Data Analysis in Evol/Evol - HW Week4

Jing-Yi Lu

Reference R code see: github.com/jingyilu/Data-analysis-ecoevo

Questions on the normal distribution (Week3)

1. **Suppose I toss a coin 10 times and it comes up heads 9x. Set up the test that this is a fair coin, and test it using (1) the binomial distribution, and (2) the normal distribution.**

H_0: The coin is fair (Probability of heads = 0.5)

H_1: The coin is not fair (Probability of heads $\neq$ 0.5)

(1) test by binomial distribution (null distribution with 10 trials and probability equals to 0.5)

P-value = 0.0215 < 0.05.

We can reject the null hypothesis, and the coin is not fair.

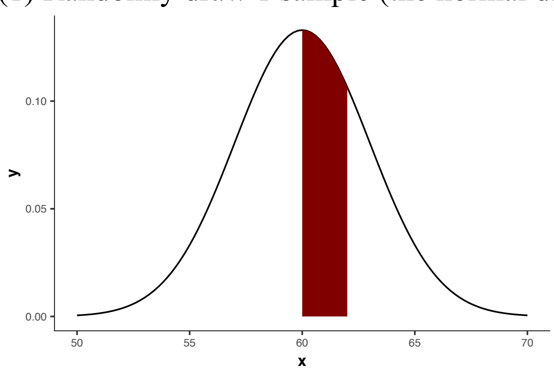
(2) test by normal distribution (null distribution with mean = 5 and s.d. = $\sqrt{10 \times 0.5 \times 0.5}$)

P-value = 0.0114 < 0.05.

We can reject the null hypothesis, and the coin is not fair.

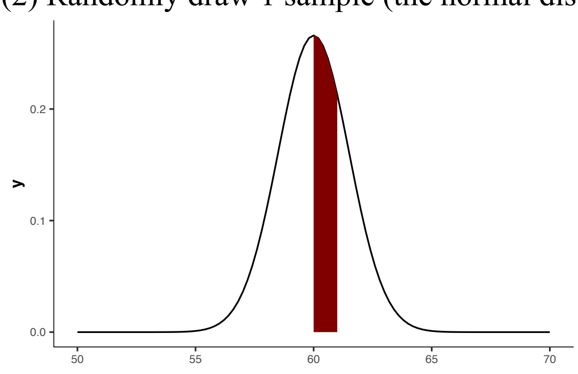
2. **Suppose femur lengths of dinosaurs in the field museum have all been measured, and in this population the mean is 60cm and variance 9cm. Compute the probability that a bone I randomly draw lies between 60cm and 62cm. Compute the probability that the average value of 4 bones I randomly draw lie between 60cm and 61cm. In each case, first sketch the probability distribution and shade the area you need to calculate.**

(1) Randomly draw 1 sample (the normal distribution with mean = 60, s.d. = $3/\sqrt{1}$))



Probability = 0.2475.

(2) Randomly draw 1 sample (the normal distribution with mean = 60, s.d. = $3/\sqrt{4}$))



Probability = 0.2475.

p.294: 2 a,b,c only. Can you see a problem, statistically, with the way this test is set up?
a. H_0: The proportion of male struck by lightning is 0.5.
H_1: The proportion of male struck by lightning is not 0.5.
b. Mean under null hypothesis: np = 648 × 0.5 = 324.
c. Standard deviation of the null distribution: $\sqrt{npq} = \sqrt{648 \times 0.5 \times 0.5} = 12.73$.
Statistically problem: The test was set up with the assumption that each case of lightning struck was independent (with the probability of being male or female), and the case would follow a binomial distribution. However, the incident of lightning is a nonrandom sampling of an extremely small proportion of the whole population.

p.296 q.13

Mean	Standard deviation	Y	Pr under n=10	Pr under n=30
14	5	15	0.264	0.137
15	3	15.5	0.299	0.181
-23	4	-22	0.215	0.085
72	50	45	0.956	0.998

p.324 q.24 compare this test in the way it is set up, to the p.294, q.2 test.
The test had 15 randomly sampled participants and examined if the tendency differed from the expectation of a normal distribution.
P.294, q.2 test treated each person as independent trials with the probability of male/female, therefore used normal approximation of binomial distribution to test the hypothesis.

p.356 q.7
a. mean difference: 33%; 95% Confidence interval: (21.01, 44.99).
b. H_0: The female mimicry has no effect on the proportion of body coverage ($\mu_{mimicc} = \mu_{nonmimics}$)
H_1: The female mimicry has effect on the proportion of body coverage ($\mu_{mimicc} \neq \mu_{nonmimics}$)
t = 5.478, *P*-value = 4.949×10^{-7}.
We can reject the null hypothesis that $\mu_{mimicc} = \mu_{nonmimics}$, and the female mimicry has effect on the proportion of body coverage.
Assumptions made for two-sample t test:
1. Random sampling.
2. The standard deviation of both populations is the same, and both normally distributed.

Compute the 90% confidence intervals for the mean, for a sample height of 25 people taken from a population whose sample mean $\bar{X}$ = 60 and sample standard deviation, s = 5.
Standard error = $5/\sqrt{25}$ = 1, d.f. = 25-1
Confidence interval = (60-1.71×1, 60+1.71×1) = (58.29, 61.71)
Interpret the confidence interval
We have 95% confidence that the true mean of the population will fall into this interval. That is, if we do independent sampling, the interval will capture the mean in 95% of the trials.
Compute the 95% confidence intervals for the same data. Based on this confidence interval, would you reject the null hypothesis that the mean value in the population is μ = 62?
95% Confidence interval = (60-2.06×1, 60+2.06×1) = (57.94, 62.06)
No, I will not reject the null hypothesis that the mean value μ = 62 as the 95% CI includes 62.