Escola Politècnica Superior

Course: 20305 Mathematics III - Statistics

Type of activity

	Workshop	Quiz	Final Exam	Final Project
Gradable			X	
Non gradable				

Specific skills that are worked on

To develop the capacity for the resolution of mathematical problems that arise in engineering and the ability for applying statistical knowledge (GIN2: CBF01; GEEI: E1; GMAT: E, E45, E47, E48).	X

Generic skills that are worked on

To develop skills in analysis and sinthesis (GIN2: CTR01; GEEI: T1; GMAT: TG2).	X

Date: June 11th, 2018

Instructions

Print your name clearly (last name, first name) in the space provided at the top of ALL pages that you turn in.
Problems must be turned in separately. Different problems cannot share the same page.
In order to get partial credit you have to show supporting work and reasoning skills.
The exam has 4 problems, partial points are given.
The maximum time to complete the exam is 3 hours.
This is a closed-book, closed notes examination; you may consult statistical tables and cheat sheet. You may use a calculator if you wish. However, cell phones are not permitted for use in any way.
Academic misconduct (fraud, inappropriate communication between examinees, impersonation, ...) will be dealt with severely.

Problem 1 *(3.0 POINTS)*
The temperature at Mount Everest affects the behaviour of satellite radios. We know that the probability that the temperature on Mount Everest is below 0°F is 0.4*; the probability that the temperature on Mount Everest is between 0°F and 32°F is* 0.4*; and the probability that the temperature on Mount Everest is is above 32°F is* 0.2*. A team of explorers climbing Mount Everest have two identical looking satellite radios, we will call them* $Radio_A$ *and* $Radio_B$*.* $Radio_A$ *is sensitive to the temperature and works* 50% *of the time given that the temperature is below 0°F,* 80% *of the time given that the temperature is between 0°F and 32°F, and* 100% *of the time given that the temperature is above 32°F.* $Radio_B$ *is insensitive to the temperature and works* 90% *of the time regardless of the temperature.*

a) *Identify and name the relevant events of the problem and their probabilities.*

The temperature T *is important to this problem. Let's define the following events:* T_1 = *"the temperature on Mount Everest is below 0°F",* $P(T_1) = 0.4$*;* T_2 = *"the temperature on Mount Everest is between 0°F and 32°F",* $P(T_2) = 0.4$*;* T_3 = *"the temperature on Mount Everest is above 32°F",* $P(T_3) = 0.2$*. The events* T_1*,* T_2 *and* T_3 *form a partition.*

Another important aspect is whether the radio works, let's define the event W = *"the radio works". To simplify notation we understand that* W_A = *"*$radio_A$ *works" and* W_B = *"*$radio_B$ *works". The probabilities given by the problem are:* $P(W_A|T_1) = 0.5$*,* $P(W_A|T_2) = 0.8$*,* $P(W_A|T_3) = 1.0$ *and* $P(W_B|T_1) = P(W_B|T_2) = P(W_B|T_3) = P(W_B) = 0.9$*.*

These probabilities could also be written using a more complex condition, $P[W|(A\cap T_1)] = 0.5$*,* $P[W|(A\cap T_2)] = 0.8$*,* $P[W|(A\cap T_3)] = 1.0$ *and* $P[W|(B\cap T_1)] = P[W|(B\cap T_2)] = P[W|(B\cap T_3)] = P[W|B] = 0.9$*.*

b) *What is the probability that radio$_A$ works at a given moment?*

This question only takes into account radio$_A$, considering the partition formed by events T_1, T_2 and T_3 using total probability:

$$P(W_A) = P(W_A|T_1) \cdot P(T_1) + P(W_A|T_2) \cdot P(T_2) + P(W_A|T_3) \cdot P(T_3) =$$

$$= 0.5 \cdot 0.4 + 0.8 \cdot 0.4 + 1.0 \cdot 0.2 = 0.2 + 0.32 + 0.2 = 0.72$$

c) *On a really cold day with temperature -25°F, I pick randomly one of the two radios and find that it doesn't work. What is the probability that I picked radio$_A$?*

This question only considers temperatures below 0°F (event T_1) and takes into account both radios. Since radios are identical and chosen randomly, the events $A =$ "choose radio$_A$" and $B =$ "choose radio$_B$" form a partition and are equally probable, $P(A) = P(B) = 0.5$. Using Bayes' theorem:

$$P(A|\overline{W}) = \frac{P(\overline{W}|A) \cdot P(A)}{P(\overline{W})} = \frac{P(\overline{W}|A) \cdot P(A)}{P(\overline{W}|A) \cdot P(A) + P(\overline{W}|B) \cdot P(B)}$$

Introducing the probability of the complement of an event,

$$P(A|\overline{W}) = \frac{(1 - P(W|A)) \cdot P(A)}{(1 - P(W|A)) \cdot P(A) + (1 - P(W|B)) \cdot P(B)} =$$

$$P(A|\overline{W}) = \frac{(1 - 0.5) \cdot 0.5}{(1 - 0.5) \cdot 0.5 + (1 - 0.9) \cdot 0.5} = \frac{5}{6}$$

Another team of explorers climbing Mount Everest have five identical satellite radios of type Radio$_B$.

d) *Find the probability that at a given moment none of the radios work. Find the probability that at a given moment at least one radio works.*

The following questions take into account 5 satellite radios of type radio$_B$ and whether they work or not. Recall that radio$_B$ is insensitive to the temperature and works 90% of the time. Let's define a discrete random variable $Y =$ "number of radios that work", that take values $Success$ (radio works) or $Failure$ (radio doesn't work). The random variable follows a binomial distribution $Y \sim \mathcal{B}(n = 5, p = 0.9)$. The probability distribution function is $P(Y = y) = \binom{n}{y} p^x (1-p)^{n-x}$. Therefore,

$$P(\text{"none of the radios work"}) = P(Y = 0) = \binom{5}{0} 0.9^0 (1 - 0.9)^5 = 0.00001$$

e) *At a given moment, what's the probability that more than half of the radios work?*

$$P(\text{"more than half of the radios work"}) = P(Y > \frac{5}{3}) = P(Y \geq 3) =$$

$$= \binom{5}{3} 0.9^3 (1 - 0.9)^{5-3} + \binom{5}{4} 0.9^4 (1 - 0.9)^{5-4} + \binom{5}{5} 0.9^5 (1 - 0.9)^{5-5} =$$

$$= 0.0729 + 0.32805 + 0.59049 = 0.99144$$

f) At a given moment, what is the expected number of working radios?

The question is about the expected value of the binomial random variable,

$$E(Y) = n \cdot p = 5 \cdot 0.9 = 4.5$$

Problem 2 *(3.0 POINTS)*
Researchers are experimenting with compounds used to bond Teflon to steel. The compounds currently in use require an average drying time of 3.0 *minutes with a standard deviation of* 0.7 *minutes. Assume that this drying time follows a normal distribution.*

a) Find the probability that the drying time is more than 4.0 *minutes.*

First of all, let's write down the information given. $X =$ *"drying time required to bond Teflon to steel (in min)",* $X \sim \mathcal{N}(\mu = 3.0, \sigma = 0.7)$.

$$P(X > 4.0) = P(Z > \frac{4.0 - 3.0}{0.7}) = 1 - P(Z \leq 1.43) = 1 - 0.9236 = 0.0764$$

b) Find the probability that the drying time is between 2.0 *and* 3.5 *minutes.*

$$\begin{aligned} P(2.0 < X < 3.5) &= P(Z < \frac{3.5 - 3.0}{0.7}) - P(Z < \frac{2.0 - 3.0}{0.7}) = \\ &= P(Z < 0.71) - P(Z < -1.43) = P(Z < 0.71) - [1 - P(Z < 1.43)] = \\ &= 0.7611 - [1 - 0.9236] = 0.6847 \end{aligned}$$

To obtain high quality products, three layers of Teflon are applied independent and consecutively once each drying period has ended.

c) Find the probability that the total drying time for high quality products is above 11.0 *minutes.*

The total drying time to bond three layers of Teflon is $T = X_1 + X_2 + X_3$. *These layers are applied independently. Therefore,* T *is normally distributed and* $\mu_T = 3.0 + 3.0 + 3.0$, $\sigma_T = \sqrt{0.7^2 + 0.7^2 + 0.7^2}$.

$$T \sim \mathcal{N}(\mu_T = 9.0, \sigma_Y = 1.212)$$

$$P(T > 11.0) = 1 - P(Z < \frac{11.0 - 9.0}{1.212}) = 1 - P(Z < 1.65) = 1 - 0.9505 = 0.0495$$

d) The manufacturing process is automatic so that it stops if the total drying time is too small (less than 7.5 *minutes) or too large (more than* 10.5 *minutes). Find the probability that the manufacturing process does not stop.*

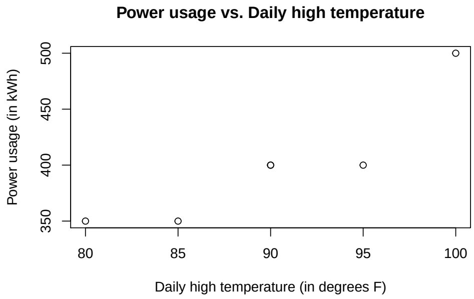
$$\begin{aligned} P(\text{"does not stop"}) &= P(7.5 < T < 10.5) = \\ &= P(Z < \frac{10.5 - 9}{1.212}) - P(Z < \frac{7.5 - 9.0}{1.212}) = \\ &= 2 \cdot P(Z < 1.24) - 1 = 2 \cdot 0.8925 - 1 = 0.7850 \end{aligned}$$

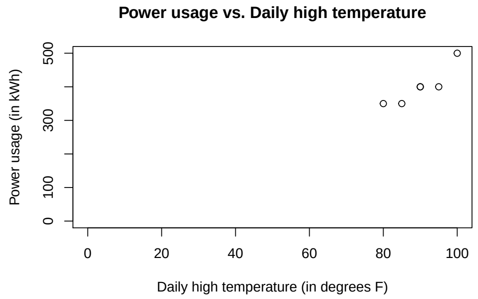
Problem 3 *(2.0 POINTS)*
A power company wants to study the impact of air conditioners on electricity usage during the summer in Palma. Let X *be day time high temperatures in degrees Fahrenheit. Let* Y *be the daily total power use in a small neighborhood in kWh. Over the summer the following data are collected on* 6 *different days:*

x_i	*85*	*90*	*100*	*80*	*95*	*90*
y_i	*350*	*400*	*500*	*350*	*400*	*400*

a) *Draw a scatter plot. From the plot, what can you say about the behavior of the data?*

The plot of the data is





We can see that the data is positively correlated, as day high temperatures increase, so does daily total power use.

b) *Find the equation of the fitted regression line.*

In order to find the fitted regression line, we first calculate means, variances and covariance.

$$\overline{x} = \frac{85+90+100+80+95+90}{6} = 90$$

$$\overline{y} = \frac{350+400+500+350+400+400}{6} = 400$$

$$s_x^2 = \frac{(85-90)^2+(90-90)^2+(100-90)^2+(80-90)^2+(95-90)^2+(90-90)^2}{6-1} = 50$$

$$s_y^2 = \frac{(350-400)^2+(400-400)^2+(500-400)^2+\cdots+(400-400)^2}{6-1} = 3000$$

$$s_{xy} = \frac{(350-400)(85-90)+(400-400)(90-90)+\cdots+(400-400)(90-90)}{6-1} = 350$$

Using these values, we calculate the parameters of the line of best fit:

$$\widehat{\beta}_1 = \frac{s_{xy}}{s_x^2} = \frac{350}{50} = 7$$

$$\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \cdot \overline{x} = 400 - 7 \cdot 90 = -230$$

The equation of the regression line is:

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x = -230 + 7x$$

c) *Estimate the power usage of the neighborhood for a 110°F day. Is the estimation reliable?*

The predicted value for power usage is:

$$\widehat{y} = -230 + 7 \cdot 110 = 540$$

The value $x = 110°F$ falls out of the range of the values used to find the regression line, and the estimation is questionable. The value of the correlation coefficient will tell us whether the estimation is reliable within the range of x.

$$r = \frac{s_{xy}}{s_x \cdot s_y} = \frac{350}{\sqrt{50} \cdot \sqrt{3000}} = 0.9037$$

The value of r is close to 1, therefore, the regression line adjusts well to the data.

Problem 4 *(2.0 POINTS)*

a) *An electrical engineer wants to study the mean melting point of a certain metal alloy used in soldering. Based on his knowledge of the population standard deviation, the engineer computes that a minimum sample size of 50 is needed if he wants to be 95% confident that the error between the sample mean and the population mean will not exceed $\pm 2° C$. What should the minimum sample size be if he wants to be 95% confident that the error between the sample mean and the population mean will not exceed $\pm 1° C$?*

Let's write down the information given by the problem. X = "melting point of metal alloy (in $°C$)". The problem states that the engineer has knowledge of the population standard deviation (σ), we can use the normal sampling distribution.

$$\mu - \overline{x} = \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \leq e$$

A minimum sample size of 50 gives a maximum error $e = 2$:

$$n = 50 \geq \left(\frac{z_{\alpha/2}\sigma}{2}\right)^2$$

$$\left(z_{\alpha/2}\sigma\right)^2 = 50 \cdot 2^2 = 200$$

The minimum number of samples required to achieve a maximum error of $e = 1$ is given by:

$$n \geq \left(\frac{z_{\alpha/2}\sigma}{1}\right)^2 = \frac{200}{1^2} = 200$$

If the error is halved (changes from 2 to 1) then n will be quadrupled.

NOTE: We do not need to know the value of σ for this computation.

b) *Another electrical engineer wants to compare the mean melting points for two different metal alloys used in soldering. He collects a sample of size $n_1 = 36$ for the melting point of alloy 1 and finds that $\overline{x_1} = 185° C$. He also collects a sample of size $n_2 = 64$ for the melting point of alloy 2 and finds that $\overline{x_2} = 185° C$. Assume that the population standard deviation for the melting point of the first alloy is $\sigma_1 = 3° C$ and the population standard deviation for the melting point of the second alloy is $\sigma_2 = 4° C$. Compute a 99% confidence interval for the difference of the means $\mu_1 - \mu_2$. What can you say about the comparison of the melting point means?*

Let's write down the information given by the problem. X_i = "melting point of metal alloy i (in $°C$)". The problem states that the engineer has knowledge of the population standard deviation (σ), we can use the normal sampling distribution for the difference of sample means.

Sample data: $\overline{x_1} = 185° C$; $n_1 = 36$; $\sigma_1 = 3° C$; $\overline{x_2} = 185° C$; $n_2 = 64$; $\sigma_2 = 4° C$.

The confidence interval for the difference of population means is:

$$(\overline{x}_1 - \overline{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{x}_1 - \overline{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

The critical score for a 99% confidence level is:

$$P(Z < z_{\alpha/2}) = 1 - \frac{0.01}{2} = 0.995; \qquad z_{\alpha/2} = 2.575$$

$$(185 - 185) - 2.575\sqrt{\frac{3^2}{36} + \frac{4^2}{64}} \leq \mu_1 - \mu_2 \leq (185 - 185) + 2.575\sqrt{\frac{3^2}{36} + \frac{4^2}{64}}$$

$$-1.82 \leq \mu_1 - \mu_2 \leq 1.82$$

Since the interval contains the value 0, *the melting point of both alloys can be equal.*