

Index of 7-L

Page	Title
1	Practical information
2	Outline of statistical analysis (revisited)
3	Inference for non-normal data
4	2 paired samples
5	2 paired samples: Visual receptive field
6	2 independent samples – Introduction
7	Exercise 7.40
8	2 independent samples – Equal variances
9	2 independent samples – General
10	2 independent samples: Parasite data
11	Inference for proportions – Overview
12	Choice of methods for proportions
13	Inference for 1 proportion – Details
14	1 proportion: Aphid drops
15	Inference for 2 independent proportions
16	2 independent proportions: Echinacea for common cold
17	Summary notes

PRACTICAL INFORMATION

Major news – home assignments:

- # 1: returned today: general impression good; solution posted at webpage; only minor common stumbling points (outlier for weight gain, decimals in worksheet),
- # 2: to be posted on Monday, deadline one week after,
- # 4: will you have own data by mid-November? if yes, start looking at/for them... (see project guidelines at homepage).

Today's lecture:

- lots of examples (data sets!),
- inference for one and two samples (continuous data):
 - i) a single sample (no assumption of known σ),¹
 - ii) two independent samples,
 - iii) two dependent (paired) samples,
- inference for one and two proportions:²
 - * similar breakdown of designs as i), ii), and iii),
 - * mix of familiar z -type inference and new approaches,
- Summary Worksheets from S: Chapters 7 and 8.

¹ PSLS 3e: Chapters 17-18; S: Chapter 9 (parts); IPS 7e: Sections 7.1-2.

² PSLS 3e: Chapters 19-20; S: Chapters 8-9 (parts); IPS 7e: Chapter 8.

OUTLINE OF STATISTICAL ANALYSIS (REVISITED)

- Data description,
- Statistical model,
- Estimation of model's unknown parameter(s),
 - * incl. confidence intervals and/or standard errors,³
- 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*,
- Hypothesis testing:
 - * formulate null hypothesis H_0 (model simplification) and alternative hypothesis H_a ,
 - * test statistic and associated P -value summarize our confidence *against null hypothesis*, which we may *reject* (low P) or *not reject* (high P),
- Conclusion / Presentation:
 - * summary of test results,
 - * illustrations of the implications of the final model, e.g. prediction.

³ Recall, that the standard error (SE) is the standard deviation in the distribution of the estimate, and thus an indication of the estimate's precision.

INFERENCE FOR NON-NORMAL DATA

If data show strong / moderate deviations from normality:

- remove outliers (if any), and see if it helps,
- try to transform the data, and see if situation better for transformed data,
 - * many different transformations exist, but log and square-root are most common for right-skewed data,
 - * results from transformed scale analysis should always be backtransformed to original scale:
 - backtransformed means \sim medians in original data,⁴
 - for CI's: backtransform both endpoints,
- statistical methods with no distributional assumptions
 - nonparametric statistics (next lecture),
- some procedures based on normal distribution are robust (or resistant), that is, work reasonably well even if assumptions are (mildly) violated (saved by the normality of \bar{X} in the CLT!):
 - * difficult to know exactly what is okay and when ...
 - guidelines in PSLS/IPS⁵ for t -distribution⁶ procedures:
 - $n < 15$: only if data close to normal (okay!),
 - $15 \leq n < 40$: ok unless strong skewness or outliers,
 - $40 \leq n$: also ok for clearly skewed distributions, but beware of strong outliers.

⁴ It is more difficult to get means and SEs on original scale.

⁵ Discussion in S is less detailed/satisfactory: assume normality or $n > 30$, in my view a very debatable guideline!

⁶ If σ is known, the inference is even less affected by non-normality, because it is the procedures involving s^2 that rely most strongly on the normality assumption.

2 PAIRED SAMPLES

Paired (matched, correlated) samples/observations:

- Data: $(X_1, Y_1) \dots, (X_n, Y_n)$ independent observation pairs:
 - * typical examples of pairs:
 - same individual: left–right, before–after,
 - different individuals:
twins, related or similar individuals,
 - * in experimental design terminology:
 - pairs \sim blocks (of size 2),
 - observations within pairs \sim treatments,
 - * purpose of pairs: reduce variability and impact of other (lurking) factors,
- Model and Analysis:
 - * usually work with differences: $D_i = Y_i - X_i$,
(ratios Y_i/X_i or other functions also possible),
 - * assume D_1, \dots, D_n sample from distribution (μ_D, σ_D) ,
where
$$\mu_D = ED_i = EY_i - EX_i,$$
 - * hypothesis H_0 : $\mu_D = 0$
 \sim no difference between (means of) X 's and Y 's,
 - * all methods for single sample inference apply!

2 PAIRED SAMPLES: VISUAL RECEPTIVE FIELD

- Data: Neuron's activity (# spikes/sec) at 9 recordings of both Spontaneous activity (SA) and Response (R),

Recording (i)	SA (X_i)	R (Y_i)	Difference (D_i)
1	2.5	16.7	14.2
2	7.5	20.0	12.5
...
9	17.5	10.0	-7.5

- Model: one sample (i.i.d.) of differences D_1, \dots, D_9 assumed to follow $N(\mu_D, \sigma_D)$, where $\mu_D = \mu_Y - \mu_X$ is the parameter of principal interest,

- Estimation: $\hat{\mu}_D = \bar{D} = 16.87$, $\hat{\sigma}_D = s_D = 16.40$,

- 95% Confidence interval for μ_D :

$$\bar{D} \pm t^* s_D / \sqrt{n} = 16.87 \pm 2.306 \cdot 16.40 / \sqrt{9} = 16.87 \pm 12.61,$$

- Test of $H_0: \mu_D = 0$ (corresponding to $\mu_Y = \mu_X$) against alternative $H_a: \mu_D > 0$ (corresp. to $\mu_Y > \mu_X$):

* test statistic: $t = \frac{\bar{D} - 0}{s_D / \sqrt{n}} = \frac{16.87}{12.61 / \sqrt{9}} = 3.09$,

- * P -value from t distribution with $df=8$:

$$P = P(t(8) > 3.09) = 0.007,^7$$

- * conclusion: clearly significant difference between neural activity at SA and R, and higher R activity.

⁷ The $t(8)$ -distribution table has $t_{.99}(8) = 2.896$ and $t_{.995}(8) = 3.355$, from which we would conclude that: $0.005 < P < 0.01$.

2 INDEPENDENT SAMPLES – INTRODUCTION

Two-sample setting:

- Data:

$X_1, \dots, X_{n_1} \sim$ first sample, of size n_1 ,

$Y_1, \dots, Y_{n_2} \sim$ second sample, of size n_2 .

- Model: all observations independent, and the X 's and Y 's are samples from separate distributions,
- typical example: treatment and control groups, e.g. study on parasite burdens in Lithuanian calves,
- how to distinguish from paired samples?
 - * not necessarily the same number of observations (possibly $n_1 \neq n_2$),
 - * no relation between X_1 and Y_1 , X_2 and Y_2 , etc.
 - * the X 's are interchangeable (“replications”), and the same for the Y 's.

Overview of inference based on normal distribution:

- assumptions: normal distributions $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$ for the two samples, with all parameters unknown,
- slightly different procedures depending on whether
 - (1) $\sigma_1 \neq \sigma_2$ (general situation⁸).
 - (2) $\sigma_1 = \sigma_2$ (simplest; often unrealistic assumption),⁹

⁸ Without a specific assumption about the σ 's: we could have $\sigma_1 = \sigma_2$ or $\sigma_1 \neq \sigma_2$.

⁹ Not part of VHM 801 syllabus; PLSLS/S avoid this (“pooled variance”) method.

EXERCISE 7.40

Identify statistical design as either (1) single sample, (2) matched pairs (paired sample) or (3) two independent samples:

- (a) *two independent samples*, because different groups of children and only one score from each child; the before versus after element is not part of the data collection,
- (b) *two paired samples*, because two scores are collected from each child, in random order; nor is the before versus after element part of the data collection here,
- (c) *one sample*, because only one sample (of 20 measurements) is taken,
- (d) *two independent samples*, because there is no connection between the measurements taken with the new and old method.

What about a slight variation of the design where 10 samples are taken from the specimen and each is analyzed with both the new and old method?

That would be *two paired samples*.

2 INDEPENDENT SAMPLES – EQUAL VARIANCES

- models: 1st sample: $N(\mu_1, \sigma_1)$, 2nd sample: $N(\mu_2, \sigma_2)$,
- assume $\sigma_1 = \sigma_2 = \sigma$, based on judgement¹⁰ or test¹¹,
- estimation of means:

$$\hat{\mu}_1 = \bar{X} \sim N(\mu_1, \sigma/\sqrt{n_1}), \quad \hat{\mu}_2 = \bar{Y} \sim N(\mu_2, \sigma/\sqrt{n_2})$$
- estimation of σ (from s_1 and s_2 in X - and Y -samples):

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2} \quad \text{and} \quad \hat{\sigma} = s = \sqrt{s^2},$$
 — “pooled” s^2 : a weighted average of s_1^2 and s_2^2 ,
- standard error of mean difference: $s_{\bar{X}-\bar{Y}} = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$,
- degrees of freedom: $df = (n_1-1) + (n_2-1) = n_1 + n_2 - 2$,
- confidence interval of level $(1-\alpha)$ for $\mu_1 - \mu_2$:

$$\mu_1 - \mu_2 : \bar{X} - \bar{Y} \pm t^* s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad t^* = t_{1-\alpha/2}(df)$$
- test of $H_0: \mu_1 = \mu_2$ against alternatives H_a ,
 - * test statistic: $t = (\bar{X} - \bar{Y}) / \left(s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)$,
 - * P -value from t distribution:
 - $H_a: \mu_1 \neq \mu_2$: $P = 2 \times P(t(df) > |t_{\text{obs}}|)$,
 - $H_a: \mu_1 > \mu_2$: $P = P(t(df) > t_{\text{obs}})$,
- note similarities with 1-sample procedures.

¹⁰ PSLS/IPS guideline for assuming equal standard deviations: $s_{\max}/s_{\min} \leq 2$.

¹¹ Variance tests (especially Bartlett’s test) are overly sensitive to non-normality.

2 INDEPENDENT SAMPLES – GENERAL

Similar procedure – changes in $s_{\bar{X}-\bar{Y}}$ and df:

- no assumption of $\sigma_1 = \sigma_2$: \Rightarrow more general procedure (also applicable when $\sigma_1 \approx \sigma_2$),
- estimation of means and standard deviations: separately for each sample: $\bar{X}, s_1, \bar{Y}, s_2$,
- standard error of mean difference: $s_{\bar{X}-\bar{Y}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$,
- degrees of freedom – two approaches:
 - * conservative (too low/“safe”) df: $\min(n_1 - 1, n_2 - 1)$,
 - * approximate with “terrible” formulae¹², but approximations generally considered good,
- confidence interval of level $(1 - \alpha)$ for $\mu_1 - \mu_2$:

$$\mu_1 - \mu_2 : \bar{X} - \bar{Y} \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}, \quad t^* = t_{1-\alpha/2}(\text{df})$$

- test of $H_0: \mu_1 = \mu_2$ against alternatives H_a :
 - * test statistic: $t = (\bar{X} - \bar{Y}) / \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$,
 - * P -value from t distribution:
 - $H_a: \mu_1 \neq \mu_2$: $P = 2 \times P(t(\text{df}) > |t_{\text{obs}}|)$,
 - $H_a: \mu_1 > \mu_2$: $P = P(t(\text{df}) > t_{\text{obs}})$.

¹² Minitab uses Satterthwaite method, Stata/R use Welch method; both ok to use.

2 INDEPENDENT SAMPLES: PARASITE DATA

- Data: $n_1 = 10$ and $n_2 = 9$ parasite counts of calves on infected (X) and safe (Y) pasture,
- Model: two independent samples (i.i.d.) from $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, respectively,
- Estimation: $\hat{\mu}_1 = \bar{X} = 51.2$, $\hat{\sigma}_1 = s_1 = 24.0$, and $\hat{\mu}_2 = \bar{Y} = 23.8$, $\hat{\sigma}_2 = s_2 = 17.6$,
- some difference in estimated standard deviations, so we are *not* going to assume that $\sigma_1 = \sigma_2$,

- Confidence interval with confidence level 95%:

$$\begin{aligned} \mu_1 - \mu_2 &: \bar{X} - \bar{Y} \pm t^* \sqrt{s_1^2/n_1 + s_2^2/n_2}, \\ &= 27.4 \pm 2.12 \sqrt{24.0^2/10 + 17.6^2/9} = 27.4 \pm 20.3, \end{aligned}$$

where $t^* = t_{.975}(16) = 2.12$ (df computed by software),

- Test of $H_0: \mu_1 = \mu_2$ against alternative $H_a: \mu_1 \neq \mu_2$:

$$* \text{ test statistic: } t = \frac{\bar{X} - \bar{Y}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} = 2.86,$$

- * approximate P -value from t distribution with $df = 16$:

$$P = 2 \times P(t(df) > 2.86) = 0.011.$$

- * conclusion: significant difference between parasite burdens on infected and safe pastures; that is, parasite levels are lower on safe pasture.

INFERENCE FOR PROPORTIONS – OVERVIEW

Basic assumption: binomial setting

\Rightarrow binomial distrib. $B(n, p)$ for number of “successes”, where n = number of “trials” and p = probability of success.

Same 3 fundamental designs:

- *one sample* — one binomial distribution, parameter of interest is p ,
- *two independent samples* — two binomial distributions, parameter of interest *here* is $p_1 - p_2$,
- *two paired samples* — not in textbooks, but discussed as a sign test (Session 8).¹³

Statistical inference:

- estimation: always use sample proportions,
- several approaches for confidence intervals:
 - * classical¹⁴ approx. of $B(n, p)$ by $N(np, \sqrt{np(1-p)})$,
 - * “plus four” (Wilson) adjustment of classical approach,
 - * “exact” based on binomial distrib. (only 1 sample),
- several approaches for tests:
 - * classical¹⁴ z -test approximation,
 - * exact based on binomial or other distributions.

¹³ Two paired samples often lead to McNemar’s test or to κ (kappa)-calculations.

¹⁴ “Classical” refers to methods based on the standard normal (z) distribution.

CHOICE OF METHOD FOR PROPORTIONS

Issue: several methods exist for both CI and test across all designs \Rightarrow we need guidelines to choose a good method.

Principle: choice of method should be based on data dimensions, with separate guidelines for different inferential procedures (CI and test) and for different designs.

PSLS guidelines:¹⁵

Design	Method	Conditions (all must be met)
1-sample (n, \hat{p})	classical CI	$n\hat{p} \geq 15; n(1-\hat{p}) \geq 15$
	“plus four” CI	$n \geq 10$
	“exact” CI	no conditions
$H_0 : p = p_0$	z -test	$np_0 \geq 10; n(1-p_0) \geq 10$
	exact test	no conditions
2-sample indep. ($n_1, \hat{p}_1, n_2, \hat{p}_2$)	classical CI	$n_1\hat{p}_1 \geq 10; n_1(1-\hat{p}_1) \geq 10;$ $n_2\hat{p}_2 \geq 10; n_2(1-\hat{p}_2) \geq 10$
	“plus four” CI	$n_1 \geq 5; n_2 \geq 5$
$H_0 : p_1 = p_2$ (combined \hat{p})	z -test	$n_1\hat{p} \geq 5; n_1(1-\hat{p}) \geq 5;$ $n_2\hat{p} \geq 5; n_2(1-\hat{p}) \geq 5$
	exact test	no conditions

- * $n\hat{p} \sim$ no. of positives; $n(1-\hat{p}) \sim$ no. of negatives,
- * 1-sample exact procedures are based on binomial distribution; CI is conservative (not exact), test is exact,
- * 2-sample Fisher’s exact test not in textbooks (Session 9).

¹⁵ Same guidelines as in IPS 7e; coverage in S is too limited to be of practical use.

INFERENCE FOR 1 PROPORTION – DETAILS

- Data: X = number of “successes” in a binomial setting.
- Model: $X \sim B(n, p)$.
- Estimation: $\hat{p} = X/n$, $SE_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n}$.
- Confidence intervals for p with confidence level $1 - \alpha$:
 - * classical approx.: $\hat{p} \pm z^* SE_{\hat{p}}$, $z^* = z_{1-\alpha/2}$,
 - * plus four approx.: $\tilde{p} \pm z^* SE_{\tilde{p}}$ (same formula, but add 2 successes and 2 failures!; Agresti & Coull, 1998),
 - * “exact”: based on binomial distribution (software),

evaluation: exact: always conservative (too wide); classical: may be very poor (too narrow) in small samples; plus four: generally good approximation,
- Test of $H_0: p = p_0$ (where p_0 is a known value), against either one- or two-sided alternative H_a ,
 - * classical, approximate z -test:
 $z = (\hat{p} - p_0) / \sqrt{p_0(1-p_0)/n} \approx N(0,1)$ under H_0 ,
 - * exact: based on binomial distribution, e.g.,
 - $H_a: p > p_0$: $P = P(X \geq X_{\text{obs}})$,
 - $H_a: p \neq p_0$: $P = 2 \min\{P(X \geq X_{\text{obs}}), P(X \leq X_{\text{obs}})\}$,¹⁶

evaluation: exact generally preferable but almost the same in large samples.

¹⁶ Other formulae exist, but this is the simplest one; see Exercise 8.85 (solution).

1 PROPORTION: APHID DROPS

Aphid landings on their feet or back:¹⁷

- Data: 19 out of 20 aphids dropped at height 20 *cm* landed on their feet,
- Model: binomial setting $\sim B(20, p)$, where p = probability of feet landing,
- Estimation: $\hat{p} = 19/20 = 0.95$, $SE_{\hat{p}} = \sqrt{\frac{0.95 \cdot 0.05}{20}} = 0.049$,
- 95% CI for p :
 - * classical: $0.95 \pm 1.96 \times 0.049 = (0.854, 1.046)$,
 - * plus four: $0.875 \pm 1.96 \times 0.068 = (0.743, 1.007)$,¹⁸
 - * “exact” (Minitab/Stata): $(0.751, 0.999)$,
- Test of $H_0: p = 0.5$ against $H_a: p > 0.5$:
 - * classical: $z = (0.95 - 0.5) / \sqrt{\frac{0.5(1-0.5)}{20}} = 4.025$,
and $P = P(Z > 4.025) = .000028$,
 - * exact using $B(20, 0.5)$ and software/formula:
 $P = P(X \geq 19) = P(X = 19) + P(X = 20) = .000020$,
- Conclusion: clear evidence of non-random landings.

Conclusions on methodology:

- “plus four” and exact CIs similar and preferable,
- exact test preferable (but z -test not too far off).

¹⁷ Pea aphids drops were videotaped after release; Ribak et al. (2013), *Current Biology* **23**, R102-103; Example 19.6 of PSLS 3e.

¹⁸ $\tilde{p} = (19+2)/(20+4) = 0.875$, and $SE(\tilde{p}) = \sqrt{0.875(1-0.875)/24} = 0.068$.

INFERENCE FOR 2 INDEPENDENT PROPORTIONS

- Data: X and Y = number of “successes” in two independent binomial settings.
- Model: $X \sim B(n_1, p_1)$ and $Y \sim B(n_2, p_2)$, X and Y independent,
- Estimation:

$$\hat{p}_1 = X/n_1, \hat{p}_2 = Y/n_2, D = \hat{p}_1 - \hat{p}_2,$$

$$SE_D = \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}.$$

- Confidence interval for $p_1 - p_2$ with conf. level $1 - \alpha$:
 - * classical approx.: $\hat{p}_1 - \hat{p}_2 \pm z^* SE_D$, $z^* = z_{1-\alpha/2}$,
 - * plus four approx.: $\tilde{p}_1 - \tilde{p}_2 \pm z^* SE_{\tilde{D}}$ (same formula, but add 1 success and 1 failure in each sample!),

evaluation: classical: may be very poor in small samples; plus four: generally good approximation,
- Test (approximate) of $H_0: p_1 = p_2 (= p)$, against either one- or two-sided alternative H_a ,
 - * estimate common value p : $\hat{p} = (X + Y)/(n_1 + n_2)$ – total number of successes / total number of trials,
 - * “pooled” standard error under H_0 :

$$SE_{D_p} = \sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$
 - * test statistic: $z = (D - 0)/SE_{D_p} = (\hat{p}_1 - \hat{p}_2)/SE_{D_p}$,
 - * P -values from $N(0,1)$ the usual way.

2 INDEP. PROPORTIONS: ECHINACEA FOR COMMON COLD

Development of cold in Echinacea and control groups:¹⁹

- Data: after exposure, 88 out of 103 persons in control group, and 44 out of 48 in a treatment group developed a cold,

- Estimation:

$$\begin{aligned} \text{tx : } \quad \hat{p}_1 &= 44/48 = 0.917, & \text{SE}_{\hat{p}_1} &= 0.040, \\ \text{control : } \quad \hat{p}_2 &= 88/103 = 0.854, & \text{SE}_{\hat{p}_2} &= 0.035, \\ \text{diff : } \quad \hat{p}_1 - \hat{p}_2 &= 0.062, & \text{SE}_{\hat{p}_1 - \hat{p}_2} &= 0.053, \\ & \tilde{p}_1 - \tilde{p}_2 = 0.052, & \text{SE}_{\tilde{p}_1 - \tilde{p}_2} &= 0.055,^{20} \end{aligned}$$

- 95% CI for $p_1 - p_2$:

- * classical: $0.062 \pm 1.96 \times 0.053 = (-0.041, 0.166)$,

- * plus four: $0.052 \pm 1.96 \times 0.055 = (-0.056, 0.160)$,

- Test of $H_0: p_1 = p_2$ against $H_a: p_1 \neq p_2$:

- * classical: $z = (\hat{p}_1 - \hat{p}_2)/\text{SE}_{D_p} = 0.062/0.058 = 1.075$, and $P = 2 \times P(Z > 1.075) = 0.282$,

- * alternative methods → Session 9.

Conclusions:

- only little difference between confidence intervals despite violation of guideline for classical method,
- P -value so large that we can be confident there is no evidence against H_0 (despite violation of guideline); observed effect is in the *opposite direction* of what one might have hoped.

¹⁹ Experimental study on efficacy of Echinacea product against common cold; Turner et al. (2005), *New England J. Med.* **353**, 341-348.; PSLS 3e Exercise 20.3.

²⁰ Calculations: $\tilde{p}_1 = (44+1)/(48+2) = 0.90$; $\tilde{p}_2 = (88+1)/(103+2) = 0.8476$, and $\text{SE}(\tilde{p}_1 - \tilde{p}_2) = \sqrt{(0.90 \cdot (1-0.90))/50 + 0.8476 \cdot (1-0.8476)/105} = 0.055$.

SUMMARY NOTES

Key words and concepts:

- designs involving 1 and 2 samples (any distribution):
 - * 1-sample, 2 independent samples, 2 paired (dependent, correlated) samples,
 - * 2 paired samples \rightarrow 1-sample for differences,
- choice of method/assumption for 2 independent normal distribution samples:
 - * equal variances assumed: pooled variance estimate, ratio of standard deviations ≥ 2 rule,
 - * no variance assumption: df determined by software,
- proportion data modelled by binomial distributions,
- CI methods for proportions:
 - * classical (based on z -distribution), “plus four”, “exact” (1-sample only),
 - * choice between methods based on n and \hat{p} ,
- test methods for proportions:
 - * classical (based on z -distribution), exact (based on binomial or other distributions),
 - * choice between methods based on n and \hat{p} .