Economics 2740
Department of Economics
University of Guelph

Hypothesis tests when the data is not normally distributed

Inference when data is not normal

1 .Examples of non-normal data

a .Categorical data (discrete)

- Which brand of cereal?
- Who did you vote for?

b .Much financial data

- stock market returns,
- exchange rate returns.

Implications for Reference

A .If n is small this causes a problem:

- **Z** is not **N (0,1)**
- **t** is not t **(n-1)**
- **Z, t** tests don’t work

Implications for Inference

B . If n is large - “saved” by the central limit theorem.

- Z Approx N (0,1)
- t Approx t (n-1) Approx N (0,1)
- our tests work “approximately”

How Large is Large?

- Depends on how far from normality
- A rough rule of thumb:

  i) **$n < 30$** n small (red light)
  conduct test only if data is normal

  ii) **$n>150$** n large (green light)
  conduct test with confidence

  iii) **$30<n<150$** grey zone (yellow light)
  conduct test with caution

Case 3: Data may be not normal

Assumptions

n – large σ^2 – unknown

$x_1, x_2, \ldots x_n \sim$ i.i.d. (m, σ^2)

PDF of the data is unkown.

Cont’d. Case 3 (non-normal data).

B. Same as case 1 but now test approximate.

- Reject Ho: $\mu \le \mu_0$ if $Z = (\bar{X} - \mu_0)/(\sigma/n^{(1/2)}) > Z\alpha$
- In case I: $\alpha =$ P(reject when $\mu = \mu_0$)
- In case III: α Approx = P(reject when $\mu = \mu_0$)
- An approximation

Explanation and Summary

Case I	Case III
$x_1, x_2, \ldots x_N$ ~i.i.d. N (m,s^2) E[x_i]=m Var(x_i)=s^2	$x_1, x_2, \ldots x_N$ ~ i.i.d. (m,s^2) E[x_i]=m Var(x_i)=s^2
$\bar{X}$ ~ N (m,s^2/n) Exact.	By central Limit Theorem: $\bar{X}$ » N (m,s^2/n)
Z= ($\bar{X}$ - m$_0$)/(s/n$^{(1/2)}$) ~ N (0,1), exact.	Z= ($\bar{X}$ - m$_0$)/(s/n$^{(1/2)}$) » N (0,1), approximate.
H_A : m ³ m$_0$ Prob (Z>Za, m= m$_0$) = a Exact.	H_A : m ³ m$_0$ Prob (Z>Za, m= m$_0$) » a Exact.

Case 4. Data may not be normal and variance unknown.

- **Assumptions:**
  - ➢n is large,
  - ➢σ^2 – unknown,
  - ➢PDF unknown
- use t= $(\bar{X} - m_0)/(s/n^{(1/2)})$ like before
- both t (n-1) & N (0,1) Provide valid approximations
  - T Approx ~ t (n-1) Approx ~ N (0,1)
  - Use t-critical values

Explanation

- Large n => s^2 approx = σ^2
- t= $(\bar{X} - m_0)/(s/n^{(1/2)}) \approx Z = (\bar{X} - m_0)/(\sigma/n^{(1/2)})$
- $\approx N(0,1)$
- also large n => $t(n-1) \approx N(0,1)$
- => $t \approx t(n-1)$
- Same Procedure as in Case 2: conduct t-test
- Difference in interpretation: $\alpha \approx$ P (Type I Error): Only an approximate test.