Stat/For/Hort 571 — Final, Fall 98 — Brief Solutions

1. (a) There are $k = 4$ treatments and $n_1 = \cdots = n_4 = 7$ observations per treatment. The ANOVA table is

Source	df	SS	MS
Trt	3	811.21	270.40
Error	24	1047.30	43.64
Total	27	1858.51	

The observed F = MSTrt/MSErr = 270.40/43.64 = 6.20. The reference is $F_{3,24}$ and we have $0.001 < \text{p} - \text{value} < 0.005$. There is strong evidence against the null that the population means are the same for all 4 treatments.

(b) The contrast is $\bar{y}_{\text{ctrl}} - \frac{1}{3}(\bar{y}_{\text{glu}} + \bar{y}_{\text{fru}} + \bar{y}_{\text{suc}})$; thus $\lambda_C = 1, \lambda_G = -1/3, \lambda_F = -1/3, \lambda_S = -1/3$. The sample value for the contrast is 9.90; the standard error is

$$s_p\sqrt{\frac{1}{7} + \frac{1}{9}\left(\frac{1}{7} + \frac{1}{7} + \frac{1}{7}\right)} = 2.883;$$

note $s_p = \sqrt{\text{MSErr}} = 6.606$. Since $P(T_{24} \geq 2.064) = 0.025$, the 95% confidence interval is given by

$$9.90 \pm 2.064 \times 2.883 = 9.90 \pm 5.95 = (3.95, 15.85).$$

The range of plausibility for the difference between the control treatment and the average of the 3 sugar treatments is from 3.95 to 15.85. Note that a difference of 0 is not a plausible value.

(c) Two population means are found to be significantly different with a comparison-wise error rate of 0.05 if the corresponding sample means differ by more than LSD $= 2.064 s_p\sqrt{2/7} = 7.288$. Thus:

55.7	56.4	64.0	68.6
+glucose	+fructose	+sucrose	control
______	______	______	______

Treatments not connected by an underline are significantly different.

2. (a) The slope is estimated by

$$\hat{b}_1 = \frac{\sum x_i y_i - (\sum x_i \sum y_i)/n}{\sum x_i^2 - (\sum x_i)^2/n} = -0.0653;$$

the intercept is estimated by $\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x} = 8.967$.

(b) First, note that SSError = SSTotal − SSRegression, where SSTotal $= \sum y_i^2 - (\sum y_i)^2/n$ and SSRegression $= \hat{b}_1(\sum x_i y_i - (\sum x_i \sum y_i)/n)$. Therefore, SSError = 0.8213. Since dfError $= n - 2 = 4$, the underlying regression variance σ_e^2 is estimated by $s_e^2 = \text{SSError}/4 = 0.205$. Note for dfError = 4, $P(V^2 \geq 11.14) = P(V^2 \leq 0.48) = 0.025$, so a 95% confidence interval for σ_e^2 is given by

$$\frac{\text{SSError}}{11.14} \leq \sigma_e^2 \leq \frac{\text{SSError}}{0.48} = 0.074 \leq \sigma_e^2 \leq 1.71.$$

(c) The standard error of $\hat{b}_1$ is

$$\frac{s_e}{\sqrt{\sum x_i^2 - (\sum x_i)^2/n}} = 0.007221;$$

therefore $t = (\hat{b}_1 - (-0.1))/0.007221 = 4.81$. Since the reference T-distribution has 4 df, the p-value is between 0.002 and 0.01. There is strong evidence that b_1 is different from -0.1.

3. (a) Since there are two hypotheses of interest, to maintain an experiment-wise error rate of 0.05, each hypothesis should be tested at level 0.025. Since the normal approximation to the binomial is justified, both null hypotheses can be tested by

$$z = \frac{\hat{p} - 0.5}{\sqrt{0.5 \times 0.5/100}} = \frac{\hat{p} - 0.5}{0.05}.$$

For the green coin, $z = -2.6$; for the red coin, $z = 2$. The corresponding p-values are 0.0094 and 0.0456. Thus, H_0 is rejected for the green coin but not for the red coin.

(b) The observed counts are given by

	A	B	C
Healthy	12	5	23
Diseased	45	58	57

The expected counts are given by

	A	B	C
Healthy	11.4	12.6	16.0
Diseased	45.6	50.4	64.0

Because the expected values are all > 5, the χ^2 approach is appropriate. The test statistic χ^2 is given by $\sum_{\text{allcells}}(\text{exp} - \text{obs})^2/\text{exp} = 9.598$ on 2 df. The p-value is between 0.005 and 0.01. There is strong evidence that the three fungicides are not equally effective.

4. (a) *False.* Since one well is dug in the uplands and one in the valley, this does not give a good indication of contamination *overall* in the two regions; there is no measure of variability within each region. Furthermore, the weekly readings from these wells may be correlated.

(b) *True.* The weight of a randomly selected mouse is predicted by $\bar{y}$. The variance of $\hat{Y}_p$ is $\sigma^2(1+1/n)$, where the "1/16" comes from the variability of $\bar{y}$ and the "1" comes from the uncertainty in the new observation. Thus, the standard error of $\hat{Y}_p$ is $s\sqrt{1 + 1/16} = 20.61$. Note that $P(T_{15} \geq 2.131) = 0.975$ and $43.9 = 2.131 \times 20.61$.

5. We need to find n so that $P((\hat{p}_A - \hat{p}_B) > 0.12 | p_A = 0.8, p_B = 0.6) = 0.95$. Note that $\text{Var}(\hat{p}_A - \hat{p}_B) = (0.8 \times 0.2)/n + (0.6 \times 0.4)/n = 0.4/n$.

The picture leads to the equation

$$\frac{0.12 - 0.2}{\sqrt{0.4/n}} = -1.645 \text{ (since } P(Z \geq -1.645) = 0.95).$$

Solving for n gives $\sqrt{n} = \sqrt{0.4}(1.645)/0.08 \doteq 13.005$ so that $n = 169.13$. Rounding up results in $n = 170$.

Grade Distribution:

```
90's   10    mean=63.8; median=63
80's   22
70's   28
60's   28
50's   25
40's   23
below 14
```