ILRST/STSCI 2100 Disc 6: Chapter 7 Random Variable and Probability Distribution

Today: (1) Review calculating E(X) and var(X) for discrete random variable (2) Normal distribution (3) QQ Plot on Minitab

1 Expected value and variance of random variables

1.1 Discrete

Expected value: $E(X) = \mu = \sum_x x * f(x)$

Variance: $Var(X) = \sigma^2 = \sum_x (X - \mu)^2 f(x)$

Standard deviation: $SD(X) = \sqrt{Var(X)}$

You are either given the p.m.f or you must first generate it.
Then, generate the table:

x	f(x)	$x * f(x)$	$(x-\mu)^2$	$(x-\mu)^2 f(x)$
0	.1	0	$(0-1.5)^2$	$(0-1.5)^2 * .1$
1	.3	.3	$(1-1.5)^2$	$(1-1.5)^2 * .3$
2	.6	1.2	$(2-1.5)^2$	$(2-1.5)^2 * .6$

The sum of the $x * f(x)$ column gives the expected value, and the sum of the $(x-\mu)^2 f(x)$ column gives the variance.

So $E(X) = \mu = 0 + 0.3 + 1.2 = 1.5$

Note: The sum of the f(x) column must be equal to 1.

1.2 Continuous

Finding E(X) and Var(X) requires the use of integrals and is out of the scope of this course.

2 Normal distribution

The normal distribution is a bell-shaped, uni-modal, symmetric curve that is centered at μ. Its density function is defined by the mean and variance. Random variables that follow a normal distribution can be standardized to have a mean of zero and a standard deviation of 1 by calculating the z-score.

Often, $z \sim N(0, 1)$, so probabilities involving z, can be found using the standard probability table (found in Appendix of textbook). The probabilities in this table give the area to the left of the z-score that you are looking at.

Ex.
$P(Z \leq 1.2) = 0.885$ and $P(Z \geq 1.2) = 1 - P(Z \leq 1.2) = 0.115$

Because the normal distribution is symmetric, $P(Z \leq -a) = P(Z \geq a)$

Ex. $P(Z \leq 1.2) = P(Z \geq -1.2)$

Using absolute values:

Recall if $|x| < a$, then $x < a$ and $x > -a$. So, $P(|Z| \leq a) = P(Z < a) + P(Z > -a)$

Application to percentile:

Percentile can be interpreted as "area". So if we are interested in the 90th percentile, we want to find z^* such that $P(Z < z^*) = .90$. Then, z^* is 1.28. This z^* then has to be converted back into the context of X, by $x = \mu + \sigma * z^*$

3 QQ plot

You will not have to be able to generate the graph by hand, but you are expected to be able to interpret and generate graphs on Minitab.

If the data points follow the unit line (a line that goes through the origin and has slope of one), then the data is likely to be normally distributed.

On Minitab: Graphs → Probability Plot → Simple Single Y Variable