STATISTICS FOR ENGINEERS
ASSIGNMENT VII ANSWERS
OCTOBER 15, 2010

Problem 7.1

...Last week, you found that two of the ten tested faces failed the test. What is the probability that Bakugan Industries supplied you only ten bad faces in that week's batch of 1000?

Solution: We are asked to calculate $\mathbb{P}\left[X = 2\right]$ given that the probability of selecting a single bad watch face is $p = 10/1000$. This is a quintessential hypergeometric problem. The probabilities are not independent, there is a fixed number of items selected from a fixed number, and there are just two types involved (good and defective). Thus, to calculate the probability for this discrete distribution, we calculate $\mathcal{H}(x = 2; N = 1000, M = 10, n = 10)$. This gives us:

$$\begin{aligned}\mathbb{P}\left[X = 2\right] &= \mathcal{H}(2; 1000, 10, 10) \\ &= \frac{\binom{10}{2}\binom{990}{8}}{\binom{1000}{10}} \\ &\approx 0.0038\end{aligned}$$

Thus, we can be pretty sure that Bakugan Industries sent you more than just 10 bad faces in this shipment.

Note: All display calculators can actually do this calculation, because all display calculators have a nCr function located somewhere in the menu.

⋆ ⋆ ⋆

There is an approximation to this: binomial. Let us use the binomial approximation here. To use this, we need to determine x, n, and p. Here, $x = 2$ and

$n = 10$, since we want the probability that we get 2 defective faces out of the 10 we draw. Finally, $p = 0.01$, since we are testing the hypothesis that there really are 10 bad in the shipment of 1000.

Thus, since our binomial distribution is discrete, we have

$$\begin{aligned}\mathbb{P}\left[X = 2\right] &= \mathcal{B}(2; 10, 0.01)\\ &= \binom{10}{2}(0.01)^2(0.99)^8\\ &=\approx 0.0042\end{aligned}$$

This is not the same as the exact answer, but it is close.

⋆⋆⋆

There is a second approximation method: the Poisson. The Poisson requires one parameter, the expected value. Here, if Bakugan is correct, then the expected number of defective faces you should get in a sample of 10 faces is $10\frac{10}{1000} = 0.10$. Thus, using the Poisson approximation gives us

$$\begin{aligned}\mathbb{P}\left[X = 2\right] &= \mathcal{P}(2; \lambda = 0.1)\\ &= \frac{\lambda^x e^{-\lambda}}{x!}\\ &= \frac{0.10^2 e^{-0.10}}{2!} = 0.0045\end{aligned}$$

Thus, the conclusion is the same, if the final probability is not.

◇

Problem 7.2

A start-up polling firm polled 500 people about their choice for the upcoming gubernatorial election in Oklahoma. Of those people, 165 stated they would vote for Jari Askins and 188 said they would vote for Mary Fallin. The rest were undecided voters.

What is the probability that more people in Oklahoma support Fallin over Askins?

Solution: This is a direct application of comparing two proportions (*one-tailed*). First, we note that the sample support for Askins is $\tilde{p}_A = 0.3307$; for Fallin, $\tilde{p}_F = 0.3765$ (remember, $\tilde{n} = 502$ here). From this, we calculate the confidence interval (leaving z_α as it is for now).

$$\tilde{p}_A - \tilde{p}_F \pm z_\alpha \sqrt{\frac{\tilde{p}_A(1-\tilde{p}_A)}{\tilde{n}_A} + \frac{\tilde{p}_F(1-\tilde{p}_F)}{\tilde{n}_F}}$$

Setting one interval endpoint equal to zero and solving for z_α gives us

$$0.0458 = z_\alpha \sqrt{\frac{0.2213}{502} + \frac{0.2347}{502}}$$

Thus, we have $z_\alpha \approx 1.5196$. Checking our z-table, we get $\alpha = 0.0643$. Remember, this is a *one-sided* hypothesis; we want to know the probability that there is *more* support for Fallin.

In other words, we are *not* sure at the $\alpha = 0.05$ level that Fallin has a higher level of support in Oklahoma than Askins. ◇

Problem 7.3

A recent final examination contained two questions. Oddly enough, the distribution of scores on the first question was distributed $Q_1 \sim \mathcal{N}(\mu = 50, \sigma^2 = 5)$, and the distribution of socres on the second question was distributed $Q_2 \sim \mathcal{N}(\mu = 45, \sigma^2 = 7)$. In the class, what was the probability that a student scored more than 100 points on the examination?

Solution: We first note that the sum of two Normal distributions is also a Normal distribution. Here,

$$Q_1 + Q_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) = \mathcal{N}(\mu = 95, \sigma^2 = 12)$$

As such, we calculate the z-score ($z = \frac{100-95}{\sqrt{12}} = 1.443$) and the attendant probability, $\mathbb{P}[Z > 1.443] = 1 - \mathbb{P}[Z \leq 1.443] = 1 - 0.9251 = 0.0749$. Thus, the probability that a student scored more than 100 on the examination was about 0.0749.

Alternately, if we had access to a computer, we could have skipped the z-score calculation and directly calculated $\mathbb{P}[X > 100] = 1 - \mathbb{P}[X \leq 100] = 0.0749$. $\diamond$

Problem 7.4

The total vote for Candidate Pan in each district of Neverland follows a lognormal distribution, with $\mu = 3$ and $\sigma^2 = 4$. What is the average district vote for Pan? What is the variance of the district vote for Pan? If we define W as the natural logarithm of the district vote ($W = \ln[X]$), what is the probability that W is greater than 4 ($\mathbb{P}[W > 4]$)?

Solution: We are given that $X \sim lognormal(\mu = 3, \sigma^2 = 4)$. As such, we know the expected district vote for Pan is $\mathbb{E}[X] = \mu_X = e^{\mu+\sigma^2/2} = e^5 = 148$. We also know the variance in the district vote is $\mathbb{V}[X] = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{14} - e^{10} = 1,180,578$.

We also know that $W \sim \mathcal{N}(\mu = 3, \sigma^2 = 4)$. Thus, the last part reduces to a check on the z-table, with $z = \frac{4-3}{\sqrt{4}} = 0.500$. $\mathbb{P}[W > 4] = 1 - \mathbb{P}[W \leq 4] = 1 - 0.6915 = 0.3085$. $\diamond$

Problem 7.5

Using the `crime.csv` dataset, is there a difference in violent crime rates between the South and the West in 1990? If so, which has a higher violent crime rate?

Solution: To perform this test, I used a t-test of independent samples, without assuming the variances in the samples are equal. The average for the South is $\bar{x}_S = 766.5$; for the West, $\bar{x}_W = 404.9$. The degrees of freedom are 24.792. Thus, going to the tables, we need to use $\nu = 24$, to be conservative. However, as we are using the computer for this, we can use the actual estimated degrees of freedom. The variance of the violent crime rate in the South is $s_S^2 = 267913$; for the West, $s_W^2 = 68500$. Using this information, we find our test statistic is $t = 2.4941$, which corresponds to $p = 0.0197$.

The results suggest that the South had a higher violent crime rate in 1990 (rate = 767) than did the West (rate = 405). These results are significant at the $\alpha = 0.05$ level ($t = 2.4941; p = 0.01966$).

The command I used in `R` was

```
t.test(vcrime90[census4=="South"],vcrime90[census4=="West"])
```

If, on the other hand, you did the calculations for 2000, then you get the same conclusion, just different values ($t = 3.0913, \nu = 24.764, p = 0.004877$). $\diamond$