SAMPLING DISTRIBUTIONS

A. *Sampling Distributions Appropriate for* $\bar{x}$

Case I

(a) Sampling from a normal population

(b) μ is known, σ is known

(c) any sample size, n

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Case II

(a) Sampling is done from a population whose distribution is unknown or non-normal

(b) μ is known, σ is known or unknown

(c) sample size is large ($n \geq 30$)

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \text{if } \sigma \text{ is known}$$

$$z = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \qquad \text{if } \sigma \text{ is unknown}$$

Case III

(a) Sampling from normal population

(b) μ is known, σ is unknown

(c) sample size is small ($n<30$)

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \qquad \text{with } v = n - 1 \text{ (degrees of freedom)}$$

B. *Sampling Distributions Appropriate for* $\bar{x}_1 - \bar{x}_2$

Case I

(a) Sampling from normal population

(b) σ_1 and σ_2 are known

(c) any sample size, n_1 and n_2

$$z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}}$$

Case II

(a) Distribution of sampling population is unknown, may be normal or non-normal

(b) σ_1 and σ_2 are known or unknown

(c) n_1 and n_2 are ≥ 30

$$z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}}} \quad \text{if } \sigma_1 \text{ and } \sigma_2 \text{ are known}$$

$$z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}}} \quad \text{if } \sigma_1 \text{ and } \sigma_2 \text{ are unknown}$$

Case III

(a) Sampling from normal population

(b) σ_1 and σ_2 are unknown but $\sigma_1{}^2 = \sigma_2{}^2$

(c) n_1 and n_2 are <30

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with } v = n_1 + n_2 - 2 \quad \text{(degrees of freedom)}$$

$$S_p^{\,2} = \frac{(n_1-1)S_1^{\,2} + (n_2-1)S_2^{\,2}}{n_1+n_2-2} \qquad \text{(pooled variance)}$$

Case IV

(a) Sampling population are normal

(b) σ_1 and σ_2 are unknown but $\sigma_1^{\,2} \neq \sigma_2^{\,2}$

(c) n_1 and n_2 are <30

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^{\,2}}{n_1} + \dfrac{S_2^{\,2}}{n_2}}}$$

$$v = \frac{\left[\dfrac{S_1^{\,2}}{n_1} + \dfrac{S_2^{\,2}}{n_2}\right]^2}{\dfrac{\left(\dfrac{S_1^{\,2}}{n_1}\right)^2}{n_1-1} + \dfrac{\left(\dfrac{S_2^{\,2}}{n_2}\right)^2}{n_2-1}} \qquad \text{rounded to the nearest integer}$$

C. Sampling Distribution Appropriate for S^2

Case I

(a) n<30

(b) sampling population is normal

$$X^2 = \frac{(n-1)S^2}{\sigma^2} \qquad v = n - 1$$

Case II

(a) $n \geq 30$

(b) any sampling population

$$z = \frac{S - \sigma}{\dfrac{\sigma}{\sqrt{2n}}}$$

D. Sampling Distribution for the Ratio of Two Variances S_1^2 / S_2^2

$$F = \frac{\sigma_2^2}{\sigma_1^2}\left(\frac{S_1^2}{S_2^2}\right) \qquad v_1 = n_1 - 1 \text{ and } v_2 = n_2 - 1$$

$$F_{1-\alpha(v_1,v_2)} = \frac{1}{F_{\alpha(v_2,v_1)}}$$

E. Sampling Distribution Appropriate for $\hat{p}$

Case I

(a) Any sampling population

(b) n<30

Use binomial distribution

Case II

(a) Any sampling population

(b) n≥30

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

F. Sampling Distribution Appropriate for $\hat{p}_1 - \hat{p}_2$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \quad \text{if } p_1 \text{ and } p_2 \text{ are known}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{if } p_1 \text{ and } p_2 \text{ are unknown}$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{(pooled estimate of proportion)}$$

MEASURES OF AVERAGE AND VARIABILITY

	Ungrouped Data	*Grouped Data*
Mean	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$
Median	use the rules you have learned in finding for the median in ungrouped data	$X_m = L + \left[\frac{\frac{n}{2} - F_{(m-1)}}{f_m} \right] W$
Standard Deviation	$S = \sqrt{\frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)}}$	$S = \sqrt{\frac{n\sum f_i x_i^2 - \left(\sum f_i x_i\right)^2}{n(n-1)}}$

GOODNESS OF FIT TEST

$$x^2 = \frac{(o_i - e_i)^2}{e_i} \qquad v = k - 1 - m$$

LINEAR REGRESSION AND CORRELATION

$$\hat{y} = a + bx$$

$$a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}$$

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$r = b\sqrt{\frac{S_{xx}}{S_{yy}}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

ANALYSIS OF VARIANCE

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Computed F
Regression	SSR	1	SSR	$F=SSR/S^2$
Error	SSE	n-2	$S^2=SSE/(n-2)$	
Total	SST	n-1		

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$$

$$SST = S_{yy}$$

$$SSR = bS_{xy}$$

$$SSE = SST - SSR$$