Math 4397/6397
Problem Set 10, due Thursday, Nov 19

Problem 1. The computations are similar for each group of cigarette consumption, as an illustration, calculations are presented for 1-4 cigarette consumption group versus non-smokers.

	1-4	0
LC	49	7
Control	91	61

OR $= \frac{49\times61}{91\times7} = 4.69$
ln(OR) $=$ log$(4.69) = 1.55$
var(ln(OR)) $= \frac{1}{49} + \frac{1}{61} + \frac{1}{91} + \frac{1}{7} = 0.191$
se(ln(OR)) $= \sqrt{0.191} = 0.437$
95% confidence interval for ln(OR): $1.55 \pm 1.96 \times 0.437 = (0.69, 2.40)$
95% confidence interval for OR: $(e^{0.69}, e^{2.40}) = (2.00, 11.05)$

The table below displays the estimated OR, ln(OR) and 95% confidence intervals for the OR and ln(OR).

Cigarette Consumption Daily	OR	ln(OR)	Var(ln(OR))	95% CI for ln(OR)	95% CI for OR
0	-	-	-	-	-
1-4	4.69	1.55	0.191	(0.69, 2.40)	(2.00, 11.05)
5-14	7.31	1.99	0.163	(1.20, 2.78)	(3.32, 16.15)
15-24	9.50	2.25	0.164	(1.46, 3.04)	(4.29, 20.91)
25-49	16.08	2.78	0.169	(1.97, 3.59)	(7.21, 36.23)
50+	17.86	2.88	0.234	(1.93, 3.83)	(6.91, 46.06)

From this, we conclude that there is a positive association between smoking and lung cancer (since all the estimated odds ratios are greater than 1). There is also evidence in the data to suggest that the odds of lung cancer increase with daily cigarette consumption.

Problem 2. The data are as follows:

	Died	Alive	Total
Trt	2	13	15
Control	4	15	19
Total	6	28	34

The null and alternative hypotheses are:

$$H_0 : p_{trt} = p_{control} \quad H_A : p_{trt} \neq p_{control}$$

All possible tables with the above fixed margins and their corresponing probabilities are given below:

table 1:
$$\begin{array}{cc|c} 0 & 15 & 15 \\ 6 & 13 & 19 \\ \hline 6 & 28 & 34 \end{array} \qquad P(table1|H_0) = \frac{\binom{6}{0}\binom{28}{15}}{\binom{34}{15}} = 0.020174$$

table 2:
$$\begin{array}{cc|c} 1 & 14 & 15 \\ 5 & 14 & 19 \\ \hline 6 & 28 & 34 \end{array} \qquad P(table2|H_0) = \frac{\binom{6}{1}\binom{28}{14}}{\binom{34}{15}} = 0.12969$$

table 3:
$$\begin{array}{cc|c} 2 & 13 & 15 \\ 4 & 15 & 19 \\ \hline 6 & 28 & 34 \end{array} \qquad P(table3|H_0) = \frac{\binom{6}{2}\binom{28}{13}}{\binom{34}{15}} = 0.302609$$

table 4:
$$\begin{array}{cc|c} 3 & 12 & 15 \\ 3 & 16 & 19 \\ \hline 6 & 28 & 34 \end{array} \qquad P(table4|H_0) = \frac{\binom{6}{3}\binom{28}{12}}{\binom{34}{15}} = 0.327826$$

table 5:
$$\begin{array}{cc|c} 4 & 11 & 15 \\ 2 & 17 & 19 \\ \hline 6 & 28 & 34 \end{array} \qquad P(table5|H_0) = \frac{\binom{6}{4}\binom{28}{11}}{\binom{34}{15}} = 0.173555$$

table 6:
$$\begin{array}{cc|c} 5 & 10 & 15 \\ 1 & 18 & 19 \\ \hline 6 & 28 & 34 \end{array} \qquad P(table6|H_0) = \frac{\binom{6}{5}\binom{28}{10}}{\binom{34}{15}} = 0.042425$$

table 7:
$$\begin{array}{cc|c} 6 & 9 & 15 \\ 0 & 19 & 19 \\ \hline 6 & 28 & 34 \end{array} \qquad P(table7|H_0) = \frac{\binom{6}{6}\binom{28}{9}}{\binom{34}{15}} = 0.003721$$

The tables in favor of the alternative (higher mortality in control group) are tables 1-3, so the p-value is $p = 0.0202 + 0.130 + 0.303 = .45$, meaning we retain the null-hypothesis.

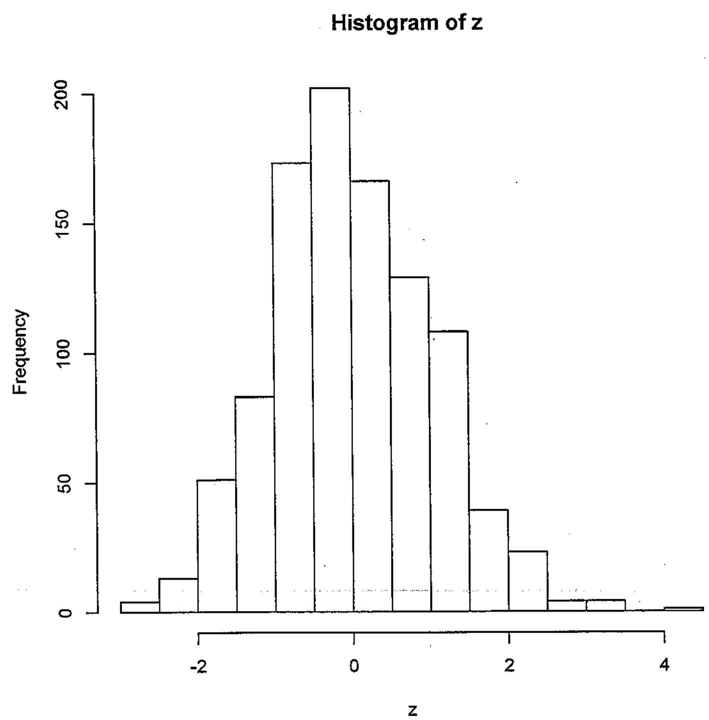
Problem 3. a. $f(x) = \sqrt{x}$ therefore $f'(x) = 1/2\sqrt{x}$. Hence the standard error is

$$\frac{1}{2\sqrt{\hat{p}}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{1}{2}\sqrt{\frac{1-\hat{p}}{n}}$$

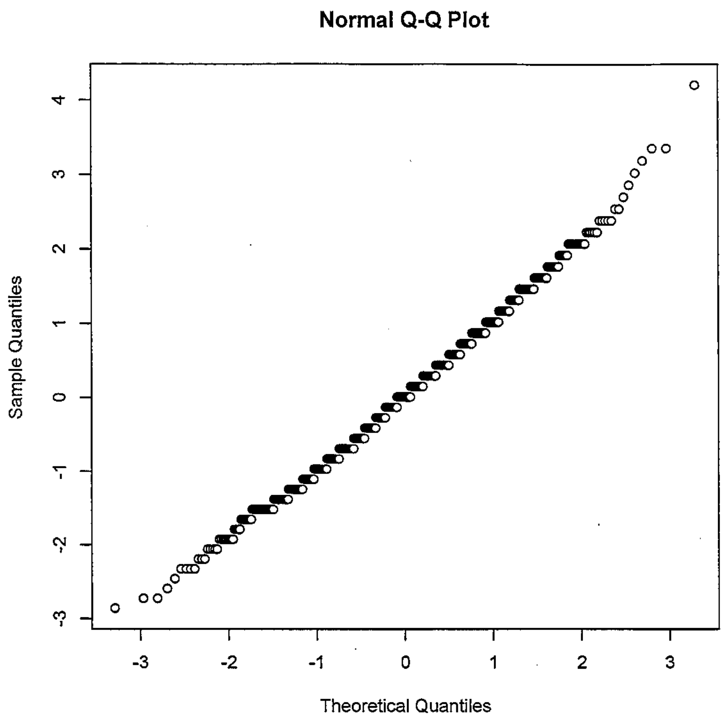
b. The simulation should result in a standard normal random variable when properly normalized:

```
n <- 200
p <- .5
x <- rbinom(1000, size = n, prob = p)
phat <- x / n
fphat <- sqrt(phat)
fp <- sqrt(p)
sefphat <- sqrt((1 - phat) / n) / 2
z <- (fphat - fp) / sefphat
hist(z)
qqnorm(z)
```

We obtain the two plots



and



which confirm that the distribution is approximately normal.

Problem 4.

- For the 10 Hz case, we have the following contingency table:

	No Change	Change	Total
Control	1	11	12
Convulsant	6	6	12
Total	7	17	24

We use the Fisher's exact test.

```
A=matrix(c(1,6,11,6),nrow=2)
> A
     [,1] [,2]
[1,]    1   11
[2,]    6    6

> fisher.test(A)
```

The output of the Fisher's test is:

```
        Fisher's Exact Test for Count Data

data:  A
p-value = 0.06865
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.001822543 1.127544004
sample estimates:
odds ratio
 0.1009016
```

- For the 40 Hz case, we have the following contingency table:

	No Change	Change	Total
Control	0	12	12
Convulsant	2	10	12
Total	2	22	24

The output of the Fisher's exact test in this case is:

```
        Fisher's Exact Test for Count Data

data:  A
p-value = 0.4783
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.000000 5.248663
sample estimates:
odds ratio
         0
```

- The difference in proportions of times when we see a change is statistically significant (at least 90% confidence) for the 10 Hz case. The difference in the 40 Hz case is not significant.