

Normal distribution and normal approximation to binomial

In this lab you will learn how to compute normal distribution probabilities and use the normal distribution as an approximation to binomial.

Compute normal probabilities:

Suppose that the height X of female UCLA students follows the normal distribution with mean $\mu=62$ inches and standard deviation $\sigma=4$ inches. Using Stata find the probability that a randomly selected female UCLA student is taller than 71 inches.

- *First you calculate the Z value. The Stata command is:*

```
. zcalc 71 62 4
```

z-score for sample observations

$$z = \frac{(X - \mu)}{s} = \frac{(71 - 62)}{4} = 2.25$$

So, the Z value is 2.25, and we want to find $P(X > 71)$ which is equivalent to $P(Z > 2.25)$.

- *Then you compute the probability. The Stata command is:*

```
. display normprob(2.25)
```

And you get...

```
.98777553
```

This is the probability that Z is less than 2.25: $P(Z < 2.25) = 0.98777553$. Compare this probability with the probability that you get from the standard normal distribution table in the textbook. Therefore $P(Z > 2.25) = 1 - 0.98777553 = 0.0122247$. So the probability that a female UCLA student is taller than 71 inches (or the proportion of female UCLA students that are taller than 71 inches) is $P(X > 71) = 0.0122$ or 1.22%.

Use the above commands to find other probabilities for the examples we did in class.

Normal approximation to binomial:

Suppose that a fair coin is flipped 10 times. Let X represent the number of tails we observe in these 10 tosses. Then X is a random variable that follows the binomial distribution with $n=10$, $p=0.5$. The values that X can take are 0,1,2,3,4,5,6,7,8,9,10. Of course you can easily compute these probabilities using the binomial formula. You can also find them using the table of binomial probabilities in the textbook. Make sure that you know both ways!

The probability distribution of X ($n=10$, $p=0.5$) can be retrieved as follows:

```
. use http://www.stat.ucla.edu/~nchristo/binomial.dta
```

Here it is:

```
. list
```

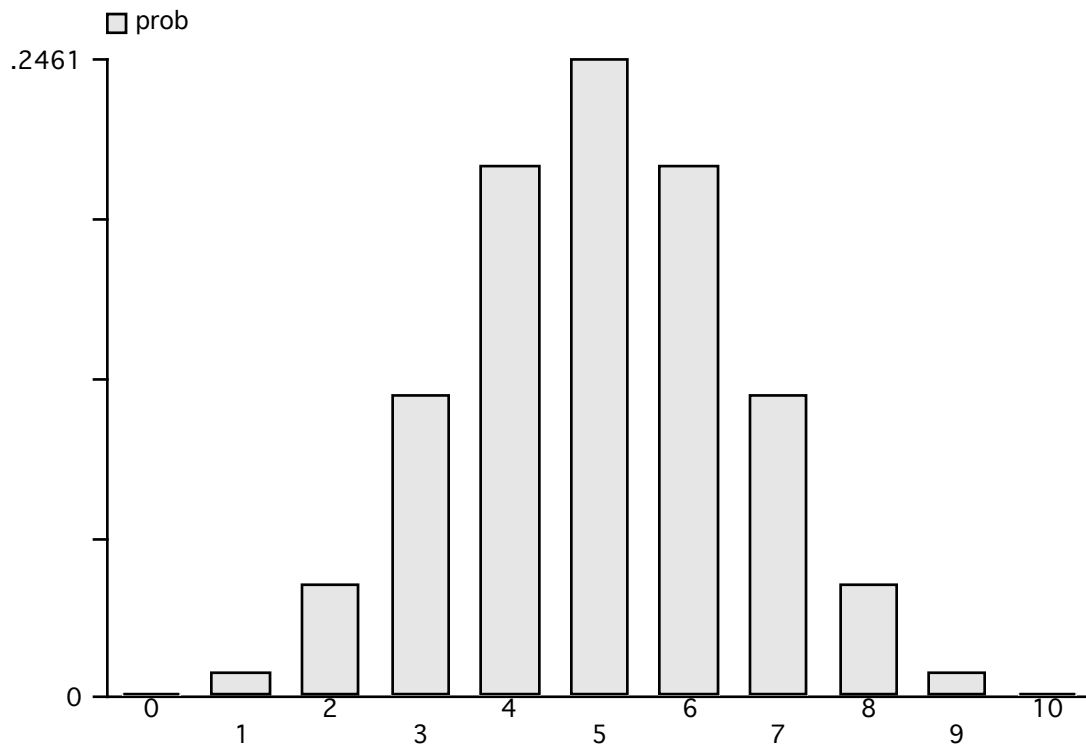
	x	prob
1.	0	.001
2.	1	.0098
3.	2	.0439
4.	3	.1172
5.	4	.2051
6.	5	.2461
7.	6	.2051
8.	7	.1172
9.	8	.0439
10.	9	.0098
11.	10	.001

The second column gives you the number of successes and the third column the corresponding probability. Let's construct the histogram for this distribution. We are going to use the following command:

```
. graph prob, bar by(x)
```

We can approximate probabilities using the normal distribution. The approximation is good if $np \geq 5$ and $n(1-p) \geq 5$. This can be achieved if n is large (then p can be small), or if p is close to 0.5 (then n can be small).

Here is the histogram first:



Unfortunately Stata does not eliminate the gaps between the bars, but don't worry about it!

- *Find the exact probability that in these 10 tosses of the coin you observe at least 2 tails (exact means that you should use the binomial distribution and not an approximation).*
- *Use the normal distribution to approximate the probability of at least 2 tails (do not forget the continuity correction factor).*
- *What is the error of the approximation (the difference between the exact and approximate probability)?*
- *Why is the approximation good?*

Now, retrieve the following probability distributions (one at a time) and answer the following questions:

```
. use http://www.stat.ucla.edu/~nchristo/binomial0.dta
. use http://www.stat.ucla.edu/~nchristo/binomial1.dta
. use http://www.stat.ucla.edu/~nchristo/binomial2.dta
. use http://www.stat.ucla.edu/~nchristo/binomial3.dta
. use http://www.stat.ucla.edu/~nchristo/binomial4.dta
. use http://www.stat.ucla.edu/~nchristo/binomial5.dta
```

(binomial n=10, p=0.05)

(binomial n=20, p=0.05)

(binomial n=20, p=0.10)

(binomial n=20, p=0.50)

(binomial n=20, p=0.90)

(binomial n=20, p=0.95)

- *Construct the histogram for each one of these distributions.*
- *In which of the above distributions will the normal approximation not be good?*
- *For n=10, p=0.05: Find the exact $P(X \geq 1)$ – using the binomial distribution, and then approximate it – using the normal distribution and compare.*

And as usual...

Good luck!