1. Agresti 1.1

   (a) Nominal

   (b) Ordinal

   (c) Interval

   (d) Nominal

   (e) Ordinal

   (f) Nominal

2. Agresti 1.2

   a.) X=number of correct answers, $X \sim Binom(n = 100, \pi = 0.25)$

   b.) $E[X] = n \times \pi = 100(0.25) = 25$; $Var[X] = n \times \pi \times (1 - \pi) = 100(0.25)(1 - 0.25) = 18.75$;

Yes, 50 correct responses would be surprising, since 50 is
z = (50 − 25)/4.33 = 5.8 standard deviations above the mean of a distribution that is approximately normal.

   c.) The distribution of $(n_1, n_2, n_3, n_4)$ is multinomial$(\pi_1, \pi_2, \pi_3, \pi_4)$ where $\pi_j = 0.25$.

   d.) $E[n_j] = n \times \pi_j = 0.25$; $Var[n_j] = n \times \pi_j \times (1 - \pi_j) = 100(0.25)(1 - 0.25) = 18.75$; $Cov(n_j, n_k) = -n \times \pi_j \times pi_k = -100(0.25)(0.25) = -6.25$; $Corr(n_j, n_k) = \frac{Cov(n_j, n_k)}{Var(n_j) \times Var(n_k)} = \frac{-6.25}{18.75} = -0.333$.

3. Agresti 1.4

a.) X=number of games with no bullets fired, $P(X=0) = \binom{n}{k}(\pi)^k(1-\pi)^{n-k} = \binom{6}{0}(\frac{1}{6})^k(1-\frac{1}{6})^{6-0} = \frac{5}{6}^6 = 0.3349$ ,Which follows the geometric distribution.

b.) Y=number of games until bullet fires,

$$P(Y=1) = \frac{1}{6}$$

$$P(Y=2) = \frac{5}{6} \times \frac{1}{6}$$

$$P(Y=3) = \frac{5}{6}^2 \times \frac{1}{6}$$

$$P(Y=4) = \frac{5}{6}^{4-1} \times \frac{1}{6}$$

$$\vdots$$

$$P(Y=n) = \frac{5}{6}^{y-1} \times \frac{1}{6}$$

This is because each round is independent and every previous round has no bullet fire with probability $\frac{5}{6}$.

4. Agresti 1.5 Let $\pi$ be the proportion who say yes, $N = 587 + 636$, and $H_0 : \pi = 0.5$. I used R to do the score test, but will show how to calculate $z_s$:

$$z_s = \frac{\hat{\pi} - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} = \frac{\frac{587}{587+636} - 0.5}{\sqrt{\frac{0.25}{587+636}}} = -1.401$$

$$P(z_s^2 = \chi_1^2 > -1.401) = 0.1612$$

π ± 1.96 * sqrt(π * (1-π) / N) = 0.479 ± 0.028

the CI: $(0.452, 0.508)$. We fail to reject the null hypothesis; we do not have sufficient evidence to conclude that the majority of people think that it should be possible

for a pregnant woman to obtain a legal abortion if she is married and does not want any more children. We are 95% confident the true proportion falls between 45.2% and 50.8% (since 0.5 is contained within this interval, that is why we failed to reject $H_0$.

```
## EX1 Problem 1.5 ##
y <- 587
n <- 587+636
## score test
test_score <- prop.test(y, n, conf.level=0.95, correct=F)
##The confidence interval is computed by inverting the score test.
```

5. Agresti 1.6

   a.) LRT $= 2[y\log(\frac{y}{.5n}) + (n-y)\log(\frac{n-y}{(1-0.5)n})] = 2[25\log(\frac{25}{0.5(25)})] = 34.7$

   b.) $\chi^2 = \frac{(25-25)^2}{25} + \frac{(0-25)^2}{25} = \frac{-25^2}{25} = 25$

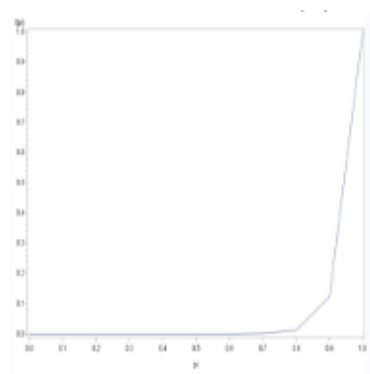
   c.) $z_w = \frac{\hat{\pi}-\pi_0}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}}$ is undefined since $\hat{\pi} = 0$, and then $z_w = \frac{-0.5}{0} = \infty$

6. Agresti 1.7

   a.)

$$L(\pi) \propto \log \pi^2 0(1-\pi)^{20-0}$$

$$\propto \log \pi^{20}$$



This likelihood function is not quadratic since it is to the twentieth power.

b.) $\hat{\pi} = 1$ maximizes $L(\pi)$. $z_w = \frac{\hat{\pi}-\pi_0}{\sqrt{\hat{\pi}(1-\hat{\pi})/n}} = \frac{1-0.5}{\sqrt{\frac{1(0)}{20}}}$, which is not defined. Additionally, $\hat{\pi} \pm 1.96\sqrt{\frac{\pi(1-\pi)}{20}} = 1 \pm 1.96(0) = (1, 1)$, which does not make sense.

c.)

$z = (1.0 - .5)/\sqrt{.5(.5)/20} = 4.47$, $P < .0001$. Score CI is (0.839, 1.000).

d.)

$$\begin{aligned} LRT &= 2[20 \log \frac{20}{.5(20)} + (20-20) \log \frac{20-20}{.5(20)}] \\ &= 2(20) \log \frac{20}{10} \\ &= 27.7 \end{aligned} \tag{2}$$

This test statistic has one degree of freedom.

```
stat<-2*20*log(2)

stat

1-pchisq(stat,1)
```

The p-value is $1.3978 \times 10^{-7}$. We reject the null hypothesis and find the same conclusion as in 1.7c. From the example on page 16 in the textbook I constructed

the confidence interval: We know the upper bound is 1 and the lower bound is $\exp \frac{3.84}{2\times 20} = 0.908$.

e.) I constructed the exact binomial test using R. We find p-value $= 1.907 \times 10^{-6}$, so we reject the null hypothesis and conclude similarly to 1.7c.

7. Agresti 1.9 For this problem, I used a $\chi^2$ goodness-of-fit test to see is 3:1 was the true ratio of green to yellow seedlings.

$$\chi^2_{GOF} = \Sigma_{j=1}^2 \frac{(n_j - \mu_j)^2}{\mu_j} = \frac{(854 - 827.25)^2}{827.25} + \frac{(249 - 275.75)^2}{275.75}$$

where the $\mu_j$'s were computed by $n \times 0.75$ or $\times 0.25$, $n = 1103$. Using R,

```
chisq.test(c(854,249),p=c(.75,.25))
```

Then the P-value is 0.063, providing moderate evidence against the null (if we set $\alpha = 0.05$, it fails to reject null hypothesis whereas if we set $\alpha = 0.10$, it rejects the null and conclude the alternative).

8. Question 8 From the question, we know $n = 100$ and we want to test $H_0 : \pi = 0.5$ versus $H_A : \pi > 0.5$ where $\pi$ is the proportion of women who improve on the new drug. I used a score test in R and found p-value=0.02275. We reject the null hypothesis; we have sufficient evidence to conclude the new drug is better. We are 95% confident that the true proportion of women who improve on the new drug is between 51.7% and 100%.

```
prop.test(60,100,p=.5,alternative="greater",conf.level=0.95,correction=F)
```

9. Agresti 1.17

   a.) $Y \sim Binomial(n, \pi)$, $E[Y] = n\pi$, $Var[Y] = n\pi(1-\pi)$.

   b.)

*Proof.*

$$
\begin{aligned}
Var[Y] &= \sum Var[Y_i] + 2\sum Cov[Y_i, Y_j] \\
&= n\pi(1-\pi) + 2\rho\pi(1-\pi) \\
&> n\pi(1-\pi)
\end{aligned}
$$

□

c.) **Theorem:** Suppose $P(Y_i = 1|\pi) = \pi \forall i$, but $\pi \sim g(\text{ffl})$ on [0,1] having mean $\rho$ and positive variance. WTS $Var[Y] > n\rho(1-\rho)$.

*Proof.*

$$
\begin{aligned}
Var[Y] &= E[Var[Y|\pi]] + Var[E[Y|\pi]] \\
&= nE[\pi] - nE[\pi^2] + n^2 Var[\pi] \\
&= n\rho - n(Var[\pi] + \rho^2) + n^2 Var[\pi] \\
&= n\rho - n\rho^2 + (n^2 - n)Var[\pi] \\
&= n\rho(1-\rho) + (n^2 - n)Var[\pi] \\
&> n\rho(1-\rho)
\end{aligned}
$$

□

10. Agresti 1.29

a.)

*Proof.*

$$L(\theta) = n_1 \log \theta^2 + n_2 \log[2\theta(1-\theta)] + n_3 \log(1-\theta)^2$$

$$= 2n_1 \log \theta + n_2 \log 2\theta + n_2 \log(1-\theta) + 2n_3 \log(1-\theta)$$

$$\begin{aligned}
\frac{\partial L(\theta)}{\partial \theta} &= \frac{2n_1\theta}{\theta^2} + \frac{2n_2}{2\theta} - \frac{n_2}{1-\theta} - \frac{2n_3}{1-\theta} \\
&= \frac{2n_1 + n_2}{\theta} - \frac{n_2 + 2n_3}{1-\theta} = 0 \\
&\implies \frac{2n_1 + n_2}{\theta} = \frac{n_2 + 2n_3}{1-\theta} \\
&\implies (2n_1 + n_2)(1-\theta) = (n_2 + 2n_3)\theta \\
&\implies 2n_1 + n_2 - 2n_1\theta n_2\theta = n_2\theta + 2n_3\theta \\
&\implies 2n_1 + n_2 = n_2\theta + n_2\theta + 2n_1\theta + 2n_3\theta \\
&\implies \hat{\theta} = \frac{2n_1 + n_2}{2n_1 + 2n_2 + 2n_3}
\end{aligned}$$

□

b.)

*Proof.*

$$\begin{aligned}
\frac{\partial^2 L(\theta)}{\partial\theta^2} &= \frac{\partial}{\partial\theta}[\frac{2n_1\theta}{\theta^2} + \frac{2n_2}{2\theta} - \frac{n_2}{1-\theta} - \frac{2n_3}{1-\theta}] && \text{from a.)} \\
&= 2n_1 + n_2\frac{-1}{\theta^2} - (n_2 + 2n_3)\frac{-1}{(1-\theta)^2} \\
&= \frac{n_2 + 2n_3}{(1-\theta)^2} - \frac{(2n_1 + n_2)}{\theta^2}
\end{aligned}$$

Then,

$$\frac{-\partial^2 L(\theta)}{\partial\theta^2} = \frac{(2n_1 + n_2)}{\theta^2} - \frac{n_2 + 2n_3}{(1-\theta)^2}$$

Then,

$$\begin{aligned}
E[\frac{-\partial^2 L(\theta)}{\partial\theta^2}] &= \frac{2(n\theta^2) + n2\theta(1-\theta)}{\theta^2} - \frac{2n\theta(1-\theta) + 2n(1-\theta)^2}{(1-\theta)^2} \\
&= 2n + \frac{2n}{\theta} - 2n - \frac{2n\theta}{1-\theta} + 2n \\
&= \frac{2n(1-\theta) - 2n\theta^2 + 2n(1-\theta)\theta}{\theta(1-\theta)} \\
&= \frac{2n - 2n\theta - 2n\theta^2 + 2n(\theta - \theta^2)}{\theta(1-\theta)} \\
&= \frac{2n - 2n\theta - 2n\theta^2 + 2n\theta + 2n\theta^2}{\theta(1-\theta)} \\
&= \frac{2n}{\theta(1-\theta)}
\end{aligned}$$

The asymptotic variance is the inverse of $E[\frac{-\partial^2 L(\theta)}{\partial\theta^2}]$, which is $\frac{\theta(1-\theta)}{2n}$. Therefore, the asymptotic standard error is $\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{2n}}$. □

c.) The expected counts for the genotypes are $n\theta^2, 2n\theta(1-\theta)$, and $n(1-\theta)^2$ since the counts are multinomially distributed. We could test these expected values compared to the observed $n_1, n_2$, and $n_3$ using $\chi^2_{GOF}$ with $df = (3-1) - 1 = 1$.