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)

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

```
## [1] 0.67
```

The probability that at least 40 of them have at least 1 checkup is 0.67.
(c) 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.
(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 11.04. The standard deviation of it would be 3.32($sd = \sqrt{Var(x)}$).

Problem 2

(a)

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

```
## [1] 0.062
```

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] 0.089
```

The probability that tornado will happen exactly 3 tims is 0.09.
(c)

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

```
## [1] 0.85
```

The probability that tornado will happen more than 3 tims is 0.85.

Problem 3

(a)

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

```
## [1] 0.19
```

The probability of a selected American man (20-29) with systolic blood pressure above 137.0 is 0.19.
(b)

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

```
## [1] 0.019
```

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 0.02.
(c)

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

```
## [1] 130
```

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 130.07.

Problem 4

(a)

```
#z statistic
qnorm(0.975,0,1)
```

```
## [1] 2
```

```
mu_lower = qnorm(0.025,80,10)
mu_upper = qnorm(0.975,80,10)
mu_lower
```

```
## [1] 60
```

```r
mu_upper
```

```
## [1] 100
```

For sample size of 40 (>30), the sampling distribution would follow: $\overline{X} \sim N(80, 10)$. In this case, standard error(SE) is $\sigma_{\overline{X}} = \frac{\sigma^2}{n} = 10$.
So, the confidence interval(CI) for the population mean pulse rate would be shown as $P(\overline{X} - 1.96\sigma_{\overline{X}} < \mu < \overline{X} + 1.96\sigma_{\overline{X}}) = 0.95$, among which CI is $60.4 < \mu < 99.6$.
(b) The confidence interval calculated above means that, for most (95%) of the random variables, $\overline{X}$ will fall within +/- 1.96SE of the true mean μ.
(c) 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 $z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{80-70}{10} = 1$.
Criteria:
- Reject H_0: if $|z| > z_{1-0.05}$,
- Fail to reject H_0: if $|z| \leq z_{1-0.05}$.

```r
print(qnorm(0.95,0,1))
```

```
## [1] 1.6
```

$z_{1-0.05} = z_{0.95} = 1.64$.
Cause $z < z_{1-0.05}$, we fail to reject H_0.
With significant level $\alpha = 0.01$, we cannot reject H_0 and the mean pulse of young women suffering from fibromyalagia is equal to 70.