Homework 02
Joseph Blubaugh
jblubau1@tamu.edu
STAT 659-700

1.6

a) If $n = 3$ and $n_1 + n_2 + n_3 = 3$, then $n_1 = 3 - n_2 - n_1$ If n_1, n_2 are known then the sum of those two components minus n is n_1.

b) For (n_1, n_2, n_3), There are 9 total possible combinations with $n = 3$. $(3, 0, 0), (2, 1, 0), (1, 1, 1),$ $(0, 1, 2), (0, 0, 3), (1, 2, 1), (1, 0, 2), (2, 0, 1), (0, 3, 0)$

c)

	n.1	n.2	n.3	Total
n	1	2	0	3
pi	0.25	0.5	0.25	1
e	0.75	1.5	0.75	3

$X^2 = \frac{(1-.75)^2}{.75} + \frac{(2-1.5)^2}{1.5} + \frac{(0-.75)^2}{.75} = 1$

```
1 - pchisq(1, 2)
```

```
[1] 0.6065307
```

The multinomial probability that $(n_1, n_2, n_3) = (1, 2, 0) = .6$

d) $n_1 = N(.75, .43)$

1.20

a)

```
A = c(8, 7, 6, 6, 3, 4, 7, 2, 3, 4); B = c(9, 9, 8, 14, 8, 13, 11, 5, 7, 6)

## A
(wald.a = mean(A) + c(-1, 1) * 1.96 * sqrt(mean(A)/10))
```

```
[1] 3.614071 6.385929
```

```
(score.a = mean(A) + (1.96^2 / (2*10)) + c(-1, 1) * 1.96/sqrt(10) *
  sqrt(mean(A) + (1.96^2 / (4*10))))
```

```
[1] 3.792904 6.591256
```

```
## B
(wald.b = mean(B) + c(-1, 1) * 1.96 * sqrt(mean(B)/10))
```

```
[1]  7.140581 10.859419
```

```
(score.b = mean(B) + (1.96^2 / (2*10)) + c(-1, 1) * 1.96/sqrt(10) *
  sqrt(mean(B) + (1.96^2 / (4*10))))
```

```
[1]  7.322766 11.061394
```

b)

```
## A
mean(A); var(A)
```

```
[1] 5
```

```
[1] 4.222222
```

```
## B
mean(B); var(B)
```

```
[1] 9
```

```
[1] 8.444444
```

The poisson distribution has a mean and variance that are equal and both treatments have a variance that is relatively close to the treatment mean.

```r
Z.a = (var(A)/mean(A) - 1) * sqrt(9 / 2)
Z.b = (var(B)/mean(B) - 1) * sqrt(9 / 2)

1 - pnorm(Z.a); 1 - pnorm(Z.b)
```

```
[1] 0.6292937

[1] 0.5520909
```

Both Z statistics are reasonably close to zero so we do not reject that the treatments are from the poisson distribution.

1.21

N	0	1	2	3	4	5	6	7	8	9
Y	15	18	8	22	10	19	14	14	18	12
E	15	15	15	15	15	15	15	15	15	15
X.2	0	0.6	3.267	3.267	1.667	1.067	0.06667	0.06667	0.6	0.6

Because our pvalue is greater than .05, we do not reject that the probabilities are equal.

```r
Chi.Sq = sum(lotto$X.2)
1 - pchisq(Chi.Sq, 9)
```

```
[1] 0.2622488
```

1.22

Maize	Green	Golden	Green-Striped	Golden-Green-Striped
Count	773	221	238	69
Pi	0.5625	0.1875	0.1875	0.0625
Expected	731.8	243.9	243.9	81.3
Chi.Sq	2.320	2.150	0.143	1.861

Because our pvalue is greater than .05, we do not reject that the probabilities are different than specified.

```r
Chi.Sq = sum(maize$Chi.Sq)
1 - pchisq(Chi.Sq, 3)
```

```
[1] 0.09069354
```

1.23

Because the parameter of π is estimated we lose an additional degree of freedom, however we still do not reject that probabilities are different than stated because our pvalue is greater than .05

```
mle = .0357

plant = data.frame(
  Type.of.Plant = c("Starchy green", "Starchy white", "Sugary green", "Sugary white"),
  Number = c(1997, 906, 904, 32),
  Pi = round(c(.25*(2 + mle), .25*(1 - mle), .25*(1 - mle), .25*mle), 3)
  )

plant$Expected = round(with(plant, sum(Number) * Pi), 1)
plant$Chi.Sq = round(with(plant, (Number - Expected)^2 / Expected), 3)

pandoc.table(t(plant))
```

Type.of.Plant	Starchy green	Starchy white	Sugary green	Sugary white
Number	1997	906	904	32
Pi	0.509	0.241	0.241	0.009
Expected	1954.1	925.2	925.2	34.6
Chi.Sq	0.942	0.398	0.486	0.195

```
Chi.Sq = sum(plant$Chi.Sq)

1 - pchisq(Chi.Sq, 2)
```

[1] 0.3640369

2.1

a) $P(C|-) = 1/4$, $P(+|\bar{C}) = 2/3$
b) sensivity = $1/3$, specificity = $3/4$
c) Cell joint probabilities

	C	C.bar	Total
Test +	0.0033	0.0267	0.03
Test -	0.0067	0.9633	0.97
Total	0.01	0.99	1

d) The marginal probablity of testing positive is .03
e) The probablity of having the disease conditional have testing positive is .11

2.2

a) Sensitivity is the probability of correctly testing positive, $P(Y = 1|X = 1)$ and specificity is the opposite, $P(Y = 2|X = 2)$ so $\pi_i = P(Y = 1|X = 1) = 1 - \pi_2$

b) $.86 * .01 = .0086$

c) The chances of having the disease is very small and sense the sensitivity and specificity are relatively high, the join probabilities are essentially equal to the margin probabilities

	C	C.bar	Total
Test +	0.0086	0.12	0.1286
Test -	0.0014	0.87	0.8714
Total	0.01	0.99	1

2.7

a) The odds ratio is not a probability. The correct interpretation is the odds of survival for females is 11.4 times greater than males where the odds ratio is equal to $\frac{(P(FemaleSurival)/(1-P(FemaleSurvival)))}{P(MaleSurvival)/(1-P(MaleSurvival))} = 11.4$

b)

$$FemaleOdds/MaleOdds = 11.4 \rightarrow 2.9/x = 11.4 \rightarrow MaleOdds = .254$$

$$FemaleOdds = 2.9 = x/(1-x) \rightarrow 2.9/(1+2.9) \rightarrow P(FemaleSurvival) = .743$$

$$MaleOdds = .254 = x/(1-x) \rightarrow .254/(1+.254) \rightarrow P(MaleSurvival) = .202$$

c) $.743/.202 = 3.68$

2.8

a) $(.847/(1-.847))/(.906/(1-.906)) = .57$
b) $.906 - .847 = .059$ The difference is 6%, not 60%

2.11

a)

	Lung.Cancer	Heart.Disease
Smoker	0.0014	0.00669
Non-Smoker	1e-04	0.00413

a) The odds of dieing from either Lung Cancer or Heart Disease are lower for Non-Smokers. Also the odds ratio of dieing from lung cancer is much higher for smokers than it is for non-smokers.

i. Difference in proportions of Smokers who died from Lung Cancer or Heart Disease: $.0014 - .00669 = -.00529$
Difference between non-smokers: $.0001 - .00413 = -.00403$

ii. Relative Risk of dieing of a smoker dieing from lung cancer compared to heart disease: $.0014/.00669 = .209$
For a non-smoker: $.0014/.00413 = .338$

iii. Odds Ratio of a smoker dieing from lung cancer to heart disease: $\frac{(.0014/(1-.0014))}{(.0069/(1-.0069))} = .201$
Odds ratio for non-smoker: $\frac{(.0001/(1-.0001)}{(.00413/(1-.00413)))} = .0241$

b) Lung Cancer: $1/(.0014 - .0001) = 770$, Heart Disease: $1/(.00669 - .00413) = 391$ Lung Cancer has the more significant reduction even though the probability of dieing from Heart Disease is higher.

2.12

a)

	Aspirin	Placebo	Total
Heart Attack	198	193	391
No Heart Attack	19736	19749	39485
Total	19934	19942	39876

b) P(Heart Attack on Aspirin) = $198/19934 = .00993$
P(Heart Attack on Placebo) = $193/19942 = .00967$
Odds Ratio: $(.00993/(1 - .00993)/(.00967/(1 - .00967))) = 1.027$

c) Because the 95% Confidence Interval crosses one we can conclude that there is not enough evidence that the odds differe for the treatments.

$$log(1.027) \pm 1.96\sqrt{\frac{1}{198} + \frac{1}{19736} + \frac{1}{193} + \frac{1}{19749}} \rightarrow .0266 \pm 1.96(.101) = (-.171, .224)$$
$$exp(-.171, .224) = (.842, 1.25)$$

2.15

a)

```
exp(
log(.0094 / .0171) + c(-1, 1) * 1.96 *
  sqrt( ((1 - .0094)/(11037*.0094)) + ((1 - .0171)/(11034*.0171)) )
)
```

```
[1] 0.4332371 0.6974898
```

b) The Wald interval is slightly wider and shifted to the right compared to the Agresti-Caffo and Newcombe intervals.

```
data heart;
input group $ attack $ count;
cards;
placebo yes 185
placebo no 10845
aspirin yes 104
aspirin no 10933
;

proc freq order=data data=heart; weight count;
tables group*attack/nocol nopct relrisk riskdiff(cl=(newcombe ac wald));
title "Heart Attack and Aspirin";
run;
```

	lwr	upr
Agresti-Caffo	0.0043	0.0104
Newcombe	0.0044	0.0104
Wald	0.005	0.011

c) 1 / (.0171 - .0094) = 130

1.23b

```
((1997 * .509) + (906 * .241) + (904 * .241) + (32 * .009)) / (1997 + 906 + 904 + 32)
```

```
[1] 0.3784764
```

2.2b

$(.86 * .01)/(.86 * .01 + .88 * (1 - .01)) = .0097$