Stat/For/Hort 571 — Midterm II, Fall 98 — Brief Solutions

1. Let p_1 and p_2 be the rates of mutation for the first and second procedures, respectively. We would like to test

$$H_0 : p_1 = p_2 \quad \text{vs.} \quad H_A : p_1 \neq p_2.$$

The estimates are $\hat{p}_1 = 9/358 \doteq 0.0251$ and $\hat{p}_2 = 16/339 \doteq 0.0472$. The test statistic is given by

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}},$$

where $\hat{p}$ is the pooled estimate of $p = p_1 = p_2$ (under the null) given by $(9+16)/(358+339) \doteq 0.0359$; thus $n_1 = 358$ and $n_2 = 339$. And $n_1\hat{p}_1 = 9, n_1(1-\hat{p}_1) = 349, n_2\hat{p}_2 = 16, n_2(1-\hat{p}_2) = 323$ so that the normal approximation is justified. The test statistic is:

$$z = -\frac{0.0221}{0.0141} = -1.57$$

with Z as the reference distribution. The p-value is $2P(Z \geq 1.57) = 2 \times 0.0582 = 0.1164$. There is very weak evidence against the null; we do not reject the null at levels commonly used in practice, such as $\alpha = 0.05$ or $\alpha = 0.10$.

2. (a) There are $k = 5$ varieties (treatments) and $n_1 = \cdots = n_5 = 6$ observations per treatment with a total of $N = 30$ observations. We have dfTrt $= k - 1 = 4$ and dfTot $= N - 1 = 29$ so that dfErr $= 25$. We compute SSErr by

$$5s_1^2 + \cdots + 5s_5^2 = 5 \times 2.427 = 12.135.$$

Noting that MS = SS/df, we obtain the ANOVA table:

Source	df	SS	MS
Legume	4	17.093	4.27325
Error	25	12.135	0.4854
Total	29	29.228	

The observed F = MSTrt/MSErr = $4.27325/0.4854 \doteq 8.80$. The reference is $F_{4,25}$. Since $P(F_{4,25} \geq 6.49) = .001$, the p-value is $< .001$; thus there is very strong evidence against the null that the population means of the five varieties are equal.

(b) Note that MSErr $= s_p^2$ is the pooled variance estimate for σ^2 and dfErr $= 25$. Let V^2 be the chi-squared distribution with 25 df and note that $P(V^2 \leq 13.12) = P(V^2 \geq 40.65) = .025$. A 95% CI is given by

$$\frac{25s_p^2}{40.65} \leq \sigma^2 \leq \frac{25s_p^2}{13.12}$$

or $0.299 \leq \sigma^2 \leq 0.925$.

3. We need to find n so that $P(\bar{Y} \geq 26.8|\mu = 28) = 0.9$. Note that $\text{Var}(\bar{Y}) = (4/\sqrt{n})^2$.

The pictures lead to the equation

$$\frac{26.8 - 28}{4/\sqrt{n}} = -1.28.$$

Solving for n gives $\sqrt{n} = 4(1.28)/1.2 \doteq 4.267$ so that $n = 18.20$. Rounding up results in $n = 19$.

4. (a) *True.* The test statistic is given by

$$\frac{\bar{y} - \mu}{s/\sqrt{n}}.$$

Since the values of $\bar{y}$ and s are such that we reject at $\alpha = 0.05$ when $n = 9$ we know that the test statistic must be larger than 2.306 in absolute value. Since $\bar{y}$ and s are the same when $n = 36$, the test statistic is now even larger in absolute value because we divide by a smaller number in the test statistic (the standard error is smaller). The test statistic certainly exceeds 2.306 and therefore also 2.030 in absolute value (note $P(T_{35} \geq 2.030) = 0.025$).

(b) *False.* The design is paired. Although we have lost the "after" data corresponding to dogs 6 and 7, the remaining data, viewed as two samples are **not** independent. In particular, the "before" and "after" data on dogs 1-5 are not independent of each other and thus, the assumptions for the independent-sample analysis are not met. Probably the best approach is to drop dogs 6 and 7 and use a paired-t analysis on dogs 1-5.

(c) *False.* One of the assumptions of the t-test is that observations form an independent random sample from a **single** population. For this problem, this is unreasonable since the leaves from the same tree are likely to be more similar than leaves from different trees. Alternatively, one could state that the leaves from a given tree are not independent of one another. Thus the t-test based on viewing the data as a single sample of size 90 is inappropriate.

5. (a) Since σ^2 is known, the reference distribution is Z. Note that $P(Z \geq 1.96) = .025$ so that the half width of the 95% CI is $1.96\sigma\sqrt{1/n_1 + 1/n_2}$. Since 1.96σ is common to all three allocations, we compute

	$\sqrt{1/n_1 + 1/n_2}$
(I)	.632
(II)	.645
(III)	.714

The **first** choice gives the shortest interval.

(b) When σ^2 is unknown, the reference distribution is T, with $n_1 + n_2 - 2$ df. The half width of the 95% CI is

$$T_{\text{df},.025}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Since s_p is assumed to be the same for all three allocations, we calculate

	df	$T_{\text{df},.025}$	$T_{\text{df},.025}\sqrt{1/n_1 + 1/n_2}$
(I)	8	2.306	1.458
(II)	13	2.160	1.394
(III)	100	1.984	1.417

The **second** choice yields the shortest interval.

Grade Distribution

```
  100:1
90-99:29
80-89:37          n = 149
70-79:28          mean = 74.83, median = 78
60-69:27          quartiles = 64, 88
50-59:11           s = 16.1
  <50:16
```