Index of 11-L: Multifactorial analysis

Practical Information

Schedule:

- fourth home assignment (statistical reporting in papers) underway, due Thursday, Apr 2,
- oral exam scheduled for Friday, April 17, 1:30-3:30pm.

Today's lecture — introductions[1] to two-way ANOVA and multiple regression:

- two-way ANOVA[2] — first extension beyond single predictor models,
  - impact on outcome of two factors (categorical variables) studied simultaneously → new concept of interaction (and additivity),[3]
- multiple regression = regression with more than one x-variable,[4]
  - model looks similar to simple linear regression, but
    - different interpretation of parameters,
    - new issue: selection of x-variables,
- both models fall within framework of (general) linear models with many aspects of analysis similar to what we have seen: least square estimation, confidence intervals and tests, residuals, R-square, estimation/prediction. . . ,
- analysis now completely by statistical software, though some explicit formulas exist.

[1] The textbook chapters are very brief — some more details in lecture, full discussion in VHM 8020 and 8120 courses.
[2] PSLS 3e Supplement: Sections 26.4-6 (in course syllabus!); S: not covered; IPS 7e: Chapter 13.
[3] Some topics skipped compared to earlier years: Friedman's non-parametric test, ANOVA without replication.
[4] PSLS 3e Supplement: Sections 28.1-7; S: 10.4; IPS 7e: Chapter 11 — note: not in course syllabus.

Data Examples for Two-Way ANOVA and Multiple Regression

Energy expenditures in Burkina Faso:[5]

- mean energy expenditures in farming families, divided by gender and season,
- summarized data (no raw data available):

Energy expenditure (calories per day)		Gender	
		men	women
Season	dry	2310	2320
	wet	3460	2890

Phosphorus levels in tomato plants: (PSLS 3e, Example 26.12)

- 3 levels of nitrogen fertilizer ($0/28/160$ kg/ha) applied to two genotypes of tomato plants (wild/mutant, wild being susceptible to Mycorrhizal fungi),
- genotypes $\sim$ blocks, 6 replicates per treatment$\times$block.

Test scores for 224 computer science major students in one year:[6]

- Y_i = grade point average (GPA) for first three semesters,
- x_{i1}, x_{i2}, x_{i3} = high school grades in math (HSM), science (HSS), English (HSE),
- x_{i4}, x_{i5} = scholastic aptitude test scores, math part (SATM), verbal part (SATV),
- x_{i6} = sex ($1=$men, $2=$women).

for i^{th} student, $i = 1, \ldots, 224$.

[5] Based on Payne: Nutrition adaptation in man: social adjustments and their nutritional implications, in Blaxter & Waterlow (eds.): *Nutrition Adaptation in Man.*, Libbey, London, 1985.

[6] CSdata included with earlier IPS textbooks; Campbell & McCabe (1984), *Communications of the ACM*, 1108–1113.

Factorial Designs

Factor (categorical, explanatory variable, e.g. treatment/control):

- grouping of observations into categories/levels, either by symbols (e.g. letters, roman numbers) or numbers,
- usually, it does not matter if factors are coded by numbers or symbols: use most natural coding, and if factors are coded numerically, check df to ensure their modelling as a grouping.[7]

Several factors in the same design? — Yes!, in good designs it is possible to separate effects of different factors from each other ⇒

- cheaper (less experimental units) than in separate experiments,[8]
- possible to study combined effect of several factors,
- increased scope of the study/experiment,

and analyzing multi-factorial data by each factor separately: is generally wrong and only gives valid results if at most one factor is of importance.

Two types of randomization for factorial experiments:

- completely randomized design,
- (randomized) block design.

[7] The software may misunderstand the factor as continuous and estimate a slope.
[8] The advantage arises e.g. if assessment of nitrogen effects can be done on wild and mutant tomato plants combined, in additive models introduced later.

Notation for Two-Way ANOVA

Data layout and notation:

observations X_{ijk}		column (C) factor $\sim j$ 1	$\ldots$	j	$\ldots$	J
	1	$X_{111},\ldots,X_{11n_{11}}$	$\ldots$	$X_{1j1},\ldots,X_{1jn_{1j}}$	$\ldots$	$X_{1J1},\ldots,X_{1Jn_{1J}}$
row (R)	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
factor	i	$X_{i11},\ldots,X_{i1n_{i1}}$	$\ldots$	$X_{ij1},\ldots,X_{ijn_{ij}}$	$\ldots$	$X_{iJ1},\ldots,X_{iJn_{iJ}}$
$\sim i$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	I	$X_{I11},\ldots,X_{I1n_{I1}}$	$\ldots$	$X_{Ij1},\ldots,X_{Ijn_{Ij}}$	$\ldots$	$X_{IJ1},\ldots,X_{IJn_{IJ}}$

- $X_{ijk}=k$th observation in group defined by row factor R$=i$ and column factor C$=j$,[9]
  - $i=1,\ldots,I$, and $I=$ number of levels of R/rows,
  - $i=j,\ldots,J$, and $J=$ number of levels of C/columns,
  - $k=1,\ldots,n_{ij}$, and $n_{ij}=$ number of obs. in (i,j)th group,
- let also $N=\sum_{ij} n_{ij}$ the total number of obs., and $\bar{X}=\sum_{ijk} X_{ijk}/N$ the overall mean,
- terminology: the dataset/design
  - is balanced, if all groups equally large ($n_{11}=\ldots=n_{IJ}$, similar to one-way ANOVA), otherwise unbalanced (also here not necessarily a problem),
  - is complete, if all $I\cdot J$ groups present, otherwise incomplete (difficult design, avoid if possible!),
  - has replication, if some of the n_{ij}'s >1, otherwise no replication (all $n_{ij}=1$).

[9] IPS uses the notation: A=row factor, B=column factor.

One-Way ANOVA for Two-Way Factorial

In a two-way design with replication, if we focus only on the grouping from the row and column factors ($I \cdot J$ groups) and otherwise forget about row and column factors $\Rightarrow$ 1-way ANOVA for combined factor with $I \cdot J$ levels:

- Model:

$$X_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad \text{for } \varepsilon_{ijk}\text{'s i.i.d. and } \sim \mathrm{N}(0, \sigma),$$

  and where μ_{ij}'s are group (population) means,

- Estimation:

$$\hat{\mu}_{ij} = \bar{X}_{ij\cdot} \text{ (combined group means)},$$
$$\hat{\sigma}^2 = s_p^2 = \textstyle\sum_{ij} \frac{n_{ij}-1}{N-IJ}\, s_{ij}^2 = \mathrm{MSE}, \text{ and } \mathrm{DFE} = N - IJ,$$

  where s_{ij} = sample standard deviation in group (i, j),

- ANOVA table:

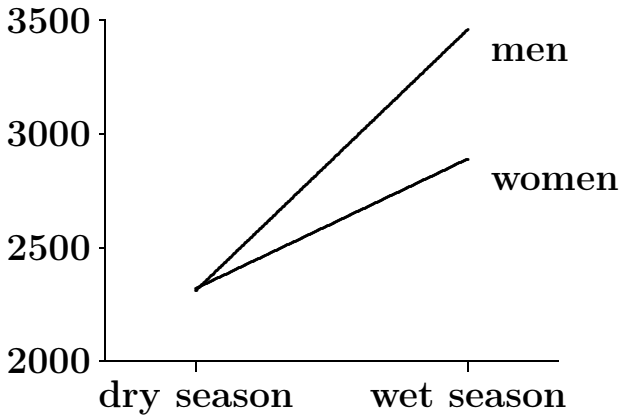
Source	DF	SS	MS	F
Groups	$IJ-1$	$\sum_{ij} n_{ij}(\bar{X}_{ij}-\bar{X})^2$	SSG/DFG	MSG/MSE
Error	$N-IJ$	$\sum_{ijk}(X_{ijk}-\bar{X}_{ij})^2$	SSE/DFE	
Total	$N-1$	$\sum_{ijk}(X_{ijk}-\bar{X})^2$		

- Problem: analysis does not directly give information about row and column factors separately $\Rightarrow$ need to decompose (split up) model's group terms.

Decomposing a Two-Way Table of Means I

Example: Energy expenditures in Burkina Faso:

Energy expend. (calories)		Gender men	Gender women	Mean
Season	dry	2310	2320	2315
	wet	3460	2890	3175
Mean		2885	2605	2745



Different ways to look at the data:

(i) four separate groups,

(ii) two gender groups for each season, or (iii) two season groups for each gender,

(iv) (overall level), two season groups, two gender groups, association between gender and season.

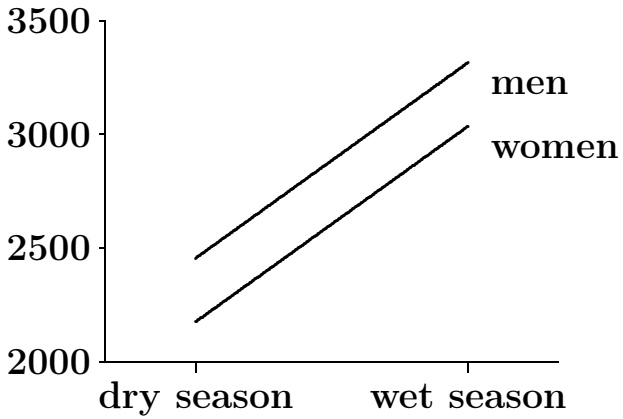
Decomposition of means $\sim$ (iv):

$\bar{X}$ overall mean	2745	2745	140	−140	$\bar{X}_{.j} - \bar{X}$ gender effect
	2745	2745	140	−140	
$\bar{X}_{i.} - \bar{X}$ season effect	−430	−430	−145	145	$\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}$
	430	430	145	−145	

Decomposing a Two-Way Table of Means II

Modified energy expenditures in Burkina Faso:

Energy expend. (calories)		Gender men	Gender women	Mean
Season	dry	2455	2175	2315
	wet	3315	3035	3175
Mean		2885	2605	2745



Decomposition of means ∼ (iv):

$\bar{X}$ overall mean

$$\begin{bmatrix} 2745 & 2745 \\ 2745 & 2745 \end{bmatrix}$$

$\bar{X}_{.j} - \bar{X}$ gender effect

$$\begin{bmatrix} 140 & -140 \\ 140 & -140 \end{bmatrix}$$

$\bar{X}_{i.} - \bar{X}$ season effect

$$\begin{bmatrix} -430 & -430 \\ 430 & 430 \end{bmatrix}$$

$\bar{X}_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}$

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Comparison of two variants of the data:

- same overall level, same overall (average) effects of gender and season,
- modified data: parallel lines ⇒ additive effects — same effect of one factor at all levels of other factor(s),
- original data: non-parallel lines ⇒ non-additive effects, or interaction between the factors gender and season.

Interaction and Additivity

Interaction — some other words:

- synergism or antagonism (depending on the type of interaction),
- epistasy (genetics),
- covariation and association.[10]

Interaction between two factors:

- the main effects provide an incomplete description, i.e.: the combined effect of two factors is not predictable from the isolated effect of each of them when examined separately,
- the effect of the first factor depends on the level of the second factor (or vice versa) — "it depends...",
- no additivity between factors $\Rightarrow$ interaction and additivity are opposites,
- non-parallel lines ($\Rightarrow$ parallel lines $\sim$ additivity).

[10] Both of these terms can be misunderstood as referring to the relation between the two factors themselves, instead of to their combined impact on an outcome.

Two-Way ANOVA Models

Basic model (in two equivalent formulations):

$$X_{ijk} = \mu_{ij} + \varepsilon_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

where the random terms ("errors") ε_{ijk} are i.i.d. and $\sim \mathrm{N}(0, \sigma)$.

New parameters and interpretations:[11]

- μ = overall mean,
- α_i = "main effect" of ith row group,
- β_j = "main effect" of jth column group,
- γ_{ij} = "interaction effect" of (i, j)th group (only meaningful with replication).

Overview of models:[12] (where R $\sim$ row factor, and C $\sim$ column factor)

EX_{ij}	Model formula	Interpretation	Corresponding model
μ_{ij}	(μ+) R+C+R*C	interaction R*C	one-way ANOVA for R×C
$\mu + \alpha_i + \beta_j$	(μ+) R+C	additivity	new model
$\mu + \alpha_i$	(μ+) R	no effect of C	one-way ANOVA for R
$\mu + \beta_j$	(μ+) C	no effect of R	one-way ANOVA for C
μ	(μ)	no effects	one-sample analysis

[11] Technical note: it is necessary to put some restrictions on α's, β's and γ's (otherwise too many parameters).
[12] Final models, i.e. after disregarding non-significant terms

Analysis of Two-Way ANOVA (with Replication)

- same steps as in one-way ANOVA: estimation, model check, ANOVA table with F-tests, contrasts and/or graphical presentation,[13]
- more rows in ANOVA table, because of more tests.

Two-way ANOVA table:

Source of variation	Degrees of freedom	Sum of squares	Mean square	Hypothesis/ F-statistic
Row factor R	$I-1$	$\sum_{ij} n_{ij}(\bar{X}_{i\cdot}-\bar{X})^2$	SSR/DFR	H_0: no row eff. $F=$MSR/MSE
Column factor C	$J-1$	$\sum_{ij} n_{ij}(\bar{X}_{\cdot j}-\bar{X})^2$	SSC/DFC	H_0: no column eff. $F=$MSC/MSE
Interaction R×C	$(I-1)(J-1)$	SSG−SSR−SSC	SSRC/DFRC	H_0: no interaction $F=$MSRC/MSE
Error	$N-IJ$	$\sum_{ijk}(X_{ijk}-\bar{X}_{ij})^2$	SSE/DFE	
Total	$N-1$	$\sum_{ijk}(X_{ijk}-\bar{X})^2$		

- test for interaction first ("read ANOVA table from bottom to top"): if significant, base conclusions on $\hat{\mu}_{ij}$'s ($\bar{X}_{ij\cdot}$'s) alone (using pairwise comparisons or contrasts),[14]
- additive model estimation and interpretation: separately for row and column factors (based on respective tests, and row and column means).

[13] Also, as in all linear models, the error standard deviation σ is estimated as $\sqrt{\text{MSE}}$ with DFE degrees of freedom.
[14] In the presence of strong interaction, tests for main effects are often not meaningful!

Summary of Two-Way ANOVA for Tomato Data

Statistical model:

$$X_{ijk} = \mu_{ij} + \varepsilon_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

where $i = 1, 2, 3$ (nitrogen: 0, 28, 160), $j = 1, 2$ (mutant, wild), and $k = 1, \ldots, 6$, or as a **model formula**:

$$\ln(\text{Phosphorus}) = \text{N} + \text{Type} + \text{N*Type} + \text{Error}.$$

ANOVA table:

Source	DF	SS	MS	F	P-value	
Nitrogen	2	0.9171	0.4586	28.4	<0.0005	strong significance for Nitrogen
Genotype	1	3.9654	3.9654	246	<0.0005	strong significance for Genotype
Interaction N∗G	2	0.0536	0.0268	1.66	0.21	interaction non-significant
Error	30	0.4843	0.0161			
Total	35	5.4204	$s = \sqrt{\text{MSE}} = 0.127$			

Presentation (on log-scale): (using $t^* = t_{.975}(30) = 2.042$)

	Nitrogen			Genotype	
statistic	0	28	160	Mutant	Wild
$\bar{X}_{i\cdot}$ or $\bar{X}_{\cdot j}$	-1.040	-1.223	-1.431	-1.563	-0.900
95% CI	$\pm t^* s/\sqrt{12} = \pm 0.075$			$\pm t^* s/\sqrt{18} = \pm 0.061$	
$\text{LSD}_{0.95}$	$t^* s\sqrt{2/12} = 0.106$			N/A	

Conclusions: no evidence of interaction (on log-scale!),
clear difference between genotypes (Wild > Mutant),
clear differences between nitrogen levels: decreasing with nitrogen.

Exercises 13.3 and 13.4

Exercise 13.3: (response, factors, number of replications, I, J, N)

(a) response = number of hours of sleep "on a typical night",
factors: smoking categories ($I = 3$), gender ($J = 2$),
$n_{ij} = 120$, and $N = 720$,

(b) response = strength of concrete specimens,
factors: mixtures ($I = 4$), cycles of freezing and thawing ($J = 3$),
$n_{ij} = 2$, and $N = 24$,

(c) response = score on final exam,
factors: teaching methods ($I = 3$), student's subject of study ($J = 2$),
$n_{ij} = 7$ and $N = 42$.

Exercise 13.4: (sources and degrees of freedom)

(a) smoking categories (DF = 2), gender (DF = 1), interaction (DF = 2), error (DFE = 714) and total (DFT = 719),

(b) mixtures (DF = 3), cycles (DF = 2), interaction (DF = 6), error (DFE = 12) and total (DFT = 23),

(c) teaching methods (DF = 2), study subject (DF = 1), interaction (DF = 2), error (DFE = 36) and total (DFT = 41).

Multiple Linear Regression Model

Data (such as CSdata) with one outcome Y and multiple predictors $x_1, x_2, \ldots$,

- x_j can be quantitative or categorical, but here we discuss only quantitative predictors,
- x_j can be explanatory or response, but response variables are considered as fixed for the modelling.

Purpose: use x-variables to predict the outcome (say GPA), hoping that prediction will be "valid" (i.e., meaningful) for a wider population (of students).

Alternative purpose: examine "effect" of x-variables on GPA (sign, strength, significance of effects), but recall that inferring causal effects from observational data is always difficult...

Multiple linear regression model[15] (with two predictors for a start):

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad \text{or} \\ \text{GPA}_i &= \beta_0 + \beta_1 \text{HSM}_i + \beta_2 \text{HSS}_i + \varepsilon_i, \end{aligned}$$

where the errors $\varepsilon_1, \ldots, \varepsilon_{224}$ are i.i.d. and $\sim \text{N}(0, \sigma^2)$,

- same setup as for simple linear regression, but more predictors.

[15] Linear, because parameters are only added and multiplied by constants, not necessarily because all predictors enter in the form $\beta \times x$.

Model Assumptions and Interpretations

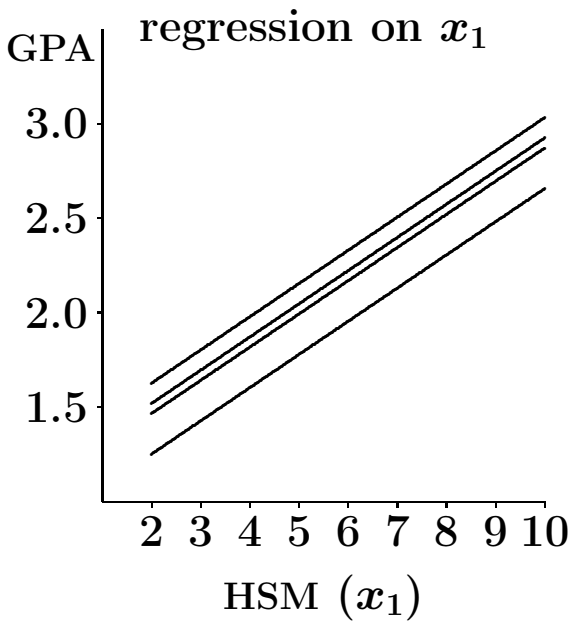
Model assumptions:

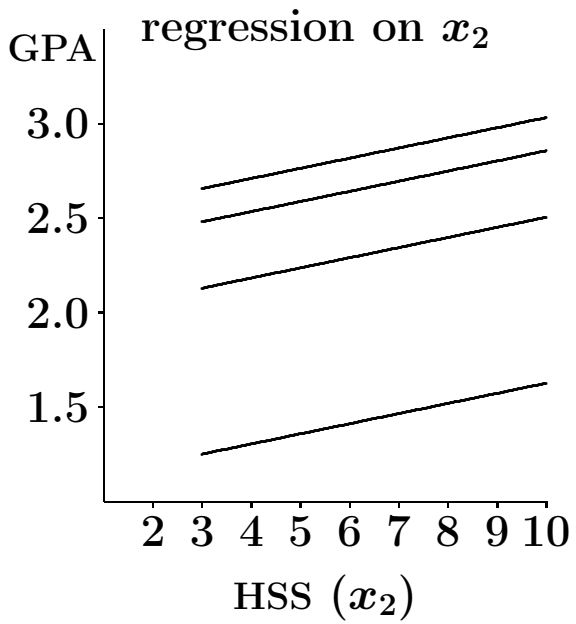
- independence, normality, variance homogeneity of ε_i's,
- linear relation:

$$\mathrm{E}(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2},$$
$$\hat{Y} = 0.740 + 0.176\,\mathrm{HSM} + 0.054\,\mathrm{HSS},$$

  - intercept β_0 (0.740) $\sim$ value for $x_1 = 0$ and $x_2 = 0$,
  - assumes a linear "effect" of x_1 on Y (for fixed x_2),
  - assumes a linear "effect" of x_2 on Y (for fixed x_1),
  - assumes additive "effects" of x_1 and x_2 (parallel lines $\rightarrow$ graph).

Fitted graphs of separate regressions (with other variable fixed at values: min, Q1, median, and Q3/max):





Multiple Linear Regression Analysis

Methods almost the same (as in simple linear regression):

- estimation by least squares method (minimizing squared deviations between observed and predicted values),
- test of simple hypothesis H_0: $\beta_j = b$ (for b known) in the "usual way" (using estimate, SE, reference t-distribution (DFE)),
- DFE $= n - p - 1$, where n=# observations, p=# x-variables in model,
- confidence and prediction intervals in the "usual way" (using estimate, SE, t^*; for predictions also new values of all x's),
- analysis of variance (ANOVA) table:
  - F-test is for hypothesis H_0: all $\beta_j = 0$ (except β_0), against alternative H_a: some $\beta_j \neq 0$ (not necessarily all) — now different than t-tests for individual β's,
  - row for Regression (Model) gives variation (SSM) and degrees of freedom (DFM) accounted for by model,
  - R^2 = SSM/SST $\sim$ proportion of variance explained by model, or squared correlation between observed and fitted values,
  - s^2 = MSE and DFE still in row for Residual (Error),
- model checking using the residuals: plot now also against each x-variable to assess linearity.

Interpretation of Regression Coefficients

Each regression coefficient (i.e., the β_j for x_j) must be viewed and interpreted in the presence of other predictors — and usually changes if the model changes!

- illustration by data example:
  - $\hat{\beta}_2 = 0.054\ (.034)$ for HSS in model with HSM,
  - $\hat{\beta}_1 = 0.151\ (.029)$ for HSS in model without HSM, i.e. in simple linear regression (sometimes called the crude or univariate estimate for HSS),

  $\Rightarrow$ size (possibly also the sign) of effects, SE's and significance may change,
- proper interpretation for β_2:
  - $\sim$ "effect" of x_2, when x_1 has been accounted for,
  - $\sim$ additional "effect" of adding x_2 to model with x_1.

But when and why does the "effect" of one predictor depend on the others (being included or not)?

- it happens (strongly)[16] when the x-variables are (strongly) associated/dependent; this we can assess by their correlations, e.g. in data example: $\text{Corr}(x_1, x_2) = 0.58$,
- it happens because dependent x's explain some of the same variation in Y,
- not the same as interaction, because the model is still assumed additive.

[16] The term collinearity is used for a situation with strong dependence among x-variables; not discussed further here.

Model Building / Variable Selection

Given an outcome $\boldsymbol{Y}$ and a set of predictors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$: how to find a good model? — we usually only want to include the "important" predictors. . . ,

- not necessarily easy to find "best" or even a good model,
- can rarely be done automatically (by some rules or procedures),
- basically, an exploratory process,
- interest is in the most succinct and parsimonious model,[17]
- but some guidelines exist:
  - start with simple associations between $\boldsymbol{Y}$ and each of the predictors (shows which of the predictors are important on their own),
  - compute also correlations among the $\boldsymbol{x}$'s (to see if any of them are strongly correlated, thereby likely to cause trouble),
  - when no. of predictors $\ll$ no. of observations: analysis should start with "full model" including all (important) predictors (and possibly also interactions),
  - from a satisfactory "full model", remove in a stepwise manner non-significant predictors to achieve a final model with only significant predictors.
  - model checking should be based on the full model (possibly repeated for final model).

[17] One might think this to be the model with highest R^2, however adding variables always $\Rightarrow$ more variation explained and higher R^2.

Summary of Analysis for CSDATA Example

- correlations with GPA: range $0.44 - 0.11$ (HSM − SATV), all significant except last one ($P = 0.09$),
- correlations between x's: moderate correlations ($0.45 - 0.58$) among HS-variables and among SAT-variables, less between HS-variables and SAT-variables,
- simple t-test for comparison of GPA between sexes: non-significant,
- full model:
  - HSM strongly significant, all others non-significant,
  - 14 standardized residuals outside $(-2, 2)$ (≈ 11 expected),
  - no single, extreme residuals (most extreme: -3.05),
  - residual plots look satisfactory (not great),
- table of model reductions:

Model	F	P	R^2	s	reduction t	reduction P
full: $x_1 - x_6$	9.72	<0.001	21.2%	0.701	–	–
$x_1 - x_5$	11.7	<0.001	21.1%	0.700	0.29	0.77
$x_1 - x_4$	14.5	<0.001	21.0%	0.699	-0.69	0.49
x_1, x_3, x_4	19.1	<0.001	20.7%	0.699	0.88	0.38
x_1, x_3	27.9	<0.001	20.2%	0.700	1.22	0.22
x_1	52.3	<0.001	19.1%	0.703	1.75	0.08

- final model with predictors HSM and HSE; estimated equation:

$$\widehat{\text{GPA}} = 0.624 + 0.183 \cdot \text{HSM} + 0.0607 \cdot \text{HSE}.$$

Model Checking for ANOVA and Multiple Regression

Model assumptions:
same as for previous models (independence, normality, variance homogeneity, mean relation $\sim$ linearity, equal means within groups, additivity...).

ANOVA models with multiple factors and multiple regression models:

- often few replications $\Rightarrow$ difficult to check model assumptions separately for each group,
- use instead residuals ("observed $-$ expected") to check model assumptions, similar to previous models:
  - variance homogeneity:
    plot standardized residuals against model's fitted/expected values and look for unequal variances across range of fittes values,[18]
  - normal distribution:
    normal probability plot of standardized residuals,
  - outliers:
    look for extreme standardized residuals (with same rules as previously),
  - other model violations:
    plot standardized residuals against data order (if applicable) or other variables.

[18] In a multifactorial ANOVA with replication, the one-way ANOVA tools: i) max/min rule, ii) variance test, still apply, when groups are defined by combinations of all factors.

Overview One-Way & Two-Way ANOVA

- Data description:
  - statistics: group[19] means and standard deviations,
  - graphs: box-plot for groups[19], interaction plot (two-way),
- Statistical model:

$$\begin{aligned} \text{one-way}: \quad X_{ij} &= \mu + \alpha_i + \varepsilon_{ij}, \\ \text{two-way (replication)}: \quad X_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \\ \text{two-way (no replication)}: \quad X_{ij} &= \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \end{aligned}$$

- Model checks — residuals plots and/or
  - equal variance: i) max/min rule, ii) variance test,
  - normality: normal plots/tests,[20]
- Statistical analysis:
  - estimation of pooled (or error) standard deviation,
  - hypothesis testing of overall effects → ANOVA table,
  - (two-way with replication): interaction significant/"substantial"?:
    - yes: interaction plot, one-way ANOVA for groups[19],
    - no: row and column factors assessed separately,
  - pairwise comparisons between group means for significant effects (based on CI, LSD or t-tests): unadjusted or adjusted to simultaneous error level of 0.05,
- Presentation: group means with SE or CI + significance indications.

[19] In a two-way design with replication, groups refer to the combined groups formed by row and column factors.
[20] With ample replication: observations within groups[19]; With limited/no replication: (standardized) residuals across all groups.

Notes on Home Assignment III

Main struggle: statistical models and hypotheses![20]

- response variables: dry feed and exercise answers,
- explanatory variable: cat's group (because *not* a simple random sample from any population of cats; it's purposely selected samples with the number of animals determined by experimentator/design (ratio of cases and controls is 2:1) — a *case-control study*, e.g. Supplementary Exercise 9.62),

1) model (for each cat group) is a single multinomial (type II) with 4 categories, and hypothesis is *independence* (or no association),

2) model is 2 independent multinomials (type I) with 4 categories each, and hypothesis is *homogeneity* (or same probabilities for the categories in the 2 samples),

3) model (for each variable: dry feed or exercise) is 2 independent binomials (type I), and hypothesis is *homogeneity* (or comparison of 2 independent proportions)
   - cat's disease status is not! an outcome,
   - estimation of (disease) incidence, relative risk or risk differences is not possible,
   - yes, we can make statements about *factor* $\Rightarrow$ *disease*, but this is really epi-theory (and not expected here beyond intuitive interpretations).

[20] Even if the (X^2) test statistic remains the same, the model determines how we most meaningfully present and interpret our results.

Further observations:

- most of you used Fisher's exact test (because of low expected values), and you managed well,
  - applies to both models (I and II),
  - is available also for larger tables than 2×2 (e.g. in Stata),
  - recommend to use as *supplementary* to X^2-statistic, but only when needed,
  - the most critical effect of low expected values violating the X^2-test guidelines is too small P-values, so clear non-significance should not be affected,
- summarizing distributions/estimates caused some trouble:
  - they follow the logic of models and testing,[21]
  - one particular bad idea popped up repeatedly: to compare p and $1-p$ (e.g., prob. of high and low exercise levels),
    - the estimates are absolutely not independent (e.g., $\frac{31}{50}$ and $\frac{19}{50} = 1 - \frac{31}{50}$), and two-sample inference (including to compare CIs) is nonsense,
    - in order to assess which is larger (if of interest!), we should look at $H_0 : p = 0.5$, e.g. by inspecting of the CI,
  - also, what to do after a significant test result:
    - if you have established a difference of sorts, illustrate that difference by the relevant estimates,
    - if you have established dependence in a 2-way table, illustrate by differences in conditional distributions.

[21] Maybe worthwhile revisiting in preparation for exam and your statistics after VHM 801...

Summary Notes

Key words and concepts for two-way ANOVA and multiple regression:

- multifactorial designs:
  - advantages over single factor designs: larger scope, allows assessment of combined effects, potentially more efficient,
  - characterizations: factors, factorial notation (e.g. $\mathbf{2 \times 2}-$design), balancedness, completeness, replication,
  - interaction: non-parallel curves, non-additive effects, effect of one factor depends on another factor; opposite of additivity,
- multiple regression:
  - interpretation (and values) of regression coefficients depend on other predictors in model, association between predictors,
  - additivity assumption for multiple predictors (most commonly),
  - model building/variable selection,
- analysis: ANOVA table, $\boldsymbol{F}$-tests for different hypotheses: testing interaction first, interaction plot, model checking by residuals, post-ANOVA analysis using LSD-values and pairwise comparisons (possibly with Bonferroni adjustments), $\boldsymbol{t}$-tests for individual regression coefficients.