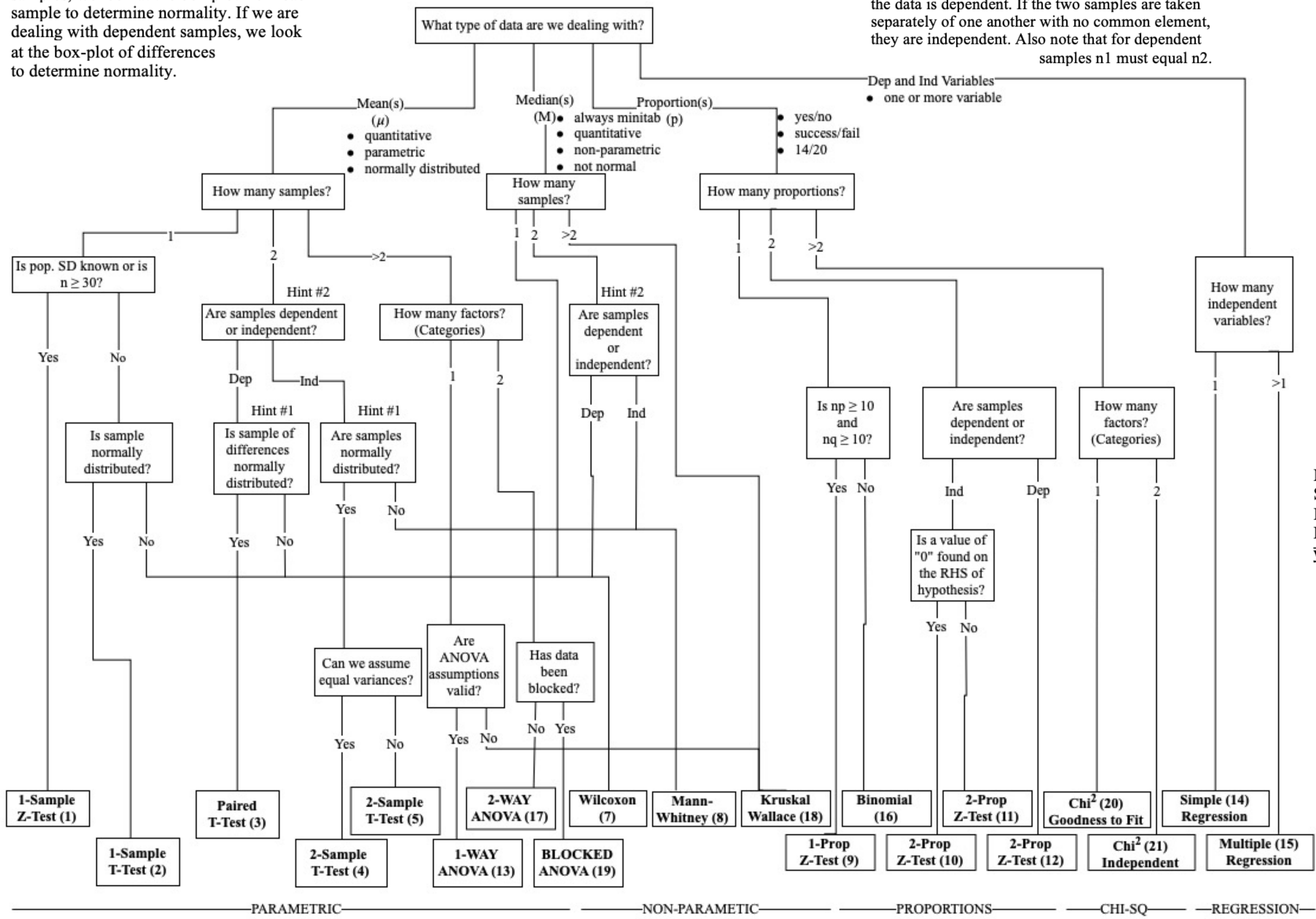
Hint #1 – If we are dealing with independent samples, we look at the box plot from each sample to determine normality. If we are dealing with dependent samples, we look at the box-plot of differences to determine normality.

Hint #2 –If each observation from the first sample is paired with an observation in the second sample, the data is dependent. If the two samples are taken separately of one another with no common element, they are independent. Also note that for dependent samples n1 must equal n2.



[1]FPCF – If a sample(s) is more than 5% of the population we must apply the FPCF. Therefore, if n/N>0.05, we multiply the denominator by √((N-n)/(N-1))

standardizing residuals means scale in standard deviations

If $F_{stat} > F_{crit}$ the p-value is $< \alpha$ from F-table. If $F_{stat} < F_{crit}$ the p-value is $> \alpha$ from F-table.

Common Z_{crit} (Z_α) Values

Alpha	($Z\alpha$) "<" in Ha 1-Tailed	($Z\alpha$) ">" in Ha 1-Tailed	($Z\alpha$) "≠" in Ha 2-Tailed
0.2	-0.84	0.84	1.28
0.1	-1.28	1.28	1.645
0.05	-1.645	1.645	1.96
0.02	-2.055	2.055	2.325
0.01	-2.325	2.325	2.575
0.005	-2.575	2.575	2.81

20

Chi Sq-Goodness to Fit

Step 1 – calculate the total of observed counts
Step 2 – Derive expected values as Described in the question
Step 3 – Calculate Chi^2 for each row
Step 4 – Add the Chi^2 from each row to get your Chi^2_{stat}
Step 5 – Finish hypothesis test

Ho: Data follows described dist.
Ha: Data follows some other dist.

$$Chi^2{}_{stat} = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

Reject Ho if $|Chi^2_{stat}| > |Chi^2_{crit}|$ or $p < \alpha$.

Table Headings for calculations:
| Group | Observed | Expected | Chi² |

21

Chi Sq-Test of Independence

Step 1 – Calculate all row totals, column totals, grand total
Step 2 – Calculate expected value for each cell (e)
Step 3 – Calculate the Chi^2 value for each cell
Step 4 – Add the Chi^2 values from all cells to get Chi^2_{stat}
Step 5 – Finish hypothesis test

Ho: Factor A and Factor B are independent
Ha: Factor A and Factor B are not independent

$$Chi^2{}_{stat} = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \qquad e_{ij} = \frac{(\text{i row total})(\text{j column total})}{\text{total sample size}}$$

Reject Ho if $|Chi^2_{stat}| > |Chi^2_{crit}|$ or $p < \alpha$.

o = observed
e = expected
df = r-1

Manual Simple Regression Calculations:

Slope = $b_1 = r*(s_x/s_y)$
Intercept = $b_0 = \bar{y} - b_1\bar{x}$
Regression equation:
$\bar{y} = b_0 + b_1\bar{x}$

r = coefficient of correlation
s_x = standard dev. of x
s_y = standard dev. of y
$\bar{x}$ = mean of x values
$\bar{y}$ = mean of y values

14

Simple Regression (see box 15 for assumptions)

Test #1-Significance of Model
Ho: $B_1 = 0$
Ha: $B_1 \neq 0$
$F_{stat} = MS_{reg}/MS_{err}$
Reject Ho if $|F_{stat}| > |F_{crit}|$ or $p < \alpha$.
$df_n = df_{reg}$ $df_d = df_{err}$

Test #1-Significance of Variable in Model
Ho: $B_n = 0$
Ha: $B_n \neq 0$ $T_{stat} = \frac{b_n - B_n}{s_{bn}}$
df = n-k-1
Reject Ho if $|T_{stat}| > |T_{crit}|$ or $p < \alpha$.
all numbers come from minitab

15

Multiple Regression
or Simple Regression
Assumptions
1) Normality
2) Equal variance
3) Independence (see other side under ANOVA)
4) Linear Model (see below)

Test #1-Significance of Variable in Model
Ho: $B_1 = B_2 = B_3 \ldots = B_n = 0$
Ha: $B_1 = B_2 = B_3 \ldots = B_n \neq 0$
$F_{stat} = MS_{reg}/MS_{err}$
Reject Ho if $|F_{stat}| > |F_{crit}|$ or $p < \alpha$.
$df_n = df_{reg}$ $df_d = df_{err}$

Test #1-Significance of Variable in Model
Ho: $B_n = 0$
Ha: $B_n \neq 0$ $T_{stat} = \frac{b_n - B_n}{s_{bn}}$
df = df_e
Reject Ho if $|T_{stat}| > |T_{crit}|$ or $p < \alpha$.
-All numbers come from minitab
-This test can be repeated for each independent variable.
CI (coefficient) = $b_n \pm T_{\alpha/2} * s_{bn}$

CI = fit $\pm T_{\alpha/2} * SE_{fit}$ (interval for average of all observations) df_{err} = n-k-1 (used for CI and PI)

CI = fit $\pm T_{\alpha/2} * \sqrt{SE_{fit}{}^2 + S^2}$ (interval for 1 observation)

-Regression equation given in output as $\hat{y} = c + B_1x_1 + B_2x_2 + \ldots + B_nx_n$
-Residual = actual value – value predicted by model by subbing independent variables into the regression equation
(actual-fit)
-The best model will be the one with the highest R^2_{adj}, the lowest S, the lowest number of independent variables and the lowest mallows CP

Variance Inflation Factor (VIF)
VIF = $1/(1-R^2)$

Multicollinearity is a problem caused by highly correlated independent variables.
It is a problem if VIF > 10.

Linear Model Assumption (Residual vs. Fit Graph)

Linear Model (good) Non- Linear Models (bad)

IMPORTANT FORMULAS
$R^2 = 1-(SS_{err}/SS_{tot})$
$R^2_{adj} = 1-(MS_{err}/(SS_{tot}/df_{tot})$
$s = \sqrt{(MSE)}$ **MS = SS/df**
coefficient of determination

$df_{reg} = k$
$df_{err} = n-k-1$
$df_{tot} = n-1$
k = # of independent variables

Correlation
-Measure the relationship between 2 variables
-Always between 1 and -1
-The closer to 1 or -1 the stronger the relationship

Ho: p = 0
Ha: p ≠ 0

Ho: p = 0
Ha: p > 0

$$T_{stat} = \frac{r}{\sqrt{\frac{1-r^2}{n-1}}}$$

df = n-2
- "r" is calculated by minitab
- Reject Ho if $|T_{stat}| > |T_{crit}|$
*see 1-tailed rules