Statistics One

Lecture 9
The Central Limit Theorem

1

Two segments

- Sampling distributions
- Central limit theorem

2

Lecture 9 ~ Segment 1

Sampling distributions

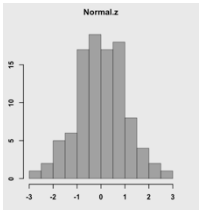
3

Review of histograms

- Histograms are used to display distributions
- For example, the body temperature of a random sample of healthy people

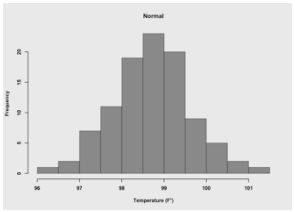
4

Review of histograms



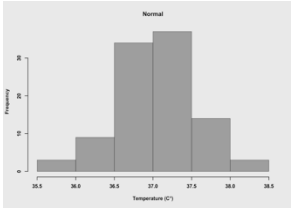
5

Review of histograms



6

Review of histograms



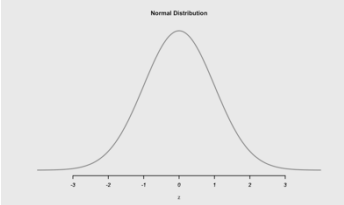
7

Review of histograms

- If a distribution is perfectly normal then the properties of the distribution are known

8

The normal distribution



9

The normal distribution & probability

- This allows for predictions about the distribution
  - Predictions aren't certain
  - They are *probabilistic*

10

The normal distribution & probability

- If one person is randomly selected from the sample, what is the probability that his or her body temperature is less than Z = 0?
  - Easy, p = .50

11

The normal distribution & probability

- If one person is randomly selected from the sample, what is the probability that his or her body temperature is greater than Z = 2? (100 F°, 38 C°)?
  - p = .02

12

The normal distribution & probability

- If this sample is healthy, then no one should have a fever
- I detected a person with a fever
- Therefore, this sample is not 100% healthy

13

Sampling distribution

- A distribution of sample statistics, obtained from multiple samples
  - For example,
    - Distribution of sample means
    - Distribution of sample correlations
    - Distribution of sample regression coefficients

14

Sampling distribution

- It is hypothetical
  - Assume a mean is calculated from a sample, obtained randomly from the population
  - Assume a certain sample size, N
  - Now, assume we had multiple random samples, all of size N, and therefore many sample means
  - Collectively, they form a *sampling distribution*

15

Sampling distribution & probability

- If one sample is obtained from a normal healthy population, what is the probability that the sample mean is less than Z = 0?
  - Easy, p = .50

16

Sampling distribution & probability

- If one sample is obtained from a normal healthy population, what is the probability that the sample mean is greater than Z = 2 (100 F°, 38 C°)?
  - p = .02

17

Sampling distribution & probability

- If this population is healthy, then no one sample should have a high mean body temperature
- I obtained a very high sample mean
- Therefore, the population is not healthy

18

Sampling distribution

- A distribution of sample statistics, obtained from multiple samples, each of size N
  - Distribution of sample means
  - Distribution of sample correlations
  - Distribution of sample regression coefficients

19

END SEGMENT

20

Lecture 9 ~ Segment 2

The Central Limit Theorem

21

Central Limit Theorem

- Three principles
  - The mean of a sampling distribution is the same as the mean of the population
  - The standard deviation of the sampling distribution is the square root of the variance of sampling distribution $\sigma^2 = \sigma^2 /N$
  - The shape of a sampling distribution is approximately normal if either (a) N >= 30 or (b) the shape of the population distribution is normal

22

NHST & Central limit theorem

- Multiple regression
  - Assume the null hypothesis is true
  - Conduct a study
  - Calculate B, SE, and t
  - t = B/SE

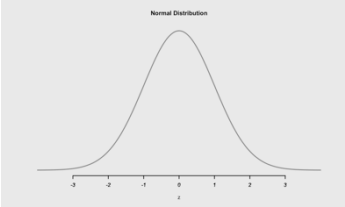
23

NHST & Central limit theorem

- Multiple regression
  - If the null hypothesis is true (B=0), then no one sample should have a very low or very high B
  - I obtained a very high B
  - Therefore, Reject the null hypothesis

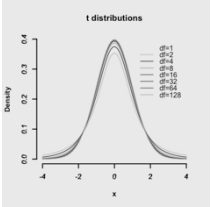
24

The normal distribution



25

The family of t distributions



26

NHST & Central limit theorem

- Multiple regression
  - Assume the null hypothesis is true
  - Conduct a study
  - Calculate B, SE, and t
  - t = B/SE
  - p-value is a function of t and sample size

27

NHST & the central limit theorem

- Multiple regression
  - If the null hypothesis is true (B=0), then no one sample should have a very low or very high B
  - I obtained a very high B
  - Therefore, Reject the null hypothesis
  - Very high and very low is $p < .05$

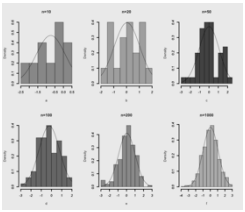
28

NHST & the central limit theorem

- Remember that sampling error, and therefore standard error, is largely determined by sample size

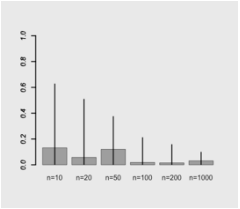
29

Sampling error and sample size



30

Sampling error and sample size



31

Central Limit Theorem

- Three principles
  - The mean of a sampling distribution is the same as the mean of the population
  - The standard deviation of the sampling distribution is the square root of the variance of sampling distribution $\sigma^2 = \sigma^2 /N$
  - The shape of a sampling distribution is approximately normal if either (a) N >= 30 or (b) the shape of the population distribution is normal

32

END SEGMENT

33

END LECTURE 9

34