Stock]	**0.351**	0.0481	7.29	<.0001
server[AVNI]	0.3438	0.0900	3.82	0.0001
server[DORIT]	**-0.199**	0.0900	-2.21	0.0275
etc.	…	…	…	…

- Hence (for example) the estimated LogServiceTime for Dorit to handle a Stock call is 4.823 + .351 + (-.199) = **4.975**
- The t-ratios and P-values here test that the individual coefficients are 0. This is not usually an interesting null hypothesis to test.
- You can find CI^s the usual way with the "Save Columns" drop-down.

Results; Model Validation

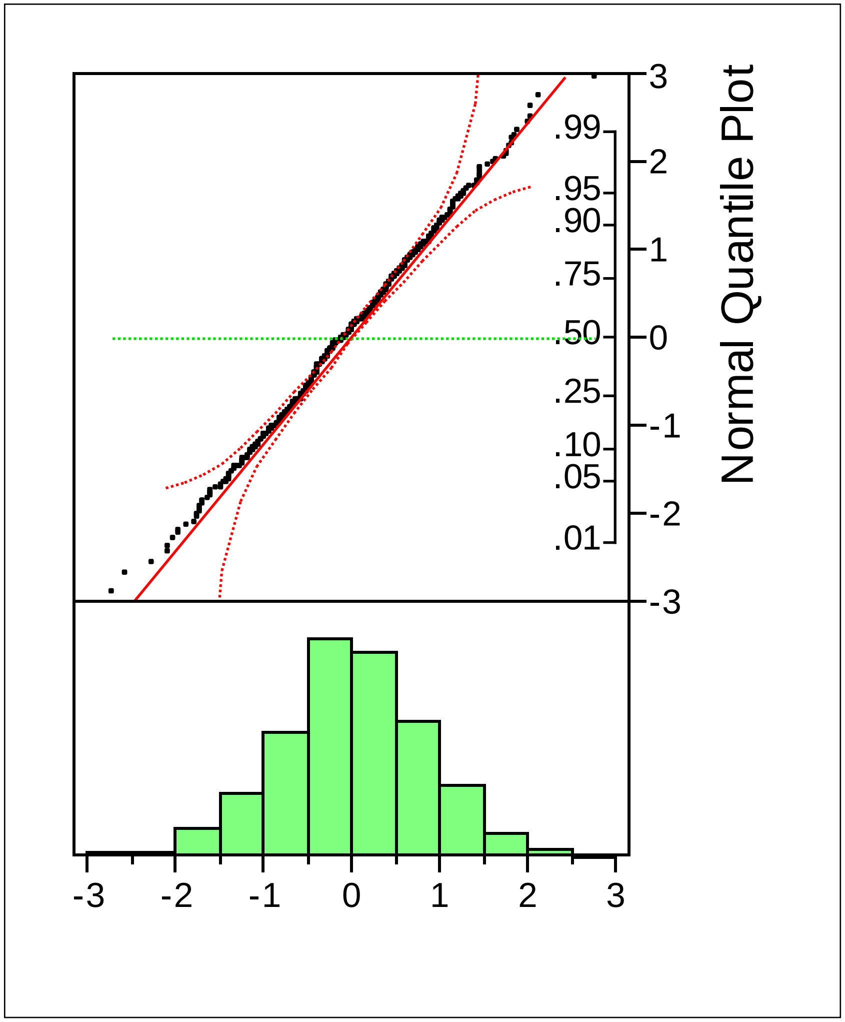
- The residual plot provides a visual check for homoscedasticity (though it can be hard to read because of superimposed points)
- Saving the residuals and then looking at the normal quantile plot of their distribution provides a good check for normality of residuals
- A good check of whether the additive model is suitable is provided by looking at the model with interactions and seeing whether the interaction effects are significantly non-zero (If so, the additive model is "Rejected" as a null hypothesis versus the model with interactions).
- The next two pages show the result of the first two investigations. Then we'll discuss the model with interactions.

Results; Residual Plot



- There is no evident heteroscedasticity here; and that fact provides a visual confirmation that the homoscedasticity assumption is probably OK.
- *[There's also a formal test of homoscedasticity that can be used here, that we won't describe. The result is to 'fail to reject' homoscedasticity.]*

Results; Normal Quantile Plot of Residuals



- This shows an **extremely good agreement** with normality**!**
- It is because of (many) plots like this that we concluded that <u>Service times are lognormal</u> for the Israeli call center.
- *[When we perform a similar 2-way analysis using ordinary ServiceTime this type of residual plot comes out strongly skewed and non-normal.]*

Results; Notes (optional)

- *You can find confidence intervals for individual means or for prediction CI^s from the drop down menu. For example, the 95% Prediction interval for the LogServiceTime for Dorit to handle a Stock call is (3.33, 6.63). This is of course a pretty wide interval, which reflects the fact that the Root Mean Square error here is still pretty large (It's 0.832 from the Summary of Fit table; and this together with the estimate of 4.975 on our p. 13 suggests we should find prediction CIs of about 4.98 ± 2×.83,* ***and this is what we get from the formal procedure*** *in JMP.)*
- *This CI corresponds to a 95% Prediction CI in terms of Service Times of*

$$\left(e^{3.33}, e^{6.63}\right) = \left(27.9,\, 757.5\right).$$

2 of Dorit's 25 Stock calls fell outside of this interval; which is about par for a 95% interval.

- *There is also a way to find multiple confidence intervals here for the effect of the Server, after controlling for the Type of Service. (But this is not automatic in JMP.) It turns out that 95% intervals of this sort show that Avni is significantly longer-winded than Dorit and Sharon, as with the analysis ignoring Service Type;*
- ***And <u>also</u> Roth and Meir are significantly longer-winded than Sharon****.*

General Model: With Interactions

- As before, we have observations Y_{ijk} corresponding to K observations at "level" i, j of two factors.
- Now we include a possible effect of the "interaction" between the two factors, after controlling for the overall (additive) effect of the two factors.
- The interaction "effect" parameter for combination i, j is labeled γ_{ij}.
- The full model for $\mu_{ij} = E\left(Y_{ijk}\right)$ is thus

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

- The model still has the same type of normality, homoscedasticity, and independence assumptions as before.

Estimation of the Means

- In this model the means μ_{ij} can take any numerical values
- Hence the best estimate of μ_{ij} is the ij^{th} cell mean

$$\hat{Y}_{ij} = \overline{Y}_{ij\bullet} \triangleq K_{ij}^{-1} \sum_{k=1}^{K_{ij}} Y_{ijk}$$

- This value minimizes

$$SSE = \sum_{i,j,k} \left(Y_{ijk} - \hat{Y}_{ij} \right)^2$$

- The estimates of $\mu, \alpha_i, \beta_j, \gamma_{ij}$ are then found (in JMP) by solving $\hat{Y}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_{ij}$.
  - *This is done subject to the convenient, minimal set of side conditions*

$$0 = \sum \alpha_i = \sum \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} .$$

Tests and DF^s

- The ANOVA full-model test has $H_0 : \mu_{ij} = \mu$ vs the alternative $H_a : \mu_{ij}$ are any numbers that are not **all** equal.
- Hence the F-statistic in the ANOVA table has

$$\text{DF} = IJ - 1.$$

- It is also possible to test the null hypotheses

$H_{0A} : \alpha_i = 0 \ \forall i$ (no Factor A effect) & $H_{0B} : \beta_j = 0 \ \forall j$ (no Factor B effect)

as well as

$$H_{0Inter} : \gamma_{ij} = 0 \ \forall i, j.$$

- Each of these null hypotheses involves a test ‘controlling for the other variables’ in the usual fashion
- The DF^s are as given in the formulas in red, below

Example: Analysis with Interactions

- Use the Fit Model platform. Choose as Model Effects: 'CallType' & 'Server' & also "cross" 'CallType' & 'Server'

[Watch in lecture how this is done!]

- Here is the ANOVA table:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	**IJ-1**=23	96.32	4.19	6.12
Error	**n-IJ**=576	393.84	0.68	Prob > F
C. Total	**n-1**=599	490.16		<.0001

- Note the DF values.
- As usual,

$$\text{Mean Square} = \text{Sum of Squares}/\text{DF, and}$$

$$F = \text{MSM}odel/MSE\text{, with 23 \& 576 DF.}$$

- We **Reject** H_0:No model effect; since P-value is <.0001

Tests for the effect types

Effect Tests

Source	DF	Sum of Squares	F Ratio	Prob > F
CallType	**I-1**=2	56.19	41.09	<.0001
server	**J-1**=7	24.26	5.07	<.0001
server*CallType	**IJ-I-J+1**=14	15.87	1.66	0.0604

- Note the DF values. Note also that these add to the previous DFModel.
- The Effect Sums of Squares have the usual interpretation: They are decrease in SSE when going from the model without the Effect type to the complete model with all types of effects.
- As usual, $F = \frac{\text{Sum-of-Squares of Effect}}{\text{DF of Effect}} \Big/ \text{MSE}$
- Here, the F for testing CallType has 2 and 576 DF^s; etc.
- At $\alpha = .05$ we **reject** H_{0A} and H_{0B}. We **fail to reject** $H_{0,\text{Inter}}$.

Estimates of Mean Service Times

- Overall, the interaction effects are not quite statistically significant.
- SO we could justifiably decide to just use the additive model from before.
- OR we could still use the full model. If we did that the **Expanded Estimates** table says

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.823	0.0338	142.88	0.0000
CallType[Stock]	0.351	0.0477	7.35	<.0001
server[DORIT]	-0.199	0.0893	-2.23	0.0263
server[DORIT]*CallType[Stock]	-0.423	0.1263	-3.35	0.0009

Hence we estimate the mean for Dorit to handle a Stock call as

$$4.823 + 0.351 + (-0.199) + (-0.423) = 4.552.$$

*The estimates for the CallType and Server effects are **equal** in the no interaction model because the model is **balanced**. Otherwise estimates may differ.*