P-Value

- To report the result of a hypothesis-testing analysis is to simply say whether the null hypothesis was rejected at a specified level of significance. This type of statement is somewhat inadequate because it says nothing about whether the conclusion was a very close call or quite clear cut.

- P-value is a quantity that conveys much information about the strength of evidence against H_0 and allows an individual decision maker to draw a conclusion at any specified level α.

- The P-value (*observed significance level*) is the probability, under the null hypothesis, that the test statistic is more ***extreme*** than the observed statistic.

What P-Values are not

- The P-value is not the probability that H_0 is true.
- The P-value is not Type I Error α.
- The P-value is not the significance level.
- The P-value is not Type II Error β

Comparison Between P-value and Type I Error α

- P-value=P(Test Statistic is more extreme than observed Test Statistic Value under Null Hypothesis)
- Type I Error=P(Test Statistic falls into Rejection Region under Null Hypothesis)

Remarks

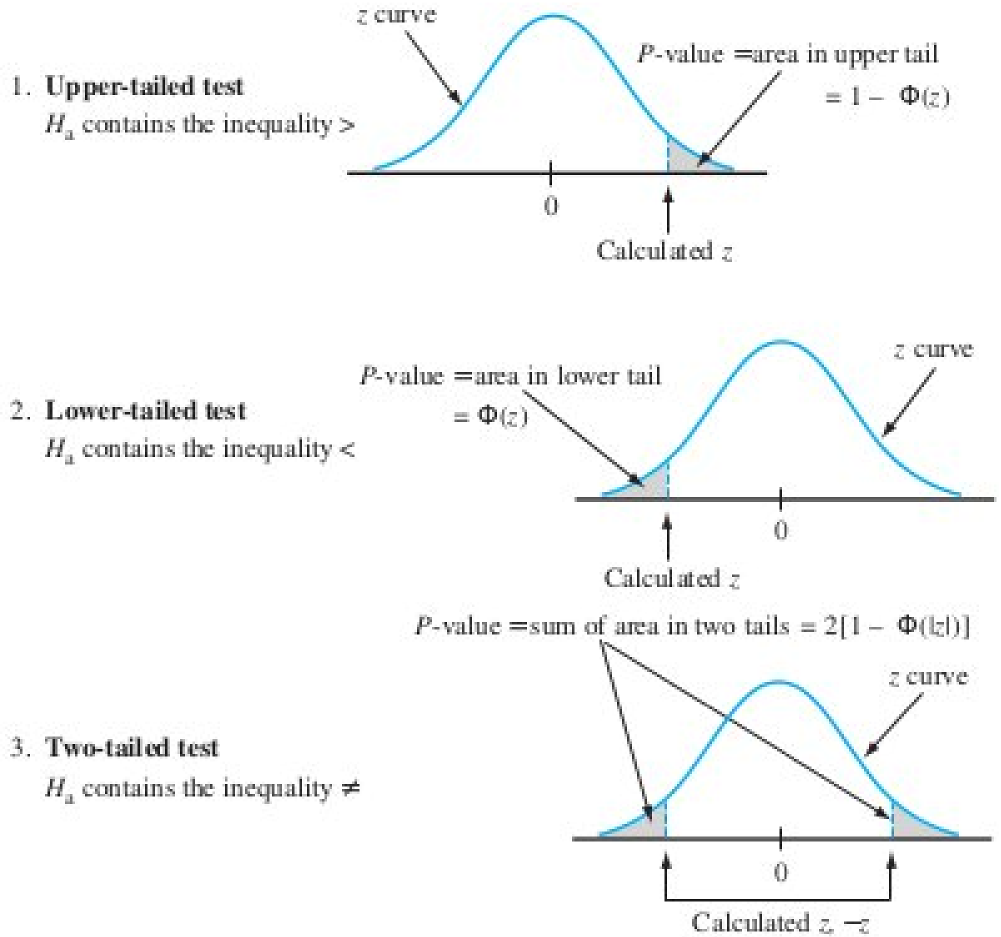
- **The smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis.**
- P-values can be seen as a more flexible procedure of Hypothesis Testing. The practical advantage is that it is easier to switch to a test of different significance level
- The decision rule based on P-values

> **Decision rule based on the *P*-value**
>
> Select a significance level α (as before, the desired type I error probability). Then
>
> $$\text{reject } H_0 \text{ if } P\text{-value} \leq \alpha$$
> $$\text{do not reject } H_0 \text{ if } P\text{-value} > \alpha$$

- The P-value is the smallest significance level α at which the null hypothesis can be rejected.

P-values and Tails

- Like Rejection Region, P-values are also related to the type of test we are concerning, uppper-tailed, lower-tailed or two-tailed.



Two sample tests

- A new drug is claimed to significantly reduce the blood pressure for high blood pressure patients. What kind of tests can we use to verify the claim?

- A new drug is claimed to perform much better in terms of reducing blood pressure than an old drug. What kind of tests can we use to verify the claim?

Things to cover

- As in the one sample testing problem, we will cover the following cases:
  1. Two normal populations with known variance.
  2. Two populations with unknown distribution and large sample size.
  3. Two normal populations with unknown variance.
  4. Two population proportions with large sample size.
  5. Tests about variances. (NOT required.)

- Basic assumptions for comparing population means:
  1. $X_1, X_2, \ldots, X_m$ is a random sample (i.i.d.) from a population with mean μ_1 and variance σ_1^2.
  2. $Y_1, Y_2, \ldots, Y_n$ is a random sample (i.i.d.) from a population with mean μ_2 and variance σ_2^2.
  3. The X and Y samples are independent of one another.

Test statistics

- Since we are comparing the population means, a natural test statistic to use would be the difference of two sample means. Because of independence we have,

$$\begin{aligned} E(\bar{X} - \bar{Y}) &= \mu_1 - \mu_2 \\ Var(\bar{X} - \bar{Y}) &= \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} \end{aligned}$$

Case I: normal, known variance

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

Test statistic: $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{\sigma_1^{\,2}}{m} + \dfrac{\sigma_2^{\,2}}{n}}} \sim \mathrm{N}(0,1)$

vs Alternative Hypothesis:

$H_a : \mu_1 - \mu_2 > \Delta_0$, reject if $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{\sigma_1^{\,2}}{m} + \dfrac{\sigma_2^{\,2}}{n}}} > Z_\alpha$

$H_a : \mu_1 - \mu_2 < \Delta_0$, reject if $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{\sigma_1^{\,2}}{m} + \dfrac{\sigma_2^{\,2}}{n}}} < -Z_\alpha$

$H_a : \mu_1 - \mu_2 \neq \Delta_0$, reject if $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{\sigma_1^{\,2}}{m} + \dfrac{\sigma_2^{\,2}}{n}}} < -Z_{\alpha/2}$ or $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{\sigma_1^{\,2}}{m} + \dfrac{\sigma_2^{\,2}}{n}}} > Z_{\alpha/2}$

Questions

- How to compute P-value for case I?
- How to compute type II errors for case I?
- In a balanced design, derive the sample size calculation formula (for alternative “>”):

$$m = n = \frac{(\sigma_1^2 + \sigma_2^2)(Z_\alpha + Z_\beta)^2}{(\Delta' - \Delta_0)^2}$$

Case II: large sample

$H_0 : \mu_1 - \mu_2 = \Delta_0$

Test statistic: $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} \sim A\mathrm{N}(0,1)$

vs Alternative Hypothesis:

$H_a : \mu_1 - \mu_2 > \Delta_0$, reject if $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} > Z_\alpha$

$H_a : \mu_1 - \mu_2 < \Delta_0$, reject if $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} < -Z_\alpha$

$H_a : \mu_1 - \mu_2 \neq \Delta_0$, reject if $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} < -Z_{\alpha/2}$ or $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} > Z_{\alpha/2}$

Questions

- How to construct confidence interval for $\mu_1 - \mu_2$ in case II?

Case III: normal, unknown variance

$H_0 : \mu_1 - \mu_2 = \Delta_0$

Test statistic: $\dfrac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}}} \sim t_\nu$, ν is the df of the t-distribution and it's approximately estimated

by the sampled data: $\nu = \dfrac{\left(\dfrac{s_1^2}{m} + \dfrac{s_2^2}{n}\right)^2}{\dfrac{(s_1^2/m)^2}{m-1} + \dfrac{(s_2^2/n)^2}{n-1}}$, and round ν town to the nearest integer.

Case III cont.

vs Alternative Hypothesis:

$$H_a : \mu_1 - \mu_2 > \Delta_0 \text{, reject if } \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} > t_{\alpha,\nu}$$

$$H_a : \mu_1 - \mu_2 < \Delta_0 \text{, reject if } \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} < -t_{\alpha,\nu}$$

$$H_a : \mu_1 - \mu_2 \neq \Delta_0 \text{, reject if } \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} < -t_{\alpha/2,\nu} \text{ or } \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} > t_{\alpha/2,\nu}$$

Questions

- How to compute P-values of the test?
- How to construct confidence interval for $\mu_1 - \mu_2$ in case III?
- What if we know that $\sigma_1^2 = \sigma_2^2$?

  The *pooled estimator* of $\sigma^2 = \sigma_1^2 = \sigma_2^2$ is given by

$$S_p^2 = \frac{m-1}{m+n-2} \cdot S_1^2 + \frac{n-1}{m+n-2} \cdot S_2^2$$

Case IV

$H_0 : p_1 - p_2 = 0$

Test statistic: $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}}$, $\hat{p} = \dfrac{m}{m+n}\hat{p}_1 + \dfrac{n}{m+n}\hat{p}_2$ (the *weighted* average of $\hat{p}_1$

and $\hat{p}_2$)

Case IV cont.

vs Alternative Hypothesis:

$H_a: p_1 - p_2 > 0$, reject if $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} > Z_\alpha$

$H_a: p_1 - p_2 < 0$, reject if $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} < -Z_\alpha$

$H_a: p_1 - p_2 \neq 0$, reject if $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} > Z_{\alpha/2}$ or $\dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{m}+\dfrac{1}{n}\right)}} < -Z_{\alpha/2}$

Paired t-test

- As in the previous example, the data is paired, the two scores (before and after) recorded for each individual are dependent, but the between individuals the pairs are independent.

- Thus in order to test H_0: $\mu_1 - \mu_2 = 0$, one has to look at the difference of each pair. The problem eventually becomes a one sample t-test problem.

Dependence Between Two Samples

- Two independent sample. $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ are independent, e.g. SAT scores of students from two different high schools.
- Two dependent sample. $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ are dependent, e.g. Math scores and Physics scores of students who are taking both these exams, in which there is natural pairing of values.

Example

In a study, six river locations were selected (six experimental objects) and the zinc concentration (mg/L) determined for both surface water and bottom water at each location.

The six pairs of observations are displayed in the accompanying table. Does the data suggest that true average concentration in bottom water exceeds that of surface water?

	Location					
	1	**2**	**3**	**4**	**5**	**6**
Zinc concentration in bottom water (x)	.430	.266	.567	.531	.707	.716
Zinc concentration in surface water (y)	.415	.238	.390	.410	.605	.609
Difference	.015	.028	.177	.121	.102	.107

Example Cont'd

At first glance, the data appears to be little difference between the x and y samples.

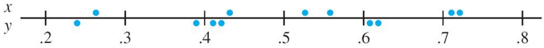


Figure: observations not identified by location

From location to location, there is a great deal of variability in each sample. It looks any differences between the samples can be attributed to location variability.

Example Cont'd

However, when the observations are identified by location, a different view emerges. At each location, bottom concentration exceeds surface concentration.

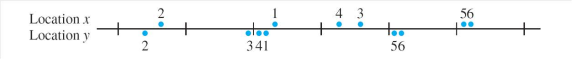


Figure: observations identified by location

This is confirmed by the fact that all $x - y$ differences are positive.

A correct analysis of this data focuses on these differences.

Assumptions

- The data consists of n independently selected pairs $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$.
- $E(X_i) = \mu_1$, $E(Y_i) = \mu_2$.
- Let $D_1 = X_1 - Y_1$, ..., $D_n = X_n - Y_n$ be the differences within pairs.
- Assume D_i's are normally distributed as $N(\mu_D, \sigma_D^2)$

The Paired t Test

- We are again interested in making an inference about the difference $\mu_D = \mu_1 - \mu_2$. An natural estimator will be $\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = \bar{X} - \bar{Y}$.
- $E(\bar{D}) = E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2 = \mu_D$
- For paired data $\bar{X}$ and $\bar{Y}$ are no longer independent, so $Var(\bar{X} - \bar{Y}) \neq Var(\bar{X}) + Var(\bar{Y})$.
- Since the D_i's constitute a normal random sample (of differences) with mean μ_D, hypotheses about μ_D can be tested using a one-sample t test.
- That is, to test hypotheses about $\mu_1 - \mu_2$ when data is paired, form the differences $D_1, D_2, \cdots, D_n$ and carry out a one-sample t test (based on $n-1$ df) on these differences.

The Paired t Test

$D = X - Y$ is the difference between observations within a pair. $\mu_D = \mu_1 - \mu_2$. $\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$ and $s_D = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})$ are sample mean and sample sd of d_i's.

- Null hypothesis: $H_0 : \mu_D = \Delta_0$
- Test statistic value: $t = \frac{\bar{d}-\Delta_0}{s_D/\sqrt{n}}$

Alternative Hypothesis	**Rejection Region for Level α Test**
$H_a : \mu > \mu_0$	$t \geq t_\alpha(n-1)$
$H_a : \mu < \mu_0$	$t \leq -t_\alpha(n-1)$
$H_a : \mu \neq \mu_0$	$t \geq t_{\alpha/2}(n-1)$ or $t \leq -t_{\alpha/2}(n-1)$

A P-value can be calculated as was done for one sample t test.