STA 4322/5328, Spring 2013
Testing Statistical Hypotheses

Instructor: Dr. Nikolay Bliznyuk

April 19, 2017

Motivating Examples

Goal: to decide whether observed data support the (null) hypothesis of interest, H_0. If not, reject the null hypothesis in favor of alternative hypothesis, H_A.

Examples

- flip a coin 10 times, decide if the outcomes support the null hypothesis $\mathrm{H}_0 : p = 1/2$. Here, one can take the alternative hypothesis $\mathrm{H}_\mathrm{A} : p = p_\mathrm{A} > 1/2$. E.g. if 10 successes are observed, it makes sense to reject H_0 in favor of H_A.
  Q: This event can happen even if H_0 is true. In this case, what is the probability that H_0 is rejected incorrectly.
- determine if a new drug is more effective than a placebo.
- Gender bias/discrimination in admissions to UC Berkeley: grad school.

  A hypothetical example: suppose the overall admission rate (across all colleges/schools) was 46% for male and 36% for female applicants. Does this suggest gender bias?

Berkeley Grad School Admission Example

	MALE		FEMALE	
Overall admission rate across all schools/colleges	46%		36%	
Look at admission by school /college	# of applicants	% admitted	# of applicants	% admitted
Engineering	**900**	50%	**400**	60%
Arts	**100**	10%	**600**	20%

This is known as Simpson's paradox: although the admission rate is higher for female applicants in every school/college, the overall admission rate is higher for males, because the proportions of male and female applicants to each college are not equal.

A numerical example

A light-bulb has expected lifetime of 1000 hours, but the one you bought burned down in under 50 hours. How likely is that manufacturer's claim is true? Let $X \sim \text{Exp}(\beta)$, the rv for the lifetime of the light-bulb.

$$\mathrm{H}_0 : \beta = 1 : \text{“manufacturer's claim valid”}$$
$$\mathrm{H}_\mathrm{A} : \beta = \beta_\mathrm{A} < 1 : \text{“manufacturer's claim is invalid”}$$

Ingredients of hypothesis tests

1. Null hypothesis: $\mathrm{H}_0 : \theta = \theta_0$
2. Alternative hypothesis: $\mathrm{H}_\mathrm{A} : \theta : \theta_\mathrm{A} \neq \theta_0$
   Possibilities:
   - $\mathrm{H}_\mathrm{A} : \theta = \theta_\mathrm{A} > \theta_0$
   - $\mathrm{H}_\mathrm{A} : \theta = \theta_\mathrm{A} < \theta_0$
   - $\mathrm{H}_\mathrm{A} : \theta = \theta_1 < \theta_0$ or $\theta = \theta_2 > \theta_0$.

   A hypothesis is called *simple* if it consists of a single point (parameter value). Otherwise it is a *composite* hypothesis.
3. Test statistic (TS): a statistic that is used to test H_0. (Recall: a statistic cannot depend on unknown quantities.)
4. Rejection region (RR): a set of values of TS for which H_0 is rejected.
5. Conclusion whether we reject H_0 or not. Reject H_0 if $\mathrm{TS} \in \mathrm{RR}$. Else do not reject (basically, “accept”, but not quite).

Probabilities of Type I and II errors

		Nature	
		H_0 is true	H_A is true
Statistician	Reject H_0	Type I error	No error.
	Accept /do not reject H_0	No error	Type II error.

Level of significance,

$$\begin{aligned}\alpha &= \Pr(\text{Type I error}) = \Pr(\mathrm{TS} \in \mathrm{RR} \mid \mathrm{H}_0 \text{ is true}) \\ &= \Pr(\text{reject } \mathrm{H}_0 \text{ when } \mathrm{H}_0 \text{ is true}).\end{aligned}$$

$$\begin{aligned}\beta &= \Pr(\text{Type II error}) = \Pr(\text{accept } \mathrm{H}_0 \text{ when } \mathrm{H}_A \text{ is true}) \\ &= \Pr(\mathrm{TS} \notin \mathrm{RR} \mid \mathrm{H}_A \text{ is true}).\end{aligned}$$

A numerical example

A light-bulb has expected lifetime of 1000 hours, but the one you bought burned down in under 50 hours. Let $X \sim \text{Exp}(\beta)$, the rv for the lifetime of the light-bulb.

beta=1/rate

$$\mathrm{H}_0 : \beta = 1 : \text{"manufacturer's claim valid"}$$
$$\mathrm{H}_\mathrm{A} : \beta = \beta_\mathrm{A} < 1 : \text{"manufacturer's claim is invalid"}$$

Q1: For the test statistic X, let the RR be $[0, 0.25]$. Compute the probabilities of type I and II errors.

```
pexp(0.25, rate=1) # alpha = 0.2211992
beta = Pr(X > 0.25 | X ~ Exp(beta_A));
need to know a concrete num. value of beta_A
```

Q2: What is the RR for the test statistic X if the desired probability of type I error is $\alpha = 0.05$?

```
> qexp(0.05, rate=1)
[1] 0.05129329
```

Relationships between RR, α and β

$$\boxed{\mathrm{H_0, H_A, TS, RR}} \longrightarrow \text{can find } \alpha, \beta$$

$$\boxed{\mathrm{H_0, H_A, TS}, \alpha} \longrightarrow \text{can find } \mathrm{RR}, \beta$$

$$\boxed{\mathrm{H_0, H_A, TS}, \beta} \longrightarrow \text{can find } \mathrm{RR}, \alpha$$

Neyman-Pearson testing framework:

- Specify $\mathrm{H}_0, \mathrm{H}_\mathrm{A}, \mathrm{TS}$ and α.
- Determine the RR.
- H_0 and H_A are setup in a way that the goal is to reject H_0.

Q: Why? **A**: "Do not try to fix something that is not broken." Unless the "alternative treatment" (under H_A) is significantly superior to the existing one (under H_0), do not change the existing regime.

Heuristics of finding tests

"Finding tests" means specifying $\boxed{\mathrm{H_0}, \mathrm{H_A}, \mathrm{TS}, \mathrm{RR}}$, either directly or indirectly (see the slide *Relationships between RR, α and β*).

1. If the test statistic T *tends to be bigger under* $\mathrm{H_0}$ than under $\mathrm{H_A}$, then reject $\mathrm{H_0}$ in favor of $\mathrm{H_A}$ for sufficiently small values of T, i.e. $RR \equiv [T < c_\alpha]$, where c_α depends on $\alpha = \mathrm{Pr}(\text{Type I error})$.

2. If the test statistic T tends to be smaller under $\mathrm{H_0}$ than under $\mathrm{H_A}$, reject $\mathrm{H_0}$ in favor of $\mathrm{H_A}$ for sufficiently large values of T: $RR \equiv [T > c_\alpha]$

To determine if T "tends to be" smaller or larger under $\mathrm{H_0}$ versus under $\mathrm{H_A}$, compare the cdfs/quantiles of T under the two hypotheses (see the cdf plots).

Examples: tests using a single normal rv

Let $X \sim Normal(\mu, 1)$. Test $\mathrm{H}_0 : \mu = 0$ vs $\mathrm{H}_\mathrm{A} : \mu < 0$ (e.g., $= -2$) at the level of significance α.

Let $X \sim Normal(\mu, 1)$. Test $\mathrm{H}_0 : \mu = 0$ vs $\mathrm{H}_\mathrm{A} : \mu > 0$ (e.g., $= 2$) at the level of significance α.

Examples: one-sided tests using an iid normal sample

Let $X_1, \ldots, X_n$ be iid $Normal(\mu, \sigma^2)$, σ^2 known. Test $\mathrm{H_0} : \mu = \mu_0$ vs $\mathrm{H_A} : \mu = \mu_A < \mu_0$ at the level of significance α.

Q: is H_0 a simple hypothesis?

Examples: one-sided tests using an iid normal sample

Let $X_1, \ldots, X_n$ be iid $Normal(\mu, \sigma^2)$, σ^2 unknown. Test $\mathrm{H}_0 : \mu = \mu_0$ vs $\mathrm{H}_\mathrm{A} : \mu = \mu_A < \mu_0$ at the level of significance α.

Q: is H_0 a simple hypothesis?

Test $\mathrm{H}_0 : \mu = \mu_0$ vs $\mathrm{H}_\mathrm{A} : \mu = \mu_A > \mu_0$ at the level of signif. α.

Two-sided tests: intuition

Let $X_1, \ldots, X_n$ be iid $Normal(\mu, \sigma^2)$, σ^2 known. Test $\mathrm{H_0} : \mu = \mu_0$ vs $\mathrm{H_A} : \mu = \mu_A \neq \mu_0$ at the level of significance α.

Exact vs Approximate (Large-Sample) Tests

Suppose $P(\theta)$ is a pivot for θ. Suppose $\mathrm{H}_0 : \theta = \theta_0$ (e.g., $\theta_0 = 1$). Let $T_1 = P(\theta_0)$. **Q**: Is T_1 a valid test statistic?

Intuition for approximate tests: to determine the RR for an exact test, one needs to know the exact distribution of the TS.

Under the same circumstances as with pivots, this may not be known (or may be hard to derive). However, it is often the case that appropriately "normalized" TS is approximately $Normal(0, 1)$.

Hence, use the standard normal quantiles as cutoffs for the RR for a given nominal level of significance α. The downside is that the exact level of significance is unknown, although the approximate level of significance is α.

3 principal approaches to tests of statistical hypotheses

1. Direct approach using test statistics.

Idea: find a test statistic whose distribution is known under $\mathrm{H}_0 : \theta = \theta_0$; then determine the "kutoffs" to reject H_0.
Claim: A pivot for θ is a valid test statistic. Why?

2. Tests using a confidence interval.

Idea: construct a CI in such a way that $\mathrm{H}_0 : \theta = \theta_0$ is rejected if and only if θ_0 is not inside an appropriate CI.
Q: How to construct a CI that is "equivalent" to a given one- or two-sided test?

3. Tests using a *p-value*.

Motivation: recall the lightbulb example. What is the probability under H_0 that an independent realization of the test statistic is less than or equal to 50/1000?
Idea: compute the p-value, here
$Pr(TS \leq 0.05 | TS \sim Exp(\beta = 1)) = 0.049$. Reject H_0 if $\alpha \geq 0.049$; else do not reject H_0.

P-values: Motivation

Suppose the test rule is "reject H_0 for large values of the TS T", RR: $\{T > k_\alpha\}$. $k_\alpha = Q(1-\alpha)$ for cdf of TS under H_0.

$$\left.\begin{array}{ll} \alpha_1 = 0.05, & k_{\alpha_1} = Q(0.95) \\ \alpha_2 = 0.01, & k_{\alpha_2} = Q(0.99) \end{array}\right\} \Rightarrow \alpha_1 > \alpha_2 \text{ implies } Q(1-\alpha_1) < Q(1-\alpha_2)$$

Often we can reject H_0 at level $\alpha = 0.05$ but fail to reject H_0 at level $\alpha = 0.01$ (when $T \in (k_{\alpha_1}, k_{\alpha_2})$).

Practical concern: since different people have different desired levels of significance α, how to report the outcome/result of the test (in order to make everyone happy)?

p-value: "observed level of significance".

P-value for a right-sided test

For right-sided test with a RR $[T > k_\alpha]$, p-value is

$$\alpha_R(t) = \Pr(T \geq t \mid \mathrm{H}_0 \text{ is true}).$$

Notice $t \in \mathrm{RR}$ if and only if $\alpha_R(t) < \alpha$.

Example: $X \sim \text{Uniform}(0, \theta)$. $\mathrm{H}_0 : \theta = 1$ versus $\mathrm{H_A} : \theta = \theta_a > 1$.
Test rule: "reject H_0 for large values of X".

$$p\text{-value: } \alpha_R(t) = \Pr(X \geq t \mid \mathrm{H}_0 \text{ is true})$$

P-values: definitions

for right-sided test with a RR $[T > k_\alpha]$, p-value is

$$\alpha_R(t) = \Pr(T \geq t \mid \mathrm{H}_0 \text{ is true}).$$

Notice $t \in \mathrm{RR}$ if and only if $\alpha_R(t) < \alpha$.

p-value for left-sided test with
$\mathrm{RR} = \{T < k_\alpha\}, \alpha_L = \Pr(T \leq t \mid \mathrm{H}_0 \text{ is true})$.

p-value for a 2-sided test: $\mathrm{RR} = \{T < Q(\frac{\alpha}{2})\} \cup \{T > Q(1 - \frac{\alpha}{2})\}$.
Reject H_0 for extremely small or extremely large values of TS T.
p-value: $\alpha(t) = 2\min\{\alpha_L(t), \alpha_R(t)\}$

Notice that $t \in \mathrm{RR}$ if and only if $\alpha(t) < \alpha$.