Statistics and hypothesis testing

Week 1

Question 1

You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle?

Answer

- All say yes: all three lie or three say the truth, otherwise, there would be different answers.
- P("all say the truth") = (2/3)^3 = 8/27
- P("all lie") = (1/3)^3 = 1/27
- P("all yes") = 1/27 + 8/27 = ⅓
- Out of these numbers, there is 8/27/(⅓) = 8/9 chance it's actually raining.

Question 2

You have two coins. One of is fair and the other is biased and comes up heads with probability 3/4. You randomly pick coin and flip it twice, and get heads both times. What is the probability that you picked the fair coin?

Answer

• F: denote the event you picked the fair coin
B: denote the event you picked the biased coin
D: observed two heads in two tosses
• We want to know $P\ (F\ |D)$
$P\ (F\ |D)$ = $P\ (D\ |\ F)\ P\ (F)$ / $P\ (D)$
$P\ (D\ |F)$ = 1/ 4
$P\ (F)$ = 1/ 2
• What is $P\ (D)$:
$P\ (D)$= $P\ (F)\ P\ (D\ |F)$ + $P\ (B)\ P(D\ |B)$
$P\ (F)$ = $P\ (B)$ = 1 / 2
$P\ (D\ |F)$ = 1 / 4
$P\ (D\ |B)$ = 9 / 16
$P\ (D)$ = 1 / 8 + 9 / 32 = 13 / 32
$P\ (F\ |D)$ = $P\ (D\ |F)\ (P\ |\ F)$ / $(P\ |\ D)$

$P(F|D)$ = (1 / 8) / (13 / 32) = 4/ 13

Question 3

Provide a simple example of how an experimental design can help answer a question about behavior. How does experimental data contrast with observational data?

Instruction:
Difference between observational data and experimental data.

Answer

- You are researching the effect of music-listening on studying efficiency
- You might divide your subjects into two groups: one would listen to music and the other (control group) wouldn't listen anything!
- You give them a test
- Then, you compare grades between the two groups

Differences between observational and experimental data:

- Observational data: measures the characteristics of a population by studying individuals in a sample, but doesn't attempt to manipulate or influence the variables of interest

- Experimental data: applies a treatment to individuals and attempts to isolate the effects of the treatment on a response variable

Observational data: find 100 women age 30 of which 50 have been smoking a pack a day for 10 years while the other have been smoke free for 10 years. Measure lung capacity for each of the 100 women. Analyze, interpret and draw conclusions from data.

Experimental data: find 100 women age 20 who don't currently smoke. Randomly assign 50 of the 100 women to the smoking treatment and the other 50 to the no smoking treatment. Those in the smoking group smoke a pack a day for 10 years while those in the control group remain smoke free for 10 years. Measure lung capacity for each of the 100 women.

Analyze, interpret and draw conclusions from data.

Question 4

In a study of emergency room waiting times, investigators consider a new and the standard triage systems. To test the systems, administrators selected 20 nights and randomly assigned the new triage system to be used on 10 nights and the standard system on the remaining 10 nights. They calculated the nightly median waiting time (MWT) to see a physician. The average MWT for the new system was 3 hours with a variance of 0.60 while the average MWT for the old system was 5 hours with a variance of 0.68. Consider the 95% confidence interval estimate for the differences of the mean MWT associated with the new system. Assume a constant variance. What is the interval? Subtract in this order (New System - Old System).

答案

Confidence interval for the difference of the means assuming equal variances:

$$(new\ -\ old)\ \pm t \times sp\ \times \sqrt{(\frac{1}{n_1})\ + (\frac{1}{n_2})}$$

Degree of freedom for student t test is 10 + 10 - 2 = 18

The critical value drawn from t distribution with 18 degree of freedom with p value of 0.05 is 2.1 (We can get this number in R using: qt(0.975, df = 18))

Pooled variance = $\sqrt{\frac{0.6 * 9 + 0.68 * 9}{10+10-2}} = 0.8$

Then 95% confidence interval is [-2 - 2.1 * 0.8 * $\sqrt{\frac{1}{10} + \frac{1}{10}}$, -2 + 2.1 * 0.8 * $\sqrt{\frac{1}{10} + \frac{1}{10}}$] = [-2.75, -1.25]