2019R1 Discrete Data Analysis (STAT5107) Assignment 1

Yiu Chung WONG 1155017920

```
set.seed(5107);
```

1.

a. nominal
b. ordinal
c. interval
d. nominal
e. ordinal
f. nominal
g. ordinal

2.

Variance of binomial distribution is

$$\sigma^2 = n\pi(1-\pi)$$

- When π is close to zero, everything is multiplied by a value close to zero, hence σ^2 is small.
- When π is close to one, everything is multiplied by (1 - somehting close to one), which is close to zero. Hence σ^2 is also small.
- When π is close to `0.5` , `n` is multiplied by `0.5 * (1 - 0.5)`, which is close to 0.25, which is bigger than the values above.

A smaller variance means more precise estimate of π and vice versa.

3.

H_0: $\pi = 0.5$ H_1: $\pi \neq 0.5$

```
yes_count <- 842;
no_count <- 982;
n <- yes_count + no_count;
H0 <- 0.5;
alpha <- .05

pie_hat <- yes_count / n;

z_s <- (pie_hat - H0) / sqrt(H0*(1-H0)/n);

p_value <- pnorm(q = z_s)

CI <- pie_hat + c(1, -1) * qnorm(alpha/2) * sqrt(pie_hat*(1-pie_hat)/n)
```

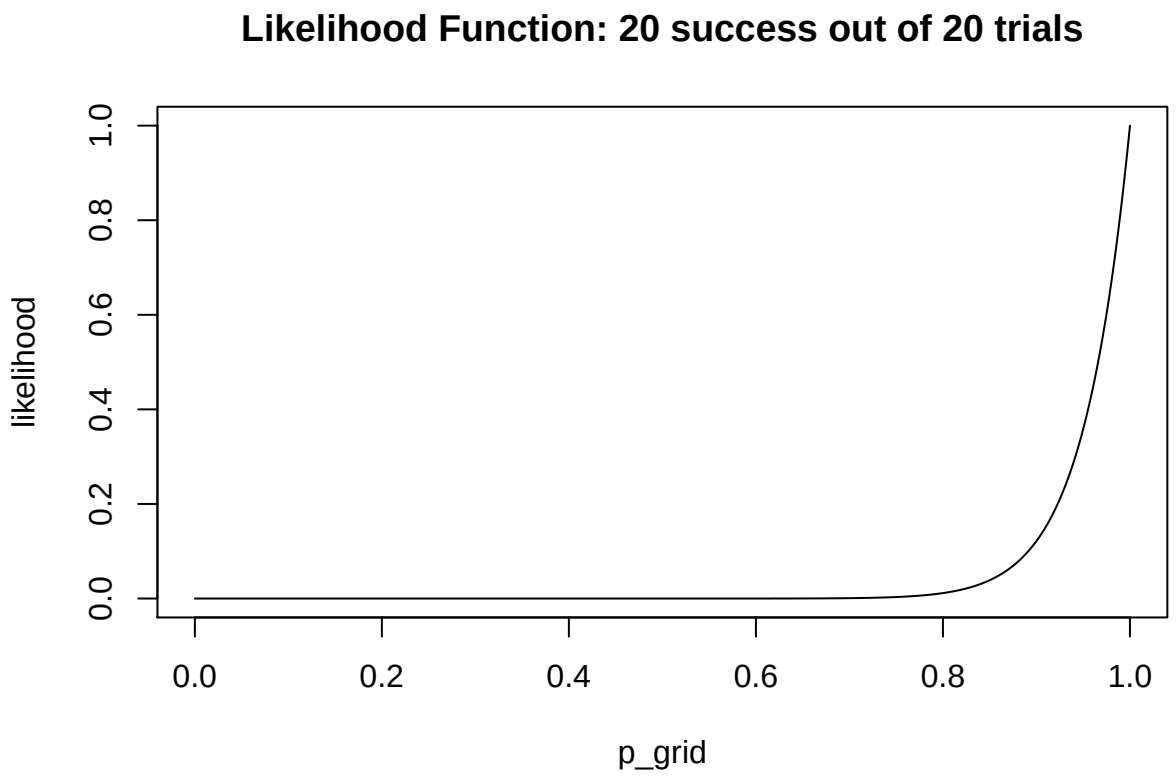
The p-value is 5.2263376×10^{-4} which is below the two tale cut off at 0.05. There is enough evidence to reject H_0 and favor H_1. 95% confidence interval: 0.4387446, 0.484501

4a.

```
n_better <- 20;
n_trial <- 20;
n_simulate <- 1e4;
H0 <- 0.5;
pie_hat <- n_better/n_trial;

p_grid <- seq(from=0 , to=1 , length.out=n_simulate);

likelihood <- dbinom(x = n_better, size = n_trial, prob=p_grid);
plot(p_grid, likelihood, main = "Likelihood Function: 20 success out of 20 trials", type = 'l');
```



The maximum likelihood estimate of π is

$$success/no.trial = 20/20 = 1$$

4b.

```
se <- sqrt(pie_hat * (1 - pie_hat) / n_trial)
CI <- pie_hat + c(1, -1) * qt(alpha/2, df = n_trial - 1) * se;
z_W <- (pie_hat - H0) / se;
```

Wald test statistic is inf. The 95% Wald confidence interval for π is between 1 and 1.

4c.

```
se <- sqrt(H0 * (1 - H0) / n_trial);
z_S <- (pie_hat - H0) / se;
CI <- pie_hat + c(1, -1) * qt(alpha/2, df = n_trial - 1) * se;
cutoff <- qt(alpha/2, df = n_trial - 1, lower.tail = F);
```

Score test statistic is 4.472136, which is greater than the cutoff at 2.0930241. There is evidence to rehject H_0. The 95% Score confidence interval for π is between 0.7659928 and 1.2340072.

4d.

```
L_H0 <- dbinom(n_better, n_trial, prob = H0, log = TRUE);
L_pie_hat <- dbinom(n_better, n_trial, prob = n_better/n_trial, log = TRUE);
z_L <- -2 * (L_H0 - L_pie_hat);
cutoff <- qchisq(.95, 1, lower.tail = T);
upper_bound <- 1 - exp(cutoff/(2*L_H0));
```

The confidence interval is between zero and 0.1293814; The log likelihood statistic is 27.7258872, which is greater than the cutoff at 3.8414588. At 0.05 α level, we conclude that there is evidence to suggest the data DO NOT follow the null hypothesis.

6.

```
deaths <- c(109, 65, 22, 3, 1, 0);
count <- 0:5;
lamb <- mean(rep(count, times=deaths));
expected <- dpois(count, lambda = lamb) * sum(deaths);

#merge cases fewer than 5 observations
deaths_merged <- deaths[1:3];
deaths_merged[3] <- sum(deaths[-c(1, 2)]);
expected_merged <- expected[1:3];
expected_merged[3] <- sum(expected[-c(1, 2)]);
count <- 0:2;

chi_squared <- sum( ((deaths_merged - expected_merged)^2)/expected_merged );

df <- length(count) - 1 - 1;

p_value <- pchisq(chi_squared, df = df, lower.tail = F);
```

Average death is 0.61 deaths per year. The chi-square statistic is 0.0634512; the p-value of the alternative hypothesis of the data fitting to a Poission distribution is 0.8011221. At 0.05 α level, we conclude that there is no real evidence to suggest the data DO NOT follow a Poisson distribution.