Grant MacEwan University

STAT 252 – Dr. Karen Buro

Formula Sheet

Descriptive Statistics

- Sample Mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Sample Variance: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - \frac{(\Sigma_{i=1}^{n} x_i)^2}{n}}{n-1}$
- Sample Standard Deviation: $s = \sqrt{\text{Sample Variance}} = \sqrt{s^2}$

Estimation

Parameter	Estimator	SE(Estimator)	Approximate Confidence Interval
μ	$\bar{x}$	$\frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
$\mu_1 - \mu_2$	$\bar{x}_d$	$\frac{\sigma_d}{\sqrt{n}}$	$\bar{x}_d \pm t_{\alpha/2}\frac{s_d}{\sqrt{n}}$

t-Test Statistics

- Test Statistic for 1-sample t-Test concerning μ if σ is unknown

$$t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \qquad df = n - 1$$

- Test Statistic for two–sample t-Test for Comparing Two Population Means:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \qquad df = \min(n_1 - 1, n_2 - 1)$$

- Test Statistic for paired t-Test for Comparing Two Population Means:

$$t_0 = \frac{\bar{x}_d - d_0}{\left(\frac{s_d}{\sqrt{n}}\right)} \qquad df = n - 1$$

One-way Analysis of Variance (1-way ANOVA)

- Total Sum of Squares: $TotalSS = \sum(x_{ij} - \bar{x})^2 = \sum x_{ij}^2 - \frac{(\sum x_{ij})^2}{n}, \quad df_{Total} = n - 1$ where $\bar{x}$ is the overall mean, from all k samples and $n = n_1 + n_2 + \ldots + n_k$.
- Sum of Squares for treatments: $SST = \sum n_i(\bar{x}_i - \bar{x})^2, \quad df_T = k - 1$
  with $\bar{x}_i$ = sample mean of observations in sample i.
- Sum of Squares for Error: $SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2, \quad df_E = n - k$
  with s_i is the standard deviation of observation from sample i.
- ANOVA–Table

Source	df	SS	MS=SS/df	F
Treatments/Groups	$k-1$	SST	$MST = SST/(k-1)$	MST/MSE
Error	$n-k$	SSE	$MSE = SSE/(n-k)$	
Total	$n-1$	$TotalSS$		

- Analysis of contrasts in the ANOVA model:

$$\gamma = C_1\mu_1 + C_2\mu_2 \ldots + C_k\mu_k, \quad \text{with} \quad C_1 + C_2 + \ldots C_k = 0$$

  Corresponding estimator

$$c = C_1\bar{x}_1 + C_2\bar{x}_2 \ldots + C_k\bar{x}_k, \quad SE(c) = s_p\sqrt{\frac{C_1^2}{n_1} + \frac{C_2^2}{n_2} + \ldots + \frac{C_k^2}{n_k}}$$

  with the pooled estimator for σ: $s_p = \sqrt{MSE}$.

  $(1-\alpha)100\%$ confidence interval for γ: $c \pm t^* SE(c)$, $df = n - k$

  Test statistic for a t-test concerning γ: $t_0 = \frac{c}{SE(c)}, \quad df = n - k$

Multiple Comparisons (Bonferroni Procedure)

Experiment wise error rate $=\alpha$,
m = number of comparisons,
comparison wise error rate $\alpha^* = \alpha/m$,

Then the margin of error (ME) for comparing treatment i with treatment j is:

$$ME_{ij} = t_{\alpha^*/2}^{df_E}\ \sqrt{MSE}\ \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

The Randomized Block Design
k treatments, b blocks

- Total Sum of Squares: $TotalSS = \sum (x_{ij} - \bar{x})^2 = \sum x_{ij}^2 - \frac{(\sum x_{ij})^2}{n}, \quad df_{Total} = n - 1$
  where $\bar{x}$ is the overall sample mean.
- Sum of Squares for treatments: $SST = \sum b(\bar{x}_{Ti} - \bar{x})^2, \quad df_T = k - 1$
  with $\bar{x}_{Ti}$ = sample mean of observations for treatment i.
- Sum of Squares for blocks: $SSB = \sum k(\bar{x}_{Bj} - \bar{x})^2, \quad df_B = b - 1$
  with $\bar{x}_{Bj}$ = sample mean of observations for block j.
- Sum of Squares for Error: $SSE = TotalSS - SST - SSB, \quad df_E = n - k - b + 1$
- ANOVA–Table for RBD

Source	*df*	*SS*	*MS*=*SS*/*df*	*F*
Treatments	$k-1$	SST	$MST = SST/(k-1)$	MST/MSE
Blocks	$b-1$	SSB	$MSB = SSB/(b-1)$	MSB/MSE
Error	$n-k-b+1$	SSE	$MSE = SSE/(n-k-b+1)$	
Total	$n-1$	$TotalSS$		

Factorial Design (here: 2-way ANOVA)
Factor A at a levels, factor B at b levels

ANOVA–Table for 2 factorial design

Source	*df*	*SS*	*MS*=*SS*/*df*	*F*
Factor A	$a-1$	$SS(A)$	$MS(A) = SS(A)/(a-1)$	$MS(A)/MSE$
Factor B	$b-1$	$SS(B)$	$MS(B) = SS(B)/(b-1)$	$MS(B)/MSE$
Inter $A*B$	$(a-1)(b-1)$	$SS(A*B)$	$MS(A*B) = SS(A*B)/(a-1)(b-1)$	$MS(A*B)/MSE$
Error	df_E	SSE	$MSE = SSE/df_E$	
Total	$n-1$	$TotalSS$		

$df_E = n - a - b - (a-1)(b-1) + 1$

Test statistic for Model Utility Test:

$$F = \frac{\frac{(SS(A) + SS(B) + SS(A*B))}{(a-1+b-1+(a-1)(b-1))}}{MSE}, \quad df_n = a - 1 + b - 1 + (a-1)(b-1), df_d = df_E$$

Non Parametric Methods

- Sign Test:

  Test Statistic S_0 = number of measurements supporting H_a (for 2-tailed, the larger of the two).

- Wilcoxon Rank Sum Test

  Test statistic: Both sample sizes are at least 10, T_1=sum of ranks for sample 1.

$$z_0 = \frac{T_1 - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}},$$

- Wilcoxon Signed Rank Test

  Test statistic: Sample size of pairs is at least 10, T_+=sum of ranks for positive differences.

$$z_0 = \frac{T_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}},$$

- Kruskal-Wallis H Test Test statistic:

$$H_0 = \frac{12}{n(n+1)} \sum n_i(\bar{R}_i - \frac{n+1}{2})^2, \quad df = k-1$$

  $\bar{R}_i$ = mean rank for sample i.

Multiple Comparison for Kruskal Wallis Test (Bonferroni procedure)

Choose experiment wise error rate α.

1. When comparing k distributions let $m = k(k-1)/2$ and the comparison wise error rate $\alpha^* = \alpha/m$.
2. Use the Wilcoxon Rank Sum test to compare every pair of treatments at significance level α^*.

Then all significant results hold simultaneously at experiment wise error rate of α.

Partial Correlation:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{x_1x_2} r_{yx_2}}{\sqrt{1 - r^2_{x_1x_2}}\sqrt{1 - r^2_{yx_2}}}$$

Simple Linear Regression

- $SS_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$, $SS_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$ and $SS_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$
- Equation of least squares line (regression line): slope: $\hat{\beta}_1 = \left(\frac{SS_{xy}}{SS_{xx}}\right)$, intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$
- Estimate for σ: $SSE = SS_{yy} - \hat{\beta}_1 SS_{xy}$, $df_E = n-2$, $s = \sqrt{\frac{SSE}{n-2}}$
- Pearson's Correlation Coefficient: $r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$
- Coefficient of determination: $R^2 = \frac{SS_{yy} - SSE}{SS_{yy}}$
- Test statistic for model utility test: $t_0 = \frac{\hat{\beta}_1}{s/\sqrt{SS_{xx}}}$, $df = n-2$
- $(1-\alpha) \times 100\%$ CI for β_1 : $\hat{\beta}_1 \pm t^* \frac{s}{\sqrt{SS_{xx}}}$, $df_E = n-2$
- A $(1-\alpha)100\%$ CI for $E(y)$ at $x = x_p$: $\hat{y} \pm t^* \, s \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$, $df_E = n-2$
- A $(1-\alpha)100\%$ Prediction Interval for y at $x = x_p$: $\hat{y} \pm t^* \, s \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{SS_{xx}}}$, $df_E = n-2$

Multiple Linear Regression

- $SSE = \sum (y_i - \hat{y}_i)^2$, $df_E = n-k-1$
- $\hat{\sigma} = s = \sqrt{MSE} = \sqrt{\frac{SSE}{n-k-1}}$
- Adjusted coefficient of determination: $R_a^2 = 1 - \frac{n-1}{n-k-1}\left(\frac{SSE}{SS_{yy}}\right)$
- Test statistic for model utility test: $F_0 = \frac{(SS_{yy} - SSE)/k}{MSE}$ (from ANOVA table), $df_n = k$, $df_d = df_E = n-k-1$
- $(1-\alpha) \times 100\%$ Confidence Interval for a slope parameter β_i: $\hat{\beta}_i \pm t^* s_{\hat{\beta}_i}$, $df = df_E = n-k-1$
- Test statistic for test about slope parameter β_i: $t_0 = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$, $df = n-k-1$