Math 215

Chapter 7 • Sampling Distributions

- **Mean of $\overline{x}$:** $\mu_{\overline{x}} = \mu$
- **Standard deviation of $\overline{x}$:** $\sigma_{\overline{x}} = \sigma / \sqrt{n}$
- **z value for $\overline{x}$:** $z = \dfrac{\overline{x} - \mu}{\sigma_{\overline{x}}}$

(Note: If n < 30, population must be normal; otherwise it doesn't matter)

- **Population proportion:** $p = X / N$
- **Sample proportion:** $\hat{p} = x / n$
- **Mean of $\hat{p}$:** $\mu_{\hat{p}} = p$
- **Standard deviation of $\hat{p}$:** $\sigma_{\hat{p}} = \sqrt{pq / n}$
- **z value for $\hat{p}$:** $z = \dfrac{\hat{p} - p}{\sigma_{\hat{p}}}$

(Note: Necessary conditions are $np > 5$ and $nq > 5$.)

Chapter 8 • Estimation of the Mean and Proportion

- **Point estimate of $\mu = \overline{x}$**
- **Confidence interval for μ when σ is known:**

$\overline{x} \pm z\sigma_{\overline{x}}$ where $\sigma_{\overline{x}} = \sigma / \sqrt{n}$

(Note: If n < 30, population must be normal.)

- **Confidence interval for μ when σ is not known:**

$$\overline{x} \pm ts_{\overline{x}} \quad \text{where} \quad s_{\overline{x}} = s / \sqrt{n} \quad \text{and} \quad s = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}}$$

$df = n\text{-}1$

(Note: If n < 30, population must be normal.)

- **Margin of error of the estimate of μ:**

$E = z\sigma_{\overline{x}}$ or $E = ts_{\overline{x}}$

- **Sample size to estimate μ:** $n = z^2\sigma^2 / E^2$
- **Confidence interval for p for a large sample:**

$\hat{p} \pm zs_{\hat{p}}$ where $s_{\hat{p}} = \sqrt{\hat{p}\hat{q} / n}$

- **Margin of error of the estimate of p:**

$E = zs_{\hat{p}}$

- **Sample size to estimate p:** $n = z^2\hat{p}\hat{q} / E^2$

Chapter 9 • Hypothesis Tests about the Mean and Proportion

- **Critical Value Approach:**

Step 1: State the null and alternative hypotheses.
Step 2: Select the distribution to use.
Step 3: Determine the rejection and non-rejection regions. (i.e. critical value(s) of test statistic)
Step 4: Calculate the observed value of the test statistic.
Step 5: Make a decision and write a conclusion.

- **Test of hypotheses about μ when σ is known:**

$$z_{observed} = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}} \quad \text{where} \quad \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

(Note: If n < 30, population must be normal; otherwise it doesn't matter.)

- **Test of hypotheses about μ when σ is not known:**

$$t_{observed} = \frac{\overline{x} - \mu}{s_{\overline{x}}} \quad \text{where} \quad s_{\overline{x}} = \frac{s}{\sqrt{n}} \quad df = n - 1$$

(Note: If n < 30, population must be normal; otherwise it doesn't matter.)

- **Test of hypotheses about p for a large sample:**

$$z_{observed} = \frac{\hat{p} - p}{\sigma_{\hat{p}}} \quad \text{where} \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$$

Chapter 10 • Estimation and Hypothesis Testing: Two Populations

- **Confidence interval for $\mu_1 - \mu_2$ when σ_1 and σ_2 unknown but equal:**

$$(\overline{x}_1 - \overline{x}_2) \pm ts_{\overline{x}_1 - \overline{x}_2} \quad df = n_1 + n_2 - 2$$

where

$$s_{\overline{x}_1 - \overline{x}_2} = s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{and} \quad s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

- **Test of hypotheses about $\mu_1 - \mu_2$ when σ_1 and σ_2 unknown but equal:**

$$t_{observed} = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{s_{\overline{x}_1 - \overline{x}_2}} \quad df = n_1 + n_2 - 2$$

where $s_{\overline{x}_1 - \overline{x}_2}$ as in confidence interval for $\mu_1 - \mu_2$

- **Confidence interval for μ_d in paired or matched samples:**

$$\overline{d} \pm ts_{\overline{d}} \quad df = n - 1$$

where

$$\overline{d} = \frac{\Sigma d}{n} \quad \text{and} \quad s_d = \sqrt{\frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}} \quad \text{and} \quad s_{\overline{d}} = \frac{s_d}{\sqrt{n}}$$

- **Test of hypotheses about μ_d in paired or matched samples:**

$$t_{observed} = \frac{\overline{d} - \mu_d}{s_{\overline{d}}}$$

where $\overline{d}$ and $s_{\overline{d}}$ as in confidence interval for μ_d

$df = n - 1$

- **Confidence interval for $p_1 - p_2$:**

$$(\hat{p}_1 - \hat{p}_2) \pm zs_{\hat{p}_1 - \hat{p}_2}$$

where

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \quad \text{and} \quad \hat{p}_1 = \frac{x_1}{n_1}\ ;\ \hat{p}_2 = \frac{x_2}{n_2}$$

- **Test of hypotheses about $p_1 - p_2$:**

$$z_{observed} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{s_{\hat{p}_1 - \hat{p}_2}}$$

where

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\overline{p}\,\overline{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

and $\overline{p} = \dfrac{x_1 + x_2}{n_1 + n_2}$ or $\overline{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

Math 215

Chapter 11 • Chi-Square Tests

- **Goodness of fit test:**

H_0: The proportions (percentages) in the categories follow the distribution hypothesized
H_1: The proportions (percentages) in the categories do not follow the distribution hypothesized

$$\chi^2_{observed} = \Sigma\frac{(O - E)^2}{E}$$

Expected frequency of a category: $E = np$

Degrees of freedom: $df = k - 1$ where k = number of categories.

- **Contingency Tables -- Test of Independence**

H_0: The row and column variables of contingency table are independent (i.e. not related)
H_1: The row and column variables of contingency table are NOT independent (i.e. are related)

$$\chi^2_{observed} = \Sigma\frac{(O - E)^2}{E}$$

Expected frequency of a cell: $E = \dfrac{\text{(Row total)(Column total)}}{\text{Samplesize}}$

Degrees of freedom:

$df = (R - 1)(C - 1)$

where R = # of row categories and C = # of column categories in contingency table.

- **Contingency Tables – Test of Homogeneity**

H_0: The proportions of elements that belong to different categories are the same in two or more different populations
H_1: The proportions of elements that belong to different categories are NOT the same in two or more different populations

(Note: Calculations and degrees of freedom same as for test of independence)

- **Confidence interval for population variance σ^2:**

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \quad \text{to} \quad \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \quad df = n - 1$$

where $s^2 = \dfrac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n - 1}$

(Note: Confidence interval for population standard deviation is found by taking square roots of confidence interval for population variance.)

- **Test of hypotheses about σ^2:**

$H_0: \sigma^2 = \sigma_0^2$

$H_1: \sigma^2 < \sigma_0^2$ OR $H_1: \sigma^2 > \sigma_0^2$ OR $H_1: \sigma^2 \neq \sigma_0^2$

$$\chi^2_{observed} = \frac{(n-1)s^2}{\sigma^2} \quad df = n - 1$$

Chapter 12 • Analysis of Variance

$H_0: \mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$
H_1: Not all k population means are equal

$$F_{observed} = \frac{MSB}{MSW}$$

where $MSB = SSB / (k - 1)$ and $MSW = SSW / (n - k)$

$df(numerator) = k - 1$ and $df(denominator) = n - k$

$$SSB = \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \ldots\right) - \frac{(\Sigma x)^2}{n}$$

$$SSW = \Sigma x^2 - \left(\frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} + \ldots\right)$$

k = the number of different samples or treatments
n_i = the size of sample i
T_i = the sum of all the values in sample i
n = the total number of values in all samples = $n_1 + n_2 + n_3 + \ldots$
Σx = the sum of all the values in all samples = $T_1 + T_2 + T_3 + \ldots$
Σx^2 = the sums of squares of all the values in all the samples

Chapter 13 • Simple Linear Regression

- **Simple linear regression model:** $y = A + Bx + \varepsilon$
- **Estimated regression model:**

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

where $b = SS_{xy} / SS_{xx}$ and $a = \overline{y} - b\overline{x} = \dfrac{\Sigma y}{n} - b\dfrac{\Sigma x}{n}$

$$SS_{xy} = \Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}$$

$$SS_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \quad \text{and} \quad SS_{yy} = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

- **Confidence interval for B:**

$$b \pm ts_b$$

$$s_b = s_e / \sqrt{SS_{xx}} \quad \text{and} \quad s_e = \sqrt{\frac{SS_{yy} - bSS_{xy}}{n - 2}}$$

Note: t distribution has n - 2 degrees of freedom.

- **Test of hypotheses about B:**

$H_0: B = 0$

$H_1: B > 0$ OR $H_1: B < 0$ OR $H_1: B \neq 0$

$$t_{observed} = \frac{b - B}{s_b} \quad df = n - 2$$

- **Confidence interval for $\mu_{y|x}$:**

$$\hat{y} \pm ts_{\hat{y}_m} \quad \text{where} \quad s_{\hat{y}_m} = s_e\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{SS_{xx}}} \quad \text{and} \quad df = n - 2$$

- **Prediction interval for y_p:**

$$\hat{y} \pm ts_{\hat{y}_p} \quad \text{where} \quad s_{\hat{y}_p} = s_e\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{SS_{xx}}} \quad \text{and} \quad df = n - 2$$

- **Linear correlation coefficient:**

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

- **Test of hypotheses about ρ:**

$H_0: \rho = 0$

$H_1: \rho > 0$ OR $H_1: \rho < 0$ OR $H_1: \rho \neq 0$

$$t_{observed} = r\sqrt{\frac{n - 2}{1 - r^2}} \quad df = n - 2$$

- **Coefficient of determination:**

r^2 = proportion of variation in y variable that is explained by its linear relationship with the x variable

$= bSS_{xy} / SS_{yy}$