清华大学统计学辅修课程

Linear Regression Analysis

Lecture 12-
ANOVA Inference &
Two-Way ANOVA

周在莹

清华大学统计学研究中心

http://www.stat.tsinghua.edu.cn

Topic 1:
Inference

Outline

- Review One-way ANOVA
- Inference for means
- Differences in cell means
- Contrasts

Cell Means Model

- $Y_{ij} = \mu_i + \varepsilon_{ij}$

  where μ_i is the theoretical mean or expected value of all observations at level i and the ε_{ij} are iid $N(0, \sigma^2)$
- $Y_{ij} \sim N(\mu_i, \sigma^2)$ independent

Parameters

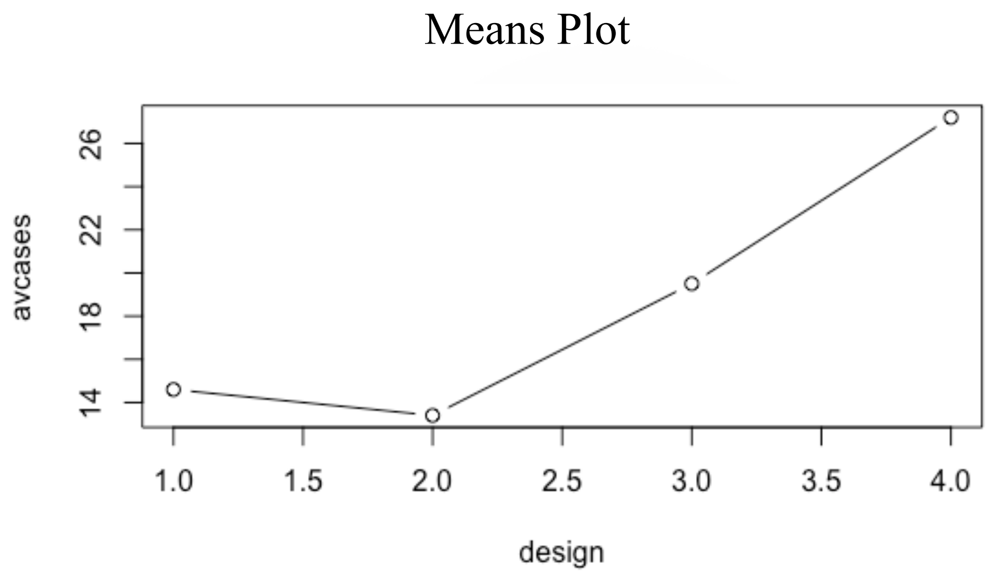
- The parameters of the model are
  - $\mu_1, \mu_2, \dots, \mu_r$
  - σ^2
- Estimate μ_i by the mean of the observations at level i,
  - $\hat{\mu}_i = \bar{Y}_{i.} = \sum_j Y_{ij}/n_i$ -level i sample mean
  - $s_i^2 = \sum_j \left(Y_{ij} - \bar{Y}_{i.}\right)^2/(n_i - 1)$ -level i sample variance
- $s^2 = \sum\left((n_i - 1)s_i^2\right)/(n_T - r)$ -pooled variance

F Test

- $F^* = \text{MSR}/\text{MSE}$
- H_0: $\mu_1 = \mu_2 = \ldots = \mu_r = \mu$ (a constant)
- H_a: not all μ_i's are the same
- Under H_0, $F^* \sim F(r-1, n_T - r)$
- Reject H_0 when F^* is large
- Common to report the P-value along with decision

Cereal Package Example

- KNNL p 676, p 685
- Y is the number of cases of cereal sold
- X is the design of the cereal package
  - there are four levels for X because there are four different designs
- i = 1 to 4 levels
- j = 1 to n_i stores with design i (n_i = 5, 5, 4, 5)



	design	N	Mean	StdDev	Minimun	Maximum
1	1	5	14.6	2.302173	11	17
2	2	5	13.4	3.646917	10	19
3	3	4	19.5	2.645751	17	23
4	4	5	27.2	3.962323	22	33

Summarize Data

- Import the data

```
> a1 = read.table("CH16TA01.txt")
> colnames(a1) = c("cases", "design", "store")
> a1$design = as.factor(a1$design)
```

- Describe the data

```
> library('plyr’)
> summaryStat <- ddply(a1, ~design, summarise,
            N = length(cases), Mean = mean(cases),
    StdDev = sd(cases),  Minimun = min(cases),  Maximum = max(cases))
> summaryStat
```

Confidence Intervals

- $\bar{Y}_{i.} \sim N(\mu_i, \sigma^2/n_i)$
- CI for μ_i is $\bar{Y}_{i.} \pm t_c s/\sqrt{n_i}$
- t_c is computed from the $t(\alpha/2,\ n_T - r)$
- Degrees of freedom larger than $n_i - 1$ because we're pooling variances together into one single estimate s
  - This is advantage of ANOVA if model assumptions appropriate

Compute Confidence Limits

- We can use this information to calculate *SSE*

	design	N	Mean	StdDev	Minimun	Maximum
1	1	5	14.6	2.302173	11	17
2	2	5	13.4	3.646917	10	19
3	3	4	19.5	2.645751	17	23
4	4	5	27.2	3.962323	22	33

- $SSE = 4(2.3)^2 + 4(3.65)^2 + 3(2.65)^2 + 4(3.96)^2 = 158.2$

```
> summaryStat = within(summaryStat,  {StdError = StdDev/sqrt(N) } )
> summaryStat = within(summaryStat,
          {Lower95CL = Mean - StdError * qt(0.975,df=N-1)})
> summaryStat = within(summaryStat,
          {Upper95CL = Mean + StdError * qt(0.975,df=N-1)})
> summaryStat
```

	design	Lower95CL	Upper95CL
1	1	11.741475	17.45853
2	2	8.871755	17.92824
3	3	15.290019	23.70998
4	4	22.280127	32.11987

There is no pooling of error in computing these CI's.
Each interval assumes different variance estimate.

CIs in R


```
> fit = lm (cases ~ 0 + design, data = a1)
> confint(fit)
```

Here is the result

	2.5 %	97.5 %
design1	11.50438	17.69562
design2	10.30438	16.49562
design3	16.03899	22.96101
design4	24.10438	30.29562

Compare with before

Lower95CL	Upper95CL
11.741475	17.45853
8.871755	17.92824
15.290019	23.70998
22.280127	32.11987

- These CI's are often narrower because more degrees of freedom (common variance)

design1	design2	design3	design4
6.191240	6.191240	6.922017	6.191240
5.717050	9.056490	8.419961	9.839747

Multiplicity Problem

- We have constructed 4 (in general, r) 95% confidence intervals
- The overall confidence level (all intervals contain its mean) is less that 95%
- Many different kinds of adjustments have been proposed
- We have previously discussed the Bonferroni adjustment (i.e., use α/r)

Familywise Error Rate (FWE总体错误率)

- Multiple Comparisons Fallacy（多重比较谬误）
- Accounting for the multiplicity of individual tests can be achieved by controlling an appropriate error rate
- The traditional/classical FWE is the probability of one or more false discoveries (falsely reject H_0)
- In <u>single-step multiple comparison procedure</u>, individual test statistics are compared to their critical values simultaneously, and after this simultaneous `joint' comparison, the multiple testing method stops
- Often single-step methods can be improved in terms of power via stepwise methods, while still maintaining control of the desired error rate

Bonferroni CIs

- Simultaneous 95% confidence limits:
- > confint(mod1, level = 1 - 0.05/4)

Here is the result

	0.625 %	99.375 %
design1	10.480212	18.71979
design2	9.280212	17.51979
design3	14.893937	24.10606
design4	23.080212	31.31979

Compare with before

2.5 %	97.5 %
11.50438	17.69562
10.30438	16.49562
16.03899	22.96101
24.10438	30.29562

The CI's become wider

Hypothesis Tests on Individual Means

- Not usually done
- A test of the null hypothesis H_0: $\mu_i = 0$
- To test H_0: $\mu_i = c$, where c is an arbitrary constant, first subtract c from all observations in each level
- Can also check confidence intervals for means

Testing Differences in means

- $\bar{Y}_{i.} - \bar{Y}_{k.} \sim N(\mu_i - \mu_k, \sigma^2/n_i + \sigma^2/n_k)$
- CI for $\mu_i - \mu_k$ is

$$\bar{Y}_{i.} - \bar{Y}_{k.} \pm t_c s(\bar{Y}_{i.} - \bar{Y}_{k.})$$

where

$$s(\bar{Y}_{i.} - \bar{Y}_{k.}) = s\sqrt{\frac{1}{n_i} + \frac{1}{n_k}}$$

Determining t_c

- There are $\binom{r}{2} = \frac{r(r-1)}{2}$ pairwise comparisons
- Handle multiplicity problem by adjusting t_c
- Many different choices are available
- Change α level (e.g., Bonferonni)
- Use different distribution

Least Significant Difference (LSD)

- Fisher's LSD test
- Simply ignores multiplicity issue
- Uses $t(n_T - r)$ to determine critical value for each comparison

Bonferroni

- Use the error budget idea
- There are $r(r-1)/2$ comparisons among r means
- So, replace for each comparison by $\alpha/(r(r-1)/2)$ and use $t(n_T - r)$ to determine critical value

Tukey

- Tukey's HSD(honestly significant difference) test
- Uses the studentized range distribution (SRD) instead of *t*. Focuses on distribution of maximum minus minimum divided by the standard deviation
- $t_c = q_c/\sqrt{2}$ where q_c is detemined from SRD
- Details are in KNNL Section 17.5 (p 746)

Scheffé

- Based on the *F* distribution
- $t_c = \sqrt{(r-1)F(1-\alpha; r-1, n_T - r)}$
- Takes care of multiplicity for all linear combinations of means, not just pairwise comparisons
- Can be used for a wide variety of data mining
- See KNNL Section 17.6 (page 753)

Multiple Comparisons

- LSD is too liberal (i.e., too many Type I errors)
- Scheffé is too conservative (very low power)
- Bonferroni is ok for small r
- Tukey is recommended for pairwise comparisons
- Other procedures exist and worthy of consideration
  - Focus on pairwise comparisons or linear combinations?
  - Focus on family-wise error rate?

Revisit Our Example

- Least significant difference (LSD) *t* test for cases
- NOTE: This test controls the Type I comparisonwise error rate, not the experimentwise error rate
- LSD Intervals ->

Alpha	0.05
Error Degrees of Freedom	15
Error Mean Square	10.54667
Critical Value of t	2.13145

design Comparison			Difference Between Means	Simultaneous 95% Confidence Limits		
4	-	3	7.700	3.057	12.343	***
4	-	1	12.600	8.222	16.978	***
4	-	2	13.800	9.422	18.178	***
3	-	4	-7.700	-12.343	-3.057	***
3	-	1	4.900	0.257	9.543	***
3	-	2	6.100	1.457	10.743	***
1	-	4	-12.600	-16.978	-8.222	***
1	-	3	-4.900	-9.543	-0.257	***
1	-	2	1.200	-3.178	5.578	
2	-	4	-13.800	-18.178	-9.422	***
2	-	3	-6.100	-10.743	-1.457	***
2	-	1	-1.200	-5.578	3.178	

Tukey

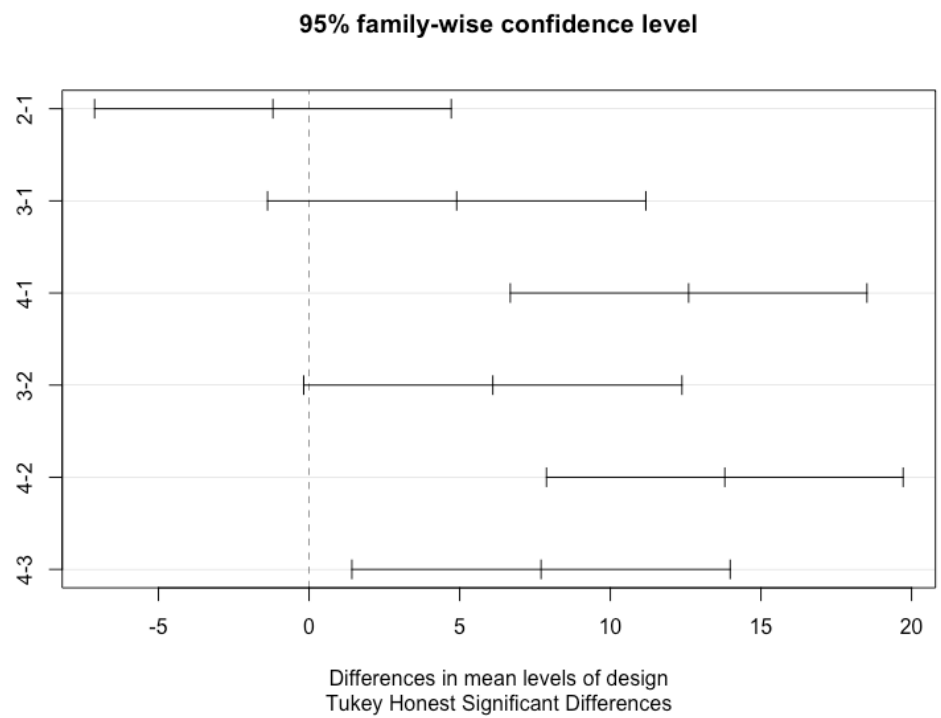
```
> mod1 <- aov(cases ~ 0 + design, data = a1)
> mod1.Tukey <- TukeyHSD (mod1, conf.level = 0.95)
> plot(mod1.Tukey, sub="Tukey Honest Significant Differences")
```

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = cases ~ 0 + design, data = a1)

$design

	diff	lwr	upr	p adj
2-1	-1.2	-7.119758	4.719758	0.9352978
3-1	4.9	-1.378852	11.178852	0.1548895
4-1	12.6	6.680242	18.519758	0.0001013
3-2	6.1	-0.178852	12.378852	0.0582866
4-2	13.8	7.880242	19.719758	0.0000368
4-3	7.7	1.421148	13.978852	0.0142180



Scheffé

- Scheffe’s Test for cases
- NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons

- Scheffé Intervals

Alpha	0.05
Error Degrees of Freedom	15
Error Mean Square	10.54667
Critical Value of F	3.28738

F(.95, 3, 15) = 3.28

$$t_c = \sqrt{(r-1)F(1-\alpha; r-1, n_T - r)}$$

=Sqrt(3*3.28) = 3.14

design Comparison			Difference Between Means	Simultaneous 95% Confidence Limits		
4	-	3	7.700	0.859	14.541	***
4	-	1	12.600	6.150	19.050	***
4	-	2	13.800	7.350	20.250	***
3	-	4	-7.700	-14.541	-0.859	***
3	-	1	4.900	-1.941	11.741	
3	-	2	6.100	-0.741	12.941	
1	-	4	-12.600	-19.050	-6.150	***
1	-	3	-4.900	-11.741	1.941	
1	-	2	1.200	-5.250	7.650	
2	-	4	-13.800	-20.250	-7.350	***
2	-	3	-6.100	-12.941	0.741	
2	-	1	-1.200	-7.650	5.250	

Pairwise Comparison in R


```
> pairwise.t.test(a1$cases, a1$design, p.adjust="none", pool.sd = T)
> pairwise.t.test(a1$cases, a1$design, p.adjust="bonferroni", pool.sd = T)
```

```
        Pairwise comparisons
    using t tests with pooled SD

data:  a1$cases and a1$design

   1       2       3
2  0.568   -       -
3  0.040   0.013   -
4  1.9e-05 6.9e-06 0.003

P value adjustment method: none
```

```
        Pairwise comparisons
    using t tests with pooled SD

data:  a1$cases and a1$design

   1       2       3
2  1.00000 -       -
3  0.23969 0.08075 -
4  0.00011 4.1e-05 0.01802

P value adjustment method: bonferroni
```

Linear Combinations of Means

- These combinations should come from research questions, not from an examination of the data
- $L = \sum c_i \mu_i$
- $\hat{L} = \sum c_i \bar{Y}_{i.} \sim N\left(L, Var(\hat{L})\right)$
  - $Var(\hat{L}) = \sum c_i^2 Var(\bar{Y}_{i.})$
  - $Var(\hat{L})$ is estimated by $s^2 \sum (c_i^2/n_i)$
- Can use our linear statistical model to construct a *t*-test of any linear combination
- $Var(\hat{L}) = MSE \sum (c_i^2/n_i)$
- Under H_0: $L = L_0$, $T \sim t(n_T - r)$

$$T = \frac{\hat{L} - L_0}{\sqrt{Var(\hat{L})}}$$

Contrasts

- Special case of a linear combination
- Requires $\sum c_i = 0$
- Example 1: $\mu_1 - \mu_2$
- Example 2: $\mu_1 - (\mu_2 + \mu_3)/2$
- Example 3: $(\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4)/2$

Contrast sum of squares

- H_0: $L = 0$
- $T = \dfrac{\sum c_i \bar{Y}_{i.}}{\sqrt{MSE \sum \left(c_i^2 / n_i\right)}} \sim t(n_T - r)$
- $T^2 = \dfrac{\left(\sum c_i \bar{Y}_{i.}\right)^2}{MSE \sum \left(c_i^2 / n_i\right)} = \dfrac{SSC/1}{MSE} \sim F(1, n_T - r)$

  where $SSC = \left(\sum c_i \bar{Y}_{i.}\right)^2 \Big/ \sum (c_i^2 / n_i)$
- $t^2(n_T - r) = F(1, n_T - r)$
- Contrast sum of squares SSC represent amount of variation due to this contrast

Multiple Contrasts

- We can simultaneously test a collection of contrasts (1 df for each contrast)
- Example 1, H_0: $\mu_2 = \mu_3 = \mu_4$
  - The F statistic for this test will have an $F(2, n_T - r)$ distribution
- Example 2, H_0: $\mu_1 = (\mu_2 + \mu_3 + \mu_4)/3$
  - The F statistic for this test will have an $F(1, n_T - r)$ distribution

Contrast and Estimate


```
> c1 <- c(.5, .5, -.5, -.5)
> c2 <- c(1, -.3333, -.3333, -.3333)
> c3 <- c(3, -1, -1, -1)
> c4 <- c(0, 1, -1,  0)
> cntrMat <- rbind("1&2 v 3&4" = c1,
                   "1 v 2&3&4" = c2,
                   "1 v 2&3&4" = c3,
                   "2 v 3 " = c4)
> library(multcomp)
> summary(glht(mod1, linfct=cntrMat))
```

```
Simultaneous Tests for General Linear Hypotheses

Fit: aov(formula = cases ~ 0 + design, data = a1)

Linear Hypotheses:
                  Estimate Std. Error t value Pr(>|t|)
1&2 v 3&4 == 0      -9.350      1.497  -6.246  <0.001 ***
1 v 2&3&4 == 0      -5.431      1.694  -3.206  0.0148 *
1 v 2&3&4 == 0     -16.300      5.083  -3.207  0.0149 *
    2 v 3 == 0      -6.100      2.179  -2.800  0.0329 *
```

F Test

- For c1, "1&2 v 3&4" , c = (.5, .5, -.5, -.5)
  - (-6.246)^2 = 39.01
  - $SSC = \left(\sum c_i \bar{Y}_{i.}\right)^2 / \sum (c_i^2/n_i)$
  - SSC <- (sum(singleMean*c))^2/sum(c^2/freq)
  - F <- SSC/MSE

Contrast	DF	SS	MS	F	P
1&2 v 3&4	1	411	411	39.01	<.0001

- For multiple comparison "2 v 3 v 4" , (0, 1, -1, 0) (0, 0, 1, -1)
  - library(car)
  - linearHypothesis(fit, rbind(c (0, 1, -1, 0), c(0, 0, 1, -1)))

```
Linear hypothesis test

Hypothesis:
design2 - design3 = 0
design3 - design4 = 0

Model 1: restricted model
Model 2: cases ~ 0 + design

Res.Df  RSS    Df Sum of Sq   F     Pr(>F)
1     17 636.13
2     15 158.20  2    477.93 22.658 2.934e-05 ***
```

Contrast	DF	SS	MS	F	P
1 v 2&3&4	1	10.29		10.29	.0059
2 v 3 v 4	2	22.66		22.66	<.0001

Last Slide

- We did large part of Chapter 17
- We used programs topic12_1.R to generate the output for today

Topic 2:
Two-Way ANOVA

Outline

- Cell means model
  - Parameter estimates
- Factor effects model
  - Parameter estimates
- ANOVA

Two-Way ANOVA Data

- The response variable Y is continuous
- There are now two categorical explanatory variables or factors
  - Treatments/groups can be classified in two ways
  - Form a two-way table

- Y is the response variable
- Factor A with levels $i = 1$ to a
- Factor B with levels $j = 1$ to b
- Y_{ijk} is the k^{th} observation in cell (i, j)
- In Chapter 19, we assume equal sample size in each cell ($n_{ij} = n$)

Bakery Example

- KNNL p 833
- Y is the number of cases of bread sold
- A is the height of the shelf display, $a = 3$
  - levels: bottom, middle, top
- B is the width of the shelf display, $b = 2$
  - levels: regular, wide
- $n = 2$ stores for each of the $3 \times 2 = 6$ treatment combinations ($n_T = 12$)

Notation

- For Y_{ijk} we use
  - i to denote the level of the factor A
  - j to denote the level of the factor B
  - k to denote the k^{th} observation in cell (i, j)
- $i = 1, \ldots, a$ levels of factor A
- $j = 1, \ldots, b$ levels of factor B
- $k = 1, \ldots, n$ observations in cell (i, j)

width\hight	1	2	3
1	47,43	62,68	41,39
2	46,40	67,71	42,46

	sales	height	width	store
1	47	1	1	1
2	43	1	1	2
3	46	1	2	1
4	40	1	2	2
5	62	2	1	1
6	68	2	1	2
7	67	2	2	1
8	71	2	2	2
9	41	3	1	1
10	39	3	1	2
11	42	3	2	1
12	46	3	2	2

CH19TA07

Model Assumptions

We assume that the response variable observations are

- Normally distributed
- With a mean that <u>may depend only on the</u> <u>levels of the factors *A* and *B*</u>
  - With constant variance
  - Independent

Cell Means Model

- $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$
  - where μ_{ij} is the theoretical mean or expected value of all observations in cell (i, j)
  - the ε_{ijk} are iid $N(0, \sigma^2)$
- This means $Y_{ijk} \sim N(\mu_{ij}, \sigma^2)$, independent
- The parameters of the model are
  - μ_{ij}, for $i = 1$ to a and $j = 1$ to b
  - σ^2

Estimates

- Estimate μ_{ij} by the mean of the observations in cell (i, j),

$$\bar{Y}_{ij.} = \sum_k Y_{ijk}/n$$

- For each (i, j) combination, we can get an estimate of the variance

$$s_{ij}^2 = \sum_k \left(Y_{ijk} - \bar{Y}_{ij.}\right)^2/(n-1)$$

- We need to combine these to get an pooled estimate of σ^2

Pooled estimate of σ^2

- In general we pool the s_{ij}^2, using weights proportional to the df, $n_{ij} - 1$
- The pooled estimate is

$$s^2 = \left(\sum_{ij}(n_{ij} - 1)s_{ij}^2\right)/\left(\sum_{ij}(n_{ij} - 1)\right)$$

- Here, $n_{ij} = n$, so

$$s^2 = (\Sigma s_{ij}^2)\,/\,(ab),$$

which is the average sample variance

Estimate in R

width\hight	1	2	3
1	45	65	40
2	43	69	44

- > summary(lm(sales ~ height * width, a))
- ▶ Cell estimates used for calculating parameter estimates
- > aggregate(sales~ height*width, a, length)
- > aggregate(sales~ height*width, a, mean)
- > aggregate(sales~ height*width, a, sd)

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     45.000      2.273  19.797 1.08e-06 ***
height2         20.000      3.215   6.222 0.000797 ***
height3         -5.000      3.215  -1.555 0.170844
width2          -2.000      3.215  -0.622 0.556718
height2:width2   6.000      4.546   1.320 0.235013
height3:width2   6.000      4.546   1.320 0.235013
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.215 on 6 degrees of freedom
Multiple R-squared:  0.9622,      Adjusted R-squared:  0.9308
F-statistic: 30.58 on 5 and 6 DF,  p-value: 0.0003384
```

Level of height	Level of width	N	sales Mean	sales Std Dev
1	1	2	45.0000000	2.82842712
1	2	2	43.0000000	4.24264069
2	1	2	65.0000000	4.24264069
2	2	2	69.0000000	2.82842712
3	1	2	40.0000000	1.41421356
3	2	2	44.0000000	2.82842712

- ▶ Commonly do not consider R-sq when performing ANOVA…interested more in difference in levels rather than the model's predictive ability

Marginal Means

- Height marginal

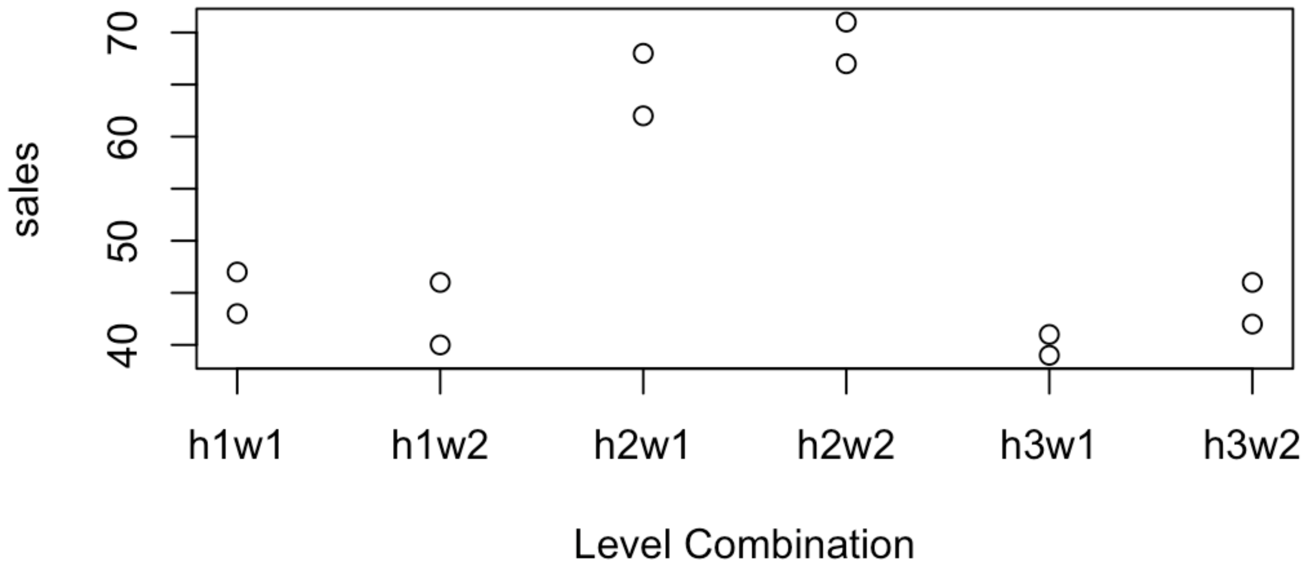
Level of height	N	sales	
		Mean	Std Dev
1	4	44.0000000	3.16227766
2	4	67.0000000	3.74165739
3	4	42.0000000	2.94392029

- Width marginal

Level of width	N	sales	
		Mean	Std Dev
1	6	50.0000000	12.0664825
2	6	52.0000000	13.4313067

Plot the Data


```
> a$cell <- with(a, paste('h', height, 'w', width, sep = ”))
> a$cell <- factor(a$cell)
> with(a, plot(as.numeric(cell), sales, xaxt = "n",
                        xlab = 'Level Combination’))
> axis(1,at=1:6, labels = levels(a$cell))
```



Questions to Consider

- Does the height of the display affect sales?
  - If yes, compare top with middle, top with bottom, and middle with bottom
- Does the width of the display affect sales?
  - If yes, compare regular and wide

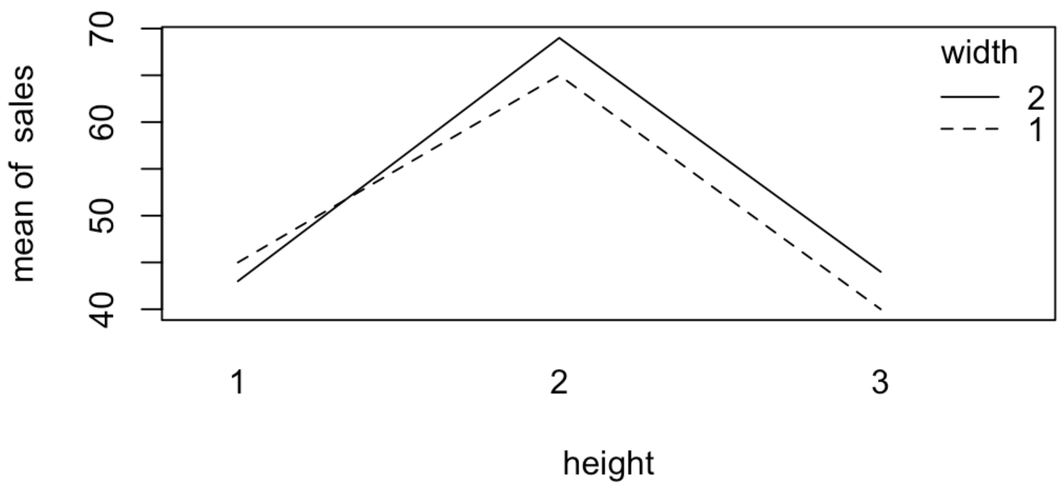
But Wait!!!

- Are these factor level comparisons meaningful?
- Does the effect of height on sales depend on the width?
- Does the effect of width on sales depend on the height?
- If yes, we have an interaction and we need to do some additional analysis

The Interaction Plot


```
with(a,
  interaction.plot(x.factor = height,
                   trace.factor = width,
                   response = sales,
                   fun = mean))
```

Plot of the Means



Factor Effects Model

- For the one-way ANOVA model, we wrote

$$\mu_i = \mu + \alpha_i$$

- Here we use

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

- Under "common" formulation
  - μ ($\mu_{..}$ in KNNL) is the "overall mean"
  - α_i is the main effect of A
  - β_j is the main effect of B
  - $(\alpha\beta)_{ij}$ is the interaction between A and B

Constraints

- $\alpha_. = \Sigma_i \alpha_i = 0$
- $\beta_. = \Sigma_j \beta_j = 0$
- $(\alpha\beta)_{.j} = \Sigma_i (\alpha\beta)_{ij} = 0$ for all j
- $(\alpha\beta)_{i.} = \Sigma_j (\alpha\beta)_{ij} = 0$ for all i

Factor Effects Model Interpretation

- $\mu = (\Sigma_{ij} \mu_{ij})/(ab)$
- $\mu_{i.} = (\Sigma_j \mu_{ij})/b$ and $\mu_{.j} = (\Sigma_i \mu_{ij})/a$
- $\alpha_i = \mu_{i.} - \mu$ and $\beta_j = \mu_{.j} - \mu$
- $(\alpha\beta)_{ij}$ is…
  - difference between μ_{ij} and $\mu + \alpha_i + \beta_j$
  - What's unexplained by completely additive model
  - $(\alpha\beta)_{ij} = \mu_{ij} - (\mu + (\mu_{i.} - \mu) + (\mu_{.j} - \mu))$
    $= \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu$

- $\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$
  - μ is the "overall" mean
  - α_i is an adjustment for level i of A
  - β_j is an adjustment for level j of B
  - $(\alpha\beta)_{ij}$ is an additional adjustment that takes into account both i and j that cannot be explained by the previous adjustments

Estimates for Factor Effects Model

- $\hat{\mu} = \bar{Y}_{...} = \sum_{ijk} Y_{ijk}/(abn)$

  $\hat{\mu}_{i.} = \bar{Y}_{i..}$ and $\hat{\mu}_{.j} = \bar{Y}_{.j.}$

- $\hat{\alpha}_i = \bar{Y}_{i..} - \bar{Y}_{...}$ and $\hat{\beta}_j = \bar{Y}_{.j.} - \bar{Y}_{...}$

- $(\widehat{\alpha\beta})_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$

For ANOVA Table: SS DF MS

- $\text{SSA} = \sum_{ijk} \hat{\alpha}_i^{\,2} = \sum_{ijk} (\bar{Y}_{i..} - \bar{Y}_{...})^2$
- $\text{SSB} = \sum_{ijk} \hat{\beta}_j^{\,2}$
- $\text{SSAB} = \sum_{ijk} \widehat{(\alpha\beta)}_{ij}^2$
- $\text{SSE} = \sum_{ijk} \left(\bar{Y}_{ijk} - \bar{Y}_{ij.}\right)^2$
- $\text{SSTO} = \sum_{ijk} \left(\bar{Y}_{ijk} - \bar{Y}_{...}\right)^2$

- $\text{df}_\text{A} = a - 1$
- $\text{df}_\text{B} = b - 1$
- $\text{df}_\text{AB} = (a-1)(b-1)$
- $\text{df}_\text{E} = ab(n-1)$
- $\text{df}_\text{T} = abn - 1 = n_T - 1$

- $\text{MSA} = \text{SSA}/\text{df}_\text{A}$
- $\text{MSB} = \text{SSB}/\text{df}_\text{B}$
- $\text{MSAB} = \text{SSAB}/\text{df}_\text{AB}$
- $\text{MSE} = \text{SSE}/\text{df}_\text{E}$
- $\text{MST} = \text{SST}/\text{df}_\text{T}$

Two-Way ANOVA: Hypotheses   *F* Statistics

- H_{0A}: $\alpha_i = 0$ for all i
- H_{1A}: $\alpha_i \neq 0$ for at least one i
- H_{0B}: $\beta_j = 0$ for all j
- H_{1B}: $\beta_j \neq 0$ for at least one j
- H_{0AB}: $(\alpha\beta)_{ij} = 0$ for all (i, j)
- H_{1AB}: $(\alpha\beta)_{ij} \neq 0$ for at least one (i, j)

- H_{0A} is tested by $F_A = MSA/MSE$;

  $df = df_A, df_E$

- H_{0B} is tested by $F_B = MSB/MSE$;

  $df = df_B, df_E$

- H_{0AB} is tested by $F_{AB} = MSAB/MSE$;

  $df = df_{AB}, df_E$

ANOVA Table

Source	df	SS	MS	F
A	a-1	SSA	MSA	MSA/MSE
B	b-1	SSB	MSB	MSB/MSE
AB	(a-1)(b-1)	SSAB	MSAB	MSAB/MSE
Error	ab(n-1)	SSE	MSE	
Total	abn-1	SSTO	MST	

- P-values
  - P-values are calculated using the F(dfNumerator, dfDenominator) distributions
  - If $P \leq 0.05$ we conclude that the effect being tested is statistically significant

Bakery Example

- NKNW p 833
- Y is the number of cases of bread sold
- A is the height of the shelf display, $a = 3$ levels: bottom, middle, top
- B is the width of the shelf display, $b = 2$: regular, wide
- $n = 2$ stores for each of the 3×2 treatment combinations

ANOVA in R

- anova(fit)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	1580.0000	316.00000	30.58	0.0003
Error	6	62.000000	10.333333		
Corrected Total	11	1642.0000			

- Note that there are 6 cells in this design…(6-1)df for model
- The interaction between height and width is not statistically significant (F=1.16; df=2,6; P=0.37)
- Given this result, move to main effects
- The main effect of height is statistically significant (F=74.71; df=2,6; P<0.0001)
- The main effect of width is not statistically significant (F=1.16; df=1,6; P=0.32)

```
Analysis of Variance Table

Response: sales
             Df Sum Sq Mean Sq F value    Pr(>F)
height        2   1544  772.00 74.7097 5.754e-05 ***
width         1     12   12.00  1.1613    0.3226
height:width  2     24   12.00  1.1613    0.3747
Residuals     6     62   10.33
```

<u>Interpretation</u>

- The height of the display affects sales of bread
- The width of the display has no apparent effect
- The effect of the height of the display is similar for both the regular and the wide widths

Additional Analyses

- We will need to do additional analyses to explain the height effect (factor A)
- There were three levels: bottom, middle and top
- We could rerun the data with a one-way ANOVA and use the methods we learned in the previous chapters

Final Slide

- We went over Chapter 19
- We used program lec12_2.r to generate the output for today