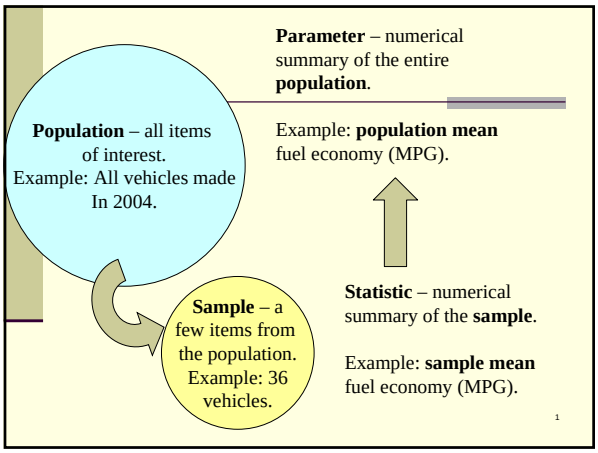
# Stat 401 B – Lecture 4




## One-sample model

$$Y = \mu + \varepsilon$$

- Y represents a value of the variable of interest
- $\mu$ represents the population mean
- $\varepsilon$ represents the random error associated with an observation

2

## Conditions

- The random error term, $\varepsilon$, is
  - Independent
  - Identically distributed
  - Normally distributed with standard deviation, $\sigma$

3

## Errors

- Model

$$Y = \mu + \varepsilon$$

- Error

$$\varepsilon = Y - \mu$$

4

## Residuals

- Estimate of error
- (Observation – Fit)
- Residual

$$\hat{\varepsilon} = Y - \overline{Y}$$

5

## Residuals

- Examine the residuals to see if the conditions for statistical inference are met.

6

## Checking Conditions

- Independence.
  - Hard to check this but the fact that we obtained the data through random sampling assures us that the statistical methods should work.

7

## Checking Conditions

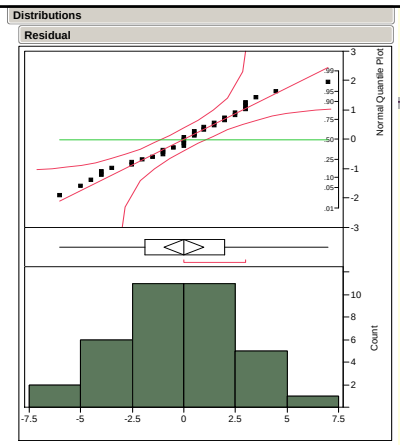
- Identically distributed.
  - Check using an outlier box plot. Unusual points may come from a different distribution
  - Check using a histogram. Bi-modal shape could indicate two different distributions.

8

## Checking Conditions

- Normally distributed.
  - Check with a histogram. Symmetric and mounded in the middle.
  - Check with a normal quantile plot. Points falling close to a diagonal line.

9




10

## MPG Residuals

- Histogram is symmetric and mounded in the middle.
- Box plot is symmetric with no outliers.
- Normal quantile plot has points following the diagonal line.

11

## MPG Residuals

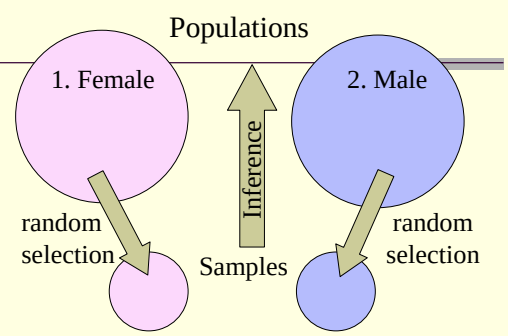
- The conditions for statistical inference appear to be satisfied.

12

## Two Independent Samples

- Question
  - In 2000, did men and women differ in terms of their body mass index?

13




14

## Two-sample model

$$Y = \mu_i + \varepsilon$$

- Y represents a value of the variable of interest
- $\mu_i$ represents the i$^{th}$ population mean
- $\varepsilon$ represents the random error associated with an observation

15

## Conditions

- The random error term, $\varepsilon$, is
  - Independent
  - Identically distributed
  - Normally distributed with standard deviation, $\sigma$

16

## Testing Hypotheses

- Question
  - In 2000, did men and women differ in terms of their body mass index, on average?

17

## Step 1 - Hypotheses

$$H_0 : \mu_1 = \mu_2 \text{ or } \mu_1 - \mu_2 = 0$$

$$H_A : \mu_1 \neq \mu_2 \text{ or } \mu_1 - \mu_2 \neq 0$$

18

## Step 2 – Test Statistic

$$t = \frac{\left(\bar{Y}_1 - \bar{Y}_2\right)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\left(27.484 - 26.868\right)}{7.544\sqrt{\frac{1}{50} + \frac{1}{50}}}$$

$$t = \frac{0.616}{1.509} = 0.408$$

$$\text{P - value} = 0.684$$

19

## Step 3 – Decision

- Fail to reject the null hypothesis because the P-value is larger than 0.05.

20

## Step 4 – Conclusion

- On average, men and women in 2000 could have had the same BMI.
- The difference between males' and females' average BMI's is not statistically significant.

21