

STAT 13, section 1, Winter 2011, UCLA Statistics
HW 5; Problem Solution

HW 5.1

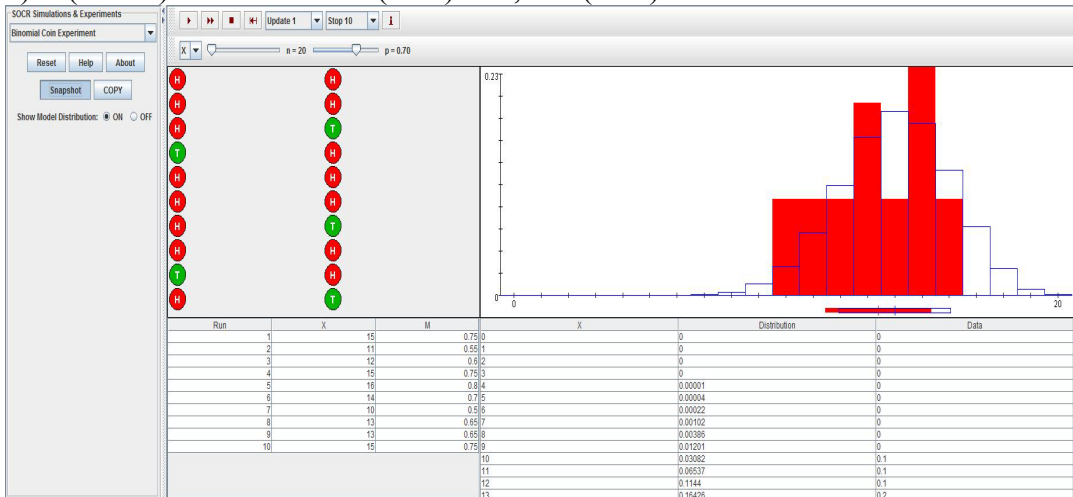
Let X denote the number of mutants in a random sample of 5, using the binomial distribution ($X \sim \text{Bin}(n=5, p=.4)$), we find that

- 1) $P(\hat{p}=0)=P(X=0)=(.6)^5=.078$
- 2) $P(\hat{p}=.1)=0$
- 3) $P(\hat{p}=.5)=0$
- 4) $P(\hat{p}=.6)=P(X=3)=6*(.4)^3*(.6)^2=.23$
- 5) $P(\hat{p}=.7)=0$
- 6) $P(\hat{p}=1.0)=P(X=5)=(.4)^5=.01$



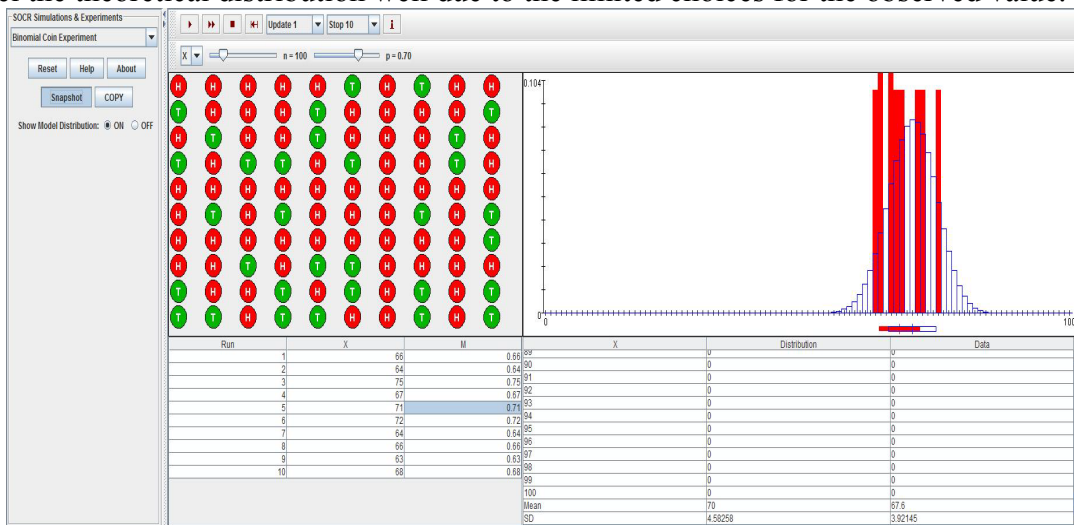
Next, using the binomial coin experiment, we first find that the empirical probabilities do not quite match the theoretical experiment. (Individual results may vary.) The snapshot below showed that the empirical distribution (red) has a larger mode but a smaller mean than that of the theoretical distribution. Due to our initial choices of a small n and the number of runs, the empirical results hardly capture the left-skewed feature, much less the actual probabilities of the underlying distribution.

For parts 2, 3 and 5, one can also use the approximation on X . In particular, $P(\hat{p}=.1)=P(X=.5)$ which means either $P(X=0)=.078$ or $P(X=1)=.259$ or the average of them .1685. $P(\hat{p}=.5)=P(X=2.5)$, then $P(X=2)=.346$ or $P(X=3)=.23$ or .288 are all acceptable. $P(\hat{p}=.7)=P(X=3.5)$. So use either $P(X=3)=.23$, or $P(X=4)=.077$ or .154.



They are supposed to be similar, under sufficient conditions.

When we increase n from 20 to 100 for each experiment while keeping the same number of experiment, the sample mean appears to be relatively closer to the theoretical mean ($n \cdot p = 70$) but the empirical distribution still does not model the theoretical distribution well due to the limited choices for the observed value.

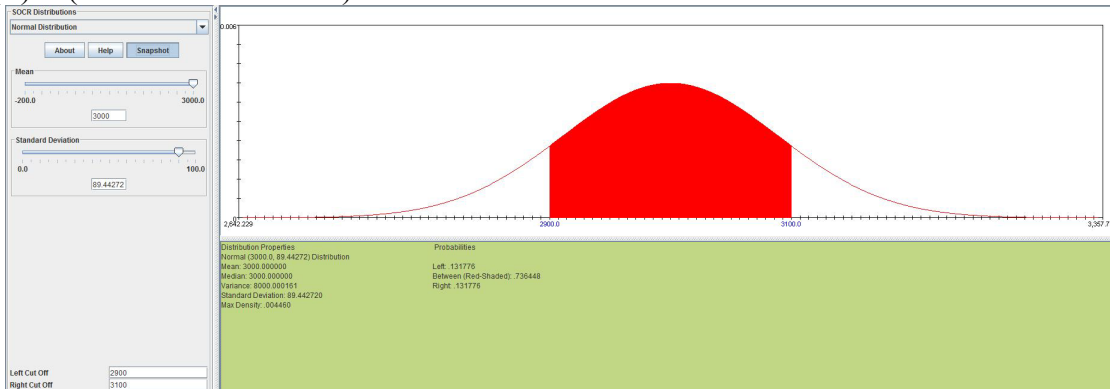


Finally, if we increase the number of experiment to 100 keeping $n=20$ per experiment, we see that the empirical results become much closer to the theoretical probabilities. In particular, we have sample mean 14.08 and s.d. 2.25 compared to the theoretical mean 14 and s.d. 2.05.

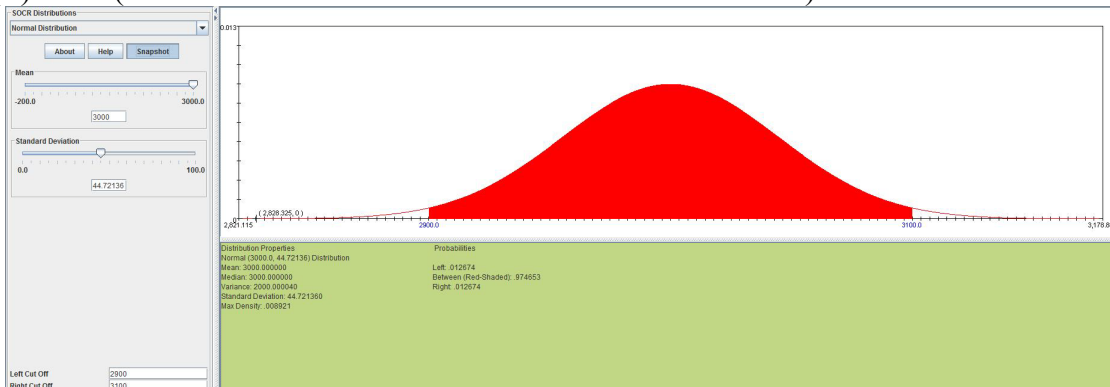
HW 5.2

By CLT, the sample mean follows a normal distribution with mean $3000mLi$, standard deviation $400/\sqrt{n}$. Let sample mean be denoted \bar{X} . We may calculate the answers using [SOCR Normal Distribution calculator](#).

1) $P(E) = P(2900 \leq \bar{X} \leq 3100) = .736$



2) $P(E) = .975$ (The standard deviation of \bar{X} is halved in this case.)



3) $P(E)$ increases as n increases. (An increase in n leads to a drop in standard deviation and therefore the same 200mLi range bears a larger impact.)

HW 5.3

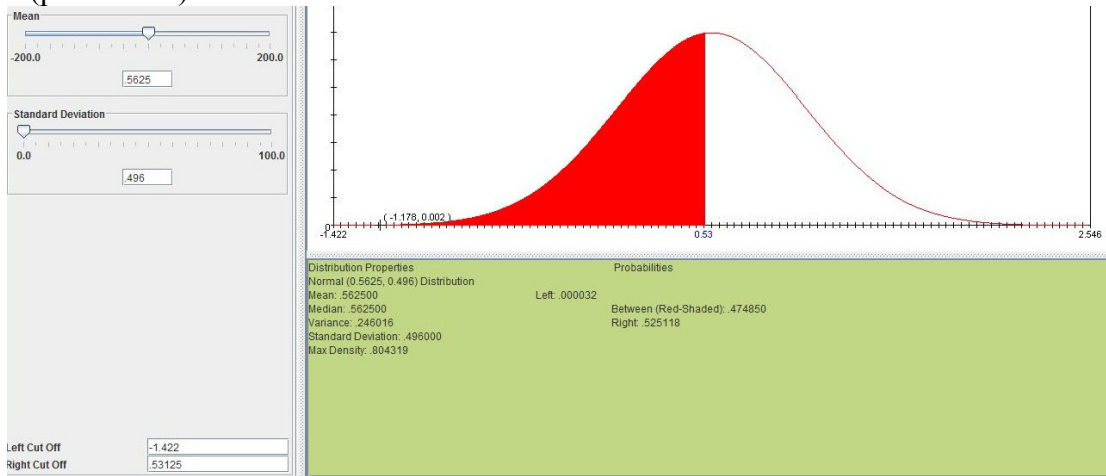
Using normal approximation to binomial, the mean for p^\wedge should be $p=9/16$ and the variance should be $p*(1-p)/n$. If we let X denote the number of purple flowers. Then X can be approximated using a binomial model with mean $n*p$ and variance $n*p*(1-p)$ under appropriate condition. Let E denote the event “ p^\wedge is closer to $1/2$ than to $9/16$ ”.

1) You do not need normal approximation in this case, partly because $n=1$ and the condition $n*p>10$ or $n*(1-p)>10$ obviously fails. Hence, X rather follows a Bernoulli distribution with $p=9/16$.

$$P(p^\wedge \text{ is closer to } 1/2 \text{ than to } 9/16) = P(E) = P(\text{no purple flower}) = P(X=0) = 7/16 = .4375.$$

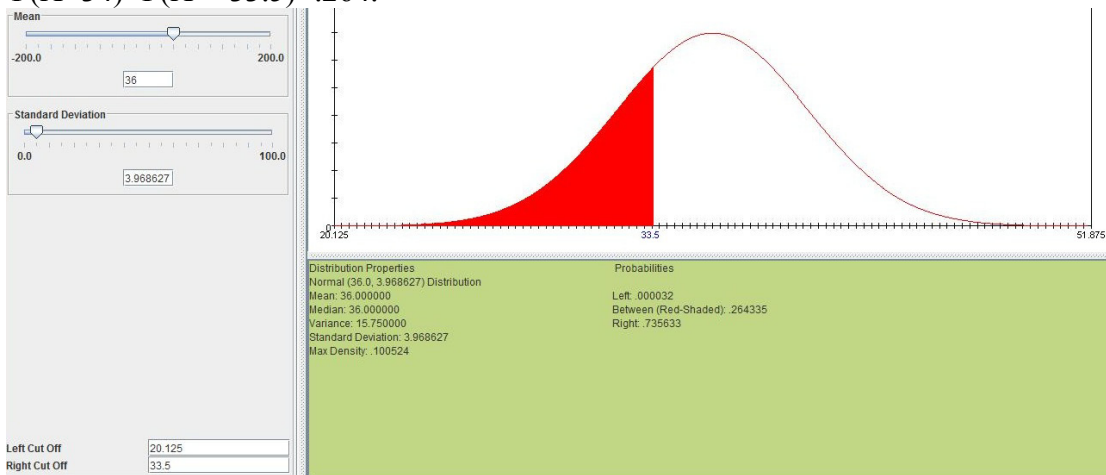
If one insists on using normal approximation on p^\wedge , it would look like

$$P(E) = P(p^\wedge < 17/32) = .475$$



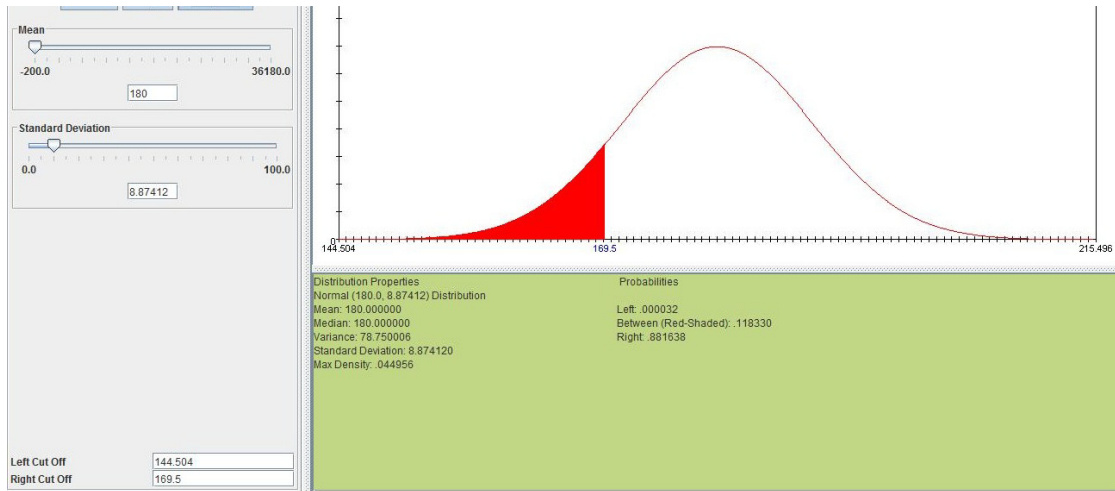
2) This time, we may use the normal approximation where X is normal with mean 36 and s.d. 3.97.

$$P(E) = P(X < 34) = P(X \leq 33.5) = .264.$$



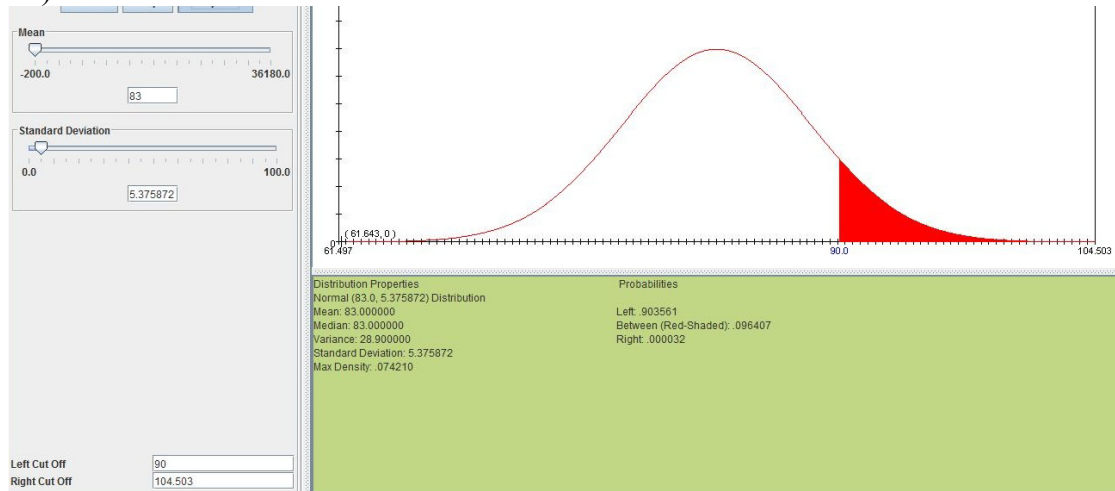
3) X is now approximately normal with mean 180 and s.d. 8.87412.

$$P(E) = P(X < 170) = P(X \leq 169.5) = .118.$$



HW 5.4

Suppose that the 10 mice are independently sampled, and therefore the sum of their weights follows a normal approximation of 83g and a variance of $10 \cdot (1.7)^2 = 28.9$, and let X be the total weight, then $P(X > 90) = .096 = 9.6\%$



HW 5.5

1) $27.8 \pm (2.9/\sqrt{6}) \cdot (t^*) = 27.8 \pm (2.9/\sqrt{6}) \cdot 2.571 = 27.8 \pm 3.04 = (24 \mu\text{g/mL}, 31 \mu\text{g/mL})$

Note that the value of t^* is found by looking up 95% ($\alpha = .025$) and $d.f. = n - 1 = 5$.

2) We are 95% confident that the (population) mean blood serum concentration of Gentamicin 1.5 hours after injection is between $24 \mu\text{g/mL}$ and $31 \mu\text{g/mL}$.

3) The general answer is “No”. This is because that a confidence interval helps us making inference on the population mean (a point estimate) and has little to do with the proportion of observations it captures, regardless how large your sample size is. It is also important that you know a 95% confidence interval does NOT mean that it captures 95% of the observations (please review the definition of C.I.). Another reason is that it also depends on the population distribution. Suppose we have a bimodal distribution where two modes are symmetric about 0 but far away from each other. If one takes a large enough sample, the \bar{X} will be expected to be around 0 and a confidence level may capture all or none of observations in this case. It does not mean that the point estimate is not reliable, but rather one ought to think twice about the usefulness of a single estimate.

HW 5.6

1) 95% C.I. = $13 \pm 2.262 * (12.4 / \text{SQRT}(10)) = (4.13 \mu\text{g/mL}, 21.87 \mu\text{g/mL})$

2) We are 95% confident that the mean difference in HBE levels between January and May is between 4.13 $\mu\text{g/mL}$ and 21.87 $\mu\text{g/mL}$.

Note: we only use the last column and therefore the d.f. = $n - 1 = 9$. In general, this problem may be treated as an example of comparing means from two independent populations.

3) From SOCR, we get (4.13, 21.87). We get exactly the same result.

