## Solutions

**Experimental design (MSA250/TMS031)** Wednesday, March 14th, 2018, 8:30 - 12:30
**Tools:** A Chalmers accepted pocket calculator with emptied memory. At the examination, sheets with statistical distributions and tables will be handed out.
**Maximum number of points:** 30p
**Limits GU:** G (15p) and VG (22p)
**Limits Chalmers:** 3:a (15p), 4:a (20p) and 5:a (25p)
Give explanations to the notation you use and motivation to your conclusions.

1. (2+2+2 p)
   (a) Explain the term "Lack of Fit" and exemplify.
   (b) Describe the term "Design Information Function".
   (c) When do you want to use a higher order model, and how can you find out that you should do that?

2. (5 p) Suppose you would like to repeat Fisher's famous tea experiment, but with the twist of improving the classical design of four cups with milk added to the tea infusion and four cups with tea infusion added to the milk. You consider the following three variants. Analyze them especially with respect to what would be the outcomes where you would reject the null-hypothesis that the tea-taster does not have the ability to tell the difference between the two kinds of tea being served. Which of the test would you suggest in the end?
   (a) You will still serve only eight cups but without the restraint that it should be four of each kind, but instead at probability $\frac{1}{2}$ pick a cup where the milk is added afterwards.
   (b) You serve 12 cups, 6 of each kind (a fact that is revealed to the taster).
   (c) You repeat the classical design ten times.

   Short answer:

   (a) You can reject the null hypothesis if the object gets at least seven correct answers with a p-value of
   $$\frac{1}{2^8} + \frac{\binom{8}{1}}{2^8} = \frac{9}{256} \approx 3.5\%.$$
   (b) You could reject the null hypothesis with p-value $\frac{37}{924} \approx 4\%$ if the tester gets at least 5 correct cups out of 6.
   (c) You can reject the null hypothesis even if there would be one or more mistakes in eight out of the ten trials with p-value
   $$1 - (\binom{10}{1}\left(\frac{69}{70}\right)^9 \frac{1}{70} + \left(\frac{69}{70}\right)^{10}) = 0.85\%,$$
   where $70 = \binom{8}{4}$, the number of all combinations of picking 4 cups out of 8.

3. (5 p) You have been given the assignment from a food magazine "Mat är allt" to give a statistical interpretation of a randomized paired design test, at five different malls in Sweden, of which soda that is most "tasty" of Coca Shmola (A) or Preppsi (B). Fourty mall visitors were randomly choosen in each mall and were asked to blindly decide which glass was more tasty. What will be your conclusion and argument?

| A | B | B | A | B | A | A | B | B | A |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 25 | 29 | 11 | 21 | 19 | 14 | 26 | 18 | 22 |

Short answer By looking at the differences $B - A$ for each mall, we get the following vector $(10, 18, 2, 12, -4)$. The question is then if we can say that Preppsi really is more "tasty" than Coca Shmola. We can choose that to be our alternative hypothesis (i.e. one-sided), or we can choose the alternative hypothesis that there is a difference (two-sided). By looking at randomized versions we get $2^5 = 32$ different configuration where only one is strictly more B-positive (and two strictly more extreme if we want to test the balanced alternative hypothesis). Now, $\frac{2}{32} \approx 0.0625$ and by performing a t-test we conclude that we can not reject the null-hypotheses with a p-value of 0.06 (or 0.12 two sided) and significance level of 5 %. Furthermore, the corresponding confidence interval is $[-3.1, 18.3]$. Hence the test is inconclusive (for both one- or two-sided) and the journalists at "Mat är allt" will not be too happy about that unfortunately.

4. (5 p) Let us have a completely randomized $2^4$ factorial design experiment with four factors A, B, C, and D. The treatment combinations (runs) are divided into four blocks by using ABC and ABD as blocking factors.

   (a) Why is blocking used in factorial designs?

   (b) Give the runs that belong to each of the four blocks?

   (c) Which additional effects (if any) are confounded with the block effect?

   (d) Is this a good choice of dividing the runs into blocks? Why/why not?

   (e) Would it have been better to choose ABCD and one of the three factor interactions or two of the two factor interactions as blocking factors? Why/why not?

   Short answer:

   (a) Blocking is used to neutralize the effect of e.g. having two people making the experiments.

   (b) Combine the runs where ABC and ABD are both - -, -+, +- or ++, four runs per block.

   (c) CD

   (d) Yes, since all main effects are free from confounding with the block effect.

   (e) No. If e.g. ABCD and ABC had been used, the main effect of D would have been confounded with the block effect. If two of the two factor interactions had been used, two (not one) of the two factor interaction effects had been confounded with the block effect.

5. (4 p) Design an eight-run fractional factorial design for an experimenter with the following five factors: temperature, concentration, pH, agitation rate and catalyst type. She tells you that she is particularly concerned about the two-factor interactions between the temperature and concentration and between catalyst type and temperature. She would like to have a design, if it is possible to construct one, with main effects unconfounded with one another. Make the design for her, tell which desing it is, and explain how you have taken her requests into account.

   Short answer:

   Let the factors be A: temperature, B: concentration, C: pH, D: agitation rate and E: catalyst type. We can e.g. suggest a $2^{5-2}$ design with resolution III, where we do a full factorial design with factors A, B and C, and generators D=AC and E=BC. The defining relation is then I=ACD=BCE=ABDE. None of the main effects is confounded with another main effect or with the interaction AB (temperature/concentration) or AE (temperature/catalyst type).

6. (5 p) Yield of four corn hybrids (A, B, C and D) are compared. 16 corn plants are planted on an area which is divided into 16 squares (4 columns and 4 rows), one plant for each square. The researcher suspects that the area is not homogeneous and would like to account for the variation due to the column and row effects.

   a) Which design would you suggest and why? Draw the design you have chosen.
   b) Give the linear model connected to your design and the hypotheses you want to test.
   c) Fill in the incomplete ANOVA table below and explain all the items in it when it is complete.

| Source | SS | df | MS | F ratio |
|---|---|---|---|---|
| Row | 0.030 | | | |
| Column | 0.827 | | | |
| Treatment | 0.427 | | | |
| Residuals | 0.129 | | | |
| Total | 1.414 | | | |

   Does the ANOVA table correspond to the design you have suggested in a)?

   d) What are your conclusions based on the ANOVA table and the F test. Give the assumptions that are needed to make these conclusions.
   e) Which design would you have if you ignored the column and row effects? How would you do the experimental set-up in this case?

   Short answer:

   a) Latin square design. The location (rows and columns) is a source for variation and needs to be taken into account.
   b) $y_{tij} = \nu + \beta_i + \gamma_j + \tau_t + \epsilon_{tij}$, where $\nu$ is the overall mean, $\beta_i$ the row effect, $\gamma_j$ the column effect, $\tau_t$ the treatment effect, and $\epsilon_{tij}$ the independent $N(0, \sigma^2)$ distributed errors. We test the null hypothesis that the four corn hybrids give the same yield on average against the hypothesis that at least one of the yield differs from the others.

c) The ANOVA table becomes

| Source | SS | df | MS | F ratio |
|---|---|---|---|---|
| Row | 0.030 | 3 | 0.010 | 0.47 |
| Column | 0.827 | 3 | 0.276 | 12.84 |
| Treatment | 0.427 | 3 | 0.142 | 6.60 |
| Residuals | 0.129 | 6 | 0.0215 | |
| Total | 1.414 | 15 | | |

d) Using 5% significance level, the critical point from F(3,6) distribution is 4.76. Since the observed F test statistics 6.60 is larger than 4.76, we can conclude (at the 5% level) that there is evidence against the null hypothesis that the corn hybrids would give on average the same yield. We have assumed that the errors are independent and $N(0, \sigma^2)$ distributed. (Column effect is also significant at the 5% level.)

e) Randomized regular ANOVA design, the plants would be randomly located among the squares. Then, the row and column effects would be included in the error term.

Good luck!

## Normal distribution and $t$-distribution

Let $X \sim N(\mu, \sigma)$. Then
$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

Let us now have a sample, $X_1, ..., X_n$, from the distribution $N(\mu, \sigma)$. ($X_i$'s are independent and have $N(\mu, \sigma)$-distribution.) The mean $\mu$ can be estimated by the sample mean
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
and the variance $\sigma^2$ by the sample variance
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

The sample mean $\bar{X}$ is $N(\mu, \sigma/\sqrt{n})$-distributed. Therefore,
$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$
If the variance is not know, it is estimated by its sample standard deviation $S$, and
$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1},$$
where $n-1$ is the number of degrees of freedom, the parameter of the $t$-distribution.

## $T$-test

Let us assume that $X \sim N(\mu_X, \sigma)$ and $Y \sim N(\mu_Y, \sigma)$, and we have a sample $X_1, ..., X_{n_X}$ from the distribution of $X$ and a sample $Y_1, ..., Y_{n_Y}$ from the distribution of $Y$. In addition, let the two samples be independent.

The test statistic of the one sample $T$ test is
$$\frac{\bar{X}-\mu_0}{S/\sqrt{n}},$$
where $\mu_0$ is the expected value under the null hypothesis.

The test statistic of the two sample $T$-test, where we test whether there is difference in means between the two samples, is
$$\frac{\bar{X}-\bar{Y}}{S_p\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}},$$

where $S_p^2 = \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}$ and $S_X^2$ and $S_Y^2$ are the variance estimators for $X$ and $Y$, respectively.

### $\chi^2$-distribution

Let $Z_1, ..., Z_n$ be a sample from $N(0,1)$-distribution. Then

$$\sum_{i=1}^{n} Z_i^2 = Z_1^2 + ... + Z_n^2 \sim \chi_n^2,$$

where $n$ is called the number of degrees of freedom, the parameter of $\chi^2$-distribution.

This implies that

$$\frac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$$

and

$$\frac{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2.$$

The sample variance can be written as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{\sigma^2}{n-1}\frac{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2},$$

where $\sum\limits_{i=1}^{n}(X_i-\bar{X})^2/\sigma^2 \sim \chi_{n-1}^2$. Therefore, $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$.

### $F$-distribution

$F$-distribution is defined as a ratio of two $\chi^2$-distributed random variables divided by their number of degrees of freedom, i.e.

$$\frac{\chi_n^2/n}{\chi_m^2/m} \sim F(n,m).$$

Let us now have two independent samples: $X_1, ..., X_n$ from $N(\mu_X, \sigma_X)$ and $Y_1, ..., Y_m$ from $N(\mu_Y, \sigma_Y)$. Furthermore, let $\bar{X}$ and $\bar{Y}$ be the sample means and $S_X^2$ and $S_Y^2$ the samples variances of the two samples, respectively. Then

$$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim \frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)} \sim F(n-1, m-1).$$

## Connection between $N(0, 1)$ and $t$-distribution

$t$-distribution is defined as a ratio of two independent random variables: a $N(0, 1)$-distributed random variable $Z$ and a square root of a $\chi^2$-distributed random variable $V$ divided by the number of its degrees of freedom $n$, i.e.

$$\frac{Z}{\sqrt{V/n}} \sim t_n.$$

Let $X \sim N(\mu, \sigma)$ and $S^2$ its estimated variance based on a sample of size $n$, and $V \sim \chi^2_n$ as above. Then,

$$\frac{X - \mu}{S} = \frac{X - \mu}{\sigma} \cdot \frac{\sigma}{S} = \frac{(X - \mu)/\sigma}{\sqrt{S^2/\sigma^2}} \sim t_{n-1}$$

since $(X - \mu)/\sigma \sim N(0, 1)$ and $(n - 1)S^2/\sigma^2 \sim \chi^2_{n-1}$.