Homework 03 Solution

Stats 13, Section 1, Spring 2013

1. Let X : Person is a mutant. Then

$$X \sim Bin(6, 0.35) \qquad \text{and} \qquad \hat{p} = \frac{X}{6}.$$

Let $y = P(X = x)$, then

$x =$	0	1	2	3	4	5	6
$\hat{p} =$	0	0.1666667	0.3333333	0.5	0.6666667	0.8333333	1.0
$y =$	$\binom{6}{0}0.35^0 0.65^6$	$\binom{6}{1}0.35^1 0.65^5$	$\binom{6}{2}0.35^2 0.65^4$	$\binom{6}{3}0.35^3 0.65^3$	$\binom{6}{4}0.35^4 0.65^2$	$\binom{6}{5}0.35^5 0.65^1$	$\binom{6}{6}0.35^6 0.65^0$
$y =$	0.07541889	0.243661	0.3280052	0.2354909	0.09510211	0.02048353	0.001838266

If we now let $T = \{0.0, 0.1, 0.5, 0.6, 0.7, 1.0\}$, then

$t =$	0.0	0.1	0.5	0.6	0.7	0.8	0.9	1.0
$P(\hat{p} = t) =$	0.07541889	0	0.2354909	0	0	0	0	0.001838266

We next do 10 runs of the binomial coin flip experiment in SOCR with $n = 20$ and $p = 0.65$ and obtain the following results:

- We can see from the plot as well as the table that the theoretical probabilities do not match empirical distribution very well.
- The values in the 2nd and 3rd columns are not close.
- We expect the distributions to be similar but with only 10 *runs* we do not necessarily expect to see these similarities. If we increase the number of *trials* from 20 to 100 we *would not* expect the distributions to be more similar.
- If we increase the number of *runs* from 10 to 100 we *would* expect the distributions to appear more similar.

2. We begin by noting that, regardless of what μ is, we would like to find the probability that

$$\begin{aligned}
P(E) &= P(\mu - 100 \leq \bar{x} \leq \mu + 100) \\
&= 2 \cdot P\left(0 \leq Z \leq \frac{(\mu + 100) - \mu}{300/\sqrt{n}}\right) \\
&= 2 \cdot P(\left(0 \leq Z \leq \frac{100\sqrt{n}}{300}\right) \\
&= 2 \cdot P(\left(0 \leq Z \leq \frac{\sqrt{n}}{3}\right)
\end{aligned}$$

- Let $n = 15$, then

$$\begin{aligned} P(E) &= 2 \cdot P\left(0 \leq Z \leq \frac{\sqrt{15}}{3}\right) \\ &= 2 \cdot P(0 \leq Z \leq 1.29) \\ &= 2 \cdot 0.4015 \\ &= 0.803 \end{aligned}$$

- Let $n = 60$, then

$$\begin{aligned} P(E) &= 2 \cdot P\left(0 \leq Z \leq \frac{\sqrt{60}}{3}\right) \\ &= 2 \cdot P(0 \leq Z \leq 2.58) \\ &= 2 \cdot 0.4951 \\ &= 0.9902 \end{aligned}$$

- As we saw above, as n increases, $P(E)$ increases at a rate of $\sqrt{n}$.

3. Let X : Number of purple flowers out of n. Then $\hat{p} = X/n$. Note that $\hat{p}$ is closer to 1/2 than 7/16 when

$$\hat{p} > \frac{1/2 + 7/16}{2} = \frac{15}{32} = 0.46875$$

and that when n is large then by normal approximation without continuity correction, we have

$$\hat{p} \sim \mathcal{N}\left(\frac{7}{16}, \sqrt{\frac{7 \cdot 9}{16n}}\right)$$

- Let $n = 1$. Then $\hat{p} = X/1 \sim Bernoulli(7/16)$ and

$$\begin{aligned} P\left(\hat{p} > \frac{15}{32}\right) &= P\left(X > \frac{15}{32}\right) \\ &= P(X = 1) \\ &= 7/16 = 0.4375 \end{aligned}$$

So the probability the misleading event occurs is 0.4375.

- Let $n = 32$. Then $\hat{p} \sim \mathcal{N}(0.4375, 0.0877)$ and

$$\begin{aligned} P(\hat{p} > 0.46875) &= P\left(Z > \frac{0.46875 - 0.4375}{0.0877}\right) \\ &= P(Z > 0.36) \\ &= 0.3594 \end{aligned}$$

So the probability the misleading event occurs is 0.3594.

- Let $n = 200$. Then $\hat{p} \sim \mathcal{N}(0.4375, 0.0351)$ and

$$\begin{aligned}
P(\hat{p} > 0.46875) &= P\left(Z > \frac{0.46875 - 0.4375}{0.0351}\right) \\
&= P(Z > 0.89) \\
&= 0.1867
\end{aligned}$$

  So the probability the misleading event occurs is 0.1867.

4. Let X : Mouse weight. Since each mouse, X_i, is approximately normal then we write

$$X_i \overset{approx}{\sim} \mathcal{N}(7.2, 1.5).$$

If we now let $Y = \sum_{i=1}^{10} X_i$, that is we define Y : Litter of 10 mice, then we find that

$$\begin{aligned}
E(Y) &= E(\sum_{i=1}^{10} X_i) \\
&= \sum_{i=1}^{10} E(X_1) \\
&= 10 \cdot 7.2 \\
&= 72
\end{aligned}$$

and

$$\begin{aligned}
Var(Y) &= Var\left(\sum_{i=1}^{10} X_i\right) \\
&= \sum_{i=1}^{10} Var(X_1) && \text{(By independence of the litters)} \\
&= 10 \cdot Var(X_1).
\end{aligned}$$

which implies $SD(Y) = 1.5\sqrt{10} = 4.7434$. Therefore,

$$Y \sim \mathcal{N}(72, 4.7434).$$

Now let T_i : Total weight of litter Y_i greater than 80g. Let $T_i = 1$ when $Y_i > 80$ and $T_i = 0$ when $Y_i < 80$. Then T is a Bernoulli distributed random variable with probability of success $p = P(Y > 80) = 0.0458$. That is,

$$T \sim Bernoulli(0.0458).$$

Finally, for n large, if we let $\hat{p}$ be the proportion of litters with weights greater than 80g, then

$$\hat{p} = \frac{\sum_{i=1}^{n} T_i}{n} \sim \mathcal{N}\left(0.0458, \sqrt{\frac{0.0437}{n}}\right)$$

To find the expected number of litters with weights in excess of 80g, we calculate

$$E(\hat{p}) = 0.0458.$$

5. Since n is small, we use a t-value. A 95% confidence interval (CI) for the population mean is

$$\begin{aligned}\bar{x} \pm t_{df,\alpha/2} \cdot \frac{s}{\sqrt{n}} &= 28.3 \pm t_{5,0.05/2} \cdot \frac{3.3}{\sqrt{6}} \\ &= 28.3 \pm 2.57 \cdot 1.3472 \\ &= 28.3 \pm 3.4623.\end{aligned}$$

   - We interpret this interval to mean that if we repeatedly sampled 6 data points, there would be 95% chance that our confidence intervals would contain the true mean.
   - It is not typical for the CI to cover most of the data points. The width of the interval will become narrower (and cover fewer data points) as the size of n increases.

6. Since n is small, we use a t-value. A 95% CI for the population mean difference would be

$$\begin{aligned}\bar{x} \pm t_{df,\alpha/2} \cdot \frac{s}{\sqrt{n}} &= 12.5 \pm t_{9,0.05/2} \cdot \frac{6.5}{\sqrt{10}} \\ &= 12.5 \pm 2.26 \cdot 2.05548 \\ &= 12.5 \pm 4.645.\end{aligned}$$

   which we can re-write as $(7.855, 17.145)$.

   - We interpret this interval to mean that if we repeatedly sampled 10 data points, there would be 95% chance that our CI would contain the true mean. Since the CI we calculated above is just one interval, we're unsure as to whether it contains the true mean or not.
   - From SOCR, we calculate the 95% CI to be $(3.8547, 21.1453)$. These results are noticeably wider our results.