BiostatisticsMidterm-tk2886

Problem 1

a). Assumptions that must be true to use the Poisson Distribution to model the number of infections per month:

1. Events occur one at a time; two or more events cannot occur exactly at the same time and location;
2. The occurrence of an event in a given period is independent of the occurrence of an event in a non-overlapping period;
3. The expected number of events during any period is constant.

b).

b) Suppose the number of infections per month follows a Poisson distribution. What is the probability that in the next month the hospital's patients will have exactly 2 unexplained infections? Include the formula and all the key steps in your calculations. *(5 points)*

Formula: $P(X=x) = f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \ x = 0, 1, 2, \ldots, n$

$\lambda = 7$, rate of unexplained infection among patients per month

Calculate: $P(X=2) = \frac{7^2 e^{-7}}{2!} = 0.0223$

We may use R code to get the same value.

```
prob = dpois(2, 7)
prob
```

```
## [1] 0.02234111
```

The probability that in the next month the hospital's patients will have exactly 2 unexplained infections is: **0.0223411**.

Problem 2

Formula: Bayes Theorem

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_i)P(B_1) + P(A|B_2)P(B_2) + \dots P(A|B_k)P(B_k)}$$

D = developed CHD

C = Initial serum cholesterol levels above 200

$P(D) = 0.25$, $P(D^c) = 1 - 0.25 = 0.75$

$P(C|D) = 0.60$

$P(C|D^c) = 0.16$

Interested in:

$$P(D^c|C^c) = \frac{P(D^c \cap C^c)}{P(C^c)} = \frac{P(C^c|D^c)\,P(D^c)}{P(C^c|D^c)P(D^c) + P(C^c|D)P(D)}$$

$P(C^c|D^c) = 1 - P(C|D^c) = 0.84$

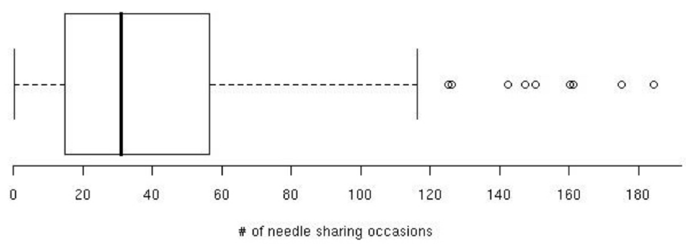
$P(C^c|D) = 1 - P(C|D) = 0.40$

$$\frac{(.84)(.75)}{(.84)(.75) + (.40)(.25)} = \frac{.63}{.63 + .1} = \boxed{0.8630 \approx 0.86}$$

The probability that a random chosen subject will not develop CHD, given that he had an initial serum cholesterol level below or equal 200 is: **0.86**.

Problem 3

Problem 3:



Researchers from New York, studying needle sharing behavior among intravenous drug users, randomly selected 500 patients from the city's two major drug detoxification clinics. At admission, each participant reported the number of their needle sharing occasions in the last year. The researchers summarized the observed data for 'number of needle sharing occasions in the last year' in the box-plot above.

a) **[Circle only one correct answer.]** Based on the box-plot of the data, the proportion of subjects whose number of needle sharing occasions is less than the sample average number of needle sharing occasions is: *(4 points)*
   i. Less than 50%
   ii. More than 50%
   iii. Exactly 50%
   iv. Exactly 100%

b) **[Circle only one correct answer.]** What is the probability that the number of needle sharing occasions for a randomly selected patient from the study is smaller than the third quartile (Q3) and larger than the first quartile (Q1)? *(4 points)*
   i. About 25%
   ii. About 50%
   iii. About 75%
   iv. About 100%

c) **[Circle only one correct answer.]** In order to report the number of needle sharing occasions per WEEK, the researchers divided all the observed data by 52 and computed new sample statistics. What is the relationship between the sample statistics for ONE WEEK and the sample statistics for ONE YEAR? *(4 points)*
   i. The 'one week' mean, median and standard deviation are 52 times larger than the 'yearly' mean, median and standard deviation
   ii. The 'one week' mean, median and standard deviation are 52 times smaller than the 'yearly' mean, median and standard deviation
   iii. The 'one week' mean and median are 52 times larger than the 'yearly' mean and median. The standard deviation will stay the same.
   iv. The 'one week' mean and median are 52 times smaller than the 'yearly' mean and median. The standard deviation will stay the same.

Problem 5

- The variable we're studying in the population is normally distributed.
- Placebo group ⇒ 40 subjects } $n \geq 30$ ✓ } Normal ✓
- Treatment group ⇒ 80 subjects } $n \geq 30$ ✓ } Normal ✓

Test Equality of the Variances:

$H_0: \sigma_1^2 = \sigma_2^2 \qquad H_A: \sigma_1^2 \neq \sigma_2^2$

$F_{stat}: \dfrac{S_1^2}{S_2^2} = \dfrac{(1.226)^2}{(0.291)^2} = 17.749 \approx 17.75$

$F_{crit}: F_{n_1-1, n_2-1, 1-\frac{\alpha}{2}} = F_{40-1, 80-1, 1-\frac{0.05}{2}}$ } 1.687 — qf(.975, 39, 79)

$F_{crit} < F_{stat}$ }
- Reject Our Null Hypothesis
- The variances are unequal.

Since, we have unequal variances, we will use Two Sample Independent t-test: (unequal variance)

$H_0: \mu_P = \mu_N \qquad H_A: \mu_P < \mu_N$

(μ_P: Placebo Patch; μ_N: Nicotine Patch)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{0.222 - 0.125}{\sqrt{\frac{(0.291)^2}{80} + \frac{(1.226)^2}{40}}} = 0.4934$$

$$\underbrace{df = \frac{\left(\frac{s_T^2}{n_T} + \frac{s_P^2}{n_P}\right)^2}{\frac{s_T^4}{n_T^2 \cdot (n_T - 1)} + \frac{s_P^4}{n_P^2 \cdot (n_P - 1)}} = 41.2}$$

- Given to us on Exam.
- Round down to the nearest Integer d'': 41

t_{crit} } $t_{d'', 1-\frac{\alpha}{2}} = t_{41, 0.975}$

qt(.975, 41) = 2.01954

$|t| \leq t_{d'', 1-\frac{\alpha}{2}}$ } $|0.4934| \leq 2.01954$ } Fail to Reject

P-value: pt(.4934, 41) = 0.6878 > $\underbrace{0.05}_{\text{alpha}}$ } Fail to Reject

Conclusion: We fail to reject the null hypothesis. We do not have enough evidence to reject the null hypothesis that there is no difference of plasma nicotine levels when applying nicotine patch versus applying placebo patch.