# Stat13 Homework 6

http://www.stat.ucla.edu/~dinov/courses_students.html

## Suggested Solutions

**HW_6_1** (25 points)

Let X = price of 100 g. X is approximately Normal.
Plan 1: T1 = 6 X
Plan 1: T2 = Y1 + Y2 + Y3 + Y4, where each Y is 1.5 X

Since X is Normal(mean = 1205, SD = 100), we have: Var(X) = 10000
E(Y) = 1.5 * 1205 = 1807.5, Var(Y) = 1.5 * 1.5 * Var(X) = 22500, SD(Y) = 150.
Y is Normal(mean = 1807.5, 150^2)

(a) E(T1) = 6*1205 = 7230. (2 points)
Var(T1) = 36 * Var(X), so SD(T1) = sqrt(36 * 100 * 100) = 600. (2 points)
T1 is Normal(mean = 7230, SD ^2= 600^2). (2point)

(b) E(T2) = 4*1807.5 = 7230. (2 points)
Var(T2) = Var(Y1 + Y2 + Y3 + Y4) = 4 *Var(Y1) = 4 * 22500, SD(T1) = 300. (2 points)
Here we assume each Y's are independent.
T2 is Normal(mean = 7230, SD^2 = 300^2). (2point)

(c) The variance of plan 1 is larger. (2 points)

(d) Plan 1:
P(T1 > 5800) =
P( Z > (5800 – 7230)/600 ) = P(Z > -2.38333) = 0.9914 (3 points)

(e) Plan 2:
P(T2 > 5800) =
P( Z > (5800 – 7230)/300 ) = P(Z > -4.77) = .999999 (3 points)

(f) Plan 1:
P(T1 < 5900) =
P( Z < (5900 – 7230)/600 ) = P(Z < -2.217) = 0.0133 (3 points)

(g) Either one of the following explanations is acceptable: (2 points)
*Explanation 1*: Probability in (e) > Probability in (d), so it is more likely to exceed $5800 using plan 2. Therefore, plan 1 is safer.
*Explanation 2:* Smaller variance yields better prediction. Plan 2 has smaller variance so it is safer.

**<u>HW_6_2</u> (10 points)**

Each die: Let X = possible outcomes, so X could take on 1, 2, 3, 4, 5, 6, 7, and 8.
P(X = 1) = P(X = 2) = …. = 1/8.

E(X) = (1 + 2 + 3 + … + 8) / 8 = 36/8 = 4.5 ( 1 point)
Var(X) = $( (1 - 4.5)^2 + (2 - 4.5)^2 + \ldots + (8 - 4.5)^2 ) / 8 = 5.25$. (1 point)

Five dice is rolled twice --

Let $X_{i,j}$ = the outcome of the ith die, at the jth time, i = 1, 2, 3, 4, 5, and j = 1, 2,3.
(This is just one way to assign subscripts.)
Then $Y = X_{1,1} + X_{1,2} + \ldots + X_{5,2} + X_{5,3}$ (there are 15 terms here.)
Note each $X_{i,j}$ is independent from another, so variance of the sum is the sum of the variance.

$m_Y$ = E(Y) = 3 * 5 * 4.5 = 67.5. ( 2 points)
Var(Y) = $Var(X_{1,1} + X_{1,2} + \ldots + X_{5,2} + X_{5,3})$ = 3 * 5 * 5.25 = 78.75.

So SD(Y) = sqrt(78.75) = 8.87412 ( 2 points)
Now we carry out the experiment 11 times:

Let $\overline{Y}$ = sample mean of five dice being rolled twice. Sample size = 11.
The by CLT, $\overline{Y}$ is Normal. To estimate the mean and SD of $\overline{Y}$ :
Mean($\overline{Y}$) = 67.5 ( 2 points)
SD($\overline{Y}$) = sqrt( Var(Y) / 11 ) = 2.675648 (2 points)

**<u>HW_6_3</u> (15 points)**

Let X = number of correctly remembered words of a mnemonics group subject.
(like “treatment”)
Let Y = number of correctly remembered words of a normal group subject.
(like “control”)

These two groups of samples are independent.

(a) For the sampled data:
mean(X) = 14.1, SD(X) = 2.468752
mean(Y) = 9.631579, SD(Y) = 3.33684 (1 point for each number)

(b) Let $\overline{X}$ = sample mean of the normal group.
Let $\overline{Y}$ = sample mean of the normal group
An estimate of the “difference in the mean” is D = $\overline{X} - \overline{Y}$ = 4.468421 ( 1point)

Sample variance of $\overline{X}$ , $S_x^2$ = 11.13450
Sample variance of $\overline{Y}$ , $S_y^2$ = 6.094737
SD(D) = $\text{sqrt}(S_x^2/19 + S_y^2/20)$ = 0.9438026 ( 1 point)

And D follows a t-distribution of DF = min(20 –1, 19 – 1) = 18.

$t_{0.975}$ = 2.101 ( 1 point)
95% CI = 4.468 ± 2.101 * 0.944 = (2.485492, 6.45135) ( 1 point)

Plain English sentence: The 95%CI does not cover zero, and this suggests that the difference, D, is significantly different from zero. ( 2 points)

(c) Approximate twice as much as old sample size.
We have: n1 = 20 and n2 = 19 are about the same. So, use n1 ≈ n2.
To find the new CI, plug in:

New n1 = 4 n1,
New n2 = 4 n2 ≈ 4 n1

Sample variance $S_x^2$ and $S_y^2$ are approximately the same as they were before plugging using 4 n1.

Then, the new CI = 0.5 old CI.

Note that $t_{0.975}$ does not change much (still about 2).
So the new sample should be four times of the old one. (4*39=156) ( 2 points)

(d) We don't know. It is possible that the 95% CI covers / doesn't cover the true mean. But if we do such experiment again and again, then ABOUT 95% of the CI's we construct will cover the true mean. (3 points)