ملخص احصاء حيوي

Biostat 109

Chapter 4 and 5

(439) Elite
Homoud
Abdullatif
Naif

Chapter 4:

4.1

There are two types of random variables:

Random variables $\begin{cases} \textit{Discrete Random Variables} \\ \textit{Continuous Random Variables} \end{cases}$

4.2:

The probability of a discrete random variable (r.v) is a table, graph, formula, or other device used to specify all possible values of the random variable along with their probabilities. Ex: number of patients visiting hospital, number of times a person had a cold in a year.

The probability distribution should always satisfy the following rules

(1) $0 \le P(X = x) \le 1$

(2) $\sum P(X = x) = 1$

Using the probability distribution of a discrete r.v we can find the probability of any event expressed in the term of r.v.X

Mean: Denoted by μ or μ_x and it is defined by: $\mu = \sum x\, P(X = x)$

Variance: Denoted by σ^2 or σ_x^2 and is defined by $\sigma^2 = \sum (x - \mu)^2 P(X = x)$

Cumulative Distributions are defined by $P(X \le x) = \sum_{a \le x} P(X = a)$ (Sum over all values $\le$x)

Combinations (n!): The number of different ways for selecting r objects from n distinct objects

$$ {}_nC_r = \frac{n!}{r!\ (n-r)!}; $$

4.3

Bernoulli Trial:

- Only 2 outcomes (success and failure)
- Discrete distribution
- Binomial distribution has *n* number of Bernoulli's trials
- To find probability we use

$$ P(X=x) = \begin{cases} {}_nC_x\ p^x\ q^{n-x} & \text{for } x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases} $$

$$ \text{Where:} \quad {}_nC_x = \frac{n!}{x!\ (n-x)!} $$

- Binomial Distribution's mean: **μ = np** (expected value)
- Binomial Distribution's variance **σ^2 = npq**
- Binomial Distribution's standard deviation **σ = √variance =√npq**

4.4

Poisson Distribution

- Discrete distribution
- Used to model discrete r.v representing number of occurrences in an interval of space or time.
- To calculate the probability, we use

$$P(X=x)=\begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; \quad for \quad x=0,1,2,3,\dots \\ 0 & ; \quad otherwise \end{cases}$$

where e = 2.71828 (the natural number).

- The Poisson Distribution's mean: $\underline{\mu=\lambda}$
- The variance is $\underline{\sigma^2=\lambda}$
- λ is the mean of the distribution

Example:

Some random quantities that can be modeled by Poisson distribution:

- No. of patients in a waiting room in an hour.
- No. of surgeries performed in a month.
- No. of rats in each house in a particular city.

4.5 Continuous Probability Distributions:

(1)The total area under the curve of $f(x) = 1$.

(2) The probability X is Between (a) and (b) = area under curve of f (x).

(3) The probability of an interval = area under curve of f(x)

Note: If **X** is continuous r.v. then:

1. $P(X = a) = 0$
2. $P(X \leq a) = P(X < a)$
3. $P(X \geq b) = P(X > b)$
4. $P(X \geq a) = 1- P(X \leq a)$
5. $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$, **(Bigger value – Lower value)**

(Equal symbol has no effect).

4.6 The Normal Distribution:

The probability density function (pdf) of the normal distribution is give by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad ; -\infty < x < \infty$$

Where (e = 2.71828) and **(π = 3.14159)**

Parameters: Mean (**μ**), Standard deviation(**σ**)

* The curve of f(x) is **symmetric** about the mean **μ**.

* Mean = Median = Mode

* Mean(**μ**) → determines the Location

* Standard deviation(**σ**) → determines the shape

* If the r.v. **X** is normally distributed with mean (**μ**) and standard deviation (**σ**) (variance **σ^2**), we write:

$$X \sim Normal(\mu, \sigma^2) \quad or \quad X \sim N(\mu, \sigma^2)$$

The Standard Normal Distribution:

Mean $\mu = 0$,   Variance $\sigma^2 = 1$

The standard normal **Random Variable** is denoted by **(Z)**

$$\mathbf{Z} \sim N(0, 1)$$

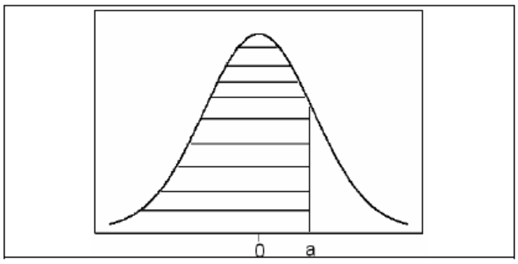
The probability density function (pdf) of Z ~ N(0, 1) is given by:

$$f(z) = n(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

Calculating Probabilities of Normal(0,1):

$$Z = \frac{X - \mu}{\sigma} \sim \text{Normal } (0,1).$$

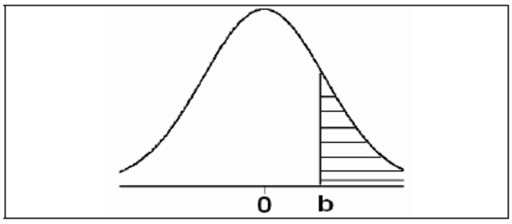
(i) $P(Z \le a) =$ From the table



(ii) $P(Z \ge b) = 1 - P(Z \le b)$

Where:

$P(Z \le b) =$ From the table

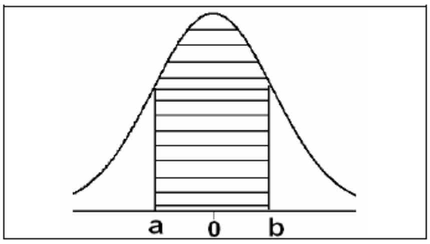


(iii) $P(a \le Z \le b) = P(Z \le b) - P(z \le a)$

Where:

$P(Z \le b) =$ from the table

$P(z \le a) =$ from the table



(iv) $P(Z = a) = 0$ *for every a.*

5.3 Distribution of the Sample Mean: (Sampling Distribution of the Sample Mean $\bar{x}$):

$\bar{x}$ = Sample mean

μ = Population mean

σ^2 = Variance

1. The mean of $\overline{X}$ is: $\mu_{\overline{X}} = \mu$.
2. The variance of $\overline{X}$ is: $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$.
3. The Standard deviation of $\overline{X}$ is call the standard error and is defined by: $\sigma_{\overline{X}} = \sqrt{\sigma^2_{\overline{X}}} = \frac{\sigma}{\sqrt{n}}$.

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$ **used with normal & non-normal with large (n)**

$$T = \frac{\overline{X} - \mu}{S / \sqrt{n}}$$ **used when σ^2 is unknown + normal distribution**

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$

Notation: degrees of freedom = df = ν

5.4 Distribution of the Difference Between Two Sample Means ($\bar{x}_1 - \bar{x}_2$):

Mean of $\bar{X}_1 - \bar{X}_2$ is: $\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$

Variance of $\bar{X}_1 - \bar{X}_2$ is: $\sigma^2_{\bar{X}_1-\bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

Standard error (standard) deviation of $\bar{X}_1 - \bar{X}_2$ is:

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\sigma^2_{\bar{X}_1-\bar{X}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- $$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

For two random samples with known variances (normal & non-normal with large sample size)

5.5 Distribution of the Sample Proportion ($\hat{p}$):

The population proportion is

$$p = \frac{N(A)}{N}$$

(p is a parameter)

The sample proportion is

$$\hat{p} = \frac{n(A)}{n}$$

($\hat{p}$ is a statistic)

The mean of the sample proportion ($\hat{p}$) is the population proportion (p); that is:

$$\mu_{\hat{p}} = p$$

The variance of the sample proportion ($\hat{p}$) is:

$$\sigma^2_{\hat{p}} = \frac{p(1-p)}{n} = \frac{pq}{n}.$$

(where q=1 −p)

The standard error (standard deviation) of the sample proportion ($\hat{p}$) is:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$$

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

For large sample size of sample Proportion.

5.6 Distribution of the Difference Between Two Sample Proportions ($\hat{p}_1 - \hat{p}_2$):

The mean, the variance and the standard error (standard deviation) of $\hat{p}_1 - \hat{p}_2$ are:

- Mean of $\hat{p}_1 - \hat{p}_2$ is:

$$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$$

- Variance of $\hat{p}_1 - \hat{p}_2$ is:

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

- Standard error (standard deviation) of $\hat{p}_1 - \hat{p}_2$ is:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

- $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

For large sample size of $\hat{p}_1 - \hat{p}_2$