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: $s = \sqrt{\text{Sample Variance}} = \sqrt{s^2}$
- Median: Order the data from smallest to largest. The median M is either the unique middle value or the mean of the two middle values.
- Lower Quartile: Order the data from smallest to largest. The lower quartile Q_1 is the median of the smaller half of the values.
- Upper Quartile: Order the data from smallest to largest. The upper quartile Q_3 is the median of the upper half of the values.
- Interquartile Range (iqr) = Upper Quartile – Lower Quartile $= Q_3 - Q_1$

Sampling Distributions

- Sampling Distribution of a Sample Mean, $\bar{x}$:

$$\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Sampling Distribution of the difference of two Sample Means, $\bar{x}_1 - \bar{x}_2$ (independent samples):

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2, \quad \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- Sampling Distribution of the difference of two Sample Means, $\bar{x}_d = \bar{x}_1 - \bar{x}_2$ (paired samples):

$$\mu_{\bar{x}_d} = \mu_1 - \mu_2, \quad \sigma_{\bar{x}_d} = \sqrt{\frac{\sigma_d^2}{n}}$$

Estimation

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

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:

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

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 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:

$$t_0 = \frac{c}{SE(c)}, \quad df = n - k$$

The Randomized Block Design (1-way ANOVA)
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	$n-1$	$TotalSS$		

The 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	$n-1$	$TotalSS$		

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

For Model Utility Test use $F = MST/MSE$, where
Mean Squares for Treatment= $MST = (TotalSS - SSE)/((a-1) + (b-1) + (a-1)(b-1))$