Index of 14-L

(new slides only)

Practical Information

Major news:

- last home assignment returned tomorrow,
- you need to notify me about your choice on midterm (by Monday),
- last lab review later today (1-2pm; usual room), and last lab tomorrow afternoon (1-4pm; computer lab).

Topics today:

- Exam: TUESDAY 13/12, 9AM-12PM, AVC 278N,
  - exam practical remarks,
  - exam assignments (types, calculations),
- Evaluation:
  - I'll leave at 10:40 to let you do the official evaluation,
  - another evaluation form tomorrow (mine),
- Review topics — next slide,
- Sample problems — pick from these:
  - final 2009: 1 (var. inference), 2 (regression), 3 (ANOVA),
  - final 2012: 1 (1-sample), 2 (two-way tables),
  - final 2013: 1 (midterm type question),
  - final 2014: 1 (descriptive stats/two-sample inference), 2 (two-way table/proportion data).
- Your questions. . .

Topics for Lecture

Review topics: pick from these!

- outline of statistical analysis (partly new slides: 14L–6/7),
- use of graphs (new slide: 14L–8),
- model choice (new slide: 14L–9),
- general statistical tests (new slides: 14L–10/11),
- degrees of freedom (new slide: 14L–12),
- overview 1-way & 2-way ANOVA (new slide: 14L–13),
- how to find P-values and percentiles ("new" slide: 6L–13),
- use of percentiles ("new" slide: 11L–15),
- interpretation of P-values ("new" slide: 13L–16),
- completely randomized design and block design (2L–9/10),
- binomial setting (4L–10),
- one- or two-sided (6L–6),
- testing by confidence interval (6L–7),
- inference for proportions — overview/single/two (7L–11,12/13/15),
- nonparametric (distribution-free) methods (8L–3)
- sign test (8L–4),
- sample size based on estimation accuracy/power (8L–15/17),
- errors of type I–II and power (8L–16),
- two-way tables: models, estimation and hypotheses (9L–8/9),
- after the ANOVA table: LSD & Bonferroni (10L–12/13),
- use of residuals for model check (12L–2/3; 13L–13).

Exam Practical Remarks

Start time and no. questions depends on your midterm choice:

* use midterm: start at 10am, 2 questions — 35%,
* drop midterm: start at 9am, 3 questions — 50%.

All aids (books and notes and calculators) are allowed, — except a computer or computer-like device (tablet or smartphone).

The assignments have equal weight, unless specified otherwise — use your time reasonably! (no extra time!)

Some hints and advices: (to use or not...)

- layout: essential requirements are
  * readability,
  * clear distinction between what is *in* the solution and what is not,

  — don't write first a draft and then a final version,
- conclusions should be part of all analyses,
- statistical model should be part of all data analysis,
- explicit calculations may prevent loss of points due to typing errors (or the like),
- errors: if you realize an error and do not have time to correct it: write what is wrong, what should have been done and how the error would affect the result.

Exam Assignments/Questions

3 assignments — possible types:

- choice of statistical model and analysis —
  - carry out analysis when calculations manageable (see below),
  - or base analysis on Minitab print + extra calculations,
  - or outline analysis if calculations not manageable,
- probability calculations manageable (see below),
- multiple choice (one or several correct answers).

Calculations by hand (calculator):

- manageable without data entry into calculator,
- no large summations,
- examples of possible calculations:
  - simple probabilities (e.g., $1-p$, $(1-p)^n$, simple binomial),
  - standardization in normal distribution,
  - normal approximation for binomial,
  - t-test and CIs (given suitable estimates),
  - backtransformation of estimates to original scale,
- examples of too complex calculations:
  - statistical analysis of normal models without calculation help (given statistics or computer output),
  - probability plots, residual plots,
  - X^2-tests and rank-based nonparametric tests,
  - power calculations.

Typical Formulas for Exam

List of formulas (non-exhaustive!) of relevance for exam:

1L: suspected outlier rule,

3L: probability rules (e.g.: addition, multiplication); means and variances for random variables,

4L: normal distrib.: calculation of probabilities and percentiles (incl. standardization, z-score);
binomial distrib.: calc. of simple prob., mean, stand.dev.,

5L: standard error (for mean),

6L: t-test and CI for 1-sample normal,

7L: t-test and CI for 2-sample normal; CI and test for 1 and 2 proportion(s),

8L: sign test; sample size for precision (1-sample normal/proportion),

9L: expected value (cell) for X^2-test,

10L: equal variance guideline; LSD-value; ANOVA table; t-test and CI for mean/difference between means,

11L: t-test and CI for regression parameters; ANOVA table; prediction,

12L: t-test for correlation,

13L: same as for 10L (one-way ANOVA).

Outline of Statistical Analysis

Data description:

- descriptive statistics: plots, tables, simple statistics,
- purpose:
  - ∗ overview of the data,
  - ∗ detect errors / "different" observations (outliers)
  - ∗ focus attention on what's relevant,

Statistical model: we formulate statistical models containing theoretical distributions and unknown parameters, in order to

- make clear the assumptions (and utilize them),
- let parameters (fixed, unknown numbers describing a population) represent issues of interest,

"All models are approximations, but some are useful" (Box).

Estimation: our information about the parameters from the observed data is summarized in statistics or estimates:

- quantities calculated from the data to give possible parameter values,
- aim of statistical methodology: obtain estimates as close to true values as possible (though *never* equal to true values),

- all estimates should be accompanied by a measure of uncertainty, such as standard error or confidence interval.

Model check:

- comparison of observed distribution and assumed theoretical distribution (using estimated parameters),
- methods: graphical (plots) or numerical (tests),
- if model unsatisfactory, *start over with new model*[1],

Hypothesis testing:

- formulate, using model parameters, null hypothesis H_0 (model simplification) and alternative hypothesis H_a,
- the test statistic and associated P-value summarize our confidence *against null hypothesis*, which we may *reject* (low P) or *not reject* (high P).

Final Model:

- simplest possible after model reduction,
- if necessary, re-estimate model parameters (+ CI).

Conclusion / Presentation:

- summary of test results,
- illustrations of implications of the final model, e.g. prediction; also, presentation of estimates (with SE or CI).

[1] For less serious violations of model assumptions, it may alternatively be reasonable to *proceed with caution*.

Graphs in Statistical Analysis

Graphs have different purposes:

- Descriptive: show the shape of the distribution in a dataset,
  - ∗ dotplot and stemplot (raw data),
  - ∗ histogram (grouped raw data),
  - ∗ boxplot (schematic for descriptive statistics),
  - ∗ scatterplot (raw data, 2 variables),

  note: small datasets best illustrated by raw data plots, and more data ⇒ more schematic/grouped plots,

- Model check: graphical assessment of one or more model assumptions,
  - ∗ normal probability plots (check normal distribution within groups/samples),
  - ∗ residual plots (check different assumptions in normal models),

- Presentation: graphical display of estimates (possibly with measure of uncertainty) from (complex) analysis,
  - ∗ group mean plots with error bars, based on model standard deviation ($\sqrt{\text{MSE}}$), (ANOVA models),
  - ∗ fitted line plots (regression).

Choice of Statistical Model

Some useful questions to ask about the data:

- purpose of study?
- what is the observational unit/experimental unit/subject?
- response/outcome[2] or explanatory/predictor variable?
- continuous or categorical (including binary) variable?
- which variables/groupings/classifications should enter into the model? — some examples:
  - a single sample (normal, binomial),
  - two independent samples (normal, binomial, multinomial),
  - several independent samples (one-way ANOVA, two-way table of counts),
  - paired observations ⇒ single sample for differences,
  - two-way classification (two-way table of counts),
- continuous variable (explanatory or response) to be used for prediction of another variable? (regression)
- two continuous response variables with no wish to predict one from the other? (correlation),
- transformation? (to achieve normal distribution, homogeneity of variance, linear relation).

[2] What defines a response variable is that it has (truly) random variation.

Multi-Purpose Statistical Tests

Some statistical tests are specific to a single situation/use, e.g. tests for normality and rank-based tests. However, tests that bear the name of a probability distribution usually have multiple uses:

Name	Use(s) in course	Instances/Versions
z-test	normal distrib. inference with known σ binomial distrib. inference	1-sample z-test 2-sample[1] z-test (barely covered) z-test for 1 and 2 proportions
t-test ("Student" t)	normal distrib. inference with unknown σ	1-sample and 2-sample[1] t-tests t-test for regression parameters t-test for correlation coefficient t-test for (contrasts and) pairwise comparisons
χ^2-test (chi-square)	inference for counts rank-based tests	2-way table tests for homogeneity and independence Kruskal-Wallis test
F-test	effects in normal distrib. models	linear regression (slope) factorial effects (ANOVA)

[1] 2 independent samples

Relations between tests:

- t-test (df) with very large df $\approx$ z-test,
- z-test squared $\sim \chi^2(\text{df}=1)$-test,
- t-test (df) squared $\sim F(\text{df}_1=1, \text{df}_2=\text{df})$-test.

General Principles of Tests and CIs

- Data $X_1, \ldots, X_n$,
- Statistical model containing a parameter μ, typically a mean parameter.
- Estimate $\hat{\mu}$ for μ, based on $X_1, \ldots, X_n$.
- Standard error $\mathrm{SE}_{\hat{\mu}}$, either
  - ∗ estimated from the data, or
  - ∗ known value (rarely realistic in practice),

  Note: in normal models we often have
  $$\mathrm{Var}(\hat{\mu}) = A\,\sigma^2 \quad \text{and} \quad \mathrm{SE}(\hat{\mu}) = \sqrt{A}\,\sigma,$$
  where σ is standard deviation in model, and A is a constant determined by the form of $\hat{\mu}$,
- Reference distribution of $\frac{\hat{\mu}-\mu}{\mathrm{SE}(\hat{\mu})} \sim$ percentiles t_p,
  Note: in normal models with estimated $\mathrm{SE}(\hat{\mu})$ the reference distribution is usually a t-distribution,
- Confidence interval $(1-\alpha)$ for μ : $\hat{\mu} \pm t_{1-\alpha/2}\,\mathrm{SE}(\hat{\mu})$,
- Test of H_0: $\mu = \mu_0$ against H_a: $\mu \neq \mu_0$,
  $$\text{test statistic} \qquad t = \frac{\hat{\mu} - \mu_0}{\mathrm{SE}(\hat{\mu})},$$
  $$P\text{-value} \qquad P = 2 \times \mathrm{P}(t \geq |t_{\mathrm{obs}}|),$$
  where $t \sim$ the reference distribution.

Degrees of Freedom

What are those "degrees of freedom" really?:

- values to distinguish between different distributions of the same type (distribution types: t, χ^2, F),
- always positive numbers: 1,2,3,...,
- (technical) can always be interpreted as differences in number of (free) parameters between two models: a comprehensive model and a reduced model,
- in normal models, a large DF for the estimate of σ corresponds to good precision (many observations),
- specific formulas exist for all standard situations:

Model	Data	Purpose	DF
single sample	$X_1, \ldots, X_n$	conf. int. t-test	$n-1$
two samples (same σ)	$X_1, \ldots, X_{n_1}$ $Y_1, \ldots, Y_{n_2}$	conf. int. t-test	n_1+n_2-2
1-way ANOVA	X_{ij}, $i = 1, \ldots, I$ $j = 1, \ldots, n_i$	conf. int. t/F-test	error: $\sum_i n_i - I$ groups: $I-1$
2-way ANOVA	X_{ijk}, $i = 1, \ldots, I$ $j = 1, \ldots, J$ $k = 1, \ldots, n_{ij}$	conf. int. t/F-test	error: $\sum_{ij} n_{ij} - IJ$ groups: $I-1$ & $J-1$ interac: $(I-1)(J-1)$
linear regression	(Y_i, x_i) $i = 1, \ldots, n$	conf. int. t-test	$n-2$
2-way table	N_{ij}, $i = 1, \ldots, I$ $j = 1, \ldots, J$	chi-square test	$(I-1)(J-1)$

Overview 1-Way & 2-Way ANOVA

- data description:
  - statistics: group[3] means + standard deviations,
  - graphs: box-plot for groups[3], interaction plot (2-way),
- statistical model:

$$\begin{aligned} \text{1-way}: \quad X_{ij} &= \mu + \alpha_i + \varepsilon_{ij}, \\ \text{2-way (repl.)}: \quad X_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \\ \text{2-way (no repl.)}: \quad X_{ij} &= \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \end{aligned}$$

- model checks — residuals plots and/or
  - equal variance: i) max/min rule, ii) variance test,
  - normality: normal plots/tests,[4]
- statistical analysis:
  - estimation of pooled (or error) standard deviation,
  - hypothesis testing of overall effects → ANOVA table,
  - (2-way repl.): interaction significant/"substantial"?:
    - *yes*: interaction plot, 1-way ANOVA for groups[3],
    - *no*: row and column factors assessed separately,
  - pairwise comparisons between group means for significant effects (based on CI, LSD or t-tests): adjusted/unadjusted to simultaneous error level of 0.05,
- presentations: group means with SE or CI + sign. indic.

[3] In a 2-way design with replication, groups refer to the combined groups formed by row and column factors.

[4] With ample replication: observations within groups[3]; With limited/no replication: (standardized) residuals across all groups.