Introduction to Hypothesis Testing

1 Intro to Statistical Hypothesis Testing

1.1 Inference about population mean(s)

1.2 Null Hypothesis (H_0): No real difference (association, effect, etc.);
→ observed difference in samples is due to chance alone

1.3 Alternative hypothesis (H_A): H_0 & H_A must account for all possible outcomes
e.g., H_0: $\mu = \mu_0$, H_A: $\mu \neq \mu_0$

1.4 State hypotheses <u>before</u> collecting data!

1.5 Typical procedure:
1.5.1 State/clarify the research question
1.5.2 Translate the question into statistical hypotheses
1.5.3 Select a significance level (α)
1.5.4 Collect data (e.g., random sample)
1.5.5 Look at, plot data; check for errors, evaluate distributions, etc.
1.5.6 Select appropriate test
1.5.7 Calculate sample(s) mean(s), standard deviation(s), standard error(s)
1.5.8 Calculate the test statistic, e.g., t_{calc}
1.5.9 Determine probability (P-value):
If H_0 true, probability of sample mean at least as far from μ as $\overline{X}$
1.5.10 If P<a, reject H_0 and accept H_A.
Otherwise indeterminate result (neither accept nor reject H_0).
1.5.11 Answer the research question

2 The Distribution of Means

2.1 Central Limit Theorem: random samples (size n) drawn from population
→ sample means will become normal as n gets large (in practice, $n \geq 20$)

2.2 Variance of sample means ↓ as ↑n: $\sigma_{\overline{X}}^2 = \dfrac{\sigma^2}{n}$ $s_{\overline{X}}^2 = \dfrac{s^2}{n}$ $s_{\overline{X}} = \dfrac{s}{\sqrt{n}}$

$\sigma_{\overline{X}}^2$ = variance of the pop. mean

$\sigma_{\overline{X}}$ = standard error

$s_{\overline{X}}$ = sample standard error

3 Types of Errors

3.1 Type I error: incorrect rejection of true null hypothesis (Probability = α)

3.2 Type II error: failure to reject false null hypothesis (Probability = β)

3.3 Two other possibilities: (1) do not reject true null hypothesis; (2) reject false null hypothesis

3.4 Significance level = probability of type I error (= α)
Must state significance level before collecting data!

3.5 In scientific communication, restrict "significant" to statistical context;
<u>never</u> use "significant" as synonym for "important" or "substantial"

3.6 Industrial statistics: α called "producer's risk" = P{reject good ones}
β called "consumer's risk", = P{accept bad ones}

4 Hypothesis Tests Concerning the Mean – Two-Tailed

4.1 Unknown σ^2: use t-distribution (t) instead of Normal (z): $t = \dfrac{\overline{X} - \mu_0}{s_{\overline{X}}}$

4.2 Performing the t-test •

4.2.1 State null (H_0) & alternative (H_0) hypotheses: e.g., H_0: $\mu = 0$, H_A: $\mu \neq 0$

4.2.2 State significance level (α); e.g., $\alpha = 0.05$

4.2.3 Define critical region; e.g., 2-tailed test: if P($|t_{calc}|$) ≤ 0.05, then reject H_0
i.e., if $|t_{calc}| \geq t_{\alpha(2),\nu}$, then reject H_0
e.g.: one sample, 2-tailed test, w/ α=0.05, and n=25 (ν=24): $t_{\alpha(2),\nu} = t_{0.05(2),24} = 2.064$

4.2.4 Determine $\overline{X}$, $s_{\overline{X}}$; e.g., $\overline{X} = 5.0$, $s_{\overline{X}} = 2.0$

4.2.5 Calculate t_{calc}: $t = \dfrac{\overline{X} - \mu_0}{s_{\overline{X}}}$; e.g., $t = \dfrac{5.0 - 0}{2.0} = 2.5$

4.2.6 Find $t_{critical}$ (= $t_{\alpha(2),\nu}$) in t-table (Zar table B.3)

4.2.7 If $t_{calc} \geq t_{critical}$, then reject H_0; otherwise do not reject H_0
e.g.,for α=0.05, and n=25 (ν=24) t = 2.5 > 2.064 → reject H_0 , conclude that $\mu \neq 0$

4.3 Cannot test hypotheses about single observation ($\nu = n$-1 = 1-1 = 0)

4.4 Assumptions of one sample t-test:
1. data are a random sample
2. sample from pop. with normal distribution

4.5 Replication:
measurements must be truly replicated; avoid pseudoreplication

5 One-Tailed Tests

5.1 Two-tailed hypotheses: H_0: $\mu = \mu_0$, H_A: $\mu \neq \mu_0$
Difference could be positive or negative

5.2 One-tailed hypotheses: H_0: $\mu \leq \mu_0$, H_A: $\mu > \mu_0$

5.3 Critical value for one-tailed test always smaller than for two-tailed (easier to get significance)
e.g., for $\alpha = 0.05$, $Z_{\alpha(1)} = 1.645$ and $Z_{\alpha(2)} = 1.960$
<u>must declare hypotheses before examining data</u>

5.4 If $t \geq t_{\alpha(1),\nu}$ then reject H_0

6 Confidence Limits of the Mean

6.1 t-distribution: indicates fraction of all possible sample means greater (or less than) t

where $t = \dfrac{\overline{X} - \mu}{s_{\overline{X}}}$

6.2 95% of all t-values occur between $t_{\alpha(2),\nu}$ and $t_{\alpha(2),\nu}$

$$P\left[-t_{0.05(2),\upsilon} \le \frac{\overline{X} - \mu}{s_{\overline{X}}} \le t_{0.05(2),\upsilon}\right] = 0.95$$

6.3 Solve for μ:

$$P\left[\overline{X} - t_{0.05(2),\upsilon} s_{\overline{X}} \le \mu \le \overline{X} + t_{0.05(2),\upsilon} s_{\overline{X}}\right] = 0.95$$

6.4 95% Confidence limits of the mean:

Lower limit: $\overline{X} - t_{0.05(2),\upsilon} s_{\overline{X}}$

Upper limit: $\overline{X} + t_{0.05(2),\upsilon} s_{\overline{X}}$

Concise statement: $\overline{X} \pm t_{0.05(2),\upsilon} s_{\overline{X}}$

6.5 General notation: 2-tailed, with sample size *n*-1, @ significance level α:

$\overline{X} \pm t_{\alpha(2),\upsilon} s_{\overline{X}}$ → e.g., 99% confidence interval

6.6. Reporting variability about the mean

In table, figure, text, must show/state (& look for) 4 things:

(1) value of mean
(2) units of measurement
(3) sample size, n
(4) measure of variability, e.g., s, s^2, $s_{\overline{X}}$, 95% CI

7 Combining Means

Welsh, A.H. et al. (1988) The fallacy of averages. *Am. Nat*. 132(2)277-288.

7.1 In general, $\mu[f(X,Y)] \ne f[\mu(X),\mu(Y)]$; $\sigma[f(X,Y)] \ne f[\sigma(X),\sigma(Y)]$

7.2 sum of random variables:

$\mu(X+Y) = \mu(X) + \mu(Y)$ $\sigma(X+Y)^2 = \sigma(X)^2 + \sigma(Y)^2 + 2\,\mathrm{cov}(X,Y)$

7.3 product of random variables:

$\mu(XY) = \mu(X)\mu(Y) + \mathrm{cov}(X,Y)$

if X and Y are independent, $\mu(XY) = \mu(X)\mu(Y)$

7.4 ratio of random variables:

$\mu(X/Y) = \mu(X)/\mu(Y) - \mathrm{cov}(X/Y,Y)/\mu(Y)$