psyc3010 lecture 2

logic and computations of factorial anova

last week: introduction to factorial designs
next week: simple effects and effect size

Blackboard

http://www.elearning.uq.edu.au/

see BB for:

- Lecture notes (before lecture, ppt & pdf)
- Tute notes (after tutes)
- Forums
- Additional material incl. course profile & practice exams (later)
- Psyc2010 practice exam (not examined _specifically_ but may include material that is also covered in psy3010, which IS examinable)

announcements

- tutorial allocations now completed – check web or 3rd year noticeboard
  - With problems, e-mail e.puhakka@psy.uq.edu.au
- full course outline available on web
- Tutes start this week !
  - First tute is immediately after class!

Revisiting assessment deadlines

- **two written assignments**
- due dates are :
  - assignment 1 → 4pm Monday September 8th
  - assignment 2 → 4pm Monday October 20th

ICEBREAKER!

last week → this week

- last week we introduced the concept of factorial designs
- we reviewed & learned important terminology and concepts:
  - Factors / independent variables, dependent variables
  - Crossed designs – A x B
  - Cell means, marginal means, grand means
  - Main effects, interaction effects, simple effects
  - How one factor qualifies or moderates the effect of another
  - Ordinal and disordinal interactions
- this week we cover factorial designs in more detail, and go over the conceptual and computational processes involved in **between-subjects factorial anova**

topics for this week

- **conceptual underpinnings of ANOVA**
  - ☞ relationship between hypotheses, variance, graphical representations of data, and formulae (one-way and two-way analyses)
- **links between *t*, one-way ANOVA, and factorial ANOVA**
- **understanding linear effects**
  - ☞ calculating residuals (error) for individual scores in factorial ANOVA
- **calculations underlying ANOVA**
- **following up interactions with plots**

anova: conceptual underpinnings

- like most statistical procedures we use, anova is all about partitioning variance
- we want to see if variation due to our experimental manipulations or groups of interest is proportionally greater than the rest of the variance (i.e., that is *not* due to any manipulations etc)
- do participants' scores (on some DV) differ from one another because they are in different groups of our study, more so than they differ randomly and due to unmeasured influences?

notation review

H_0: $\mu_1 = \mu_2$ or (more mathematically convenient) $\mu_j - \mu . = 0$

H_1: $\mu_1 \neq \mu_2$ or (more convenient) $\mu j - \mu . \neq 0$ for at least one j

for a one-way ANOVA, with j conditions:

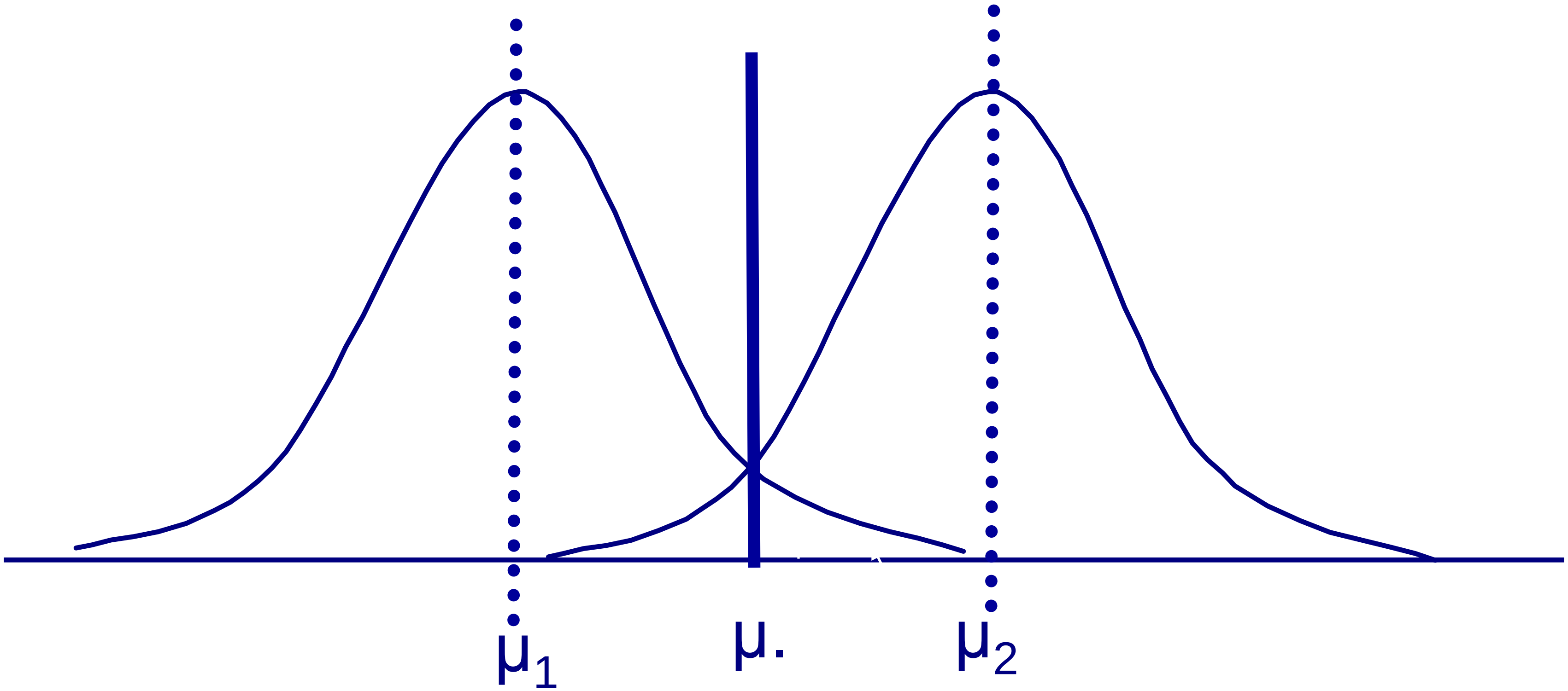
mu j = population means of group j

mu dot = population grand mean

null hypothesis = there is no between-group variance (no variability between the group means and the grand mean)

alternative hypothesis = at least one group mean is significantly different from the grand mean.

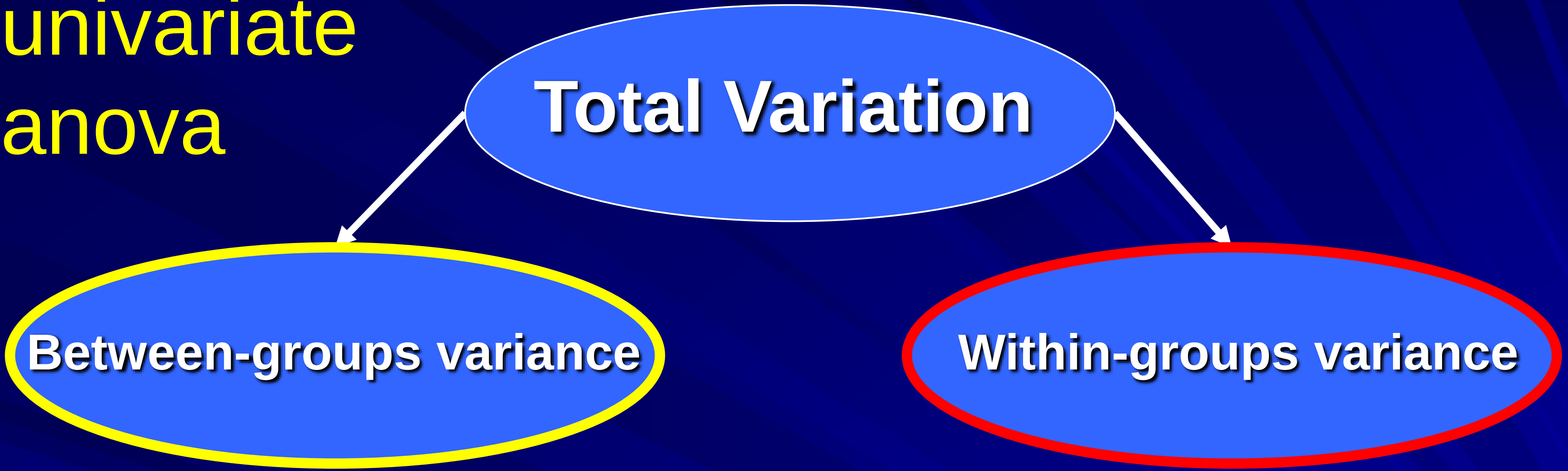
Sources of variance…



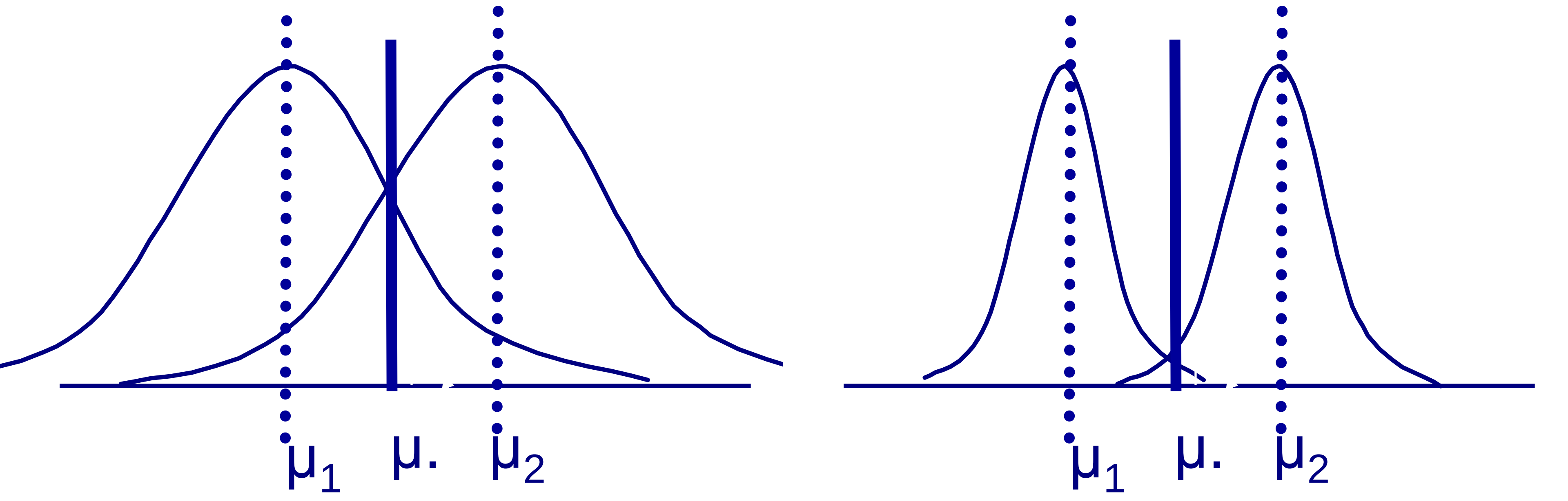
Null hypothesis : $\mu_j = \mu\,.$ [or sum(mu j – mu dot) = 0]

Alt hyp: $\mu j \neq \mu\,.$ for at least one j [or sum(mu j – mu dot) ≠ 0]

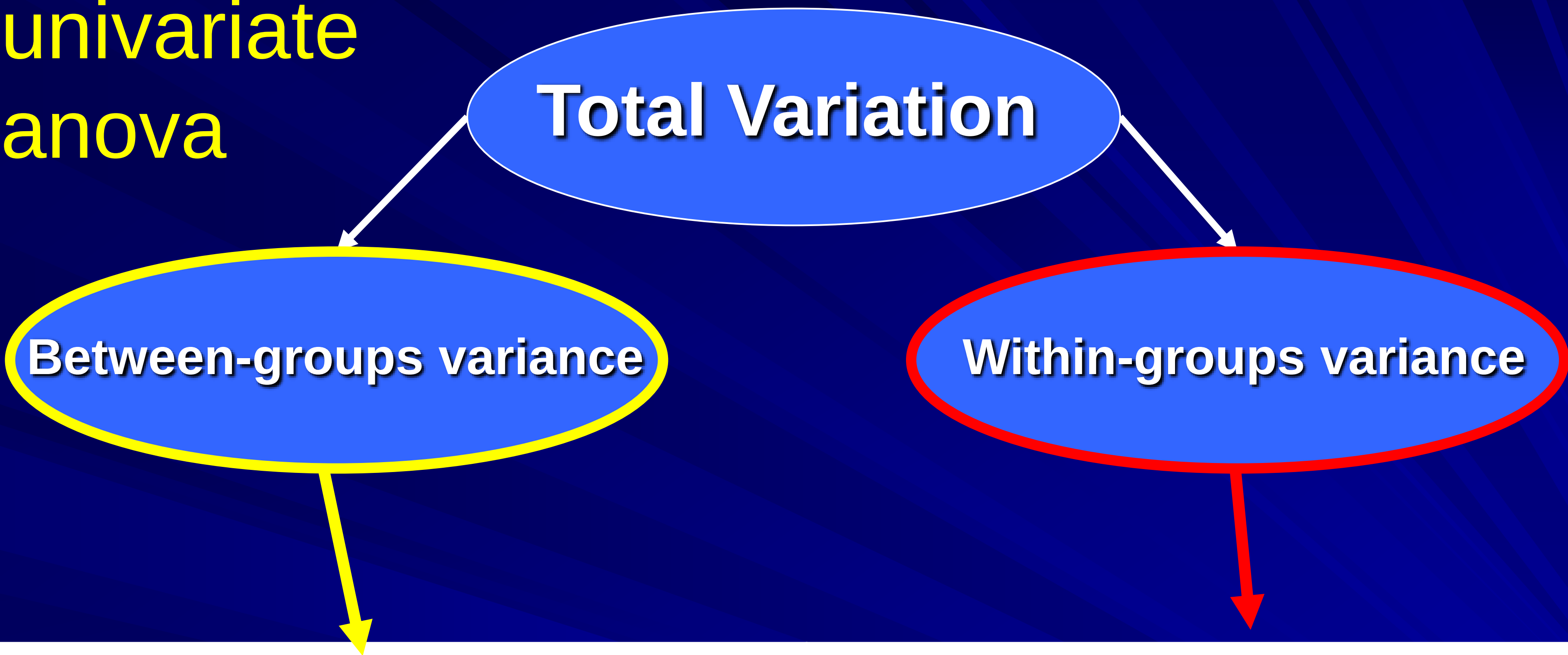
univariate anova



But need to consider not just absolute variability between groups but relative variability compared to ‘error’ variance = within-group variance



univariate anova



$$n\sum(\bar{X}_j - \bar{X}_{.})^2$$

$n\sum(X$ bar j – X bar dot$)^2$ = people per group x sum of squared differences between **group means and grand mean** = estimate of between groups variability

$$\sum(X_{ij} - \bar{X}_j)^2$$

$\sum(X$ ij – X bar j$)^2$ = **sum of squared differences** between **individual scores and group mean** = estimate of within groups variability

So what is ANOVA ?

- 1. Estimate of between-groups variability
- 2. Estimate of within-groups variability
- 3. Weight each variability estimate by # of observations used to generate the estimate ("degrees of freedom")
- Compare ratio

$$\frac{[[n\sum(\bar{X}_j - \bar{X}_{\cdot})^2] / (j-1)]}{[[\sum(X_{ij} - \bar{X}_j)^2] / [j(n-1)]]}$$

When the F ratio is > 1, the treatment effect (variability between groups) is bigger than the "error" variability (variability within groups). Or more specifically:

The sum of the squared differences between the group means and the grand mean x the number of people in each group, divided by the number of groups minus 1, is bigger than

the sum of the squared differences between the observations and the group means, weighted by the number of observations in each group minus 1 x the # of groups

hypothesis testing

***differences between 2 means – t-test* (or one-way anova)**

- ***H_0: $\mu_1 = \mu_2$***
  - the null hypothesis – no differences between treatment means
- ***H_1*: the null hypothesis is false**
  - the alternative hypothesis – there is a difference between treatment means

differences among 3+ means – one-way anova

- ***H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_j$***
  - the null hypothesis – no differences among treatment means
- ***H_1*: the null hypothesis is false**
  - the alternative hypothesis – there is at least one difference among treatment means

logic of the t-test

- **independent samples t-test:** ***'Is the difference between two sample means greater than would be expected by chance?'***

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

A ratio of the systematic variance (i.e. your experimental manipulation) to the unsystematic variance

t = $\frac{\text{observed difference between two independent means}}{\text{estimate of the standard error of the mean differences}}$

if the observed difference is similar to the difference you would typically expect between means, t = 1

if the observed difference is ***greater than*** *the difference you would typically expect between means, t > 1*

larger values of t indicate that H_0 is probably wrong

logic of univariate (one-way) anova

- **the test statistic is the F-ratio**

$$F = MS_{treat}/MS_{error}$$

A ratio of the systematic variance (i.e. your experimental manipulation or treatment) to the unsystematic variance

where $\boldsymbol{MS_{treat}}$ = index of variability among treatment means (SS_{TR}/df_{TR}) or (SS_j/df_j)

and $\boldsymbol{MS_{error}}$ = index of variability among participants within a cell, i.e. pooled within-cell variance (SS_{Error}/df_{Error})
= average of s^2 from each sample, a good estimate of σ_e^2 (population variance)

- ***if MS_{treat} is a good estimate of σ_e^2, $F = MS_{treat}/MS_{error} = 1$***
- ***if $MS_{treat} > \sigma_e^2$, $F = MS_{treat}/MS_{error} > 1$***
- **larger values of F indicate that H0 is probably wrong**

the structural model of univariate anova

$$X_{ij} = \mu + \tau_j + e_{ij}$$

for *i* cases and *j* treatments:

X_{ij}, any DV score is a combination of:

μ → ***the grand mean,***

τ_j → ***the effect of the j-th treatment*** *(μ_j - μ)*

e_{ij} → error, *averaged over all i cases and j treatments*

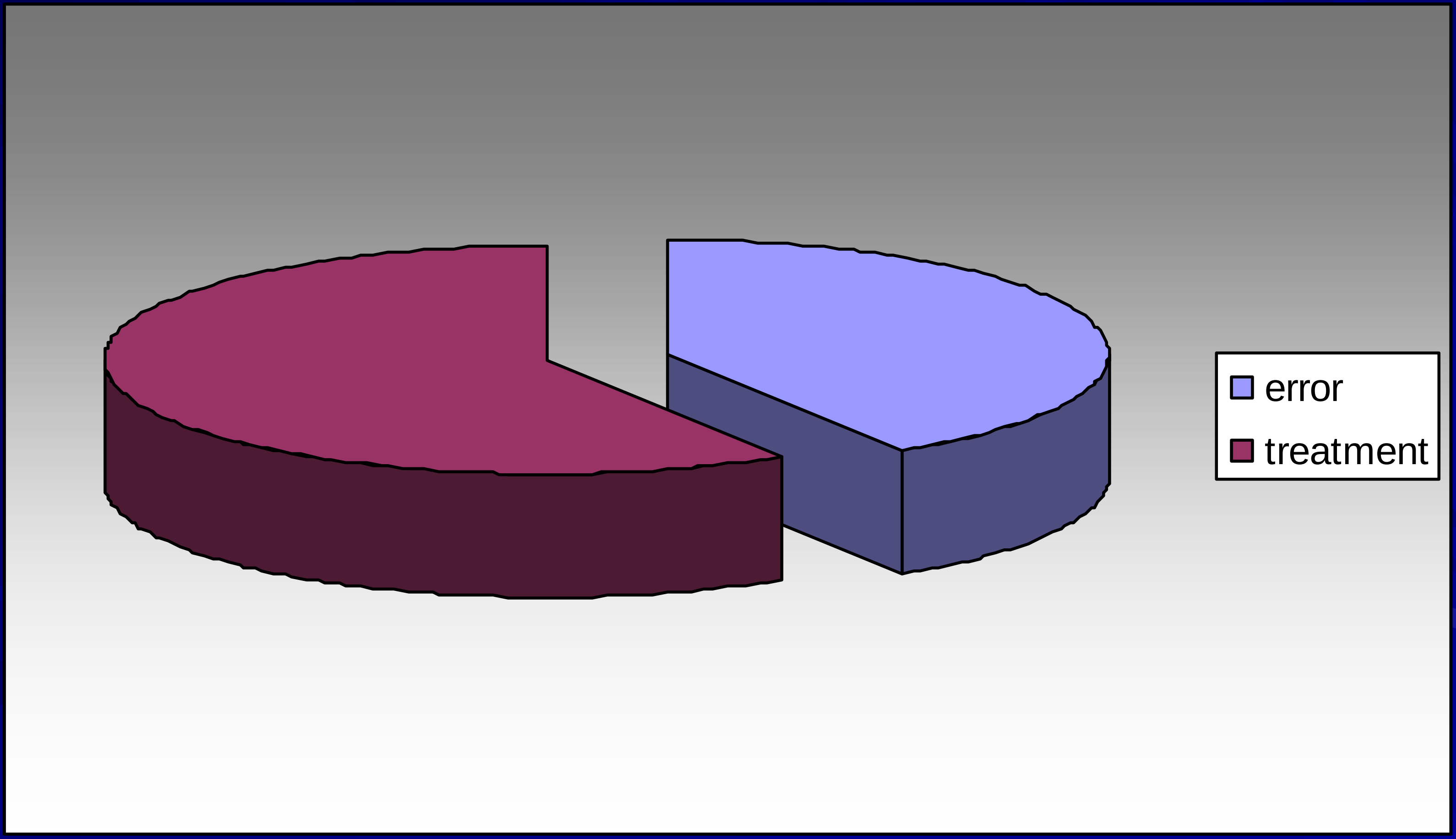
Derivation for one-way ANOVA: expected mean squares

- an expected value of a statistic is defined as the 'long-range average' of a sampling statistic
- our ***expected mean squares*** – are:
  - $E(\text{MS}_{\text{error}}) \rightarrow \sigma_e^2$
    - *i.e., the long term average of the variances within each sample (S^2) would be the population variance* σ_e^2
  - $E(\text{MS}_{\text{treat}}) \rightarrow \sigma_e^2 + n\sigma_\tau^2$
    - *where σ_τ^2 is the long term average of the variance between sample means and n is the number of observations in each group*
    - *i.e., the long term average of the variances within each sample PLUS any variance between each sample*
    - Basically - if group means don't vary then $n\sigma_\tau^2 = 0$, and so then

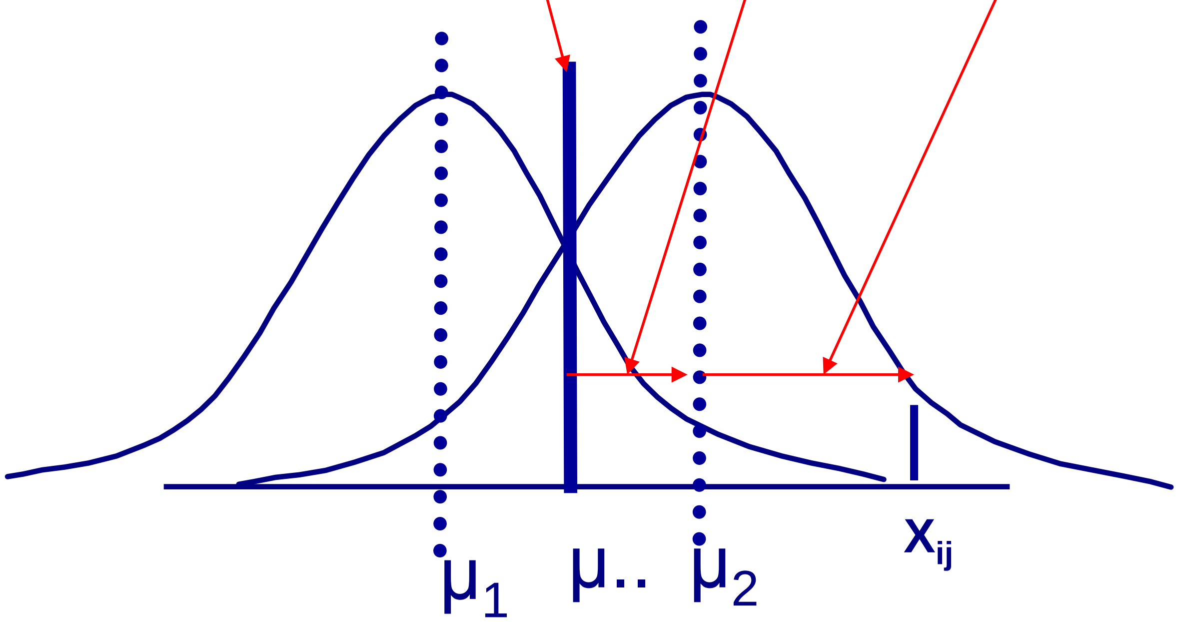
$E(\text{MS}_{\text{treat}}) = \sigma_e^2 + 0 = \sigma_e^2 = E(\text{MS}_{\text{error}}) = \sigma_e^2$

See e.g., Howell (2007) p. 303

partitioning the variance



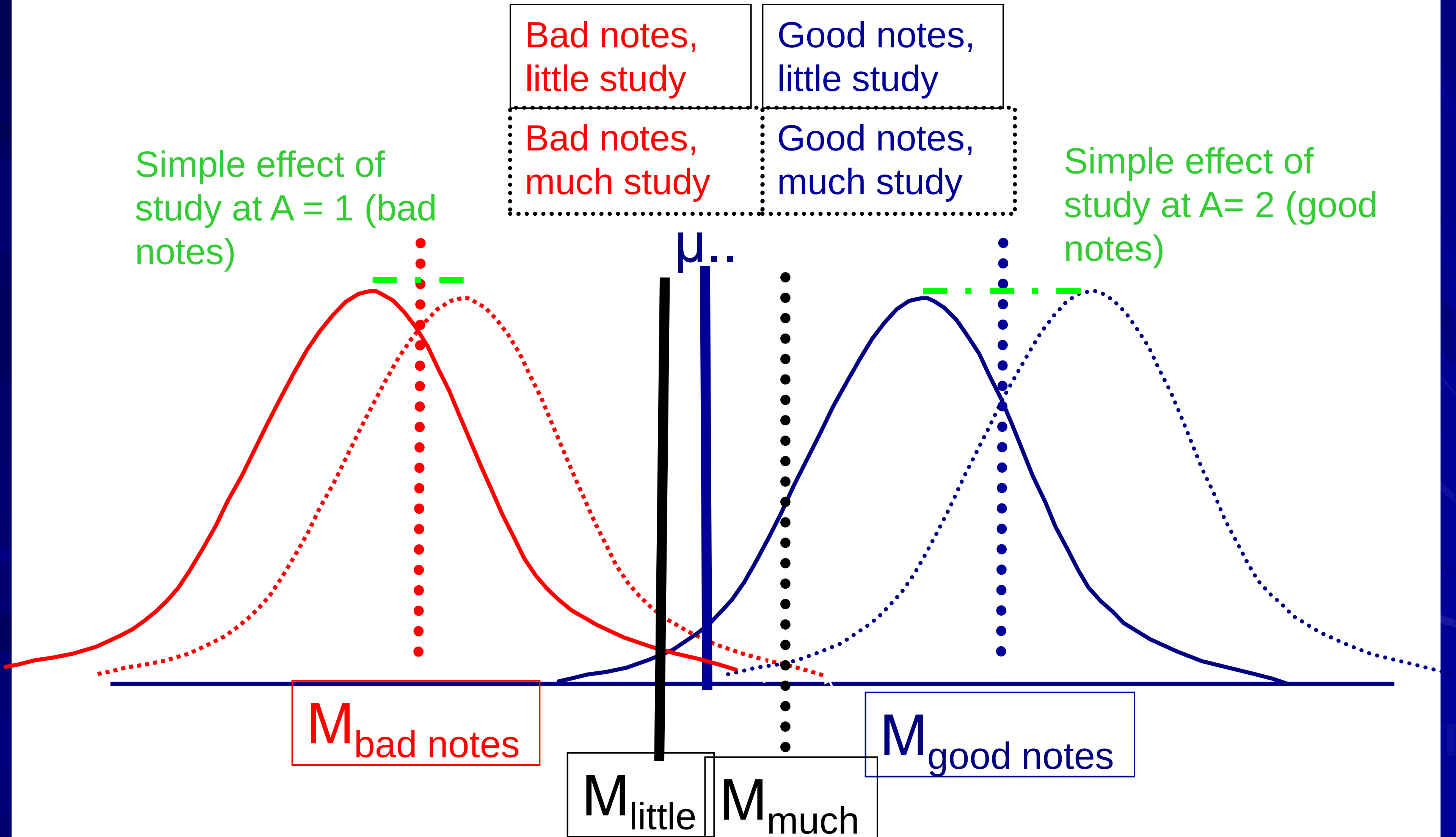
$$X_{ij} = \mu + \tau_j + e_{ij}$$

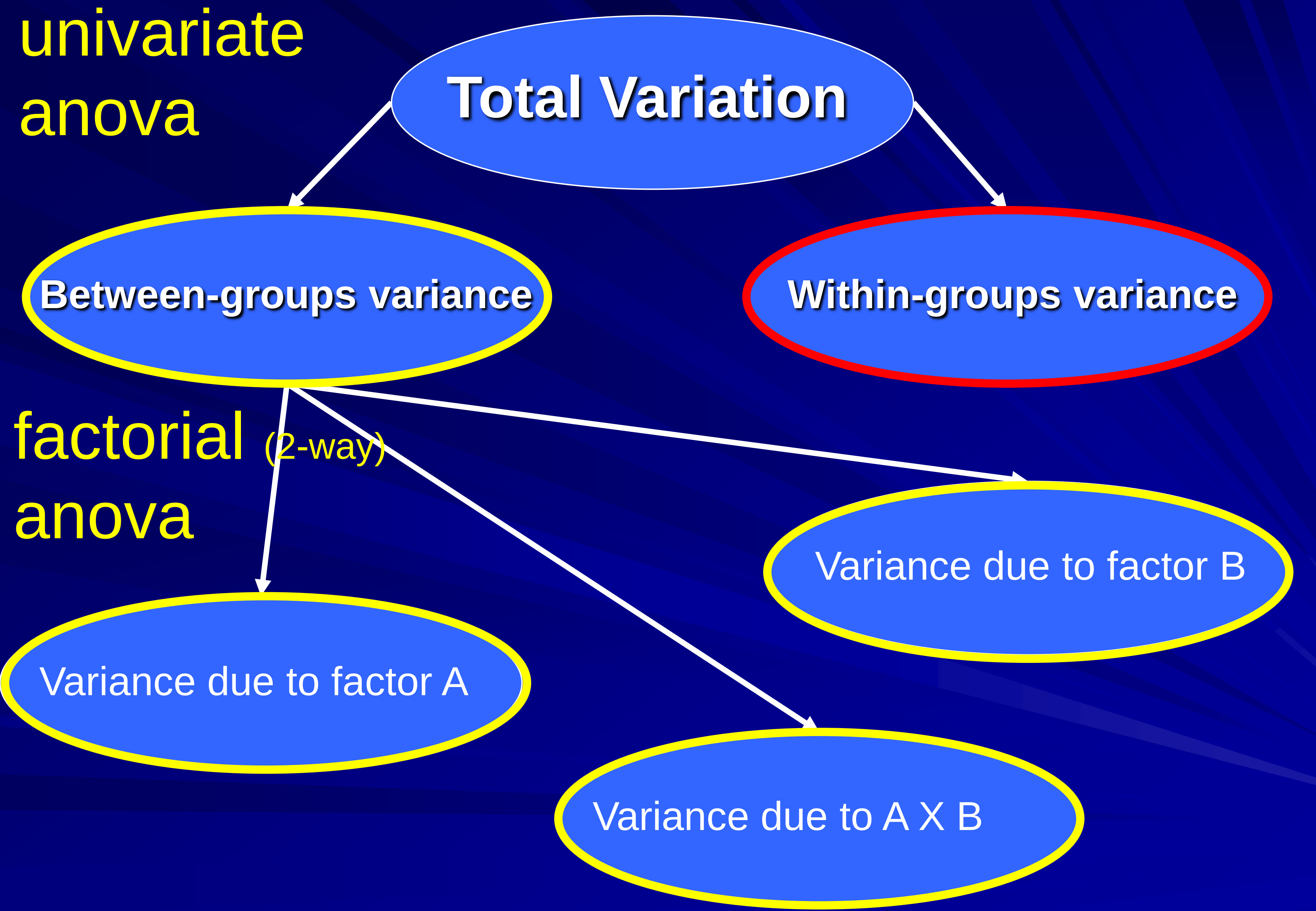


Research questions

- One-way: Is there a treatment effect (is there between-group variability)?
- Two-way:
  - Is there a main effect of A ? (Is there variability between the levels of A, averaging over the other factor? [Do the A group means differ from each other? Do the marginal means of A differ from the grand mean?])
  - Is there a main effect of B ? (Is there variability between the levels of B, averaging over the other factor? [Do the B group means differ from each other? Do the marginal means of B differ from the grand mean?])
  - Is there an A x B interaction? (Does the simple effect of A change for different B groups? Does the simple effect of B change for different A groups?) [Does the simple effect change across the levels of the other factor? Do the cell means differ from the grand mean more than would be expected given the effects of A and B?]

Sources of variance in 2 way factorial designs




univariate
anova
Total Variation
Between-groups variance
Within-groups variance
factorial (2-way)
anova
Variance due to factor B
Variance due to factor A
Variance due to A X B

So what is factorial ANOVA ?

- 1. In 2-way design, estimate between-groups variability
  - Due to main effect of first factor
  - Due to main effect of second factor
  - Due to interaction of two factors
- 2. Estimate within-groups variability
- 3. Weight each variability estimate by # of observations used to generate the estimate (“degrees of freedom”)
- Compare ratio
  - of between-groups variability among levels of A to error, B to error, and ABcells (adjusted for main effects) to error

Formulae for a 2 way between-subjects design with factors A and B

Conceptual or Definitional

Total

$$SS_{TOTAL} = \sum (X - \overline{X}_{..})^2$$

Squaring the deviation of every score from the grand mean x 1 (# of observations behind every score) = total SS

Between-Groups

$$SS_A = nb \sum (\overline{X}_{j.} - \overline{X}_{..})^2$$

$$SS_B = na \sum (\overline{X}_{.k} - \overline{X}_{..})^2$$

Squaring the deviation of the marginal means for each level of the factor from the grand mean x [n x levels of other factor (# of observations behind each factor marginal mean)] = factor SS

$$SS_{AB} = n \sum (\overline{X}_{jk} - \overline{X}_{j.} - \overline{X}_{.k} + \overline{X}_{..})^2$$

Squaring the deviation of each cell mean from the grand mean, adjusting for the 2 marginal means, x n (# of observations behind each cell mean) = factor SS

Within Cells

$$SS_{ERROR} = \sum (X - \overline{X}_{jk})^2$$

Squaring the deviation of each score from the cell mean x 1 (# of observations behind each score) = within-cell or error SS

Formulae for a 2 way between-subjects A x B design

Conceptual

Total

$$SS_{TOTAL} = \sum (X - \overline{X})^2$$

Between-Groups

$$SS_A = nb\sum (\overline{X}_{j.} - \overline{X}_{..})^2$$

$$SS_B = na\sum (\overline{X}_{.k} - \overline{X}_{..})^2$$

$$SS_{AB} = n\sum (\overline{X}_{jk} - \overline{X}_{j.} - \overline{X}_{.k} + \overline{X}_{..})^2$$

Within Cells

$$SS_{ERROR} = \sum (X - \overline{X}_{jk})^2$$

Hint: T = total = sum of X

Computational

Total

$$SS_{TOTAL} = \sum X^2 - \frac{(\sum X)^2}{N} = \sum X^2 - \frac{(T_{..})^2}{N}$$

Between-Groups

$$SS_A = \sum \frac{T_{j.}^2}{nb} - \frac{(T_{..})^2}{N}$$

$$SS_B = \sum \frac{T_{.k}^2}{na} - \frac{(T_{..})^2}{N}$$

$$SS_{CELLS} = \sum \frac{T_{jk}^2}{n} - \frac{(T_{..})^2}{N}$$

$$SS_{AB} = SS_{CELLS} - SS_A - SS_B$$

Within Cells

$$SS_{ERROR} = SS_{TOTAL} - SS_{CELLS}$$

Degrees of freedom summary

***df*total** = N – 1

***df*factor** = no. of levels of the factor – 1

df_B = b – 1
df_A = a – 1

***df*interaction** = product of *df* in factors included in the interaction

df_{BA} = (b – 1) x (a –1)

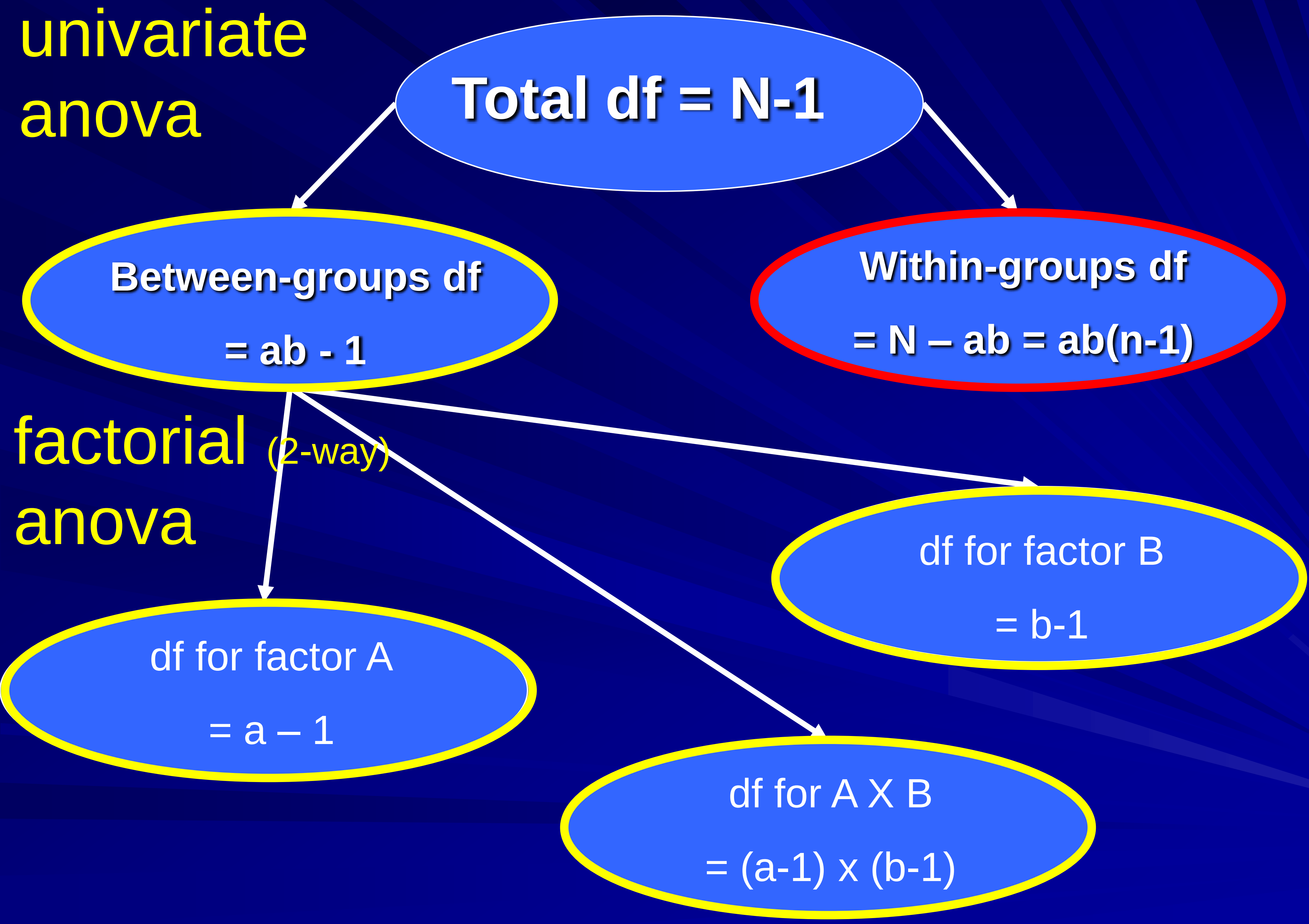
$\boldsymbol{df_{\text{error}}}$ = total no. of observations – no. of treatments
= N – ba

or *df* for each cell x no. of cells
= (n – 1) ba

For a refresher of what DFs are see:

Howell (2007) p.50

Field (2005) p.319


univariate anova
Total df = N-1
Between-groups df
= ab - 1
Within-groups df
= N – ab = ab(n-1)
factorial (2-way) anova
df for factor B
= b-1
df for factor A
= a – 1
df for A X B
= (a-1) x (b-1)

hypothesis testing

factorial anova

- **main effects** (shown for an IV with 3 levels)
  - H_0: $\mu_1 = \mu_2 = \mu_3$
    - no differences among means across levels of the factor
  - H_1: **null is false**
- **Interaction** (shown for a 2 x 3 design)
  - H_0: $\mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} = \mu_{13} - \mu_{23}$
    - if there are differences between particular factor means, they are ***constant*** at each level of the other factor (hence the parallel lines)
    - The ‘difference of the differences’ is zero
  - H_1: **null is false**

logic of factorial anova

A simple extension of one-way anova

- **the F-ratio is still the test statistic we use**

$$F = MS_{treat}/MS_{error}$$

as for univariate anova, $\boldsymbol{MS_{error}}$ = pooled variance (average s^2)

but now we have a separate $\boldsymbol{MS_{treat}}$ for each effect:

1) MS_{treat} for effect of factor A = MS_A(first main effect)
***2) MS_{treat} for effect of factor B* = MS_B** (second main effect)
3) MS_{treat} for effect of factor AB = MS_{AB} (interaction effect)

A ratio of the systematic variance of EACH EFFECT (i.e. of your experimental manipulations or treatments) to the unsystematic variance

Derivation for factorial ANOVA: expected mean squares

- $E(\mathrm{MS}_{\mathrm{error}})$
  - σ_e^2 (i.e., pooled within group variance - as for univariate anova)
- $E(\mathrm{MS}_{\mathrm{A}})$
  - $\sigma_e^2 + nb\sigma_\alpha^2$ (i.e., pooled within group variance PLUS variance between levels of A)
- $E(\mathrm{MS}_{\mathrm{B}})$
  - $\sigma_e^2 + na\sigma_\beta^2$ (i.e., pooled within group variance PLUS variance between levels of B)
- $E(\mathrm{MS}_{\mathrm{AB}})$
  - $\sigma_e^2 + n\sigma_{\alpha\beta}^2$ (i.e., pooled within group variance PLUS variance between the different combinations of A and B levels)

the conceptual model of factorial anova

$$X_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + e_{ijk}$$

for i cases, factor A with *j* treatments, factor B with k treatments, and the AxB interaction with *jk* treatments:

X_{ijk}, any DV score is a combination of:

μ → the grand mean,

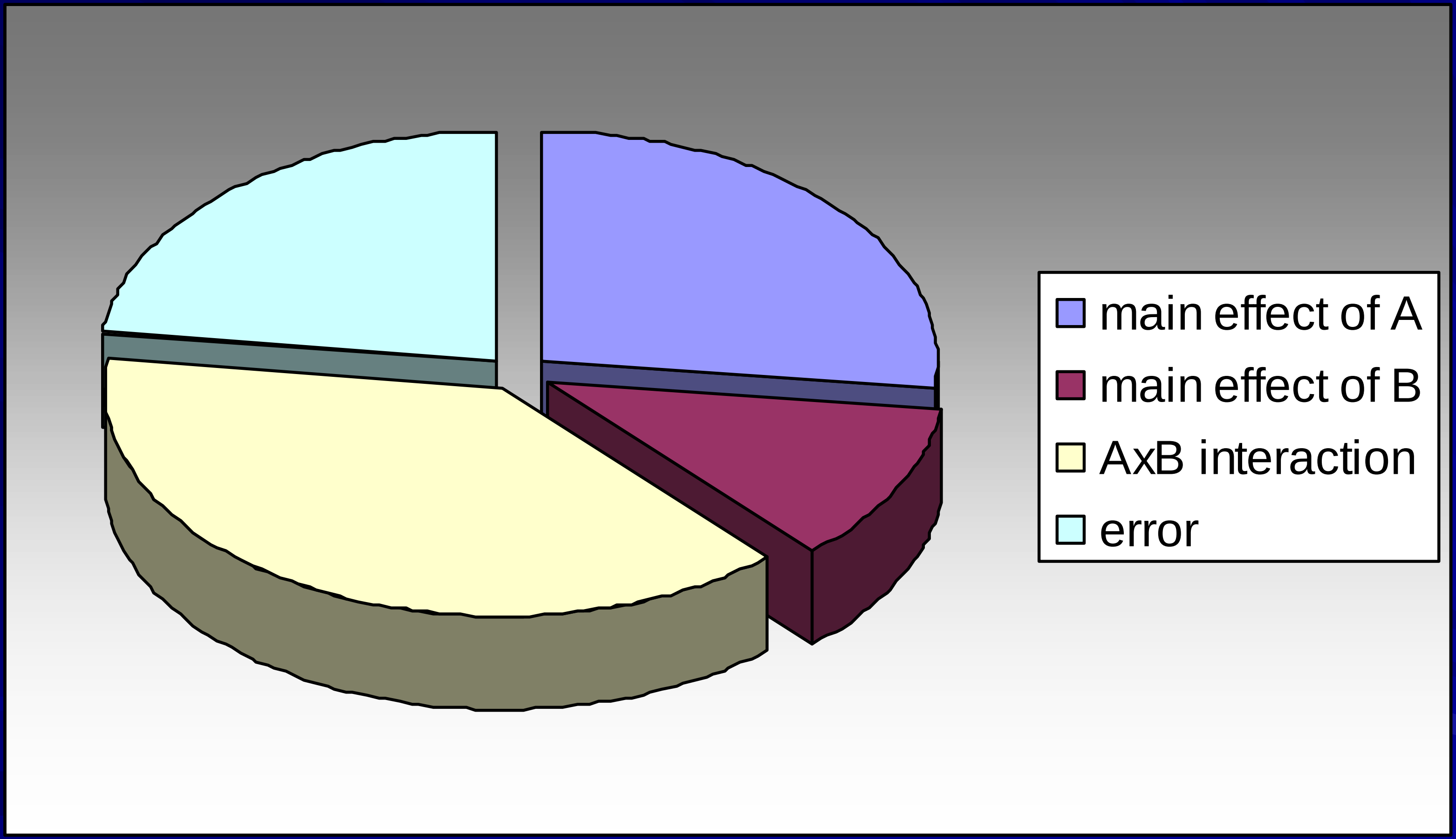
α_j → the effect of the j-th treatment of factor A (μ_{Aj} - μ),

β_k → the effect of the k-th treatment of factor B (μ_{Bk} - μ),

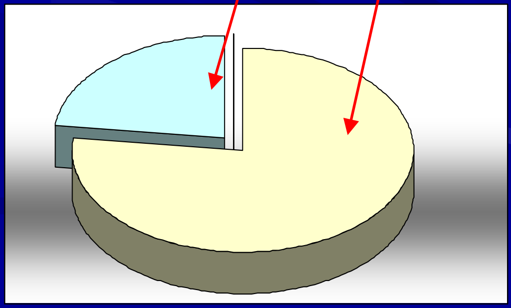
$\alpha\beta_{jk}$ → the effect of differences in factor A treatments at different levels of factor B treatments (μ - μ_{Aj} - μ_{Bk} + μ_{jk}),

e_{ijk} → error, averaged over all j treatments, k treatments and i cases

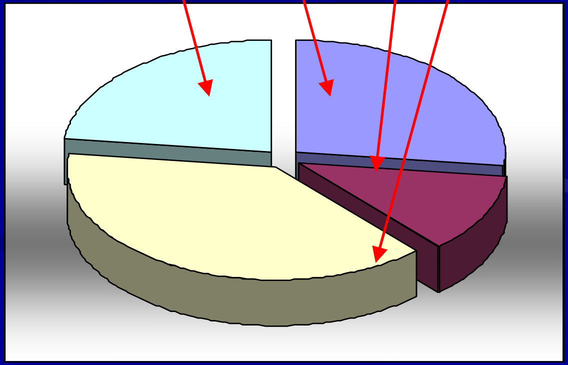
partitioning the variance



1-way (univariate) between-subjects anova

Logic	Derivation (expected mean squares)	Linear Model
Compare variance ***between*** groups with variance ***within*** groups: $\frac{\text{var between}}{\text{var within}}$ (if variance between groups is larger than variance within, group differences are probably significant)	$E(MS_{treat}) = \sigma_e^2 + n\sigma_\tau^2$ $E(MS_{error}) = \sigma_e^2$ ↓ These are what our calculations of MS_{treat} and MS_{error} are based upon (our estimates of var **between** & var **within**) $\frac{MS_{treat}}{MS_{error}} = F$	$X_{ij} = \mu + \tau_j + e_{ij}$ (the grand mean does not get a portion as a mean has no variance) 

2-way (factorial) between-subjects anova

Logic	Derivation (expected mean squares)	Linear Model
For each effect: Compare variance ***between*** means with variance ***within*** groups: var b|n (levels of) A / var w|n groups var b|n (levels of) B / var w|n groups var b|n (levels of) A x B / var w|n groups	$E(MS_A) = \sigma_e^2 + nb\sigma_\alpha^2$ $E(MS_B) = \sigma_e^2 + na\sigma_\beta^2$ $E(MS_{AB}) = \sigma_e^2 + n\sigma_{\alpha\beta}^2$ $E(MS_{error}) = \sigma_e^2$ ↓ To test each effect (i.e., main effects and the interaction): $\frac{MS_{effect}}{MS_{error}} = F$	$X_{ijk} =$ $\mu + \alpha_j + \beta_k + \alpha\beta_{jk} + e_{ijk}$ 

assumptions of anova

- **population**
  - treatment populations are ***normally distributed*** (assumption of normality)
  - treatment populations ***have the same variance*** (assumption of homogeneity of variance)
- **sample**
  - samples are ***independent*** – no two measures are drawn from the same participant
    - c.f. repeated-measures anova – more on that later in the semester
  - each sample obtained by ***independent random sampling*** – within any particular sample, no choosing of respondents on any kind of systematic basis
  - each sample has ***at least 2 observations*** and **equal *n***
- **data (DV scores)**
  - measured using a continuous scale (interval or ratio)
  - mathematical operations (calculations for means, variance, etc) do not make sense for other kinds of scales

an application of between-subjects factorial anova

- **A psychological study of creativity in complex socio-chemical environments (Field, 2000)**
- ***2 factors:***
  – three groups of participants go to the pub and have:
    - No beer, or 2 pints or 4 pints
  – half of the participants are distracted and half are not distracted (controls)

- hence, a 2 x 3 between-subjects factorial design
- ***DV:*** Creativity
  – unbiased 3rd parties rate the quality of limericks made up by each of our participants

an application of between-subjects factorial anova

research questions:

- ***Is there a main effect of alcohol consumption?***
  - does the quality of limerick you make up depend upon how many pints of beer you have had?
- ***Is there a main effect of distraction?***
  - Does the quality of limerick you make up depend on whether you were distracted or not?
- ***Is there a consumption x distraction interaction?***
  - does the effect of distraction upon creativity depend upon consumption (or does the effect of consumption upon creativity depend upon distraction)

a combination of IV levels, e.g., 0 pints and distracted, is called a ***cell***

in most between-subjects factorial designs there are ***n*** observations per cell

the number of cells multiplied by ***n*** gives you ***N***, the total number of observations

(abn = N)

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***

cell totals are calculated – these are just the sum of ***n*** observations in each cell

marginal totals for each level of each factor are also calculated – these are the sum of the corresponding cell totals – summed over levels of the ***other*** factor

the ***grand total*** is the sum of all N observations

Distraction	**Alcohol Consumption (pints)** **0**	**2**	**4**	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***

cell means are calculated – these are just the average of ***n*** observations in each cell

marginal means for each level of each factor are also calculated – these are group means averaged over the levels of the other factor

the **grand mean** is the average of all N observations

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$X_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + e_{ijk}$

so…

$X_{111} = \mu + \alpha_1 + \beta_1 + \alpha\beta_{11} + e_{111}$

where

$\alpha_1 = \mu_{A1} - \mu$
$= 63.75 - 58.33 = \mathbf{5.42}$

$\beta_1 = \mu_{B1} - \mu$
$= 56.46 - 58.33 = \mathbf{-1.87}$

$\alpha\beta_{11} = (\mu - \mu_{A1} - \mu_{B1} + \mu_{AB11})$,
$= 58.33 - 63.75 - 56.46 + 66.88 = \mathbf{5}$

therefore...

$50 = 58.33 + 5.42 – 1.87 + 5 + e_{111}$
*(and $e_{111} =$ **16.88**)*

Distraction	**Alcohol Consumption (pints) 0**	**2**	**4**	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Control	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$X_{ijk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + e_{ijk}$ so…

$X_{332} = \mu + \alpha_3 + \beta_2 + \alpha\beta_{32} + e_{332}$

where

$\alpha_3 = \mu_{A3} - \mu$

$= 46.56 - 58.33 =$ **- 11.87**

$\beta_2 = \mu_{B2} - \mu$

$= 60.21 - 58.33 =$ **1.88**

$\alpha\beta_{32} = (\mu - \mu_{A3} - \mu_{B2} + \mu_{AB32})$,

$= 58.33 - 46.56 - 60.21 + 57.50 =$ **9.06**

therefore…

$70 = 58.33 – 11.87 + 1.88 + 9.06 + e_{332}$

*(and e_{332} = **12.6**)*

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Control	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

Formulae for a 2 way between-subjects A x B design

Conceptual

Total

$$SS_{TOTAL} = \sum (X - \overline{X})^2$$

Between-Groups

$$SS_A = nb\sum(\overline{X}_{j.} - \overline{X}_{..})^2$$

$$SS_B = na\sum(\overline{X}_{.k} - \overline{X}_{..})^2$$

$$SS_{AB} = n\sum(\overline{X}_{jk} - \overline{X}_{j.} - \overline{X}_{.k} + \overline{X}_{..})^2$$

Within Cells

$$SS_{ERROR} = \sum(X - \overline{X}_{jk})^2$$

Computational

Total

$$SS_{TOTAL} = \sum X^2 - \frac{(\sum X)^2}{N} = \sum X^2 - \frac{(T_{..})^2}{N}$$

Between-Groups

$$SS_A = \sum \frac{T_{j.}^2}{nb} - \frac{(T_{..})^2}{N}$$

$$SS_B = \sum \frac{T_{.k}^2}{na} - \frac{(T_{..})^2}{N}$$

$$SS_{CELLS} = \sum \frac{T_{jk}^2}{n} - \frac{(T_{..})^2}{N}$$

$$SS_{AB} = SS_{CELLS} - SS_A - SS_B$$

Within Cells

$$SS_{ERROR} = SS_{TOTAL} - SS_{CELLS}$$

Degrees of freedom summary

***df*total** = N – 1

***df*factor** = no. of levels of the factor – 1

df_B = b – 1
df_A = a – 1

***df*interaction** = product of *df* in factors included in the interaction

df_{BA} = (b – 1) x (a –1)

$\boldsymbol{df}_{\mathbf{error}}$ = total no. of observations – no. of treatments
= N – ba

or *df* for each cell x no. of cells
= (n – 1) ba

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

Calculations

SS_{total}

-same as in univariate anova

-variability around grand mean

SS_A and ***SS_B***

- similar to SS_{treat} in univariate anova

- variability among marginal means

SS_{cells}

- variability among cell means

- caused by effect of A, B or A X B

SS_{AB} $= SS_{cells} - SS_A - SS_B$

- variability due to A x B

SS_{error} - same as univ. anova

- variability around cell mean

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$$SS_{\text{total}} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (A)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A{}^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= (1020^2 + 1035^2 + 745^2) / 16 - 163333.3$$

$$= \mathbf{3332.3}$$

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A^{\;2}}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= (1020^2 + 1035^2 + 745^2) / 16 - 163333.3$$

$$= \mathbf{3332.3}$$

$$SS_B = \frac{\sum T_B^{\;2}}{na} - \frac{(\Sigma X)^2}{N} = (1355^2 + 1445^2) / 24$$

$$-163333.3 \quad = \mathbf{168.78}$$

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	[illegible]	[illegible]	[illegible]	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A^{\ 2}}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= (1020^2 + 1035^2 + 745^2) / 16 - 163333.3$$

$$= \mathbf{3332.3}$$

$$SS_B = \frac{\sum T_B^{\ 2}}{na} - \frac{(\Sigma X)^2}{N} = (1355^2 + 1445^2) / 24$$

$$-163333.3 \quad = \mathbf{168.78}$$

$$SS_{cells} = \frac{\sum T_{AB}^{\ 2}}{n} - \frac{(\Sigma X)^2}{N} =$$

$$(535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2) / 8$$

$$- 163333.3 \quad = \mathbf{5479.2}$$

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
Distraction	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
	65	70	55	
	70	65	65	
	60	60	70	
Controls	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= (1020^2 + 1035^2 + 745^2) / 16 - 163333.3$$

$$= \mathbf{3332.3}$$

$$SS_B = \frac{\sum T_B^2}{na} - \frac{(\Sigma X)^2}{N} = (1355^2 + 1445^2) / 24$$

$$-163333.3 \quad = \mathbf{168.78}$$

$$SS_{cells} = \frac{\sum T_{AB}^2}{n} - \frac{(\Sigma X)^2}{N} =$$

$$(535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2)/8$$

$$- 163333.3 \quad = \mathbf{5479.2}$$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$$= \mathbf{5479.2} - \mathbf{3332.3} - \mathbf{168.78} = \mathbf{1978.12}$$

Distraction	Alcohol Consumption (pints) 0	2	4	Marginal Totals (B)
Distraction	50	45	30	
	55	60	30	
	80	85	30	
	65	65	55	
	70	70	35	
	75	70	20	
	75	80	45	
	65	60	40	
Cell Totals	**535**	**535**	**285**	***1355***
Cell Means	***66.88***	***66.88***	***35.63***	***56.46***
Controls	65	70	55	
	70	65	65	
	60	60	70	
	60	70	55	
	60	65	55	
	55	60	60	
	60	60	50	
	55	50	50	
Cell Totals	**485**	**500**	**460**	***1445***
Cell Means	***60.63***	***62.50***	***57.50***	***60.21***
Marginal Totals (*A*)	***1020***	***1035***	***745***	***2800***
Means	***63.75***	***64.69***	***46.56***	***58.33***

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= (1020^2 + 1035^2 + 745^2) / 16 - 163333.3 = \mathbf{3332.3}$$

$$SS_B = \frac{\sum T_B^2}{na} - \frac{(\Sigma X)^2}{N} = (1355^2 + 1445^2) / 24 - 163333.3 = \mathbf{168.78}$$

$$SS_{cells} = \frac{\sum T_{AB}^2}{n} - \frac{(\Sigma X)^2}{N} =$$

$$(535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2)/8 - 163333.3 = \mathbf{5479.2}$$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$$= 5479.2 - 3332.3 - 168.78 = \mathbf{1978.12}$$

$$SS_{error} = SS_{total} - SS_{cells}$$

$$= 8966.7 - 5479.2 = \mathbf{3487.5}$$

Summary Table

Source	df	SS	MS	F
A (cons)				
B (dist)				
AB				
Error				
Total				

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= 1020^2 + 1035^2 + 745^2 / 16 - 163333.3$$

$$= \mathbf{3332.3}$$

$$SS_B = \frac{\sum T_B^2}{na} - \frac{(\Sigma X)^2}{N} = 1355^2 + 1445^2 / 24$$

$$-163333.3 = \mathbf{168.78}$$

$$SS_{cells} = \frac{\sum T_{AB}^2}{n} - \frac{(\Sigma X)^2}{N}$$

$$= 535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2 / 8$$

$$- 163333.3 = \mathbf{5479.2}$$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$$= 5479.2 - 3332.3 - 168.78 = \mathbf{1978.12}$$

$$SS_{error} = SS_{total} - SS_{cells}$$

$$= 8966.7 - 5479.2 = \mathbf{3487.5}$$

Summary Table

Source	df	SS	MS	F
A (cons)		3332.3		
B (dist)				
AB				
Error				
Total				

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A^{\,2}}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= 1020^2 + 1035^2 + 745^2 \,/\, 16 - 163333.3$$

$$= 3332.3$$

$$SS_B = \frac{\sum T_B^{\,2}}{na} - \frac{(\Sigma X)^2}{N} = 1355^2 + 1445^2 \,/\, 24$$

$$-163333.3 = 168.78$$

$$SS_{cells} = \frac{\sum T_{AB}^{\,2}}{n} - \frac{(\Sigma X)^2}{N}$$

$$= 535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2 \,/8$$

$$- 163333.3 = 5479.2$$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$$= 5479.2 - 3332.3 - 168.78 = \mathbf{1978.12}$$

$$SS_{error} = SS_{total} - SS_{cells}$$

$$= \mathbf{8966.7} - \mathbf{5479.2} = \mathbf{3487.5}$$

Summary Table

Source	df	SS	MS	F
A (cons)		3332.3		
B (dist)		168.75		
AB				
Error				
Total				

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = \mathbf{8966.7}$$

$$SS_A = \frac{\sum T_A^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= 1020^2 + 1035^2 + 745^2 / 16 - 163333.3$$

$$= 3332.3$$

$$SS_B = \frac{\sum T_B^2}{na} - \frac{(\Sigma X)^2}{N} = 1355^2 + 1445^2 / 24$$

$$- 163333.3 = 168.78$$

$$SS_{cells} = \frac{\sum T_{AB}^2}{n} - \frac{(\Sigma X)^2}{N}$$

$$= 535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2 / 8$$

$$- 163333.3 = 5479.2$$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$$= 5479.2 - 3332.3 - 168.78 = \mathbf{1978.12}$$

$$SS_{error} = SS_{total} - SS_{cells}$$

$$= 8966.7 - 5479.2 = \mathbf{3487.5}$$

Summary Table

Source	df	SS	MS	F
A (cons)		3332.3		
B (dist)		168.75		
AB		1978.12		
Error				
Total				

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = 8966.7$$

$$SS_A = \frac{\sum T_A^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= 1020^2 + 1035^2 + 745^2 / 16 - 163333.3$$

$$= 3332.3$$

$$SS_B = \frac{\sum T_B^2}{na} - \frac{(\Sigma X)^2}{N} = 1355^2 + 1445^2 / 24$$

$$-163333.3 \quad = 168.78$$

$$SS_{cells} = \frac{\sum T_{AB}^2}{n} - \frac{(\Sigma X)^2}{N}$$

$$= 535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2 / 8$$

$$- 163333.3 \quad = 5479.2$$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$$= 5479.2 - 3332.3 - 168.78 = 1978.12$$

$$SS_{error} = SS_{total} - SS_{cells}$$

$$= 8966.7 - 5479.2 = 3487.5$$

Summary Table

Source	df	SS	MS	F
A (cons)		3332.3		
B (dist)		168.75		
AB		1978.12		
Error		3487.5		
Total				

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$$= 172300 - 163333.3 = 8966.7$$

$$SS_A = \frac{\sum T_A^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$$= 1020^2 + 1035^2 + 745^2 / 16 - 163333.3$$

$$= 3332.3$$

$$SS_B = \frac{\sum T_B^2}{na} - \frac{(\Sigma X)^2}{N} = 1355^2 + 1445^2 / 24$$

$$-163333.3 = 168.78$$

$$SS_{cells} = \frac{\sum T_{AB}^2}{n} - \frac{(\Sigma X)^2}{N}$$

$$= 535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2 / 8$$

$$- 163333.3 = 5479.2$$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$$= 5479.2 - 3332.3 - 168.78 = 1978.12$$

$$SS_{error} = SS_{total} - SS_{cells}$$

$$= 8966.7 - 5479.2 = 3487.5$$

Summary Table

Source	df	SS	MS	F
A (cons)		3332.3		
B (dist)		168.75		
AB		1978.12		
Error		3487.5		
Total		8966.7		

$$SS_{total} = \Sigma X^2 - \frac{(\Sigma X)^2}{N}$$

$= 172300 - 163333.3 = 8966.7$

$$SS_A = \frac{\sum T_A^2}{nb} - \frac{(\Sigma X)^2}{N}$$

$= 1020^2 + 1035^2 + 745^2 / 16 - 163333.3$

$= 3332.3$

$$SS_B = \frac{\sum T_B^2}{na} - \frac{(\Sigma X)^2}{N}$$

$= 1355^2 + 1445^2 / 24$

$-163333.3 \quad = 168.78$

$$SS_{cells} = \frac{\sum T_{AB}^2}{n} - \frac{(\Sigma X)^2}{N}$$

$= 535^2 + 535^2 + 285^2 + 485^2 + 500^2 + 460^2 /8$

$- 163333.3 \quad = 5479.2$

$$SS_{AB} = SS_{cells} - SS_A - SS_B$$

$= 5479.2 - 3332.3 - 168.78 = 1978.12$

$$SS_{error} = SS_{total} - SS_{cells}$$

$= 8966.7 - 5479.2 = 3487.5$

Summary Table

Source	df	SS	MS	F
A (cons)	2	3332.3		
B (dist)	1	168.75		
AB	2	1978.12		
Error	42	3487.5		
Total	47	8966.7		

MS stands for
MEAN-SQUARE

– this is a corrected variance estimate used to calculate the F-ratio

MS = SS/df

df_{factor}

= number of levels of that factor –1

- **Consumption (A) → 3 - 1 = 2**
- **Distraction (B) → 2- 1 = 1**

$df_{interaction}$

= product of *df* values for factors involved in interaction

- **AxB → 1 X 2 = 2**

df_{error}

= *N* – number of cells in the design

- **error → 48 – 6 = 42**

df_{total}

= N-1 → 48-1 =47

Summary Table

Source	df	SS	MS	F
A (cons)	2	3332.3	1666.15	
B (dist)	1	168.75	168.75	
AB	2	1978.12	989.06	
Error	42	3487.5	83.02	
Total	47	8966.7		

MS stands for ***MEAN-SQUARE***

– this is a corrected variance estimate used to calculate the F-ratio

MS = SS/df

df_{factor}

= number of levels of that factor –1

- **Consumption (A) → 3 - 1 = 2**
- **Distraction (B) → 2- 1 = 1**

$df_{interaction}$

= product of *df* values for factors involved in interaction

- **AxB → 1 X 2 = 2**

df_{error}

= *N* – number of cells in the design

- **error → 48 – 6 = 42**

df_{total}

= N-1 → 48-1 =47

Summary Table

Source	df	SS	MS	F
A (cons)	2	3332.3	1666.15	20.07
B (dist)	1	168.75	168.75	2.03
AB	2	1978.12	989.06	11.91
Error	42	3487.5	83.02	
Total	47	8966.7		

$$F = MS_{treat}/MS_{error}$$

SPSS provides significance levels (or you can look up tables)

Summary Table

Source	df	SS	MS	F	sig
A (cons)	2	3332.3	1666.15	20.07	.000
B (dist)	1	168.75	168.75	2.03	.161
AB	2	1978.12	989.06	11.91	.000
Error	42	3487.5	83.02		
Total	47	8966.7			

the results of this anova show . . .

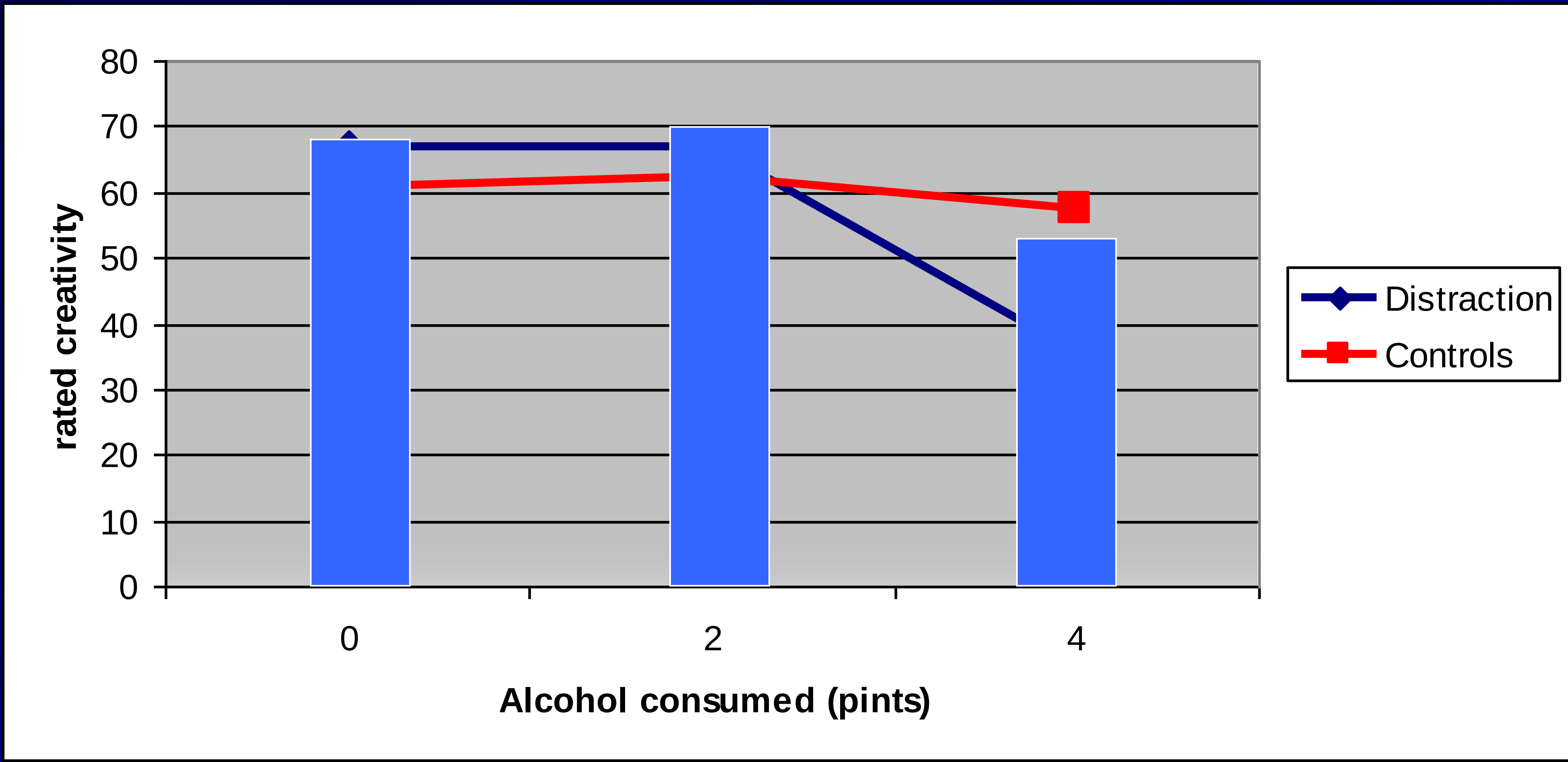
a significant ***main effect of pints consumed***

no ***main effect of distraction***

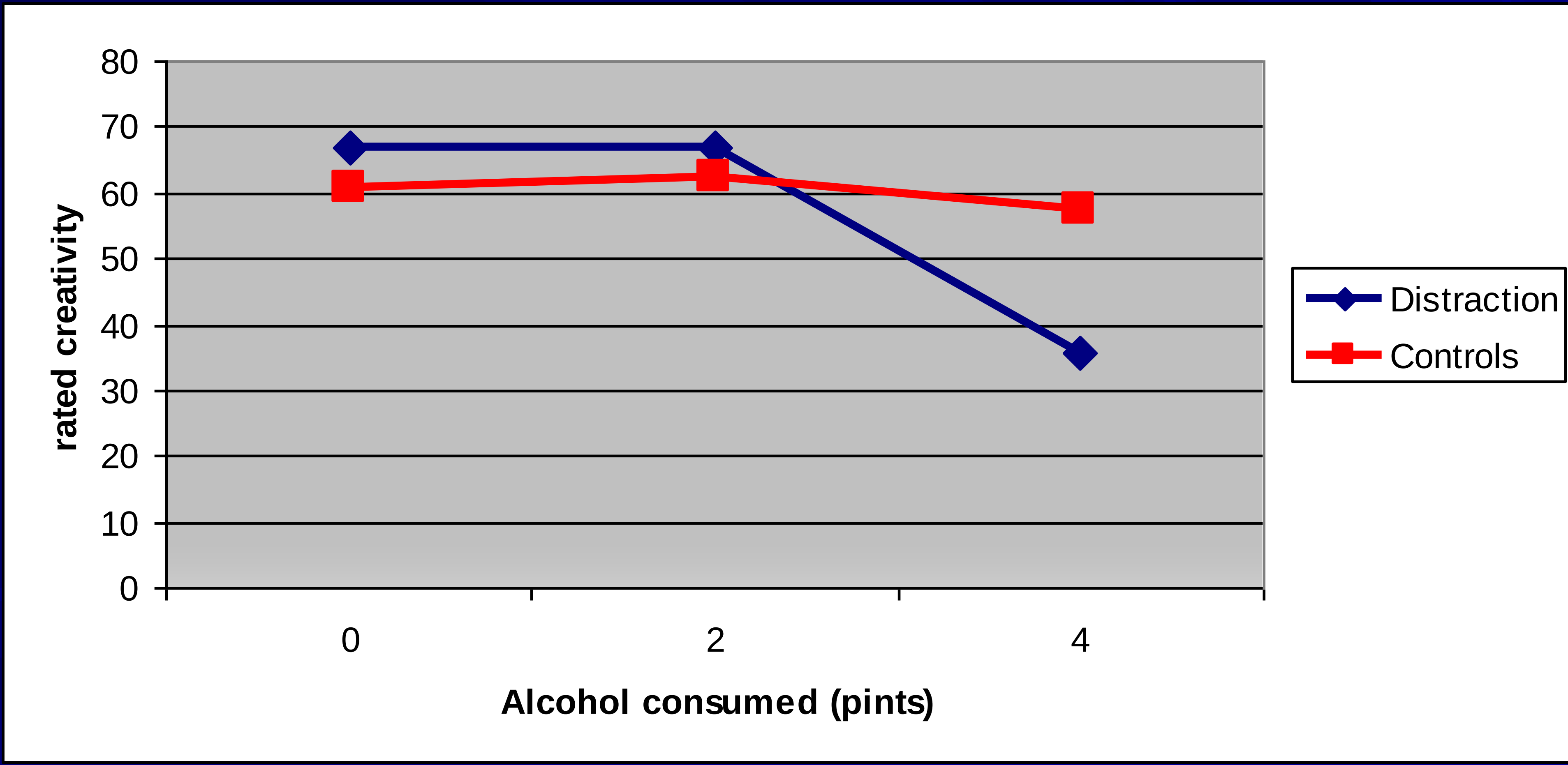
a significant ***consumption x distraction interaction***

no main effect of distraction

main effect of alcohol consumed



disordinal interaction



- next week: Simple effects & effect size
- Readings:
  - skim Howell chapter 12
  - read Howell chapter 13.4
  - Field chapter 10.1 and 10.2

- tutes this week focus on visually identifying main effects and interactions