Statistics Handout 3

Discrete Random Variable:
When the observations of a quantitative random variable can take only a finite number of values or a countable number of values, we say that the variable is a *discrete random variable.*

Continuous Random Variable:
When the observations of a quantitative random variable can take on any of the countless number of values in a line interval, we say that the variable is a *continuous random variable.*

Important properties of a normal curve

1. The curve is bell-shaped with the highest point over the mean μ.
2. It is symmetrical about a vertical line through μ.
3. The curve approaches the horizontal axis but never touches or crosses it.
4. The transition (inflection) points between cupping upward and downward occur above $\mu + \sigma$ and $\mu - \sigma$.

Empirical Rule

For a distribution that is symmetrical and bell-shaped (in particular, for a normal distribution):
Approximately 68 % of the data values will lie within one standard deviation on each side of the mean.
Approximately 95 % of the data values will lie within two standard deviations on each side of the mean.
Approximately 99.7 % of the data values will lie within three standard deviations on each side of the mean.

Some commonly used statistics and corresponding parameters

Measure	*Statistic*	*Parameter*
Mean	$\bar{x}$ (x bar)	μ (mu)
Variance	s^2	σ^2 (sigma squared)
Standard Deviation	s	σ
Proportion	$\hat{p}$ (p hat)	p

C Confidence Interval for μ (large samples, n $\geq$ 30)

$$\bar{x} - E < \mu < \bar{x} + E$$

Where $\bar{x}$ = sample mean

$E = Z_c \dfrac{\sigma}{\sqrt{n}}$ if the population standard deviation σ is known

$E = Z_c \dfrac{s}{\sqrt{n}}$ if σ is not known and we use the sample standard deviation s as an approximation for σ.

$c =$ confidence level (0 < c < 1)

Z_c = critical value

n = sample size ($n \geq 30$)

C Confidence Interval for μ (small samples, n < 30)

$$\bar{x} - E < \mu < \bar{x} + E$$

Where $\bar{x}$ = sample mean

$E = t_c \dfrac{s}{\sqrt{n}}$

c = confidence level (0 < c < 1)

t_c = critical value for confidence level c,

And degrees of freedom $d, f = n - 1$

Taken from t distribution

$n =$ sample size (small samples, $n < 30$)

$s =$ sample standard deviation

Type I and Type II Errors

	Our decision	
Truth of H_o	And if we do not reject H_o	And if we reject H_o
If H_o is true	Correct decision; no error	Type I error
If H_o is false	Type II Error	Correct decision; no error

Statistical significance

If we reject H_o, we say that the data are *statistically significant*. If we do not reject H_o, we say that the data are *not statistically significant.*

Hypothesis testing using P values from a computer or calculator

1. Establish the level of significant α of the test.
2. Determine the null and alternate hypotheses, H_o and H_1.
3. Enter information about the observed sample into the computer or calculator and look for the P value of the observed sample statistic in this computer output.
4. Compare your level of significance with the P value.

   If P value $\leq \alpha$, reject H_o.

   If P value $> \alpha$, do not reject H_o.

Independent:
We say that two sampling distributions are *independent* if there is no relation whatsoever between specific values of the two distributions.