MAT 222 FORMULA CARD

Standard score: $z = \frac{x-\mu}{\sigma}$

Standard score for sample mean $\bar{x}$: $z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Chapter 6: Introduction to Inference

Confidence interval for mean μ (σ known):

$\bar{x} \pm z^* \dfrac{\sigma}{\sqrt{n}}$,

z^*	1.645	1.960	2.576
C	90%	95%	99%

Sample size for confidence interval for μ with margin of error m:

$$n = \left[\frac{z^*\sigma}{m}\right]^2$$

z Statistic for $H_0 : \mu = \mu_0$ (σ known):

$$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$$

Chapter 7: Inference for Distributions

Standard error of $\bar{x}$: $\mathrm{SE}_{\bar{x}} = \dfrac{s}{\sqrt{n}}$

Confidence interval for mean μ (σ unknown):

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}, \qquad \text{df} = n-1$$

One-sample t Statistic for $H_0 : \mu = \mu_0$ (σ unknown):

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}, \qquad \text{df} = n-1$$

Two-sample z statistic for $H_0 : \mu_1 = \mu_2$ (σ_1, σ_2 known):

$$z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

Two-sample t statistic for $H_0 : \mu_1 = \mu_2$ (σ_1, σ_2 unknown):

$$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}},$$

df = minimum of $n_1 - 1$ and $n_2 - 1$

Two-Sample Confidence interval for $\mu_1 - \mu_2$:

$$(\bar{x}_1-\bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}},$$

df = minimum of $n_1 - 1$ and $n_2 - 1$

Pooled two-sample estimator of σ^2:

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

Pooled two-sample t statistic for $H_0 : \mu_1 = \mu_2$ when $\sigma_1 = \sigma_2$:

$$t = \frac{\bar{x}_1-\bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \qquad \text{df } = n_1+n_2-2$$

Pooled two-sample confidence interval for $\mu_1 - \mu_2$ when $\sigma_1 = \sigma_2$:

$$(\bar{x}_1-\bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \qquad \text{df } = n_1+n_2-2$$

Two-sample F statistic for $H_0 : \sigma_1 = \sigma_2$:

$$F = \frac{\text{larger } s^2}{\text{smaller } s^2}$$

Chapter 8: Inference for Proportions

Sample proportion: $\hat{p} = X/n$, X = number of "successes"

Standard error of $\hat{p}$: $\mathrm{SE}_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Confidence interval for p:

$$\hat{p} \pm z^* \mathrm{SE}_{\hat{p}}$$

z statistic for $H_0 : p = p_0$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Sample size for desired margin of error m:

$$n = \left(\frac{z^*}{m}\right)^2 p^*(1-p^*) \qquad (p^* = \text{guessed value})$$

or

$$n = \frac{1}{4}\left(\frac{z^*}{m}\right)^2 \qquad \text{(conservative approach with } p^* = 1/2\text{)}$$

Difference of sample proportions: $D = \hat{p}_1 - \hat{p}_2$

Standard error of sample difference D:

$$\mathrm{SE}_D = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Confidence interval for $p_1 - p_2$:

$$(\hat{p}_1 - \hat{p}_2) \pm z^* \mathrm{SE}_D$$

Pooled estimator of p when $p_1 = p_2$:

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

Standard error of D under $H_0 : p_1 = p_2$:

$$\mathrm{SE}_{D_p} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

z statistic for $H_0 : p_1 = p_2$:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\mathrm{SE}_{D_p}}$$

Chapter 9: Inference for Two-Way Tables

Expected cell counts:

$$\text{expected cell count} = \frac{\text{row total} \times \text{column total}}{n}$$

Chi-square test statistic:

$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

df = (# of rows – 1)(# of columns – 1)

Chapter 10: Inference for Regression

Simple Linear Regression Model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

where the ϵ_i are independent and normally distributed with mean 0 and variance σ^2.

Sample variance of x's: $s_x^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$

Sample variance of y's: $s_y^2 = \frac{1}{n-1}\sum(y_i - \bar{y})^2$

Sample correlation:

$$r = \frac{1}{n-1}\sum\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

Least-Squares Regression line: $\hat{y} = b_0 + b_1 x$,

Slope (Estimate of β_1): $b_1 = r\frac{s_y}{s_x}$

Intercept(Estimate of β_0) : $b_0 = \bar{y} - b_1\bar{x}$

Estimate of σ^2:

$$s^2 = \frac{1}{n-2}\sum e_i^2, \quad \text{where } e_i = y_i - \hat{y}_i$$

Standard error of b_0:

$$\mathrm{SE}_{b_0} = s\sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}}$$

Level C confidence interval for β_0:

$$b_0 \pm t^* \mathrm{SE}_{b_0}, \ \mathrm{df} = n-2$$

Standard error of b_1:

$$\mathrm{SE}_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$$

Level C confidence interval for β_1:

$$b_1 \pm t^* \mathrm{SE}_{b_1}, \ \mathrm{df} = n-2$$

Test statistic for $H_0 : \beta_1 = 0$:

$$t = \frac{b_1}{\mathrm{SE}_{b_1}}, \ \mathrm{df} = n-2$$

Estimate for mean response μ when $x = x^*$:

$$\hat{\mu} = b_0 + b_1 x^*$$

Standard error of $\hat{\mu}$ when $x = x^*$:

$$\mathrm{SE}_{\hat{\mu}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

Level C confidence interval for μ when $x = x^*$:

$$\hat{\mu} \pm t^* \mathrm{SE}_{\hat{\mu}}, \qquad \mathrm{df} \; = n - 2$$

Estimate for future observation of y when $x = x^*$:

$$\hat{y} = b_0 + b_1 x^*$$

Standard error of $\hat{y}$ when $x = x^*$:

$$\mathrm{SE}_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}}$$

Level C prediction interval for y when $x = x^*$:

$$\hat{y} \pm t^* \mathrm{SE}_{\hat{y}}, \qquad \mathrm{df} \; = n - 2$$

Sum of Squares

$$\mathrm{SST} = \sum (y_i - \bar{y})^2, \quad \text{(Total Sum of Squares)}$$

$$\mathrm{SSM} = \sum (\hat{y}_i - \bar{y})^2, \quad \text{(Model Sum of Squares)}$$

$$\mathrm{SSE} = \sum (y_i - \hat{y}_i)^2, \quad \text{(Error Sum of Squares)}$$

- SST=SSM+SSE
- MS $= \frac{\text{sum of squares}}{\text{degrees of freedom}}$
- $s^2 = MSE$
- $r^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = \frac{\mathrm{SSM}}{\mathrm{SST}}$

The ANOVA F test for $H_0 : \beta_1 = 0$

$$F = \frac{\mathrm{MSM}}{\mathrm{MSE}} = \frac{\mathrm{SSM/DFM}}{\mathrm{SSE/DFE}}, \; \mathrm{df} = (1, n-2)$$

Test statistic for $H_0 : \rho = 0$: $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$
df $= n - 2$.

Chapter 11: Multiple Regression

Multiple Regression Model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

Least squares estimates of $\beta_0, \beta_1, \ldots, \beta_p$:

$$b_0, b_1, \ldots, b_p$$

Estimate of σ: $s = \sqrt{\mathrm{MSE}}$.

Level C confidence interval for β_j:

$$b_j \pm t^* SE_{b_j}, \; \mathrm{df} = n - p - 1$$

Test statistic for $H_0 : \beta_j = 0$:

$$t = \frac{b_j}{\mathrm{SE}_{b_j}}, \; \mathrm{df} = n - p - 1.$$

Sum of squares SS: $\mathrm{SST} = \mathrm{SSM} + \mathrm{SSE}$

Degrees of freedom DF:

$$\mathrm{DFT} = \mathrm{DFM} + \mathrm{DFE},$$

$\mathrm{DFT} = n - 1$, $\mathrm{DFM} = p$, $\mathrm{DFE} = n - p - 1$,

Mean square model: $\mathrm{MSM} = \dfrac{\mathrm{SSM}}{\mathrm{DFM}}$

Mean square error: $\mathrm{MSE} = \dfrac{\mathrm{SSE}}{\mathrm{DFE}}$

Test statistic for $H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$:

$$F = \frac{\mathrm{MSM}}{\mathrm{MSE}}, \; \mathrm{df} = (p, n - p - 1).$$

Squared multiple correlation: $R^2 = \dfrac{\mathrm{SSM}}{\mathrm{SST}}$

Chapter 12: One-way ANOVA

One-way ANOVA model:

$$x_{ij} = \mu_i + \varepsilon_{ij},$$

for $i = 1, \dots, I$ and $j = 1, \dots, n_i$, and $N = n_1 + \cdots + n_I$.

Pooled-sample variance:

$$s_p^2 = \frac{(n_1-1)s_1^2 + \cdots (n_I - 1)s_I^2}{(n_1-1) + \cdots + (n_I - 1)} = \text{MSE}$$

Sum of squares (SS): SST = SSG + SSE

$$\begin{aligned} \text{SSG} &= \sum_{groups} n_i(\bar{x}_i - \bar{x})^2 \\ \text{SSE} &= \sum_{groups} (n_i - 1)s_i^2 \end{aligned}$$

Degrees of freedom (DF):

$$DFT = DFG + DFE,$$

where DFT $= N - 1$, DFG $= I - 1$, DFE $= N - I$.

Mean square (MS): $\text{MSG} = \frac{\text{SSG}}{\text{DFG}}$, $\text{MSE} = \frac{\text{SSE}}{\text{DFE}}$

Test statistic for $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$:

$$F = \frac{\text{MSG}}{\text{MSE}}, \ \text{df} = (I-1, N-I).$$

Coefficient of determination: $R^2 = \frac{\text{SSG}}{\text{SST}}$

Population contrast: $\psi = \sum a_i \mu_i$, where $\sum a_i = 0$

Sample contrast: $c = \sum a_i \bar{x}_i$

Standard error of c :

$$\text{SE}_c = s_p \sqrt{\sum \frac{a_i^2}{n_i}}$$

Test statistic for $H_0 : \psi = 0$:

$$t = \frac{c}{\text{SE}_c}, \ \text{df} = N - I$$

Level C confidence interval for ψ:

$$c \pm t^* \text{SE}_c, \ \text{df} = N - I.$$

Multiple Comparisons t statistic:

$$t_{ij} = \frac{\bar{x}_i - \bar{x}_j}{s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}}, \quad \text{df} = N - I$$

Simultaneous Confidence Intervals for Mean Differences:

$$(\bar{x}_i - \bar{x}_j) \pm t^{**} s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}, \quad \text{df} = N - I$$

Chapter 13: Two-way ANOVA

Factors: Two factors A and B, factor A has I levels, and factor B has J levels

Two-way ANOVA model:

$$x_{ijk} = \mu_{ij} + \varepsilon_{ijk},$$

for $i = 1, \dots, I$, $j = 1, \dots, J$ and $k = 1, \dots n_{ij}$.

Pooled-sample variance:

$$s_p^2 = \frac{\sum (n_{ij} - 1)s_{ij}^2}{\sum (n_{ij} - 1)} = \text{MSE}$$

Sum of squares (SS): SST = SSA + SSB + SSAB + SSE

Degrees of freedom (DF):

$$\text{DFT = DFA + DFB + DFAB + DFE},$$

$$\text{DFM = DFA + DFB + DFAB},$$

where

$$\begin{aligned} &\text{DFT} = N - 1, \\ &\text{DFA} = I - 1, \\ &\text{DFB} = J - 1, \\ &\text{DFAB} = (I-1)(J-1), \\ &\text{DFE} = \text{N-IJ}. \end{aligned}$$

Mean square (MS): For the factors A and B, for the interaction AB, and for the error E:

$$\text{MS} = \frac{\text{SS}}{\text{DF}}$$

Test statistic for H_0: Main effect of A is zero:

$$F = \frac{\text{MSA}}{\text{MSE}}, \quad \text{df} = (I-1, N-IJ).$$

Test statistic for H_0: Main effect of B is zero:

$$F = \frac{\text{MSB}}{\text{MSE}}, \quad \text{df} = (J-1, N-IJ).$$

Test statistic for H_0: Interaction effect of A and B is zero:

$$F = \frac{\text{MSAB}}{\text{MSE}}, \quad \text{df} = ((I-1)(J-1), N-IJ).$$