Stata Program Notes
Biostatistics: A Guide to Design, Analysis, and Discovery
Chapter 5: Probability Distributions

Note 5.1 – Finding binomial and Poisson probabilities

1. Finding binomial probabilities:

In Stata, the command **binomial(n,k,p)** can be used to find the probability of observing k or fewer successes in **n** trials when the probability of success is **p**. Assuming n=4 and p= 0.25, the probability of k ≤ 1 can be found by using the commands below.

```
Stata commands:

display binomial(4,1,0.25)
```

```
Stata output:

0.73828125
```

The probability of k = 1 can be found by:

```
Stata commands:

display binomial(4,1,0.25)- binomial(4,0,0.25)
```

```
Stata output:

0.421875
```

The calculation can be verified using the mathematical definition:

$$\Pr\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{where } k = 0, 1, 2, \ldots, n.$$

```
Stata commands:

display comb(4,1)*(0.25)^1*(1-0.25)^3
```

```
Stata output:

0.421875
```

2. Finding Poisson probabilities:

To calculate Poisson probabilities you should install the files that enable the functions **epoisson** and **cpoisson** which give exact and cumulative probabilities. For the moment, we will not make use these functions; however, we will show calculations that produce the same results. To find Poisson probabilities, we start with the Poisson probability mass function

$$\Pr\{X = x\} = \frac{e^{-\mu}\mu^{x}}{x!} \quad \text{where x = 0, 1, 2, …}$$

In Stata, we can consider using the expression **exp(-*m*)****m*^*x*** divided by **lnfactorial(*x*)**. The Stata command **lnfactorial(*x*)** returns the natural log of *n* factorial. If we exponentiate the natural log of *n* factorial, we get *n* factorial. As an example, we can evaluate the probability

$$\Pr\{X = 3\} = \frac{e^{-2}2^{3}}{3!}$$ where μ =2 as shown below.

```
Stata commands:

display exp(-2)*[(2^3)/exp(lnfactorial(3))]
```

```
Stata output:

0.18044704
```

In this case, it would have been less complicated if we had used Stata's **poisson( )** function.

```
Stata commands:

display poisson(2,3) - poisson(2,2)
```

```
Stata output:

0.18044704
```

Consider another example where x ranges from 0 to 19. The Poisson probability mass function is still

$$\Pr\{X = x\} = \frac{e^{-\mu}\mu^{x}}{x!} \quad \text{where } x = 0, 1, 2, …$$

Although we allow *x* to range from 0 to 19, let us assume that we are only interested in finding the probability of *X* being less than or equal to 18 or Pr{X ≤ 18}. We begin by calculating the probability for each value of *x* which can be found in the column named **prob**, and then calculate the cumulative probability in the next column named **cum_prob** as shown in the Stata commands below:

```
Stata commands:

set obs 20
gen x = _n -1
gen prob = exp(-35.31)*[(35.31^x)/exp(lnfactorial(x))]
gen cum_prob = sum(prob)
list
```

```
Stata output:

     +-------------------------+
```

```
     |  x      prob      cum_prob |
     |----------------------------|
  1. |  0    4.62e-16    4.62e-16 |
  2. |  1    1.63e-14    1.68e-14 |
  3. |  2    2.88e-13    3.05e-13 |
  4. |  3    3.39e-12    3.70e-12 |
  5. |  4    3.00e-11    3.37e-11 |
     |----------------------------|
  6. |  5    2.12e-10    2.45e-10 |
  7. |  6    1.24e-09    1.49e-09 |
  8. |  7    6.28e-09    7.77e-09 |
  9. |  8    2.77e-08    3.55e-08 |
 10. |  9    1.09e-07    1.44e-07 |
     |----------------------------|
 11. | 10    3.84e-07    5.28e-07 |
 12. | 11    1.23e-06    1.76e-06 |
 13. | 12    3.63e-06    5.39e-06 |
 14. | 13    9.85e-06    .0000152 |
 15. | 14    .0000248    .0000401 |
     |----------------------------|
 16. | 15    .0000585    .0000986 |
 17. | 16    .0001291    .0002276 |
 18. | 17    .0002681    .0004957 |
 19. | 18    .0005259    .0010216 |
 20. | 19    .0009773    .0019989 |
     +----------------------------+
```

Notice that the probability for 18 or fewer cases, equaling 0.001, can be found when $x = 18$ under the **cum_prob** column as shown in the Stata output above.

Note 5.2 – Creating the Poissonness plot

Using the frequency values from Table 5.4, we can create a *Poissonness plot*. Basically we are plotting { *ln(freq(x)) + ln(x!) versus x* }. This is a simple scatter plot that can be created using the Stata commands below.

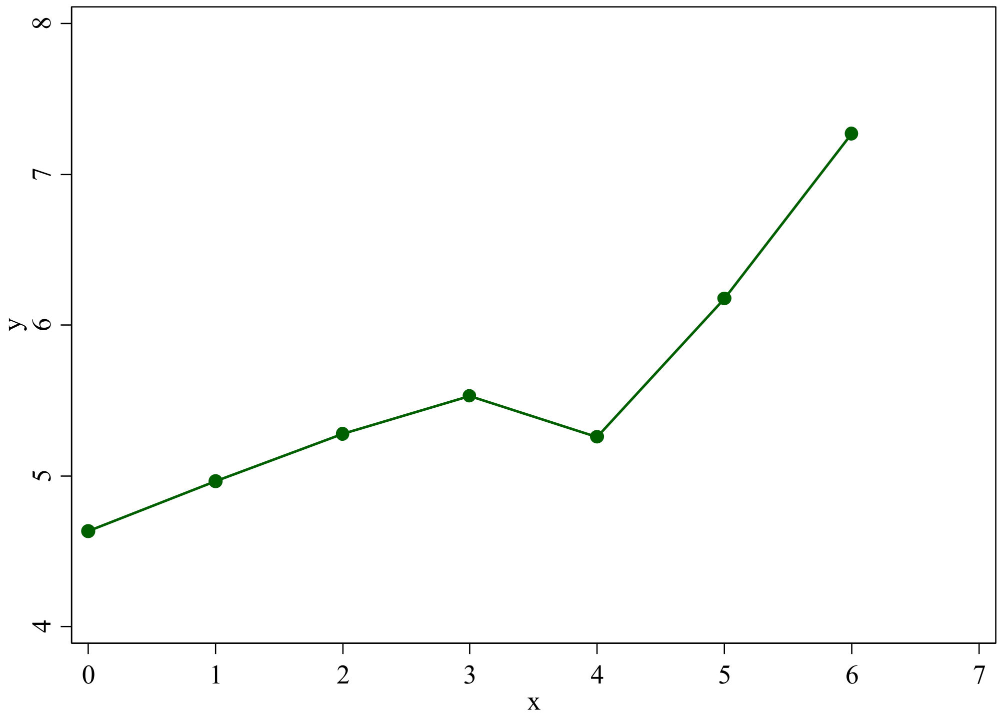
```
Stata commands:

input x freq
0 103
1 143
2  98
3  42
4   8
5   4
6   2
end

gen y = ln(freq) + lnfactorial(0) if x ==0
replace y = ln(freq) + lnfactorial(1) if x == 1
replace y = ln(freq) + lnfactorial(2) if x == 2
replace y = ln(freq) + lnfactorial(3) if x == 3
replace y = ln(freq) + lnfactorial(4) if x == 4
replace y = ln(freq) + lnfactorial(5) if x == 5
replace y = ln(freq) + lnfactorial(6) if x == 6
```

```
twoway (connected y x, lpattern(solid)), scheme(s1color) ///
       xlabel(0(1)7) ylabel(4(1)8)
```

After running the Stata commands, the plot created should look like the one below:



Note 5.3 – Finding normal probabilities

In Stata, the command **normal( )** returns the cumulative probability associated with a value from the standard normal distribution. For example if we would like to find the area under the standard normal distribution's probability density function less than or equal to a value of 1.0, we could use the command **normal(1.0)** as shown below:

```
Stata commands:

display normal(1.0)
```

```
Stata output:

0.84134475
```

In Stata, the command **invnormal( )** returns the inverse cdf value based on the standard normal distribution. For example, suppose we are interested in finding the 95th percentile of a standard normal distribution. We could use the following command **invnormal(0.95)** as shown below:

```
Stata commands:

display invnormal(0.95)
```

```
Stata output:

1.6448536
```

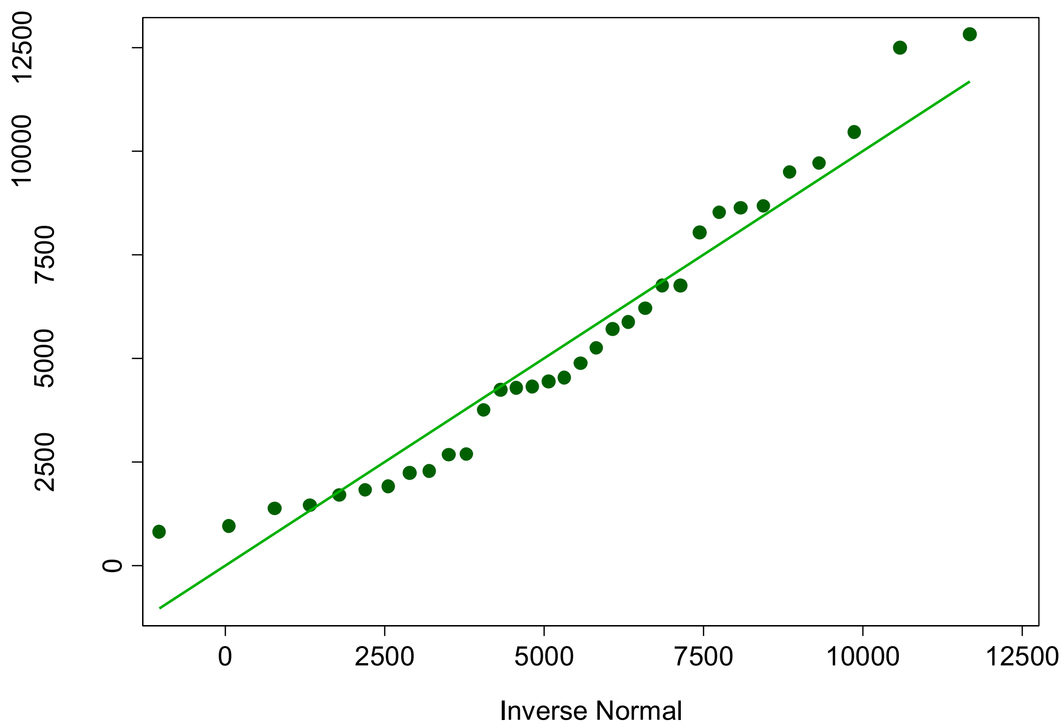
Note 5.4 – Creating a normal probability plot

We can create normal probability plots by using Stata's **qnorm** command. Using the data on vitamin A intake from 33 boys illustrated in Table 5.5, the plot created by the Stata command **qnorm** places the observed variable, vitamin A, on the vertical axis instead of the horizontal axis as shown in the text book. Below are the Stata commands used to create the normal probability plot.

```
Stata commands:

qnorm vita, ylabel(0(2500)12500) xlabel(0(2500)12500) ///
scheme(s1color) ytitle("Vitamin A (IU)")             ///
yscale(titlegap(4)) xscale(titlegap(4))
```

After running the Stata commands, the plot created should look like the one below:



How to cite this page:
Stata Program Notes for Chapter 5. Biostatistics: A Guide to Design, Analysis, and Discovery. from http://www.biostat-edu.com/ProgramNotes/ (accessed January 1, 2013).