TENTAMEN: Experimental design (March 15, 2011)

Short solutions

1) a) When we compare two dependent samples (X and Y), where we have pairwice observations (twins, pair of shoes,blood pressure before and after taking some medicin,...). We assume that observed differences $X_i - Y_i$ are a sample of $N(\mu, \sigma^2)$, where μ is the expected value for the difference and σ^2 is the variance of the difference.

   b) Sometimes it is not feasible to run all experiments in random order (and often random order would increase variance), and split-plot designs can be used. Split-plot designs allow some contrasts of interest to be estimated with great precision.

   c) Whenever possible, randomize. (You can also explain the randomization test here.)

2) a) $D = -ABC$ and therefore, $I = -ABCD$

   b) $l_A = A - BCD$, $l_B = B - ACD$, $l_C = C - ABD$, $l_D = D - ABC$, $l_{AB} = AB - CD$, $l_{AC} = AC - BD$, $l_{AD} = AD - BC$, $l_{BC} = BC - AD$, $l_{BD} = BD - AC$, and $l_{CD} = CD - AB$

   c) IV

   d) $L_A = (-12 - 16 - 11 + 18 + 10 + 20 - 15 + 11)/4 = 1.25$ is a biased estimator for the main effect of factor A if $BCD = 0$ (and the error term has expectation 0).

3) a) The ANOVA table becomes

Source	df	SS	MS	F-ratio	p-value
Soil	2	38	19.0	2.34	$0.15 > 0.05$
Residual	9	73	8.1		
Total	11	111			

   b) No, the p-value is larger than 0.05 and therefore, we cannot say that there is a significant difference between the soil preparation methods at the 5% significance level. We have assumed independent errors that are $N(0, \sigma^2)$ distributed.

   c) Do a randomized block design by having locations as blocks. Then, you would remove the error coming from the difference between the locations from the error term. (In fact, then, you would be able to reject the hypothesis that there is no difference between the methods.)

d) A 95% confidence interval for the difference of means (based on pairwice observations within locations) is $\bar{d} \pm t^{4-1}_{0.025} s_{\bar{D}} = 3.5 \pm 1.6$, where $\bar{d} = 3.5$, $t^{3}_{0.025} = 3.182$, $s_{\bar{D}} = s_D/\sqrt{4} = 0.5$ and s_D is the standard deviation of the difference $D = X_B - X_A$.

4) Let us assume that the frequency X (in %) is $N(\mu, \sigma^2)$-distributed, where $\sigma^2 = 0.25$. We want to test the hypothesis $H_0 : \mu = 1.0$ against $H_1 : \mu > 1.0$. The power $P(H_0 \text{ rejected}|H_1 \text{ true})$ is 0.99 and $\alpha = 0.05$. Then,
$P(\frac{\bar{X}-1.0}{0.5/\sqrt{n}} \geq z_{0.05}|\mu = 1.2) = P(\frac{\bar{X}-1.2}{0.5/\sqrt{n}} \geq z_{0.05} - \frac{0.2}{0.5/\sqrt{n}})$
$= P(Z \geq 1.645 - 0.4\sqrt{n}) = 0.99$,
where $Z \sim N(0, 1)$. We obtain $n = 99$ by solving the equation $1.645 - 0.4\sqrt{n} = -2.33$.

5) a) $A = \frac{\text{temperature}-90}{10}$ and $B = \frac{\text{pressure}-97.5}{22.5}$

b) $y = 62.75 - 1.05A + 0.9B$ (you get the parameter estimates by computing the main effects of the two factors)

c) Curvature $\mathcal{K} = \bar{y}_f - \bar{y}_c$, where $\bar{y}_f$ is the mean value of the response variable at the design points and $\bar{y}_c$ the mean value at the middle point. Test $H_0 : \mathcal{K} = 0$ against $H_1 : \mathcal{K} \neq 0$ by a T-test with the test statistic $\hat{\mathcal{K}}/S_{\mathcal{K}}$ which has t_{k-1} distribution. Here, $\hat{\mathcal{K}} = -5.95$, $S_{\mathcal{K}} = 0.867$, $k = 5$, and the value of the test statistic becomes -6.86. Since the test statistic is much smaller than $-t^{(4)}_{0.025} = -2.776$, the null hypothesis can be rejected at the level of significance 0.05. This means that the plane model is not good.

d) Fit a higher (second) order model, which includes the interaction term and A^2 and B^2.

e) We have assumed independent $N(0, \sigma^2)$ distributed errors.