Numerical Summaries of Data

Sample Mean

$$\bar{x} = \frac{\sum x_i}{n}$$

Weighted Mean

$$\bar{x}_w = \frac{\sum (x_i \cdot w_i)}{\sum w_i}$$

Population Mean

$$\mu = \frac{\sum x_i}{N}$$

Sample Variance

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}$$

Population Variance

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}$$

Sample Standard Deviation

$$s = \sqrt{s^2}$$

$$= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$$

Population Standard Deviation

$$\sigma = \sqrt{\sigma^2}$$

$$= \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}}$$

Range

$max - min$

Interquartile Range

$IQR = Q_3 - Q_1$

Lower Fence

$Q_1 - 1.5IQR$

Five Number Summary

$min, Q_1, median, Q_3, Umax$

Upper Fence

$Q_3 + 1.5IQR$

Elementary Probability

Probability of the Complement

$$P(A^c) = 1 - P(A)$$

General Multiplication Rule

$$P(A\ and\ B) = P(A) \cdot P(B|A)$$

General Addition Rule

$$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$$

Conditional Probability

$$P(B|A) = \frac{P(B\ and\ A)}{P(A)}$$

Discrete Random Variables

Expected Value (mean)

$$\mu_x = E(x) = \sum x_i P(x_i)$$

Standard Deviation

$$\sigma_x = \sqrt{\sum x_i^2 P(x_i) - \mu^2}$$

Binomial Distribution

$$P(x) = {}_nC_x \cdot p^x \cdot q^{n-x}, \quad \text{where} \quad {}_nC_x = \frac{n!}{x!(n-x)!}$$

The Normal Distribution

Z – score	Raw Score	Z – score for Sample Mean
$z = \dfrac{x-\mu}{\sigma}$	$x = z\sigma + \mu$	$z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

Estimation (Confidence Intervals)

ONE SAMPLE

Confidence Intervals for μ	**Confidence Intervals for p**
$\bar{x} \pm t_c \cdot \dfrac{s}{\sqrt{n}}$ where $df = n - 1$	$\hat{p} \pm z_c \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ $where\ \hat{p} = \dfrac{x}{n}$

TWO SAMPLES

Confidence Intervals for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_c \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

df = smaller of $n_1 - 1$ or $n_2 - 1$

Confidence Intervals for $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) \pm z_c \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where $\hat{p} = \dfrac{x}{n}$

MINIMUM SAMPLE SIZE

To estimate μ

$$n = \left(\frac{z_c \cdot \sigma}{E}\right)^2$$

To estimate ρ

$$n = \hat{p}(1-\hat{p})\left(\frac{z_c}{E}\right)^2$$

Without an estimate for p, use $p = 0.5$.

Hypothesis Testing

ONE SAMPLE

Hypothesis Tests for μ	**Hypothesis Tests for p**
$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$ where $df = n - 1$	$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$ $where\ \hat{p} = \frac{x}{n}$

TWO SAMPLES

Hypothesis Tests for $\boldsymbol{\mu_1, \mu_2}$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

df = smaller of $n_1 - 1$ or $n_2 - 1$

Hypothesis Tests for $\boldsymbol{p_1, p_2}$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\bar{p}(1-\bar{p})}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}, \quad \hat{p} = \frac{x}{n}$

Regression

Correlation coefficient

$$r = \frac{\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)}{n - 1}$$

OR

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2}\sqrt{n\sum y^2 - (\sum y)^2}}$$

Least-Squares Regression

$$\hat{y} = ax + b$$

where

$$a = r \cdot \frac{s_y}{s_x} = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2}$$

$$b = \bar{y} - a\bar{x}$$