Printout

Saturday, August 28, 2021 9:45 AM

Section 1

MANE 6313

Subsection 1

Week 2, Module A

Student Learning Outcome

Analyze simple comparative experiments and experiments with a single factor.

Module Learning Outcome

Review probability distributions.

Simple Comparative Experiments

- Chapter two only focuses on simple experiments to compare two conditions. Later, we will extend these techniques.
- Largely review of statistics and hypothesis testing from previous classes.
- Statistics focuses on two emphases: *Descriptive statistics* and *inferential statistics*.
- In inferential statistics we are concerned with estimation and hypothesis testing.

- Differences or fluctuations in outcomes from the same treatment are called *noise* in the results

random error

- Often this noise is referred to as *experimental error* or simply *error*
- We classify *statistical error* as those errors that arise from variation that is uncontrolled and generally unavoidable.
- The presence of noise implies that the response variable is a *random variable*

Stochastic → random

deterministic

Review of Probability Distributions

- The probability structure of a random variable is defined by its *probability distribution*.
- For a discrete random variable

$p()$ → Probability mass function

(capital) Sigma = $\sum_{i=1}^{2} x_i = x_1 + x_2$

$$
\begin{aligned}
0 \leq p(y_i) \leq & \quad 1 \text{ for all values of } y_i \\
P(y = y_i) = & \quad \underline{p(y_i)} \text{ for all values of } y_i \\
\sum_{\text{all values of } y_i} p(y_i) = & \quad 1
\end{aligned}
$$

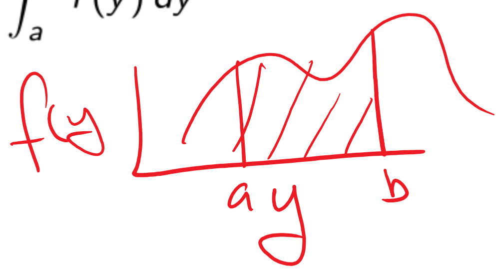
- For a continuous random variable

f() - probability density function

$$f(y) \geq 0$$

$$P(a \leq y \leq b) = \int_a^b f(y)\,dy$$

$$\int_{-\infty}^{\infty} f(y)\,dy = 1$$



μ - mu

- The mean of a probability distribution is defined to be

$$\mu = \begin{cases} \int_{-\infty}^{\infty} yf(y)\,dy & y \text{ continuous} \\ \sum_{\text{all y}} yp(y) & y \text{ discrete} \end{cases}$$

- The expected value operator is

$$E(y) \equiv \mu$$

- The variance of a probability distribution is defined to be

$$\sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\,dy & y \text{ continuous} \\ \sum_{\text{all } y} (y-\mu)^2 p(y) & y \text{ discrete} \end{cases}$$

σ - Sigma

- The variance can be written in terms of the expectation $\sigma^2 = E\left[(y-\mu)^2\right]$
- We will often use the variance operator $V(y) \equiv E\left[(y-\mu)^2\right] = \sigma^2$

Common Probability Distributions[1]

TABLE • 1 Summary of Common Probability Distributions

Name	Probability Distribution	μ Mean	σ^2 Variance	Section in Book
Discrete				
Uniform	$\frac{1}{n}, a \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a+1)^2-1}{12}$	3-5
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n, 0 \le p \le 1$	np	$np(1-p)$	3-6
Geometric	$(1-p)^{x-1} p$ $x = 1, 2, \ldots, 0 \le p \le 1$	$1/p$	$(1-p)/p^2$	3-7
Negative binomial	$\binom{x-1}{r-1} (1-p)^{x-r} p^r$ $x = r, r+1, r+2, \ldots, 0 \le p \le 1$	r/p	$r(1-p)/p^2$	3-7
Hypergeometric	$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$ $x = \max(0, n-N+K), 1, \ldots$ $\min(K, n), K \le N, n \le N$	np where $p = \frac{K}{N}$	$np(1-p)\left(\frac{N-n}{N-1}\right)$	3-8
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$	λ	λ	3-9
Continuous				
Uniform	$\frac{1}{b-a}, a \le x \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a)^2}{12}$	4-5
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $-\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma$	μ	σ^2	4-6
Exponential	$\lambda e^{-\lambda x}, 0 \le x, 0 < \lambda$	$1/\lambda$	$1/\lambda^2$	4-8
Erlang	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, 0 < x, r = 1, 2, \ldots$	r/λ	r/λ^2	4-9.1
Gamma	$\frac{\lambda x^{r-1} e^{-\lambda x}}{\Gamma(r)}, 0 < x, 0 < r, 0 < \lambda$	r/λ	r/λ^2	4-9.2
Weibull	$\frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-(x/\delta)^\beta}$ $0 < x, 0 < \beta, 0 < \delta$	$\delta\Gamma\left(1+\frac{1}{\beta}\right)$	$\delta^2\Gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2$	4-10
Lognormal	$\frac{1}{x\omega\sqrt{2\pi}} \exp\left(\frac{-[\ln(x)-\theta]^2}{2\omega^2}\right)$	$e^{\theta+\omega^2/2}$	$e^{2\theta+\omega^2}(e^{\omega^2}-1)$	4-11

Functions of Random Variables

c - constant
y - random variable

1. $E(c) = c$
2. $E(y) = \mu_y$
3. $E(cy) = cE(y) = c\mu$
4. $V(c) = 0$
5. $V(y) = \sigma^2$
6. $V(cy) = c^2V(y) = c^2\sigma^2$

For two random variables with the following properties $E(y_1) = \mu_1$, $V(y_1) = \sigma_1^2$, $E(y_2) = \mu_2$ and $V(y_2) = \sigma_2^2$

1. $E(y_1 + y_2) = E(y_1) + E(y_2) = \mu_1 + \mu_2$
2. $V(y_1 + y_2) = V(y_1) + V(y_2) + 2Cov(y_1, y_2)$ where $Cov(y_1, y_2) = E\left[(y_1 - \mu_1)(y_2 - \mu_2)\right]$
3. $V(y_1 - y_2) = V(y_1) + V(y_2) - 2Cov(y_1, y_2)$

$V(c_1 y_1 + c_2 y_2) = c_1^2 V(y_1) + c_2^2 V(y_2) + 2c_1c_2 Cov(y_1, y_2)$

Looking at 3 $c_1 = 1$, $c_2 = -1$

$V(c_1 y_1 + c_2 y_2) = c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 + 2c_1c_2 Cov(y_1, y_2)$

$\rightarrow Cov(y_1, y_2) = 0$

If y_1 and y_2 are independent, then

1. $V(y_1 \pm y_2) = V(y_1) + V(y_2) = \sigma_1^2 + \sigma_2^2$
2. $E(y_1 \cdot y_2) = E(y_1) \cdot E(y_2) = \mu_1 \cdot \mu_2$

In general note that

$$E\left(\frac{y_1}{y_2}\right) \neq \frac{E(y_1)}{E(y_2)}$$

- For two random variables to be independent we need to examine the joint distribution function $f_{y_1,y_2}(y_1, y_2)$ and the two marginal distribution functions $f_{y_1}(y_1)$ and $f_{y_2}(y_2)$
- We need to show that $\underbrace{f_{y_1,y_2}}_{\text{joint}}(y_1, y_2) = f_{y_1}(y_1) \cdot f_{y_2}(y_2)$
- For normally distributed random variables (and only normally distributed r.v.) if $Cov(y_1, y_2) = 0$ then y_1 and y_2 are independent.

Sampling

- We will use the following statistics as point estimators for μ and σ^2 respectively

Single value, fnc (Sample)

$$\hat{\mu} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

and

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

\- Properties of estimators

- Properties of point estimators
  - *Unbiased* $E(\bar{y}) = \mu$
  - *Minimum variance*
- The normal distribution is one of the most important sampling distributions

Central Limit Theorem

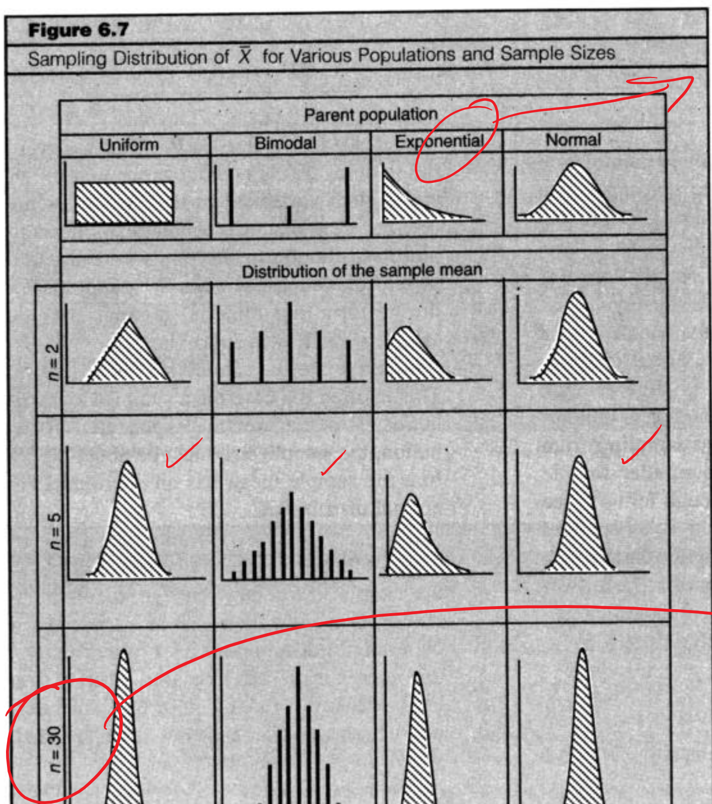
If $y_1, y_2, \ldots, y_n$ is a sequence of n independent and identically distributed random variables with $E(y_i) = \mu$ and $V(y_i) = \sigma^2$ (both finite) and $x = y_1 + y_2 + \cdots + y_n$ then

$$z_n = \frac{x - n\mu}{\sqrt{n\sigma^2}}$$

has an approximate $N(0, 1)$ distribution

- What is the implication here? any distribution becomes normally distributed as n becomes large

CLT Demonstration[2]



worst case

n = 25-30
suggested size
of a
sample