Solution to HW1

Problem 1.2

a. nominal; **b.** ordinal; **c.** Ordinal; **d.** nominal; **e.** nominal; **f.** ordinal.

Problem 1.4

(a) $Y \sim Bin(2, \pi = 0.5), P(Y = y) = \binom{2}{y} 0.5^y (1 - 0.5)^{2-y} = 0.25\binom{2}{y}$, for $y = 0, 1, 2$, which gives

$$P(Y = 0) = 0.25, P(Y = 1) = 0.5, P(Y = 2) = 0.25.$$

(b) When $\pi = 0.6$, $Y \sim Bin(2, \pi = 0.6)$ and $P(Y = y)$'s are:

$$P(Y = 0) = 0.16, P(Y = 1) = 0.48, P(Y = 2) = 0.36.$$

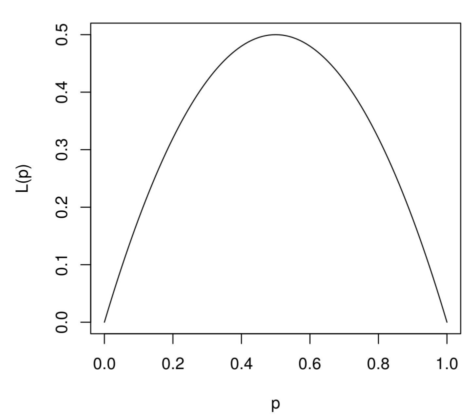
$\pi = 0.4$, $Y \sim Bin(2, \pi = 0.4)$ and $P(Y = y)$'s are:

$$P(Y = 0) = 0.36, P(Y = 1) = 0.48, P(Y = 2) = 0.16.$$

(c) The likelihood function is

$$L(\pi) = \binom{2}{1} \pi^1 (1 - \pi)^{2-1} = 2\pi(1 - \pi).$$

The likelihood function



(d) It is obvious from the likelihood plot that π maximizing $L(\pi)$ is $\widehat{\pi} = 0.5$.

Problem 1.6

(a) Since $n_1 + n_2 + n_3 = 3$, so $n_3 = 3 - n_1 - n_2$.

(b) All possible sets are: (3,0,0), (2,1,0), (2,0,1), (1,2,0), (1,1,1), (1,0,2), (0,0,3), (0,1,2), (0,2,1), (0,3,0).

(c) The probability $= \binom{3}{1,2,0} 0.25^1 0.5^2 0.25^0 = 3/16$.

(d) $n_1 \sim Bin(n = 3, \pi = 0.25)$.

Problem 1.12

(a) We have a binomial problem with $n = 25$ and observed $y = 0$ so $p = 0$. The Wald test statistic became $z = (0 - 0.5)/0$, an undefined value.

(b) The 95% Wald CI is $[0, 0]$, not believable.

(c) The score test statistic is

$$z = \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/n}} = \frac{0 - 0.5}{\sqrt{0.5 \times 0.5/25}} = -5.$$

The P-value of the score test is $P(|Z| \geq |-5|) = 5.73 \times 10^{-7}$.

(d) The score CI for π is:

$$\begin{aligned} & \left| \frac{0 - \pi_0}{\sqrt{\pi_0(1 - \pi_0)/25}} \right| \leq 1.96 \\ \Leftrightarrow \quad & \frac{\pi_0^2}{\pi_0(1 - \pi_0)/25} \leq 1.96^2 \\ \Leftrightarrow \quad & \frac{\pi_0}{1 - \pi_0} \leq 0.1536 \\ \Leftrightarrow \quad & 0 \leq \pi_0 \leq 0.133. \end{aligned}$$

(e) Adding 2 "successes" and 2 "failures" to the data, we got $y^* = 2, n^* = 29, \widehat{p}^* = 0.069$. So the modified 95% Wald CI is

$$0.069 \pm 1.96\sqrt{0.069 \times (1 - 0.069)/29} = [-0.022, 0.16] = [0, 0.16].$$

Problem 1.13

(a) The likelihood function under $H_0 : \pi = \pi_0$ is

$$L(\pi_0) = \binom{n}{y} \pi_0^y (1 - \pi_0)^{n-y}$$

For our problem, $y = 0$, $n = 25$, so $L(\pi_0) = (1 - \pi_0)^{25}$. Under $H_0 : \pi = 0.5$, $\pi_0 = 0.5$ and $L_0 = 0.5^{25}$.

(b) The likelihood function under all possible value π is

$$L(\pi) = \pi^y (1-\pi)^{n-y} = (1-\pi)^{25},$$

a monotonic deceasing function of $\pi \in [0,1]$. Hence the maximum likelihood is $L_1 = L(0) = 1$.

(c) The LRT statistic is

$$-2\log(L_0/L_1) = -2\log(0.5^{25}) = -50\log(0.5) = 34.66.$$

The P-value $= P(\chi_1^2 \geq 34.66) = 3.9 \times 10^{-9}$.

(d) For $H_0 : \pi = 0.074$, the LRT is

$$-2\log(L_0/L_1) = -2\log(1-0.074)^{25} = 3.844,$$

with the P-value $= P(\chi_1^2 \geq 3.844) = 0.05$. That is, $\pi_0 = 0.074$ is the smallest value of π_0 that we can reject at level 0.05. Therefore, the 95% LR CI is $[0, 0.074]$.

Problem 2.1: From the conditions given, we have $P(-|C) = 1/4$ for the false negative rate, and $P(+|\bar{C}) = 2/3$ for the false positive rate.

(a) The correct probabilities are $P(-|C) = 1/4$ and $P(+|\bar{C}) = 2/3$.

(b) The sensitivity is $P(+|C) = 1 - (-|C) = 1 - 1/4 = 3/4$.

(c) $P(C,+) = P(+|C)P(C) = (3/4) \times 0.01 = 0.0075$, $P(C,-) = P(-|C)P(C) = (1/4) \times 0.01 = 0.0025$, $P(\bar{C},+) = P(+|\bar{C})P(\bar{C}) = (2/3) \times (1-0.01) = 0.66$, $P(\bar{C},-) = P(-|\bar{C})P(\bar{C}) = (1/3) \times (1-0.01) = 0.33$.

(d) The marginal distribution for Y is: $P(Y=+) = P(C,+) + P(\bar{C},+) = 0.0075 + 0.66 = 0.6675$, $P(Y=-) = 1 - P(Y=+) = 0.3325$;

(e) $P(C|+) = P(C,+)/P(Y=+) = 0.0075/0.6675 = 0.011$. So if a man is diagnosed with the cancer, the chance that he actually has the cancer is still pretty low (0.011). However, this probability is 10% greater than the cancer probability without the diagnosis. So the diagnosis is still useful.

Problem 2.12:

(a) The 2×2 classification table:

		Heart Attack	
		Yes	No	
Treatment	Aspirin	198	19736	19934
	Placebo	193	19749	19942

(b) The estimated odds-ratio is:

$$\widehat{\theta} = \frac{198 \times 19749}{193 \times 19736} = 1.027.$$

The odds of having a heart attack for women receiving aspirin is 1.027 times the odds of having a heart attack for women receiving placebo. Since heart attack is a rare disease, we can interpret that women receiving aspirin is 2.7% more likely to have a heart attach than women receiving placebo.

(c) A 95% CI for $\log \theta$ is:

$$\log(1.027) \pm 1.96\sqrt{1/198 + 1/19736 + 1/193 + 1/19749} = [-0.173, 0.226].$$

So a 95% CI for θ is $[e^{-0.173}, e^{0.226}] = [0.841, 1.254]$.

(d) The estimated relative risk is:

$$\widehat{\psi} = \frac{198/19934}{193/19942} = 1.026.$$

Using SAS, a 95% CI for the relative risk is [0.843, 1.250].

Problem 2.14

(a) We cannot compare these two proportions using the methods we discussed so far since these two proportions were calculated based on the same sample, so they are correlated. If we knew the data at individual level, then we can compare them. We will discuss this kind of problem in Chapter 8 for matched data.