STATISTICS FOR BUSINESS

Solved examinations, 2019

Faculty of Economics and Business, Donostia
University of the Basque Country

Author and teacher:
Josemari Sarasola

Gizapedia

gizapedia.hirusta.io

Foreign students that took the May 20 examination:

STATISTICS FOR BUSINESS

Professor: Josemari Sarasola

Date: June 4, 2018, 10:00

1st problem: Fisher's exact test *(1.5 points)*

Among customers entering a shop, we have collected sex and whether each or them has finally bought anything. Data are the following:

Sex (↓) / Did he/she buy? (→)	Yes	No
Woman	8	4
Man	3	9

Tasks to perform:

(a) Taking 8 women that had bougth something as the pivot frequency, test if sex and buying behavior are related. Significance level: 20%.

(b) Idem taking as the pivot frequency the 3 men not buying.

Hint: $\binom{24}{12} = 2704156$

2nd problem: (i) Bernoulli process; (ii) Poisson process *(3 points)*

(i) The probability for a customer entering a shop of being finally a buyer 0.2. Customers are independent with each other. **Tasks to perform:**
(Remark: In (a), (b), (d) sections you don't have to calculate the last result: just give the expression to calculate the required probability.)

(a) Calculate the probability of being at least 10 buyers among 12 customers entering the shop.

(b) Calculate the probability of being 5 customers not buying before the first buyer. Calculate also the expected value of the number of customers not buying before the first buyer.

(c) On Saturday, there were 14 customers entering the shop and only one of them bought something. May we state that buying behavior decreased on that day? $\alpha = 5\%$.

(ii) On the other hand, time between sales is estimated to be 2 hours on average. We asume that sales occur randomly and independently. The shop is opened 8 hours a day.

(d) Calculate the probability of being at most 3 sales along 8 hours. No need to give the last result, just give the expression to calculate the probability.

(e) Calculate the probability of being at least 380 sales along 100 days. Hint: when λ is bigger than 30, we may approximate Poisson probabilities with $N(\lambda, \sqrt{\lambda})$.

3rd problem: CLT *(1.5 points)*

Water consumption a day in a village is distributed uniformly between 10 Hl. and 30 Hl. Consumptions along different days are independent.

Tasks to perform:

(a) We have 1000 Hl. What is the probability of having enough water for 45 days?

(b) Give the minimum total comsumption level along 100 days with a 99% probability?

4th problem: Goodness of fit *(2 points)*

We think that stops in a factory each day follow a discrete uniform distribution, that is to say, equality of probabilities for 0, 1, 2, 3 and 4 stops.

We have collected numbers of stops each day along 40 days. Data are the following:

No. of stops	No. of days
0	7
1	8
2	10
3	9
4	6

Tasks to perform: Test if the discrete uniform distribution is fit or right to those data. $\alpha = 10\%$.

5th problem: Parametric testing *(2 points)*

The quantity of sugar in a type of cookie must be at most 4gr on average. But we suspect it's likely to be bigger for the cookies produced along the last month. So we have collected data measuring the quantity of sugar in randomly drawn cookies that have been produce so far. Data are the following:

$$5 - 6 - 4 - 3.5 - 5 - 6.5$$

The quantity of sugar in each cookie follows allegedly a normal distribution with unknown standard deviation.

Tasks to perform:

(a) Take a decision about the problem, setting H_0, drawing the corresponding schema and performing the appropiate calculations. $\alpha = 10\%$.

(b) Now assume that standard deviation of the quantity of sugar per cookie for the population (all the cookies) is known: $\sigma = 1gr$. Calculate in such case the p-value for the test and give the resulting decision.

1st problem: Fisher's exact test *(1.5 points)*

Among customers entering a shop, we have collected sex and whether each or them has finally bought anything. Data are the following:

Sex (↓) / Did he/she buy? (→)	Yes	No
Woman	**8**	**4**
Man	**3**	**9**

Tasks to perform:

(a) **Taking 8 women that had bougth something as the pivot frequency, test if sex and buying behavior are related. Significance level: 10%.**

(b) **Idem taking as the pivot frequency the 3 men not buying.**

Hint: $\binom{24}{12} = 2704156$

The null hypothesis states that both variables are independent.

We calculate expected frequencies into parentheses; eg. for the first cell: $11 \times 12/24 = 5.5$.

Sex (↓) / Did he/she (→)	Yes	No	Totals
Woman	8 (5.5)	4 (6.5)	12
Man	3 (5.5)	9 (6.5)	12
Totals	11	13	24

As wew reject independence when the observed frequency is much bigger or smaller than the expected frerquency, the test is 2-sided, and so we have to compare the p vlaue to $\alpha/2$.

We take as pivot number 8. If there were independence, that frequence would be 5.5. So, 8 is too big, taking into account the null hypothesis, and hence que have to calculate the probability of being 8 women or *more* in that cell:

$$P[X \geq 8] = P[8, 9, 10, 11] = \frac{\binom{11}{8}\binom{13}{4}}{\binom{24}{12}} + \frac{\binom{11}{9}\binom{13}{3}}{\binom{24}{12}} + \frac{\binom{11}{10}\binom{13}{2}}{\binom{24}{12}} + \frac{\binom{11}{11}\binom{13}{1}}{\binom{24}{12}} = 0.04976636$$

As the p value is smaller than 0.1 (0.2/2), we reject the null hypothesis, that is to say, independence between the two variables, and hence we must decide that sex and buying behavior are linked.

(b)

Taken 3 as the pivot number, under independence the expected frequency would be 5.5, adn compared to it 3 is too small. So we have to calculate the probability of 3 or *less*.

We reset the table amd calculate the probability in the same way:

Sex (↓) / Did he/she buy? (→)	Yes	No	Totals
Man	3	9	12
Woman	8	4	12
Totalak	11	13	24

$$P[X \leq 3] = P[3, 2, 1, 0] = \frac{\binom{11}{3}\binom{13}{9}}{\binom{24}{12}} + \frac{\binom{11}{2}\binom{13}{10}}{\binom{24}{12}} + \frac{\binom{11}{1}\binom{13}{11}}{\binom{24}{12}} + \frac{\binom{11}{0}\binom{13}{12}}{\binom{24}{12}} = 0.04976636$$

The p value is the same as that of the previous section. As it's smaller than half the significance level, we reject independence, and decide that sex and buying behavior are related.

2nd problem: (i) Bernoulli process; (ii) Poisson process *(3 points)*

(i) The probability for a customer entering a shop of being finally a buyer 0.2. Customers are independent with each other. Tasks to perform:

(Remark: In (a), (b), (d) sections you don't have to calculate the last result: just give the expression to calculate the required probability.)

(a) Calculate the probability of being at least 10 buyers among 12 customers entering the shop.

(b) Calculate the probability of being 5 customers not buying before the first buyer. Calculate also the expected value of the number of customers not buying before the first buyer.

(c) On Saturday, there were 14 customers entering the shop and only one of them bought something. May we state that buying behavior decreased on that day? $\alpha = 5\%$.

(ii) On the other hand, time between sales is estimated to be 2 hours on average. We asume that sales occur randomly and independently. The shop is opened 8 hours a day.

(d) Calculate the probability of being at most 3 sales along 8 hours. No need to give the last result, just give the expression to calculate the probability.

(e) Calculate the probability of being at least 380 sales along 100 days. Hint: when λ is bigger than 30, we may approximate Poisson probabilities with $N(\lambda, \sqrt{\lambda})$.

(i)

(a)

X : *buying customers* $\sim B(n = 12, p = 0.2)$

$P[X \geq 10] = P[X = 10] + P[X = 11] + P[X = 12] = 0.2^{10} \cdot 0.8^2 \cdot \frac{12!}{10!2!} + 0.2^{11} \cdot 0.8^1 \cdot \frac{12!}{11!1!} + 0.2^{12}$

(b)

Not buying customers before the first buyer (X) distributes according to the geometric distribution: $G(p = 0.2)$. So:

$$P[X = 4] = (1 - 0.2)^5 \cdot 0.2$$

Expected number of not buying customers before the first buyer is:

$$\mu = \frac{q}{p} = \frac{0.8}{0.2} = 4$$

So, there are 4 not buying customers before the first buyer.

(c)

We take as H_0 the opposite of what we are thinking about the buying behavior. As we think the buying tendency has decreased, we will think in advance that it has increased: $H_0 : p \geq 0.2$.

Evidence shows that among 14 customers only 1 of them (1/14=7%) bought something, but normally they are 20%. So we think there have been *few* buying customers. Hence:

$$P[X \leq 1] = 0.2^1 \cdot 0.8^{13} \cdot \frac{14!}{1!13!} + 0.2^0 \cdot 0.8^{14} \cdot \frac{14!}{0!14!} = 0.1979121$$

As the p value is bigger than the significance level (one-sided test), we accept the null hypothesis and therefore state that there is not stronig enough evidence that the buying tendency has decreased.

(ii)

(d)

We have information about the mean time:

$$\frac{1}{\lambda} = 2h \rightarrow \lambda = \frac{1}{2} = 0.5 \ \textit{sales per hour} \rightarrow \lambda(8h) = 8 \times 0.5 = 4$$

$$P[X \leq 3] = \frac{e^{-4}4^0}{0!} + \frac{e^{-4}4^1}{1!} + \frac{e^{-4}4^2}{2!} + \frac{e^{-4}4^3}{3!}$$

(e)

$$\lambda(100d) = 4 \times 100 = 400$$

$$P(\lambda = 400) \longrightarrow N(\lambda = 400, \sqrt{\lambda} = \sqrt{400} = 20)$$

$P[X \geq 380] = P[X > 380.5] = P\left[Z > \frac{380.5 - 400}{20}\right] = P[Z > -0.975] = P[Z < 0.975] = 0.8352199$, (solved with R).

3rd problem: CLT *(1.5 points)*

Water consumption a day in a village is distributed uniformly between 10 Hl. and 30 Hl. Consumptions along different days are independent.

Tasks to perform:

(a) We have 1000 Hl. What is the probability of having enough water for 45 days?

(b) Give the minimum total comsumption level along 100 days with a 99% probability?

$\boxed{\text{(a)}}$

$$X_i : daily\ consumption \sim U(10,30) \begin{cases} \mu = \dfrac{10+30}{2} = 20 \\ \sigma^2 = \dfrac{(30-10)^2}{12} = 33.3 \end{cases}$$

We have enough with 1000 Hl when consumption along 95 days is smaller or equal to 1000Hl.

Consumption along 95 days distributes in his way, by CLT:

$$\mathbf{X} = X_1 + X_2 + \cdots + X_{95} \sim N(20 \times 45 = 900, \sqrt{33.3 \times 45} = 38.71)$$

And now we calculate the desired probability:

$$P[\mathbf{X} < 1000] = P\left[Z < \frac{1000-900}{38.71}\right] = P[Z < 2.58] = 0.99506$$

$\boxed{\text{(b)}}$

Consumption along 100 day distributes in this way:

$$\mathbf{X} = X_1 + X_2 + \cdots + X_{100} \sim N(20 \times 100 = 2000, \sqrt{33.3 \times 100} = 57.73)$$

Minimum consumption is that which is bigger than a bigger x number:

$$P[\mathbf{X} > x] = P\left[Z > \frac{x-2000}{57.73}\right] = 0.99 \rightarrow \frac{x-2000}{57.73} = -2.32 \rightarrow x = 1866.06Hl.$$

Minimum consumption will be 1866.06Hl. with a 0.99 probability.

4th problem: Goodness of fit *(2 points)*

We think that stops in a factory each day follow a discrete uniform distribution, that is to say, equality of probabilities for 0, 1, 2, 3 and 4 stops.

We have collected numbers of stops each day along 40 days. Data are the following:

No. of stops	**No. of days**
0	**7**
1	**8**
2	**10**
3	**9**
4	**6**

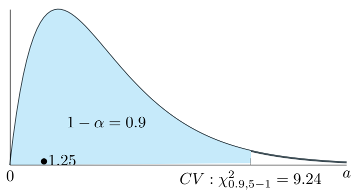
Tasks to perform: Test if the discrete uniform distribution is fit or right to those data. $\alpha = 10\%$.

We must apply the chi-square test. In this test, the null hypothesis is that the set model is right.

The model is the discrete uniform distribution, so it gives equal probabilities to all possible values:

No. of stops	Observed (O)	Prob.	Expected (E)	$\frac{(O-E)^2}{E}$
0	7	1/5=0.2	8	0.125
1	8	1/5=0.2	8	0
2	10	1/5=0.2	8	0.5
3	9	1/5=0.2	8	0.125
4	6	1/5=0.2	8	0.5
	40	1	40	$\mathbf{X}^2 = 1.25$

Let's give the critical value (CV):



As the chi-square statistic is smaller than the critical value, we accept the null hypothesis and hence state that data fit to model.

5th problem: Parametric testing *(2 points)*

The quantity of sugar in a type of cookie must be at most 4gr on average. But we suspect it's likely to be bigger for the cookies produced along the last month. So we have collected data measuring the quantity of sugar in randomly drawn cookies that have been produce so far. Data are the following:

$$5 - 6 - 4 - 3.5 - 5 - 6.5$$

The quantity of sugar in each cookie follows allegedly a normal distribution with unknown standard deviation.

Tasks to perform:

(a) Take a decision about the problem, setting H_0, drawing the corresponding schema and performing the appropiate calculations. $\alpha = 10\%$.

(b) Now assume that standard deviation of the quantity of sugar per cookie for the population (all the cookies) is known: $\sigma = 1gr$. Calculate in such case the p-value for the test and give the resulting decision.

(a)

The situation is this: normal population and unknown variance. So we have to perform the t test.

$$\overline{x} = 5;\ \hat{s} = \sqrt{\frac{(5-5)^2 + (6-5)^2 + \cdots + (6.5-5)^2}{6-1}} = 1.14$$

It seems on the basis of data that the requirement about sugar is not fulfilled. So we take the opposite as the null hypothesis:

$$H_0 : \mu < 4$$

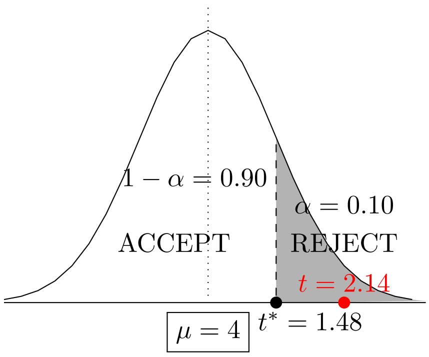
It's a one-sided test with rejection on the upper side (we reject $\mu < 5$ when $\overline{x}$ is big).

Let's calculate the t statistic:

$$t = \frac{5-4}{\frac{1.14}{\sqrt{6}}} = 2.14$$

Let's give the critical value and draw the corresponding schema to perform the test:

$t^* = t_{0.9,6-1} = 1.48$



The t statistics in othe rejection region. So we reject the null hypothesis and decide that the requirement is not fulfilled.

$\boxed{\text{(b)}}$

Now the population standard deviation is known: $\sigma = 1$. The null hypothesis is the same: $H_0 : \mu < 4$.
So, the arithmetic mean distributes in this way:

$$\overline{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) : N(4, 1/\sqrt{6} = 0.40)$$

Let's calculate the p value (the direction of the test is on the upper side, just like in the previous section):

$$p = P[\overline{x} > 5] = P\left[Z > \frac{5-4}{0.40}\right] = P[Z > 2.5] = 0.0062$$

$p = 0.0062 < \alpha$. So we reject the null hypothesis and decide that the requierement is not fulfilled.

STATISTICS FOR BUSINESS
Teacher: Josemari Sarasola
Date: July 9, 2019, 10:00 AM
Duration: 1 hour 3 quarters

1st problem: Binomial test *(1.5 points)*

According to data collected in a bank, clients that finally don't pay the loan they have contracted are 20%. In order to decrease that percentage, stronger conditions have been set up to get a loan. After deploying those conditions, 15 loans have been given, and among them only one of them has not paid the amonut of the loan. May we conclude that the new conditions have decreased the percnetage of not paying? Significance level: 5%.

2nd problem: Sign test *(1.5 points)*

We want to increase sales in a franchising firm's shops. For that purpose, we have set a new merchandising system. Below, you have the amonut sales in the shops before and after the new system (data are paired):

$$45 - 67 - 87 - 56 - 92 - 78 - 62 - 81 - 70 - 94$$

$$56 - 68 - 86 - 59 - 95 - 88 - 69 - 87 - 70 - 86$$

By means of the sign test, decide if the new merchandising system has been successful, with a 10% significance level, both

(a) calculating the critic value (or critical region).

(b) and calculating the p value.

3rd problem: Poisson processes *(1.5 points)*

Mean time (distance, better) till the next failure is 20.000km, following the exponential distribution.

(a) Calculate the probability of time till the next failure being at least 35.000km, taking as the time unit 10.000km.

(b) Calculate the probability of being 4 failures in 50.000km, using the formula for the Poisson distribution.

(c) Which is the distribution of the time till the 5th failure?

(d) Calculate the probability of making at most 80.000km till the 5th failure.

4th problem: Uniform distribution and CLT *(1.5 points)*

The distance covered by a given model of car along one year distributes following the uniform distribution in the 40.000-60.000km interval.

(a) Calculate the probability of the minimum distance in the year along 4 years being at most 50.000km.

(b) Calculate the mean number of kilometers covered in the year with the minimum number of kilometers along 4 years.

(c) Now forget the previous questions and take the CLT. How many kilometers cover at most 10 cars along 3 years with a 0.98 probability. Remark: 10 cars along 3 years, so how many uniform distributions must we sum.

(d) Idem, but how many kilometers *at least*?

5th problem: Validation *(2 points)*

We have drawn the following sample:

$$18, 37, 38, 47, 50, 18, 30, 31, 36, 17, 22, 34, 34, 40, 16, 20, 33, 25, 38, 18$$

Test by means of the runs test using the normal approximation for big samples if the sample was drawn randomly. $\alpha = 0.10$

6th problem: Parametric testing *(2 points)*

We suspect that the proportion of faulty items has increased from the normal value of 10%. To test that, we took 200 items and 30 of them were faulty.

(a) Draw a conclusion about the increasing, setting the critical value. $\alpha = 0.02$.

(b) Draw a conclusion about the increasing, calculating the p value. $\alpha = 0.02$.
