p8130_hw2_yl5508

Yifei LIU (yl5508)

2023/10/11

Problem 1

(a) $P(exactly\ 40) = \binom{56}{40} \cdot (0.73)^{40} \cdot (1-0.73)^{16} = 0.113 = 11.3\%$

(b)

```r
trial = 56
success = 39
p = 0.73
print(1-pbinom(success,trial,p))
```

```
## [1] 0.6678734
```

The probability that at least 40 of them have at least 1 checkup is 0.6678734.

(c) **(1)** The condition we should check before using Poisson appoximation to binomial is:
- n must be large (n>100): in this problem, n=56<100.
- Probability of success p should be small (p<0.01): in this problem, p is 0.73.
As a result, we should not use Poisson as an approximation to binomial.
(2) Try to use normal approximation:
- np 10: $np = 56 \cdot 0.73 = 40.88 > 10$
- n(1-p) 10: $n(1-p) = 56 \cdot (1-0.73) = 15.12 > 10$
So, we can use normal approximation to binomial. The normal approximation distribution would be $N\ (40.88, 11.04)$

```r
mu_norm = 56*0.73
sd_norm = sqrt(56*0.73*(1-0.73))
norm_eql40 = pnorm(40.5,mu_norm,sd_norm)-pnorm(39.5,mu_norm,sd_norm)
norm_leq40 = 1-pnorm(39.5,mu_norm,sd_norm)
print(c(mu_norm,sd_norm,norm_eql40,norm_leq40))
```

```
## [1] 40.8800000  3.3222884  0.1155355  0.6610668
```

For problem (a): using normal approximation to binomial, we get possibility 11.6%.
For problem (b): using normal approximation to binomial, we get possibility 66.1%.
The results shown above is quite close to the results we have calculated from binomial.

(d) Expectation of a binomial R.V. is `E[X] = np`, which is 40.88.

(e) Variance of a binomial R.V. is `Var(x) = np(1-p)`, which is 11.0376. The standard deviation of it would be 3.3222884($sd = \sqrt{Var(x)}$).

Problem 2

(a)

```
print(ppois(2,6))
```

```
## [1] 0.0619688
```

The PMF of Poisson distribution can be expressed as $P(X = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$.
For tornado happened fewer than 3 times, $P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \sum_{k=0}^{2} \frac{e^{-6} \cdot 6^k}{k!} = 0.062 = 6.2\%$.

(b)

```
print(dpois(3,6))
```

```
## [1] 0.08923508
```

The probability that tornado will happen exactly 3 tims is 0.0892351.

(c)

```
print(1-ppois(3,6))
```

```
## [1] 0.8487961
```

The probability that tornado will happen more than 3 tims is 0.8487961.

Problem 3

(a)

```
print(1-pnorm(137,128,10.2))
```

```
## [1] 0.188793
```

The probability of a selected American man (20-29) with systolic blood pressure above 137.0 is 0.188793.

(b)

```
print(pnorm(125,128,10.2/sqrt(50)))
```

```
## [1] 0.01877534
```

The population distribution of the R.V. is normal and the sample size is larger than 30. So, we can denote the sampling distribution as $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$, which is $\overline{X} \sim N(128, \frac{10.2}{\sqrt{50}})$ in this problem. The probability that sample mean will be less than 125.0 is 0.0187753.

(c)

```r
print(qnorm(0.9,128,10.2/sqrt(40)))
```

```
## [1] 130.0668
```

For sample size of 40, the sampling distribution would follow: $\overline{X} \sim N(128, \frac{10.2}{\sqrt{40}})$. The 90th percentile of the sampling distribution is 130.0668372.

Problem 4

(a)

```r
#t statistic
t_975 = qt(0.975,39)
print(t_975)
```

```
## [1] 2.022691
```

For true from population is unknown, we need to calculate the estimated standard error: $\frac{s}{\sqrt{n}} = \frac{10}{\sqrt{40}} = 1.58$. So, the confidence interval(CI) for the population mean pulse rate would be shown as $P(\overline{X} - 2.02\sigma_{\overline{X}} < \mu < \overline{X} + 2.02\sigma_{\overline{X}}) = 0.95$, among which CI is $76.81 < \mu < 83.19$.

(b) The confidence interval calculated above means that, for most (95%) of the random variables, $\overline{X}$ will fall within +/- 2.02 estimated SE of the true mean μ. The 95% CI is $(76.81, 83.19)$.

(c)

```r
#statistic t
t = abs((80-70)/(10/sqrt(40)))
t_995 = qt(0.995,39)
print(c(t,t_995))
```

```
## [1] 6.324555 2.707913
```

Using method of tests for the mean of a normal distribution with known variance.
Suppose that $\mu_0 = 70$, and $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$.
With the significant level $\alpha = 0.01$, test statistic is $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}} = \frac{80-70}{10/\sqrt{40}} = 6.32$.
Criteria:
- Reject H_0: if $|t| > t_{39,0.995}$,
- Fail to reject H_0: if $|t| \leq t_{39,0.995}$. $t_{39,0.995} = 2.71$.
Cause $|t| > t_{39,0.995}$, we can reject H_0.
Result: with significant level $\alpha = 0.01$, we should reject H_0 and the mean pulse of young women suffering from fibromyalagia isnot equal to 70. In other words, the possibility of getting a sample with a mean equivalent to 80 or more extreme than that is less than 1%, given that the population mean is 70.