May 4, 2004

**Solutions: Chapter 10** Problems 5, 6, 11, 13, 14, 37, 40

**Problem 5.**

$F_n(u) = \frac{1}{n}[1_{X_1 \le u} + \ldots + 1_{X_n \le u}]$
$F_n(v) = \frac{1}{n}[1_{X_1 \le v} + \ldots + 1_{X_n \le v}]$
$\mathrm{E}(F_n(u)) = F(u)$, $\mathrm{E}(F_n(v)) = F(v)$.

Assuming $u < v$

$$\begin{aligned}\mathrm{E}(F_n(u) \cdot F_n(v)) &= \tfrac{1}{n^2}[\textstyle\sum_{i=1}^n \mathrm{E}(1_{X_i \le u} 1_{X_i \le v}) + \sum_{i=1}^n \sum_{j \ne i} \mathrm{E}(1_{X_i \le u} 1_{X_j \le v})] \\ &= \tfrac{1}{n^2}[\textstyle\sum_{i=1}^n F(u) + \sum_{i=1}^n \sum_{j \ne i} F(u)F(v)] \\ &= \tfrac{1}{n}[F(u) + (n-1)F(u)F(v)]\end{aligned}$$

Finish by using $\mathrm{Cov}(F_n(u), F_n(v)) = \mathrm{E}(F_n(u) \cdot F_n(v)) - \mathrm{E}(F_n(u)) \cdot \mathrm{E}(F_n(v))$.

**Problem 6.**

a) Matlab commands

```
x=data vector
stairs(sort(x),(1:length(x))/length(x)    % emperical cdf
hist(x)                                   % histogram, the same as hist(x,10)
normplot(x)                               % normal probability plot
prctile(x,90)                             % 0.90-quantile
```

The distribution appears to be rather close to normal.

b) Since $\bar{x} = 14.58$ and $s = 0.78$, the one-sided 99% CI for $\mu$ is
$(-\infty, 14.58 + 2.33 \cdot 0.78) = (-\infty, 16.40)$.

Expected means

| | | |
|---|---|---|
| 1% dilution | $\mu_1 = 14.58 \cdot 0.99 + 85 \cdot 0.01 = 15.28$ | can not be detected |
| 3% dilution | $\mu_3 = 14.58 \cdot 0.97 + 85 \cdot 0.03 = 16.69$ | can be detected |
| 5% dilution | $\mu_5 = 14.58 \cdot 0.95 + 85 \cdot 0.05 = 18.10$ | can be detected |

**Problems 11, 13, 14.**

$1 - F(t) = e^{-\alpha t^\beta}$, $f(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^\beta}$, $h(t) = \alpha\beta t^{\beta-1}$.

If $\beta = 1$, then $h(t) = \alpha$ is memoryless.

If $0 < \beta < 1$, then $h(t)$ decreases with $t$ meaning that the longer you live the healthier you become.

If $\beta > 1$, then $h(t)$ increases with $t$ meaning that the older individuals die more often than the younger.

**Problem 37.**

a) The matlab commands

```
trimmean(x,10)
trimmean(x,20)
```

give $\bar{X}_{0.1} = 14.586$ and $\bar{X}_{0.2} = 14.605$.

b) (14.41; 14.75)

c) Nonparametric 90% CI for $M$ is $(X_{(k)}, X_{(60-k)})$, where $\mathrm{P}(Y < k) = 0.05$. Use the normal approximation $Y \in \mathrm{N}(29.5; (3.84)^2)$ for the null distribution $Y \in \mathrm{Bin}(59, 0.5)$ to compute $k \approx 23.2$. Thus $(X_{(k)}, X_{(60-k)}) = (X_{(23)}, X_{(37)}) = (14.41; 14.77)$.

d) Matlab commands

```
n=59; B=1000;
z=x(random('unid',n,n,B));
t1=trimmean(z,10);
t2=trimmean(z,20);
std(t1)
std(t2)
```

give the standard errors 0.1034 and 0.1004 for $\bar{X}_{0.1}$ and $\bar{X}_{0.2}$ respectively.

e) Matlab commands

```
c11=prctile(t1,5)
c12=prctile(t1,95)
c21=prctile(t2,5)
c22=prctile(t2,95)
```

give a 90% CI for $\mu_{0.1}$: $(2\bar{X}_{0.1} - c12; 2\bar{X}_{0.1} - c11) = (14.435; 14.765)$
and a 90% CI for $\mu_{0.2}$: $(2\bar{X}_{0.2} - c22; 2\bar{X}_{0.2} - c21) = (14.463; 14.784)$.

f) Matlab commands

```
iqr(x)
median(abs(x-median(x)))
```

Warning: mad(x) in Matlab stands for the mean abs. dev.

g) Matlab commands (vector z comes from the d) part)

```
q=prctile(z,75);
hist(q)
std(q)
```

give the standard error 0.1332 of the upper quartile.

**Problem 40.**
Matlab command (x = control and y = seeded data)

```
qqplot(x,y)
```

fits the line $y = 2.5x$ claiming 2.5 times more rainfall fromm seeded clouds.
Matlab command

```
qqplot(log(x),log(y))
```

fits the line $log(y) = 2 + 0.8 log(x)$ meaning a decreasing slope in the relationship $y = 7.4x^{0.8}$.