Norwegian University of Science and Technology
Department of Mathematical Sciences

Tentative solutions to TMA4255 Applied Statistics, May 25, 2012

Problem 1 Vitamin C

a) Hypotheses:
The text states that the researchers *wanted to know if there was a difference in the response measure for the two different supplements*, which means that a two-sided hypothesis should be used.

$$H_0 : \mu - \eta = 0 \text{ vs. } H_1 : \mu - \eta \neq 0$$

Assumptions:
We assume that the data are normally distributed, that is, $X_i \sim N(\mu, \sigma^2)$ and $Y_j \sim N(\eta, \tau^2)$, $i = 1, \ldots, n_1$ and $j = 1, \ldots, n_2$, and that the two samples are independent.

Equal variances:
Assuming data to be normally distributed we may test the equality of variance by performing an F-test. The null and alternative hypothesis:

$$H_0 : \sigma^2/\tau^2 = 1 \text{ vs. } H_1 : \sigma^2/\tau^2 \neq 1$$

Let S_1^2 and S_2^2 be the variances of two independent random samples of size n_1 and n_2 taken from normal populations with variances σ^2 and τ^2, respectively, then

$$F = \frac{S_1^2/\sigma^2}{S_2^2/\tau^2}$$

has an F-distribution with n_1-1 and n_2-1 degrees of freedom. Under the null $F = S_1^2/S_2^2$ and since we have a two-sided test we reject the null when $f_{obs} > f_{\alpha/2,n_1-1,n_2-1}$ or when $f_{obs} < f_{1-\alpha/2,n_1-1,n_2-1}$. We only have tables for small values for α, so we need to use the relationship

$$f_{1-\alpha,\nu_1,\nu_2} = \frac{1}{f_{\alpha,\nu_2,\nu_1}}$$

From our data we have $f_{obs} = 4.44^2/2.77^2 = 2.57$ and with $\alpha = 0.02$ (why: only tables for 0.05 and 0.01 in the textbook) we find from the tables that the critical values are $f_{0.01,9,9} = 5.35$ and $f_{0.99,9,9} = 1/f_{0.01,9,9} = 1/5.35 = 0.187$. Thus we do not reject the null and conclude that we may assume that the variances are equal. Comment: using $\alpha = 0.05$ (not in the tables of the textbook) would give a cut-off of 4.03 - and thus the same conclusion as for $\alpha = 0.02$. (Comment: there is a rule of thumb stating that the ratio of estimated variances should be less than 4 to assume equal variances, given not too small sample sizes.)

t-test:
We choose to use a pooled estimate for the variance, $S_p^2 = \frac{(n_1-1)\cdot S_1^2+(n_2-1)\cdot S_2^2}{n_1-n_2-2}$ and get $s_p = \sqrt{\frac{9\cdot 4.44^2+9\cdot 2.77^2}{10+10-2}} = \sqrt{13.7} = 3.7$. The t-test is based on the t-statistic

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} S_p}$$

which we calculate to be $t_{obs} = \frac{13.18-8}{\sqrt{\frac{1}{10}+\frac{1}{10}}\cdot 3.7} = \frac{5.18}{1.65} = 3.13$. This is a two-sided test, and using significance level $\alpha = 0.05$ we reject the null hypothesis when $t_{obs} > t_{\alpha/2,n_1+n_2-2}$ or when $t_{obs} < t_{1-\alpha/2,n_1+n_2-2}$. From the tables we find that the critical values are $t_{0.025,18} = 2.101$ and $t_{0.975,18} = -2.101$. We have observed a value more extreme than the critical values and we reject the null hypothesis.

Conclusion:
We have reason to believe that there is a difference in odontoblast cell length for the two supplements.

b) Assumptions:
The Wilcoxon rank-sum test is a so-called nonparametric test, and have no underlying parametrical distributional assumptions. However, the distributions need to be continuous and the shape of the distribution for the two populations should be equal. We test for quality of the location parameter (mean or median). The Wilcoxon rank-sum test is used for two independent samples.

When data are not normal the Wilcoxon rank-sum test should be used instead of the two-sample t-test. The Wilcoxon rank-sum test may also be used on normal data, but will then give lower power than the t-test. The Wilcoxon rank-sum test is robust towards outliers since only ranks are considered.

Wilcoxon rank-sum test.
We have equal sample sizes and choose supplement 1 to be sample 1. The sum of ranks for supplement 1 is $w_1 = 7+8+9+10+11+16+17+18+19+20 = 135$. We then need to calculate $u_1 = w_1 - n_1\cdot(n_1+1)/2 = 135-10\cdot 11/2 = 80$. Since this is a two-sided test we also need to calculate w_2 and u_2. $w_2 = (n_1+n_2)\cdot(n_1+n_2+1)/2 - w_1 = 20\cdot 21/2 - 135 =$

$210 - 135 = 75$, and $u_2 = w_2 - n_2 \cdot (n_2 + 1)/2 = 75 - 10 \cdot 11/2 = 75 - 55 = 20$. To find the critical value we look to table A.17 in the textbook, and find "two-tailed test" and $\alpha = 0.05$ and $n_1 = 10$ and $n_2 = 10$ and read off 23. Thus, reject the null hypothesis if $\min(u_1, u_2) \leq 23$. We have $u_2 = 20$ and thus reject the null hypothesis. If would also be possible to use approximate inference based on a normal approximation since both n_1 and n_2 exceeds 8.

Conclusion:
Reject the null hypothesis. We have reason to believe that there is a difference in odontoblast cell length for the two supplements.

Analyses in R:

```
> oj <- c(8.2,9.4,9.6,9.7,10.0,14.5,15.2,16.1,17.6,21.5)
> aa <- c(4.2,5.2,5.8,6.4,7.0,7.3,10.1,11.2,11.3,11.5)
> var.test(oj,aa)

F test to compare two variances

data:  oj and aa
F = 2.5701, num df = 9, denom df = 9, p-value = 0.1759
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.6383833 10.3473187
sample estimates:
ratio of variances
          2.570128

> t.test(oj,aa,var.equal=TRUE)

Two Sample t-test

data:  oj and aa
t = 3.1319, df = 18, p-value = 0.005762
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.705205 8.654795
sample estimates:
mean of x mean of y
    13.18      8.00

> wilcox.test(oj,aa)
```

```
Wilcoxon rank sum test

data:  oj and aa
W = 80, p-value = 0.02323
alternative hypothesis: true location shift is not equal to 0
```

Problem 2 Chemical yield

a) Fitted regression model:

$$\hat{y} = 181.8 + 2.146x_1 - 0.0440x_2 + 0.0007774x_1x_2$$

Estimate for σ^2?
From the fit: $S^2 = 27.2502^2 = \text{MSE} = 743$.

Hypothesis for x_2:

$$\beta_2 = 0 \text{ vs. } \beta_2 \neq 0$$

The p-value is found to be 0.678. Given that the truth is that $\beta_2 = 0$ there is a 0.678 probability to observe a test statistic T which is at least as extreme ($t_{obs} \leq -0.42$ or $t_{obs} \geq 0.42$) as what we have observed.

Good model?

- Linearity: looking at the scatter plots we see a linear trend in $x1$ and $x_1 \cdot x_2$ vs. y. We also see that x_1 and x_2 are approximately independent (this is by construction). In the plot of the studentized residuals vs. fitted value we see no clear trend, and thus may assume that linearity in the parameters of the model may be an adequate assumption.
- Covariates included in the model: Only the x_1 covariate gives a p-value below 0.05 when testing each of the covariates, and the x_2 is from the scatter plots seen to have low correlation with y. We may try to refit the model without the x_2 term, this will also change the estimated coefficients for the other covariates. In an overall level the regression is found to explain more than just the average yield level (p-value for the regression is $1.8e-10$).
- Normality of errors: looking at the qq-plot for the studentized residuals the assumption of normality seems plausible.
- Explanatory powers: the model explains 93.8% of the variability of the data, which is a high number.

Conclusion: the model seem to be adequate, but we may investigate dropping the x_2 term from the model and assess the new fit.

b) Mallow's C_p:(from the textbook)

$$C_p = p + \frac{(s^2 - \hat{\sigma}^2)(n-p)}{\hat{\sigma}^2}$$

where p=number of parameters estimated, n=number of observations, s^2= estimated variance (MSE) of model under investigation, $\hat{\sigma}^2$=estimated variance of the most complete model (Model A).

Missing C_p:

$$C_p = 3 + \frac{(885 - 743) \cdot (21 - 3)}{743} = 6.44$$

Using C_p:
We are in general looking for a small value for C_p. A rule of thumb is that we would like a model where $C_p \approx p$. A too high C_p may indicate a model that is underfitted (not explaining variability), and a too low C_p may indicate a model that is overfitting the data. By default $C_p = p$ for the model we use as the most complete model (Model A).

Compare models:
Looking first at C_p we see that model 4 and 5 are good candiates. Both are fitting 3 parameters and should have C_p around 3. If we in addition look at the R^2 for only these two models (since they have the same number of parameters fitted we may use R^2 to compare them), we see that model 5 explains 93.7% of the variance in the data whild model 4 explains 93.3%. From this I would recommend model 5. But, we also need to examine residual plots and model fit for this model in order to conclude.

BTW: data were simulated from model 5.

Problem 3 Treatment of tennis elbow

a) Hypotheses:
Let p_A, p_B and p_c be the success probabilities for the three treatments.

$$H_0 : p_A = p_B = p_C \text{ vs. } H_1 : \text{at least one pair differs}$$

We will use a χ^2-test for homogeniety, where the test statistic approximately follows a χ^2-distribution with $(c-1) \cdot (r-1)$ degrees of freedom. Here $c = 3$ and $r = 2$ (or the other way around), yielding $2 \cdot 1 = 2$ degrees of freedom.

Expected frequencies are calculated as (column totals)·(row totals)/(grand total). The table of observed and expected frequencies are as follows.

Result	A	B	C	Total
Failure	22 (26.8)	14 (27.7)	44 (25.5)	80
Success	41 (36.2)	51 (37.3)	16 (34.5)	108
Total	63	65	60	188

Showing how the Failure and A cell expected value is calculated: $80 \cdot 63/188 = 26.8$. The contribution from this cell to the test statistic is $\frac{(22-26.8)^2}{26.8} = 0.86$

The test statistic consists of 6 terms, and is given as

$$X^2 = \frac{(22-26.8)^2}{26.8} + \frac{(14-27.7)^2}{27.7} + \cdots + \frac{(16-34.5)^2}{34.5} = 36.6$$

The null hypothesis is rejected if the test statistics is larger than $\chi^2_{0.05,2} = 5.991$. Clearly, the null hypothesis is rejected.

Assumptions:
This test is an approximate test and should not be used for small sample sizes. The rule of thumb is that no cells should have expected count equal to or less than 5.

Conclusion:
There is reason to believe at at least two of the treatments have different success rates.

b) Hypotheses:
Let μ_A, μ_B and μ_c be the expected pain-free grip force for each of the the three treatments.

$$H_0 : \mu_A = \mu_B = \mu_C \text{ vs. } H_1 : \text{at least one pair differs}$$

This hypothesis can be tested using one-way analysis of variance. We need to fill in the ANOVA table (SS, MS, df, F), which can be calculated from the summary statistics.

Let $\bar{x}_A$ denote the average and s_A the standard deviation of treatment A. Ditto for treatments B and C. Let $\bar{x}$ denote the grand mean.

$$\begin{aligned} SSA &= n_A(\bar{x}_A - \bar{x})^2 + n_B(\bar{x}_B - \bar{x})^2 + n_C(\bar{x}_C - \bar{x})^2 \\ &= 63 \cdot (70.2 - 69.0)^2 + 65 \cdot (83.6 - 69.0)^2 + 60 \cdot (51.8 - 69.0)^2 \\ &= 31697 \\ SSE &= (n_A - 1)s_A^2 + (n_B - 1)s_B^2 + (n_C - 1)s_C^2 \\ &= 62 \cdot 25.4^2 + 64 \cdot 22.9^2 + 59 \cdot 23.0^2 \\ &= 104773 \end{aligned}$$

Source	SS	df	MS	F
Treatment	31697	2	15849	28
Error	104773	185	566	
Total	136470	187		

The F statistic, here observed to be 28, should be compared with the critical value $f_{0.05,2,185} = 3.0$, and we thus reject the null hypothesis.

Assumptions:
The one-way ANOVA model is

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

where the error terms are independent and normally distributed with the same variance across treatment groups.

Conclusion:
There is reason to believe that the expected pain-free grip force is not the same for all the treatments.

c) Remark: $n_1 = n_2 = n_3 = 63$ only here.

If you did b) above you may use the pooled estimate for σ, which is $\sqrt{MSE}$ from the ANOVA, $s = \sqrt{566} = 23.8$ and the same $n = 63$ for all treatment groups. But, stricktly speaking we should recalculate σ to be $s_p = \sqrt{\frac{(n_A-1)s_A^2+(n_B-1)s_B^2+(n_C-1)s_C^2}{n_A+n_B+n_3-3}} = 23.8$. So, no real change here.

Tukey's method: we construct confidence intervals for expected differences (i, j):

$$\bar{x}_i - \bar{x}_j \pm q(\alpha, 3, 185) \cdot \sqrt{s^2/n}$$

Here $q(0.05, 3, 185) = 3.32$ from Appendix A.12 from the textbook, so $q(0.05, 3, 185) \cdot s/\sqrt{n} = 3.32 \cdot 23.8/\sqrt{63} = 10.0$. This gives the following three confidence intervals at overall level 95%.

$$\begin{aligned}
\text{A vs B} &: 70.2 - 83.6 \pm 10.0 = [-23.4, -3.4] \\
\text{A vs C} &: 70.2 - 51.8 \pm 10.0 = [8.4, 28.4] \\
\text{B vs C} &: 83.6 - 51.8 \pm 10.0 = [21.8, 41.8]
\end{aligned}$$

At overall significance level 5% there are signficant differences between all treatment pairs.

Assumptions:
The same assumptions as for the one-way ANOVA need to be made. This is a post hoc test expecially tailored for performing all pairwise comparisons in the one-way ANOVA.

What is the individual confidence level used for each of the comparisons?
If we would make individual confidence intervals confidence level $(1-\delta)$ we would use a t-interval:

$$\bar{x}_A - \bar{x}_B \pm t_{\delta/2,185} \cdot \sqrt{2 \cdot s^2/n}$$

Comparing this to the Tukey interval we see that the factor $\sqrt{2}$ for the t-interval is included in the critical value q. The equation we need to solve is

$$t_{\delta/2,185} \cdot \sqrt{2} = 3.32$$
$$t_{\delta/2,185} = 3.32/\sqrt{2} = 2.348$$

We consult the table for critical values for the t-distribution and find that $t_{0.01,\infty} = 2.326$ and $t_{0.0075,\infty} = 2.432$. This means that $\delta/2$ is between 0.01 and 0.0075, meaning that the individual CI level is between $1 - 0.01 \cdot 2 = 0.98$ and $1 - 0.0075 \cdot 2 = 0.985$. The individual CI level will be in the interval [98%, 98.5%].

d) Let $\bar{X}_A$ be the mean of a random sample with physiotherapy and $\bar{X}_C$ the mean of a random sample with wait-and-see. A natural estimator for γ is

$$\hat{\gamma} = \frac{\bar{X}_A - \bar{X}_C}{\bar{X}_C}$$

We turn to first order Taylor approximations with

$$h(\bar{X}_A, \bar{X}_C) = \frac{\bar{X}_A - \bar{X}_C}{\bar{X}_C} = \frac{\bar{X}_A}{\bar{X}_C} - 1$$
$$\frac{\partial h(\bar{X}_A, \bar{X}_C)}{\partial \bar{X}_A}) = \frac{1}{\bar{X}_C}$$
$$\frac{\partial h(\bar{X}_A, \bar{X}_C)}{\partial \bar{X}_C}) = -\frac{\bar{X}_A}{\bar{X}_C^2}$$

where the random variable $\bar{X}_A$ has $\mathrm{E}(\bar{X}_A) = \mu_A$ and $\mathrm{Var}(\bar{X}_A) = \sigma_A^2/n_A$, and $\bar{X}_A$ has $\mathrm{E}(\bar{X}_C) = \mu_C$ and $\mathrm{Var}(\bar{X}_C) = \sigma_C^2/n_C$.

Define

$$h'_A(\mu_A, \mu_C) = \frac{\partial h(\bar{X}_A, \bar{X}_C)}{\partial \bar{X}_A} \mid_{\bar{X}_A=\mu_A, \bar{X}_C=\mu_C} = \frac{1}{\mu_C}$$
$$h'_C(\mu_A, \mu_C) = \frac{\partial h(\bar{X}_A, \bar{X}_C)}{\partial \bar{X}_C} \mid_{\bar{X}_A=\mu_A, \bar{X}_C=\mu_C} = -\frac{\mu_A}{\mu_C}$$

The first order Taylor approximation for two independent samples:

$$\begin{aligned}
\mathrm{E}(h(\bar{X}_A, \bar{X}_C)) &\approx h(\mu_A, \mu_C) = \frac{\mu_A}{\mu_C} - 1 = \frac{\mu_A}{\mu_C} - 1 \\
\mathrm{Var}(h(\bar{X}_A, \bar{X}_C)) &\approx (h'_A(\mu_A, \mu_C))^2 \mathrm{Var}(\bar{X}_A) + (h'_C(\mu_A, \mu_C))^2 \mathrm{Var}(\bar{X}_C) \\
&= (\frac{1}{\mu_C}))^2 \cdot \sigma_A^2/n_A + (-\frac{\mu_A}{\mu_C^2})^2 \cdot \sigma_C^2/n_C
\end{aligned}$$

Estimates using numerical values $n_A = 63$, $n_C = 60$, $\hat{\mu}_A = \bar{x}_A = 70.2$, $\hat{\mu}_C = \bar{x}_C = 51.8$, $\hat{\sigma}_A^2 = s_A^2 = 25.4^2$, $\hat{\sigma}_C^2 = s_C^2 = 23.0^2$ are as follows.

$$\begin{aligned}
\hat{\mathrm{E}}(h(\bar{X}_A, \bar{X}_C)) &\approx \frac{70.2}{51.8} - 1 = 0.355 \\
\hat{\mathrm{Var}}(h(\bar{X}_A, \bar{X}_C)) &\approx (\frac{1}{51.8})^2 \cdot 25.4^2/63 + (\frac{70.2}{51.8^2})^2 \cdot 23.0^2/60 \\
&= 0.0038 + 0.0060 = 0.0098
\end{aligned}$$

We estimate a 0.355 increase in mean pain-free grip force when comparing wait-and-see with physiotherapy, with an estimated variance of 0.0098, and hence estimated standard deviation of 0.099. Thus a 35.5% increase and a 0.99 percent point standard deviation for the increase.

If the samples were not independent this will not influence the estimated increase in mean pain-free grip force, but will change the estimated variance. To the formula under independence a last term is added.

$$\begin{aligned}
\mathrm{Var}(h(\bar{X}_A, \bar{X}_C)) \approx [h'_A(\mu_A, \mu_C)]^2 \mathrm{Var}(\bar{X}_A) + [h'_A(\mu_A, \mu_C)]^2 \mathrm{Var}(\bar{X}_C) + \\
2 \cdot h'_A(\mu_A, \mu_C) \cdot h'_C(\mu_A, \mu_C) \mathrm{Cov}(\bar{X}_A, \bar{X}_C)
\end{aligned}$$

If the dependence is positive (positive covariance between pain-free grip forces) then the estimated variance will decrease (as compared with the independence case) since the $h'_C(\mu_A, \mu_C)$ is negative, while for negative correlation the estimated variance will increase.