4: Statistical Inference and Hypothesis Testing

1

A Turkish license plate number consists of a total of 7 or 8 characters. The first two characters are fixed and represent the province. For example, 16 is the numerical prefix for Bursa. The next 5 or 6 characters can be a combination of letters and numbers with a minimum of 1 and a maximum of 3 letters. Note that only 23 letters of the Turkish alphabet are used in generating license plates.

How many possible license plates could be stamped for a province? How many license plates with no repetitions could be stamped for a province?

Sample license plates: 16 HU 6766, 06 T 8965, 34 KO 768

Answer

According to wikipedia, the letters must come before the number for licence plates in Turkey. I will make this assumption for this question as it was not given not stated prior.

There will be
$23(10^4+10^5)+23^2(10^3+10^4)+23^3(10^2+10^3)=21732700$
Possible licence plates for one province of Turkey.

Without repeats,
$\frac{23!}{22!}\left(\frac{10!}{6!}+\frac{10!}{5!}\right)+\frac{23!}{21!}\left(\frac{10!}{7!}+\frac{10!}{6!}\right)+\frac{23!}{20!}\left(\frac{10!}{8!}+\frac{10!}{7!}\right)=12333060$

2

The response of a patient to medical treatment A can be good, fair, and poor, 60%, 30%, and 10% of the time, respectively. 80%, 60%, and 20% of those that had a good, fair, and poor response to medical treatment A live at least another 5 years. If a randomly selected patient has lived 5 years after the treatment, what is the probability that s/he had a poor response to medical treatment A?

Answer

By Bayes' Law
$P(poor|lives)=\frac{P(lives|poor)P(poor)}{P(lives)}$

$P(lives|poor)=0.2$
$P(poor)=0.1$
$P(lives)=0.6(0.8)+0.3(0.6)+0.1(0.2)=0.68$

$P(poor|lives) = \frac{0.2(0.1)}{0.68} = 0.029$

3

Relentless.com, an online retailer, randomly sampled 250 customers to estimate the time they spend on their website. The results show that the 250 customers spent 20 minutes on average on the website. If the population standard deviation of the time spent on the website is 25 minutes, estimate μ, the true (population) average time spent on the website, at a 95% confidence level.

✓ **Answer**

Assuming we may approximate the distribution of time spent on the website as a normal distribution,

$time \sim T_{249}\left(20, \frac{25}{\sqrt{250}}\right)$
$t_{249}(0.05) = 1.9695$

Our 95% confidence interval will be
$20 \pm 3.114\ min$

4

A law firm wants to estimate the mean number of memoranda produced per associate. A sample of 20 associates shows that they produced an average of 20 memoranda per month with a standard deviation of 2 memoranda per month. Would it be reasonable to conclude that the population mean is 21 memoranda? What about 25?

✓ **Answer**

Assuming we may model the distribution of memoranda produced per month as a normal or student distribution, whilst something like this may be approximated better as a random-event based distribution, we may perform a t test on it.

Firstly, it would never make sense to conclude after a study that the population mean is anything other than the sample mean unless there are other factors that may contribute to the changing of this mean, which are not present in this situation. Hypothetically, if we were to test if those values were likely or possible based off our sample, we may conduct a t test to check the possibility.

$memoranda \sim T_{19}(20, 2)$
We may now construct a 95% confidence interval from this distribution.

$t_{19}(0.05) = 2.093$
$I = 20 \pm 2.093(2)$
$I = [15.814, 24.186]$

We are now 95% certain that the actual population mean will reside in our interval. Which leaves us statistically certain at 95% confidence that 21 is a possible population mean, but 25 is not.

5

During 2011 a company implemented a number of policies aimed at reducing the ages of its customers' accounts. In order to assess the effectiveness of these measures, the company randomly selects 10 customer accounts. The average age of each account is determined for the years 2010 and 2011. These data are given below.

Assuming that the population of paired differences between the average ages in 2011 and 2010 is normally distributed, could we say that the mean average account age is reduced?

Account	Average Account Age in 2011	Average Account Age in 2010	Paired Difference
1	27	35	-8
2	19	24	-5
3	40	47	-7
4	30	28	2
5	33	41	-8
6	25	33	-8
7	31	35	-4
8	29	51	-22
9	15	18	-3
10	21	28	-7

✓ **Answer**

$\bar{x} = -7$
$s = 6.12825877028341$
$X \sim T_9(\bar{x}, s)$

$t = 1.833113$ at 95% confidence

$I = (, -7 + ts] = (, 4.233]$

With our 95% one-tailed confidence interval, we cannot reject the null hypothesis that there is no change in account age, since our interval includes the value 0.

6

A vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contacts per week. (He would like to increase this figure.) As a check on his claim, $n = 36$ salespeople are selected at random, and the number of contacts made by each is recorded for a single randomly selected week. The mean and variance of the 36 measurements were 17 and 9, respectively. Does the evidence contradict the vice president's claim?

✓ Answer

We may construct a t-interval from the given data at the 95% confidence level.

$df = 35$
$\bar{x} = 17$
$s = 3$

$t_{35}(0.05) = 1.6896$ (one-tailed)

$I = [11.9312,)$

Since our test value of 15 lies within our 95% confidence interval, we cannot conclusively conclude if the average sales per week is below or above 15. Therefore, the evidence does not contradict the president's claim, but it does not support it either.

7

A normal population has a mean of 60 and a standard deviation of 12. You select a random sample of 9. Compute the probability the sample mean is greater than 63.

✓ Answer

We may calculate our sample deviation with the formula:
$s = \frac{\sigma}{\sqrt{n}}$
$s = 4$

$X \sim N(60, 4)$
$P(X > 63) = P(N > 0.75) = 0.22663$