Tentamentsskrivning i Statistisk slutledning, TM, 5p.

Tid: måndagen den 22 maj 2006 kl 14.00-18.00.
Examinator och jour: Serik Sagitov, tel. 772-5351, mob. 0736 907 613, rum MC 1420.
Hjälpmedel: Chalmersgodkänd räknare, egen formelsamling (4 sidor på 2 blad A4) samt utdelade tabeller.
För "3" fordras 12 poäng, för "4" - 18 poäng, för "5" - 24 poäng.
GU betyg: "G" - 12 poäng, "VG" - 20 poäng.

1. (5 points) A research project studied the physical properties of wood materials constructed by bonding together small flakes of wood. Different species of trees were used (aspen, birch, maple) and the flakes were made of different sizes. Some of the data are given in the table where the physical property measured was the tension modulus of elastisity in the direction perpendicular to the alignment of the flakes, in pounds per square inch (psi).

	Aspen	Birch	Maple
Flake size	308	214	272
0.015 inches	428	433	376
by 2 inches	426	231	322
Flake size	278	534	158
0.025 inches	398	512	503
by 2 inches	331	320	220

a. Explain what is a possible interaction effect here? Illustrate with a graph based on the given data set. What kind of interaction the graph suggests?
b. Apply an appropriate test to see if the interaction effect is significant. What assumptions do you make?

2. (5 points) Do piano lessons improve the spatial-temporal reasoning of preschool children? This question was examined by analyzing the change in the reasoning scores of 34 preschool children after six months of piano lessons:

Score change	-3	-2	-1	0	1	2	3	4	5	6	7	9	Tot
Number of children	1	2	1	2	1	3	5	7	2	3	5	2	34

These changes were compared with the changes of 44 children in a control group:

Score change	-6	-4	-3	-2	-1	0	1	2	3	4	5	7	Tot
Number of children	1	1	3	2	7	11	6	7	1	3	1	1	44

a. Display the data in the form of two histograms in a way that makes it easy to compare the two distributions.
b. Explain the need for a control group in the experiment. Is there any real change in the spatial-temporal reasoning for the control group during the six month period? Test an appropriate null hypothesis against a relevant alternative.

c. Find a 99% confidence interval for the difference between two mean changes. Is the difference between two means statistically significant?

d. What assumptions do you make when computing the confidence interval in c)? How realistic they are?

3. (5 points) In an experiment comparing the taste of instant versus fresh-brewed coffee each subject tastes two unmarked cups of coffee, one of each type, in random order and states which he or she prefers.

Of the 50 subjects who participate in the study, 19 prefer the instant coffee.

a. Find a 90% confidence interval for the proportion p of the population who prefer the taste of fresh-brewed coffee. What exactly does this interval say about the value of the population proportion p?

b. Given a Beta prior Beta(14,11) compute a posterior distribution of p. Explain your calculation. Using the prior and posterior $\mathrm{E}(p)$, $\mathrm{Var}(p)$ sketch the graphs for the prior and posterior pdf of p.

Recall that a Beta(a,b)-distribution has:

pdf $f(p)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}p^{a-1}(1-p)^{b-1}$, $0<p<1$, $\mu=\frac{a}{a+b}$, $\sigma^2=\frac{\mu(1-\mu)}{a+b+1}$.

c. Find a posterior mean estimate for p and explain how can one compute a 90% credibility interval for the proportion p.

4.(5 points) PTC is a compound that has a strong bitter taste for some people and is tasteless for others. The ability to taste this compound is an inherited trait. Many studies have assesed the proportions of people in different populations who can taste PTC. The following table gives results for samples from several countries.

	Ireland	Portugal	Norway	Italy
Tasters	558	345	185	402
Nontasters	225	109	81	134

Do the data provide evidence that the proportion of PTC tasters varies among the four countries? Give a complete summary of your analysis.

5. (5 points) A certain chemotherapy treatment for cancer tends to lengthen the lifetimes of very seriously ill patients and decrease the lifetimes of the least ill patients. An experiment on mice produced the following lifetimes in months

treatment group 10, 13, 15, 15, 16, 16, 17, 19

control group 05, 07, 10, 16, 19, 20, 22, 24

a. Draw two empirical survival functions. Explain what are hazard functions and how can you see them from the graphs.

b. Draw a sketch showing the qualitative behaviour of a Q-Q plot. Explain your drawing.

c. Using the given data suggest a linear model for the relationship between the survival time under the chemotherapy treatment and the survival time in the control group.

6. (5 points) The next table gives the average LSAT and GPA scores for the enetering classes of 15 American law schools in 1973. LSAT is a national test for prospective lawers, GPA is the undergraduate point average. The sample correlation coefficient for the two scores is $r = 0.776$.

School	1	2	3	4	5	6	7	8		
LSAT	576	635	558	578	666	580	555	661		
GPA	3.39	3.30	2.81	3.03	3.44	3.07	3.00	3.43		
School	9	10	11	12	13	14	15		Mean	St.dev
LSAT	651	605	653	575	545	572	594		600.3	41.8
GPA	3.36	3.13	3.12	2.74	2.76	2.88	2.96		3.095	0.244

a. Describe a non-parametric bootstrap procedure to find a sampling distribution of the statistic r. How can we use it to estimate the standard error of r?

b. Compare the noise variances σ^2 for two simple linear regression models: 1) LSAT explained by GPA, and 2) GPA explained by LSAT.

c. If a law school number 16 had a average GPA score of 2.80 what would be your prediction interval for the LSAT of this school?

Statistical tables supplied:
1. Normal distribution table
2. Chi-square distribution table
3. t-distribution table
4. F-distribution table

Partial answers and solutions are also welcome. Good luck!

ANSWERS

1a. The following table gives the **rounded** means of the six samples

	Aspen	Birch	Maple	Size means
S1, $\bar{x}$	387	292	323	334
S2, $\bar{x}$	336	455	294	362
Species means	362	374	308	348

Figure 1 indicates an interaction between the flake size and tree species. Namely the respons to birch with respect to two flakes sizes is totally different from that of aspen and maple.

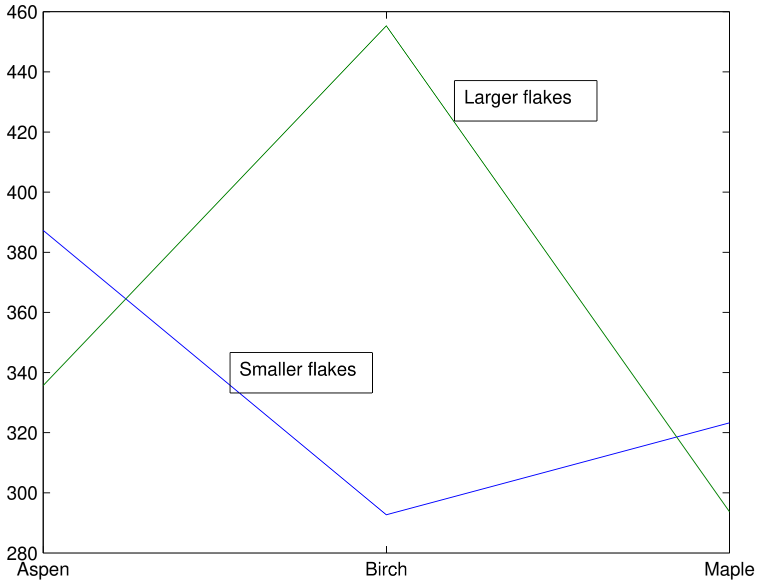

Figure 1: *Mean responses for different flake sizes*

1b. To test the null hypothesis of interaction we apply the two-way ANOVA test. With A = species and B = flake size, from the last table we compute

$$\begin{aligned} SS_{AB} &= K\sum_i\sum_j \hat{\delta}_{ij}^2 \\ &\approx 3\cdot[(387+348-334-362)^2+(336+348-362-362)^2 \\ &+ (292+348-334-374)^2+(455+348-362-374)^2 \\ &+ (323+348-334-308)^2+(294+348-362-308)^2] = 41577. \end{aligned}$$

Using the 18 data values we find

$$\begin{aligned}
\mathrm{SS_E} &\approx (308-387)^2+(428-387)^2+(426-387)^2\\
&+ (278-336)^2+(398-336)^2+(331-336)^2\\
&+ (214-292)^2+(433-292)^2+(231-292)^2\\
&+ (534-455)^2+(512-455)^2+(320-455)^2\\
&+ (272-323)^2+(376-323)^2+(322-323)^2\\
&+ (158-294)^2+(503-294)^2+(220-294)^2 = 147141.
\end{aligned}$$

The observed test statistic becomes

$$F_{AB} = \frac{\mathrm{SS_{AB}}IJ(K-1)}{(I-1)(J-1)\mathrm{SS_E}} \approx \frac{41577\cdot 3\cdot 2\cdot 2}{2\cdot 162567} = 1.7$$

which according to the F-distribution table with degrees of freedom (2,12) is not significant even at 10 percent level. We do not reject the null hypothesis and conclude that the observed interaction could easily be attributable to chance.

2a. Figure 2 gives the histograms based on relative frequences. Clearly, the scores in the piano group are generally higher than the scores in the control group.

2b. A control group is needed to balance out external factors like time passed between two score measurements in the piano group. Control sample

$$m=44,\ \sum y_i = 17,\ \bar y = 0.39,\ \sum y_i^2 = 259,\ s_y = 2.42,\ s_{\bar y} = 0.37$$

Test H_0: $\mu_y = 0$ against H_1: $\mu_y > 0$ using the large sample test for the mean. The observed test statistic $T = \frac{\bar y}{s_{\bar y}} = 1.05$ is not significant according to the normal distribution table. We conclude that the observed change in the spatial-temporal reasoning for the control group during the six month period is not statistically significant.

2c. Piano sample

$$n=34,\ \sum x_i = 123,\ \bar x = 3.62,\ \sum x_i^2 = 753,\ s_x = 3.06,\ s_{\bar x} = 0.52$$

The pooled sample standard deviation $s_p = 2.72$ and the standard error $s_{\bar x - \bar y} = 0.62$ for the point estimate $\bar x - \bar y = 3.23$ of $\mu_x - \mu_y$.

An exact 99% confidence interval for $\mu_x - \mu_y$ is

$$3.23 \pm 2.64\cdot 2.72\cdot\sqrt{\frac{78}{34\cdot 44}} = 3.23 \pm 1.65.$$

It doesn't cover zero proving that the difference between two goups is significant at 1 percent significance level.

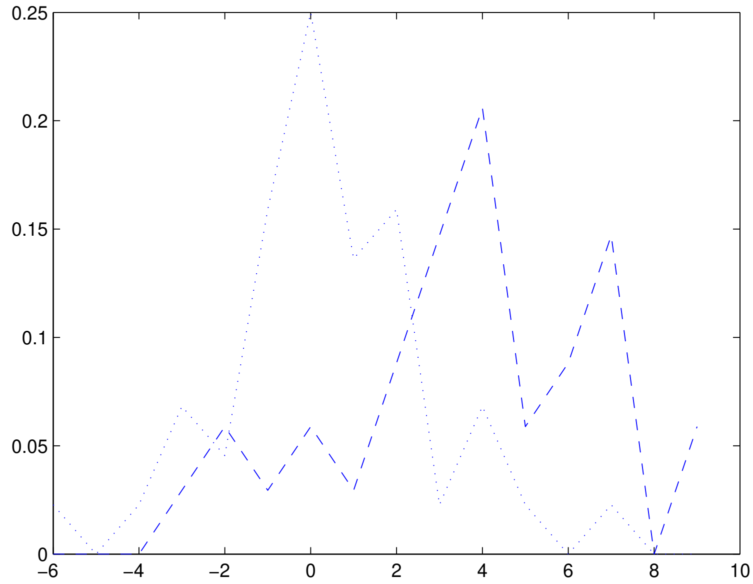

Figure 2: *Piano - dashed line, control - dotted line.*

An approximate 99% confidence interval for $\mu_x - \mu_y$ is pretty much the same

$$3.23 \pm 2.58 \cdot \sqrt{0.52^2 + 0.37^2} = 3.23 \pm 1.65.$$

2d. The key assumption is that we deal with two independent samples. Independence is relevant unless there is a hidden systematic external factor in the allocation of children to two groups.

For the exact CI formula we additionally assume that two samples are taken from two normal distributions with the same variance. The assumption of equality of variances is realistic judging from from the values of the sample standard deviations $s_x = 3.06$ and $s_y = 2.42$. As to the assumption of normality - the control group distribution looks on Figure 2 somewhat skewed to the left.

The asymptotic CI formula works because the sample sizes are fairly large. It applies without the assumption of normality.

3a. Let p be the proportion of the population who prefer the taste of fresh-brewed coffee. The sample proportion $\hat{p} = 0.62$ has the standard error $s_{\hat{p}} = 0.07$. Since the sample size $n = 50$ is large we may build an approximate 90% CI for p: 0.62 ± 0.12 or $(0.50, 0.74)$.

The interval $(0.50, 0.74)$ is an outcome of a random sampling experiment, which produces an interval covering the true value of p with probability 0.9.

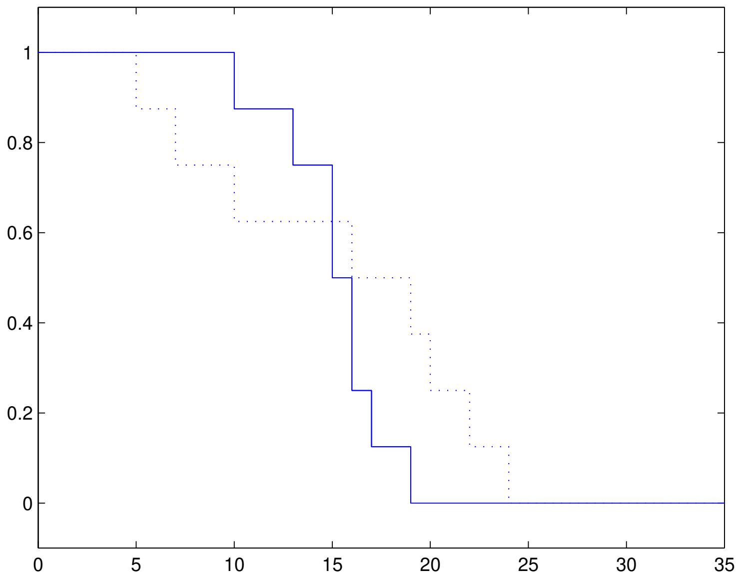

Figure 3: *Survival functions: chemotherapy - solid line, placebo - dotted line.*

3b. We are given the data distribution $X \in \text{Bin}(50, p)$ with the Beta prior distribution $p \in$ Beta(14,11). Using the fact that the Beta prior is conjugate to the Binomial distrution we compute a posterior distribution of p as Beta(45,30).

The graphs of the prior and posterior pdf look like bell curves over the interval [0,1] with

$\mu = 0.56$, $\sigma = 0.10$ for the prior pdf of p

and

$\mu = 0.60$, $\sigma = 0.06$ for the posterior pdf of p.

3d. The PME of p is $\hat{p}_{pme} = 0.60$. A 90% credibility interval of p is an interval around 0.60 containing 90% of the B(45,30) distribution. To get a picture of the credibility interval draw a bell curve with $\mu = 0.60$, $\sigma = 0.06$ and cut off 5% of the area on the left side of $\mu = 0.60$ and 5% of the area on the right side of $\mu = 0.60$. Using the normal approximation we obtain 0.60 ± 0.10. This interval is similar in value to the earlier computed CI, but remember the diffirent nature of these two interval estimates.

4. Test

H_0: the same proportion of PTC tasters for the four countries

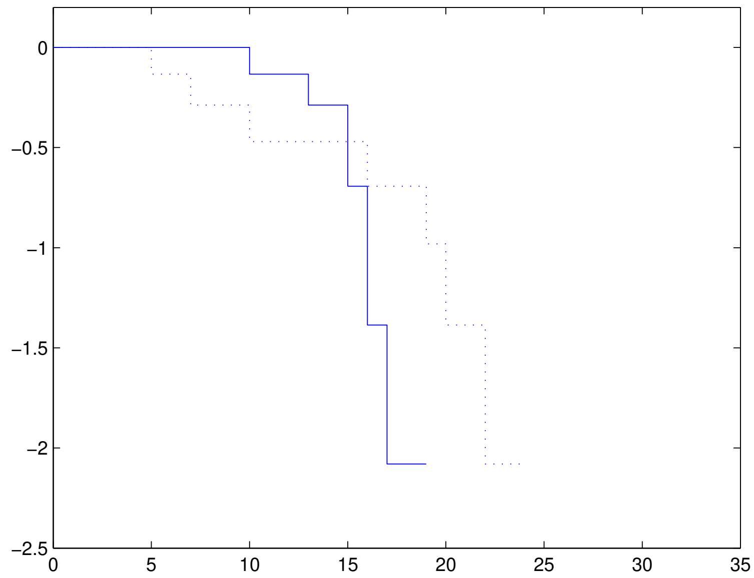

Figure 4: *Log - survival functions: chemotherapy - solid line, placebo - dotted line.*

against

H_1: the proportion of PTC tasters varies among the four countries

Use the chi-square test for homogeneity:

	Ireland	Portugal	Norway	Italy	Total
Tasters	558(572.2)	345(31.8)	185(194.4)	402(391.7)	1490
Nontasters	225(210.8)	109(122.2)	81(71.6)	134(144.3)	548
Total	783	454	266	536	2039

The observed chi-square test statistic:

X^2=0.35+0.53+0.45+0.27

+0.95+1.43+1.23+0.74=5.96

The approximate null distribution of X^2 is χ^2_3. The critical value at α=0.1 is 6.25. We can not reject H_0 even at 10% significance level.

5a. The survival funtions are shown on Figure 3. To interprete the hazard functions we should be draw the log-survival functions: Figure 4. The negative slope of the log survival function

$$-\frac{d}{dt}\log S(t) = -\frac{d}{dt}\log(1 - F(t)) = \frac{F'(t)}{1 - F(t)} = h(t)$$

will give the hazard function. From Figure 4 we clearly see that the hazard function for the treatment is smaller in the beginning and becomes larger than the control hazard function for longer treatments.

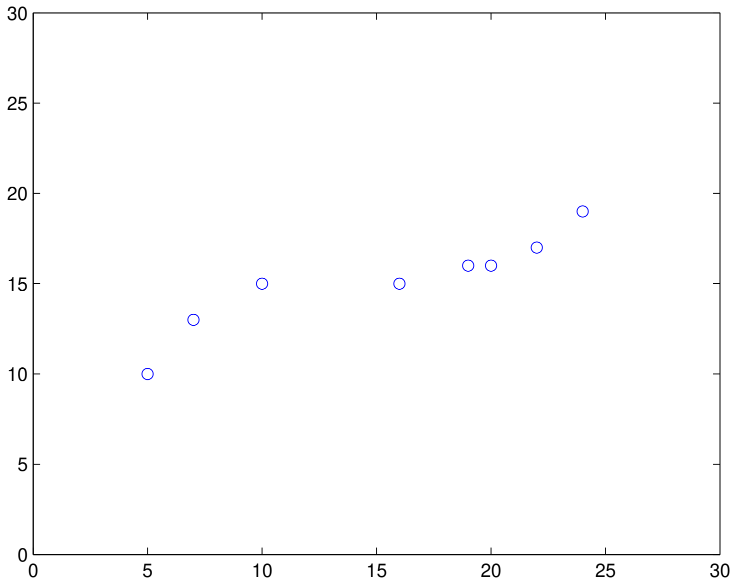

Figure 5: *QQ-plot: survival time y under chemotherapy against survival time x under placebo.*

5b. The QQ-plot is given by Figure 5.

5c. Fit by eye to the QQ-plot on Figure 5, the linear model: $Y \stackrel{d}{=} 10 + 0.4 \cdot X$ relates the marginal distribtutions of two random variables
X = the survival time under placebo treatment
Y = the survival time under the chemotherapy treatment.

6a. Draw 1000 samples of size 15 from the dataset of 15 pairs of scores without replacement. For each of these resamples compute a sample correlation coefficient r_i, $i = 1, \ldots, 1000$. The distribution of r_i, $i = 1, \ldots, 1000$ approximates the sampling distribution of the smple correlation r. Then the standard error of r is estimated by $\sqrt{\frac{1}{999}\sum_{i=1}^{1000}(r_i - \bar{r})^2}$, where $\bar{r} = \frac{1}{1000}\sum_{i=1}^{1000} r_i$.

6b. With X = LSAT and Y = GPA we have $n = 15$, $r = 0.776$, $\bar{x} = 600.3$, $s_x = 41.8$, $\bar{y} = 3.095$, $s_y = 0.244$.

Model 1: $Y = \beta_0 + \beta_1 X + \epsilon$, $\epsilon \sim N(0, \sigma^2)$. Variance estimate $s^2 = \frac{n-1}{n-2} s_y^2 (1 - r^2) = 0.026$.

Model 2: $X = \beta_0 + \beta_1 Y + \epsilon$, $\epsilon \sim N(0, \sigma^2)$. Variance estimate $s^2 = \frac{n-1}{n-2} s_x^2 (1 - r^2) = 748.6$.

6c. Applying Model 2 we estimate the expected LSAT score as : $\bar{x} + b_1(2.80 - \bar{y}) = 561$, where $b_1 = r s_x / s_y = 133$. An exact 95% prediction interval:

$$561 \pm t_{0.025,13} s \sqrt{1 + \frac{1}{n} + \frac{1}{n-1} \left(\frac{2.80 - \bar{y}}{s_y} \right)^2}$$
$$= 561 \pm 2.16 \cdot \sqrt{748.6} \cdot 1.06 = 561 \pm 63.$$