STA 113: Probability and Statistics in Engineering
Homework # 5: Solutions

0.1 Problem 1

Histogram of sample means:

```
Midpoint   Count
     1.0       6  *
     1.2      13  **
     1.4      65  *******
     1.6     239  ************************
     1.8     303  *******************************
     2.0     344  ***********************************
     2.2     279  ****************************
     2.4     183  *******************
     2.6     123  *************
     2.8      34  ****
     3.0       9  *
     3.2       2  *
```

Histogram of sample variances:

```
Midpoint   Count
     0.5      32  ****
     1.0     228  ***********************
     1.5     422  *******************************************
     2.0     368  *************************************
     2.5     286  *****************************
     3.0     124  *************
     3.5      84  *********
     4.0      31  ****
     4.5      12  **
     5.0       7  *
     5.5       3  *
     6.0       0
     6.5       1  *
     7.0       0
     7.5       1  *
     8.0       0
     8.5       1  *
```

a.) For one of my simulations, I obtained a mean of sample means = 1.99312. I obtained a variance of sample means = 0.127381. The central limit theorem states that $\mu_{\overline{y}} = \mu = 2$ and $\sigma^2_{\overline{y}} = \sigma^2/n = 2/16 = 0.125$. Given the large number of samples (1600), the values obtained should be close to the theoretical values, as they are.

b.) I did several simulations and found values of 110, 103, and 98 sample means larger than 2.5.

c.) If, under the central limit theorem, $\overline{y} \sim N(\mu, \sigma^2/n)$ (i.e., $\overline{y}$ is distributed normally with mean $= \mu$ and variance $= \sigma^2/n$), then $P[\overline{y} > 2.5] = P\left[Z > \frac{2.5-2}{\sqrt{2/4}}\right] = P[Z > 1.414] = 0.0786$.

$0.0786 * 1600$(the number of random samples) $= 125.76$. Given this large number of random samples, it's surprising that the number of sample means I found to be greater than 2.5 (110, 103, and 98) are not close to the theoretical estimation nor do they seem to be centered around 125.76. This seems to occur because a number of the sample means exactly equal to 2.5, which occurs because we only take 16 samples and the sample means take on a discrete distribution that is not necessarily modelled well by the theoretical distribution of sample means, which is continuous. Whether you include values > 2.5 or ≥ 2.5 determines whether you are over or under the theoretical value of 125.76. Dr. Wolpert's code must have put values exactly equal to 2.5 into the group coded by zeros in c5.

d.) Use the following minitab code:

```
erase c1
random 1600 into c1;
poiss 2.
histogram c1
mean (c1)
let c2 = stdev(c1)
let c3 = c2 * c2
print(c2)   # this gives the variance of the 1600 samples
```

Histogram of 1600 samples drawn from a Poisson distribution with mean of 2.

```
Midpoint   Count
       0     219  **********************
       1     420  ******************************************
       2     434  ********************************************
       3     281  *****************************
       4     155  ****************
       5      61  *******
       6      23  ***
       7       7  *
```

$\overline{y} = 1.995$
$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2 = 1.94369$

Since individual samples are drawn from a Poisson distribution with mean $= 2$, (and therefore variance $=2$), $\overline{y}$ and s^2 should both be approximately equal to 2, as they are.

No, this distribution does not look similar to the distribution of sample means. This distribution is not symmetrical and has a wider spread than the distribution of sample means. Both are centered at $y = 2$. This occurs because the sample means are distributed normally with mean=2 and variance adjusted by n. The actual samples are Poisson distributed with mean=variance=2.

0.2 Problem 2

Modify the minitab code as follows:

```
erase c1, c2
store 'ransam2'
random 100 into c3;
uniform 0 1.   ###  or "normal 0 1." or "exponential 1"
let c4=stdev(c3)
let c4=c4*c4
stack c2 c4 c2
end.
execute 'ransam2' 1000
histogram c2
```

a.) $X \sim \text{uniform}(0, 1)$

```
 Midpoint   Count
    0.060       3  *
    0.065      11  **
    0.070      57  ******
    0.075     129  *************
    0.080     263  ***************************
    0.085     237  ************************
    0.090     192  ********************
    0.095      75  ********
    0.100      25  ***
    0.105       7  *
    0.110       1  *
```

b.) $X \sim \text{normal}(0, 1)$

```
 Midpoint   Count
      0.6       2  *
      0.7      42  *****
      0.8     116  ************
      0.9     245  *************************
      1.0     265  ***************************
      1.1     173  ******************
      1.2     106  ***********
      1.3      33  ****
      1.4      13  **
      1.5       3  *
      1.6       1  *
      1.7       1  *
```

c.) $X \sim$ exponential(1)

```
Midpoint   Count
     0.4      10   *
     0.6     119   ************
     0.8     275   ****************************
     1.0     291   ******************************
     1.2     170   *****************
     1.4      80   ********
     1.6      31   ****
     1.8      14   **
     2.0       5   *
     2.2       5   *
```

At least for part b.), for normally distributed data, we could obtain the theoretical sampling distribution of s^2 based on $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, with n-1 degrees of freedom. Theoretical sampling distributions for the others are beyond the scope of the class.

0.3 Problem 3

$\mu_p = 15, \sigma_p = 5$

a.) If the hormone is not effective, then the distribution of times should have the same distribution as for the placebo. Therefore, under the central limit theorem, $\overline{y} \sim N(\mu_p, \sigma_p^2/n)$, which for this case is $\overline{y} \sim N(15, \frac{5^2}{20})$.

b.) $P[\overline{y} \leq 6] = P\left[\frac{6-\mu_p}{\sigma_p/\sqrt{n}}\right] = P\left[Z < \frac{-9}{5/\sqrt{20}}\right] = P[Z < -8.049845] \approx 0$

c.) Under the null hypothesis that melatonin has no effect, $P[\overline{y} \leq 6] \approx 0$, therefore, $\overline{y} = 5$ is unlikely. If melatonin has no effect, then $\overline{y}_m$ should be distributed normally with mean μ_p and variance σ_p^2/n. Therefore, $\overline{y}_m$ should lie in the interval (12.81,17.19) with 95 percent probability, since $P\left[\overline{y} < \frac{(12.81-15)}{5/\sqrt{20}}\right] = P[Z < -1.96] = 0.025$ and $P\left[\overline{y} > \frac{(17.19-15)}{5/\sqrt{20}}\right] = P[Z > 1.96] = 1 - 0.975 = 0.025$. Since this interval does not include $\overline{y} = 5$, this suggests that the null hypothesis is incorrect, i.e., that melatonin does decrease the amount of time until falling asleep.

0.4 Problem 4

$n = 330, p = 0.54, q = 0.46$

$n * p \geq 4$, so the approximation is likely to be good.

a.) $P[Y < 100] \approx P\left[Z < \frac{(x+.5)-np}{\sqrt{npq}}\right] = P\left[Z < \frac{(100+.5)-178.2}{9.053839}\right] = P[Z < -8.581995] \approx 0$

b.) $P[Y \geq 200] \approx P\left[Z > \frac{(x-.5)-np}{\sqrt{npq}} > \frac{(200-.5)-178.2}{9.053839}\right] = P[Z > 2.352593] = 0.0093$