Extended Answer Section

Answer these questions in the answer book(s) provided.
Ask for extra books if you need them.

1. In an effort to reduce loss of productivity a safety program was implemented in 10 industrial plants. The average weekly losses of worker hours due to accidents for the 10 plants both before and after implementation of the safety program are recorded (respectively) in the two R vectors `before` and `after`.

Using the R output below, perform an appropriate test of significance that makes normality assumptions to determine whether the apparent reduction is significant. As part of your answer you should

(a) clearly describe the statistical model being used;

(b) verify any necessary assumptions via boxplot sketch;

(c) state the hypotheses being tested in terms of model parameters;

(d) obtain the tightest bounds possible for the p-value using statistical tables;

(e) give a clear interpretation of the numerical value of the p-value.

```
> rbind(before,after)
       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
before   45   57   73   83   46   34  124   26   33    17
after    36   51   60   77   44   29  119   24   35    11
> mean(before)
[1] 53.8
> mean(after)
[1] 48.6
> sort(before)
 [1]  17  26  33  34  45  46  57  73  83 124
> sort(after)
 [1]  11  24  29  35  36  44  51  60  77 119
> sort(before-after)
 [1] -2  2  2  5  5  6  6  6  9 13
> var(before)
[1] 1027.733
> var(after)
[1] 962.9333
> var(before-after)
[1] 16.62222
```

2. An urn contains 10 balls, numbered 0,1,2,...,9. Five balls are drawn out in order in such a way that all possible sequences are equally likely, resulting in random digits $X_1, X_2, \ldots, X_5$.

(a) What is the probability (to 4 decimal places) that exactly 3 digits are less than or equal to 3 if the sampling is done

(i) with replacement;

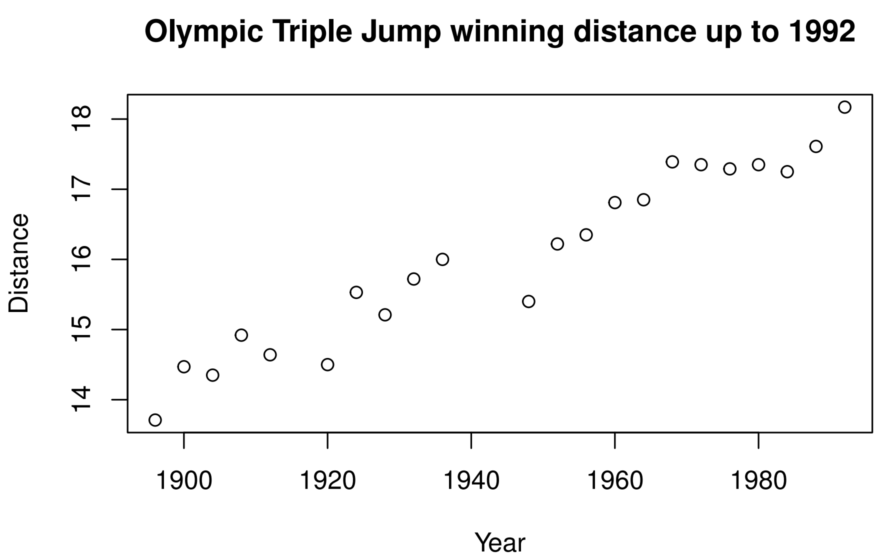
(ii) without replacement?

(b) Determine $\mu = E(X_1)$.

(c) Determine $\sigma^2 = Var(X_1)$.

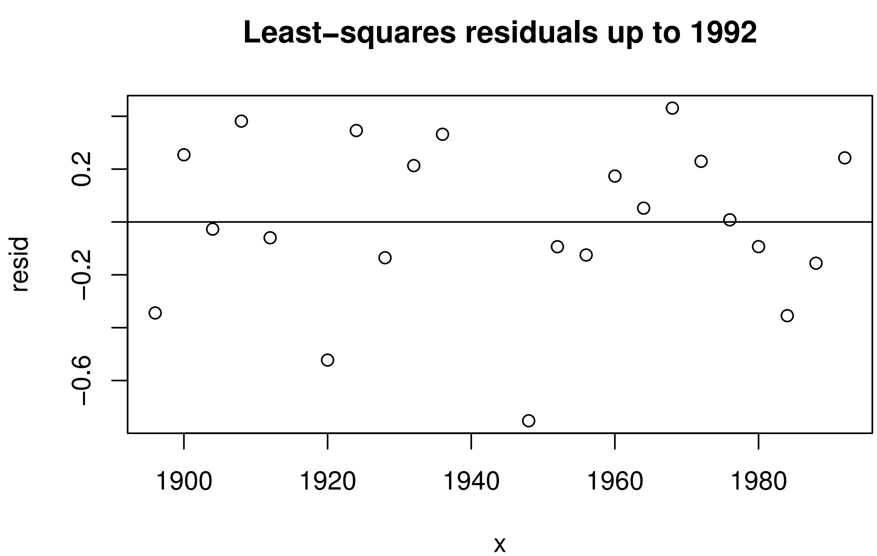
(d) If the sampling is done with replacement and $S = X_1 + \cdots + X_5$ is the sum of the random digits, compute a normal approximation with continuity correction to $P(S \leq 10)$.

3. Below appears a scatterplot of the winning distance in the Olympic triple jump event against year up to the year 1992 (note the Olympic games were not held in 1916, 1940 or 1944).



Study carefully the R commands and output below and answer the questions that follow.

```
> x=Year
> y=Distance
> Sxx=sum((x-mean(x))^2)
> Sxx
[1] 19953.45
> Sxy=sum((x-mean(x))*(y-mean(y)))
> Sxy
[1] 805.0145
> Syy=sum((y-mean(y))^2)
> Syy
[1] 34.4571
> b=Sxy/Sxx
> b
[1] 0.04034462
> a=mean(y)-b*mean(x)
> a
[1] -62.43908
> resid=y-(a+b*x)
> length(resid)
[1] 22
> sum(resid^2)
[1] 1.979089
> plot(x,resid,main="Least-squares residuals up to 1992")
> abline(h=0)
```

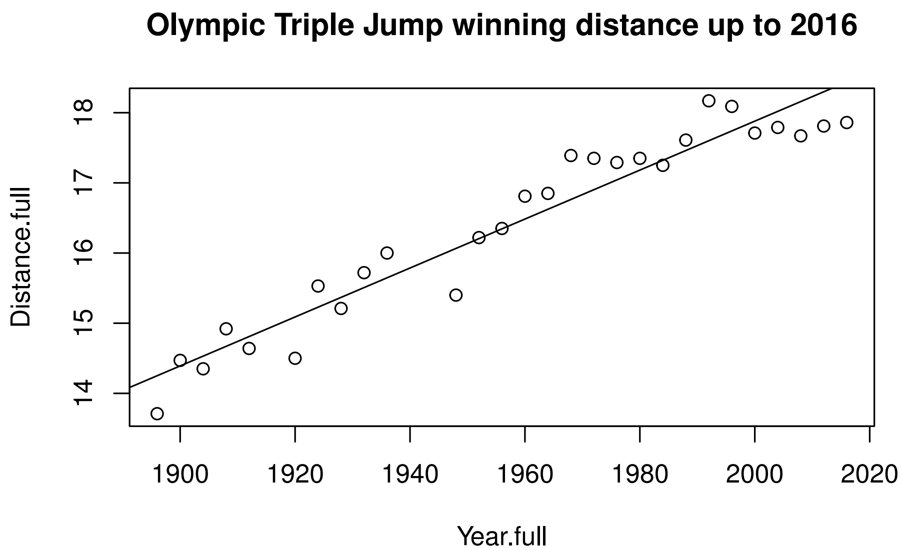


(a) Comment on the goodness of fit of the least-squares line to the data.

(b) Describe a statisical model involving normality assumptions for the data and provide a two-sided 95% confidence interval for the expected increase per year of the winning distance.

(c) The results up to 2016 are obtained and the procedure is repeated on the full dataset. Comment on the goodness of fit of the least-squares line to the full dataset.

```
> plot(Year.full,Distance.full,
+       main="Olympic Triple Jump winning distance up to 2016")
> fit=lm(Distance.full~Year.full)
> abline(fit)
```




```
> plot(Year.full,fit$resid,main="Least-squares residuals up to 2016")
> abline(h=0)
```