Solution outline
Exam 16 December 2009, KLH3004 and KLMED8004 “Medical Statistics part I”

Problem 1: Probability and illness

A=the adult is ill
C=the child is ill
$P(A)=0.3$
$P(C)=0.1$

a) Indepedence:
$P(\text{both ill})=P(A\cap C)=P(A)\cdot P(C)=0.3\cdot 0.1=0.03$
$P(\text{at least one ill})=P(A \cup C)=P(A)+P(C)-P(A\cap C)=0.3+0.1-0.03=0.37$

b) Dependence: $P(A\cap C)=0.08$
$P(\text{at least one ill})=P(A \cup C)=P(A)+P(C)-P(A\cap C)=0.3+0.1-0.08=0.32$

Under positive dependence, $P(\text{both ill})=P(A\cap C) >P(A)\cdot P(C)$, that is P(both ill) is larger than in the independent case. Thus, P(at least one ill)=0.4-P(both ill) will decrease.

Problem 2: HPV-vaccine

a)
n=total number of girls in the class
p=0.2, probability of not accepting the offer of taking the HPV-vaccine
X=the number of girls not accepting the offer of taking the HPV-vaccine

Expected number of girls not accepting the HPV-vaccine offer: $E(X)=n\cdot p=10\cdot 0.2=2$

Probability of all accepting: $P(X=0)=0.2^{(0)}\cdot(1-0.2)^{(10)}=0.8^{10}=0.107$.
Table X of Rosner, page Y= 0.1074.

We have assumed that the girls make their decision independently of eachother, and that the probability that each girls declines the HPV-vaccine is 0.2.
This is what is assumed in a binomial distribution:

i) we ask n girls independently
ii) each girl has a probability of declining the vaccine called p
iii) we count the number of girls declining the vaccine.

b)
$P(X\geq 8)=P(X=8)+P(X=9)+P(X=10)=0.2^{(8)}\cdot(1-0.2)^{(2)}+0.2^{(9)}\cdot(1-0.2)^{(1)}+$
$0.2^{(10)}\cdot(1-0.2)^{(0)}=7.3728e-05$
$+4.096e-06+1.024e-07$
$= 7.79264e-05$

or from Rosner
$P(X\geq 8)=P(X=8)+P(X=9)+P(X=10)=0.0001+0.0000+0.0000=0.0001$

This probability equals the p-value in a one-sided test of testing

H0: $p \leq 0.2$ vs H1: $p > 0.2$.
We test that the proportion of girls declining to take the vaccine is smaller than or equal to 0.2, versus the alternative hypothesis that the proportion of girls declining to take the vaccine is larger than 0.2. This is equivalent to testing the null hypothesis that the vaccine coverage is larger than or equal to 0.8, versus the alternative hypothesis that the vaccine coverage is less than 0.8.

This means that if the school is randomly selected from all schools in Norway we would reject the null hypothesis that the vaccine coverage is larger than or equal 0.8. But, our sample size is small and we could be worried that the independence assumption between the girls is not satisfied.

Problem 3: The normal distribution with multiple choice questions

a)
X=weight of newborn boys:
X is distributed as N(3.60,0.5^2) and Z=(X-3.6)/0.5 is then distributed as N(0.1)

$P(X \geq 4)=P((X-3.6)/0.5 \geq (4-3.6)/0.5)=P(Z \geq 0.8)=0.212$

b)
Y=vital capacity of 12 yrs boy
Y is distributed as N(3.0, 0.4^2) and Z=(Y-3)/0.4 is distributed as N(0,1)
Denote the unknown lower limit g.
$0.975=P(Y \geq g)=P((Y-3)/0.4 \geq (g-3)/0.4)$.
And we know that $P(Z \geq -1.96)=0.975$.
Thus, g is found as
(g-3)/0.4=-1.96, which means that g=3-1.96*0.4=2.216.

Problem 4: Traning and RCT

a)
Medians:
Group 1: (1000+1020)/2=1010 seconds
Group 2: (680+725)/2=702.5 seconds

b)
Lower quartile, upper quartile, median
The asterisk (*) denotes a possible outlier, the observed value 1415 in Group 2

c)
The observations in Group 1 and 2 are assumed independent, $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$.

$H_0 : \mu_1 = \mu_2$ or equivalent $H_0 : \mu_1 - \mu_2 = 0$.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\frac{1}{n_1 + n_2 - 2}[(n_1 - 1)s_1^{\ 2} + (n_2 - 1)s_2^{\ 2})]}}$$

$$= \frac{914.13 - 747.50}{\sqrt{\left(\frac{1}{8} + \frac{1}{10}\right)\frac{1}{8 + 10 - 2}[(8-1)165.247^2 + (10-1)257.480^2]}}$$

$$= 1.859$$

(Equation 2.19 in Rosner)

Here,

$$|t| < t_{n_1+n_2-2,1-\alpha/2} = t_{16,0.975} = 2.12$$

So we do not reject the null hypothesis at level 5%.

d)
95% confidence interval:
-27.50 to 418.76
Agrees with conclusion in c, since it contains the null hypothesis value 0.

e)

Group	x	rank
2	540	1
2	550	2
2	590	3
2	595	4
1	650	5
2	680	6
2	725	7
2	735	8
2	775	9
1	800	10
1	840	11
2	870	12
1	1000	13
1	1020	14
1	1030	15
1	1045	16
1	1160	17
2	1415	18

Rank sum Group 1: 5+10+11+13+14+15+16+17=101

Expected under H_0: $E(R_1) = n_1(n_1 + n_2 + 1)/2 = 8(8 + 10 + 1)/2 = 76$

P=0.026 implies reject H0 at level 5%

f)
The t-test and the nonparametric test give different conclusions. There is one outlier in Group 2. the nonparametric test is robust to outliers, and I would recommend a nonparametric test.

Problem 5: Power and sample size
Answer: Alternative iii).