Stat/For/Hort 571
Midterm II, Fall 2004
Brief Solutions

1. (a) TRUE — From the tables, $P(T \leq 1.533) = 1 - P(T \geq 1.533) = 0.90$ (Check the directionality!!) Similarly, $P(Z \leq 1.533) = 1 - P(Z \geq 1.533) \approx 0.938$. Alternatively, since the t distribution is known to have fatter tails than the Z, the probability that t is above 1.533 (in the right hand tail) is smaller than the corresponding probability for Z. Considering the directionality, the conclusion follows. (A picture might be helpful!!)

   (b) FALSE — The given 'reasoning' essentially is a 'dressed-up' version of the argument that scientific significance can outweigh statistical significance. In practice, one needs both. A claim about what might have occurred means practically nothing. Inference rests on the data actually obtained. (Issues regarding df are not relevant here.) The most one can say for X is that if the *potential* scientific importance is great enough, perhaps a new experiment with larger n might be conducted.

2. This is a two-independent sample situation.

   (a) Let Y_A and Y_B be the concentrations on A and B respectively and let μ_A and μ_B be the respective population means. The needed summary statistics are $\bar{y_A} = 6.1$, $s_A^2 = 0.8733$, $\bar{y_B} = 4.9$, and $s_B^2 = 0.74$. Then, due to the balanced data, $s_p^2 = (s_A^2 + s_B^2)/2 = .8067$. The CI for $\mu_A - \mu_B$ is $\bar{y_A} - \bar{y_B} \pm t * \sqrt{s_p^2(1/n_A + 1/n_B)}$. Since s_p^2 has 6 df, the appropriate t-value is 1.943. Thus, $(-0.03 < \mu_A - \mu_B < 2.43)$.

   (b) The null hypothesis is written: $H_0 : \mu_A - \mu_B = 0.3$. We notice that the value of 0.3 *is* contained within the CI above. Thus, we can conclude that the $p - value > 0.10$ for the given test.

3. The appropriate formula for the pooled variance is: $s_p^2 = (\sum_{i=1}^{k}(n_i - 1)s_i^2)/(N - k)$ where k is the number of treatments and N is the total number of observations. You are given the standard deviations so you must square them. $s_p^2 = 157.02$. There are $N - k = 46$ degrees of freedom associated with this variance. This can be thought of as the addition of the number of df for each variety $(9 + 9 + 6 + 5 + 8 + 9)$.

4. (a) The random variable, Y is the number of tigers with the bacteria present. A reasonable model (given the available information) is $Y \sim B(15, p)$. The hypotheses can be written: $H_0 : p = 0.05$ vs $H_A : p > 0.05$. Evidence against the null is obtained for 'large' observed values of Y. Using the basic definition of p-values, $p - val = P(Y \geq 2)$. This can be written as $1 - (P(Y = 0) + P(Y = 1))$. Using the binomial formula results in $p - val = 0.171$. This means that there is no meaningful evidence against the claim that the rate of occurrence of this bacteria is 0.05 or less.

   (b) The random variable Y is the number of cats with bacteria present. A reasonable model is $Y \sim B(60, p)$. $H_0 : p = 0.15$ vs $H_A : p \geq 0.15$. Since $np(= 9)$ and $n(1 - p)(= 51)$ are both larger than 5, the normal approx may be used. Let $\hat{p} = Y/60$. Then, under H_0, $\hat{p_{NA}} \sim N(.15, (.0461)^2)$. $\alpha = P(\hat{p} \geq 0.25) = P(Z \geq 2.17) = .015$.

5. (a) The general form for a CI for the difference between two proportions is: $(\hat{p_A} - \hat{p_B}) \pm Z_{\alpha/2} * \sqrt{\frac{p_A*(1-p_A)}{n_A} + \frac{p_B*(1-p_B)}{n_B}}$. Because $p_A = 0.4$ and $p_B = 0.6$, both expressions of the form $p(1 - p)$ are the same. Since $Z_{\alpha/2}$ affects all CIs equally (given the same $1 - \alpha$), we need to minimize $\sqrt{\frac{.24}{n_A} + \frac{.24}{n_B}}$. This is 0.163 for choice (1) and 0.107 for choice (2). Choice (2) has smaller width.

   (b) By looking at the expression above, the CI width is minimized when $\frac{1}{n_A} + \frac{1}{n_B}$ is as small as possible with $n_A + n_B = 100$. Based on part(a), the width is smaller for the values of n_A and n_B that are closer together. Thus the 'natural' conjecture is that the CI width can be minimized if $n_A = n_B = 50$. This is the correct answer. It is primarily for this reason that most scientific studies comparing 2 (or more) groups have equal sample sizes.

Grade Distribution

```
90-99:22
80-89:41
70-79:26          median =   79
60-69:17
50-59:13
below: 8
```