utbm
université de technologie
Belfort-Montbéliard

SQ 28 End-term exam

Monday June 25th 2007

For this test, you may use an electronic calculator and the Tables of Statistics. Length 2 hours.
The five parts are independent.

Golf is an individual game, played on a course made up of 18 holes. On every hole, the player puts his ball on a mown part, called start, and hits it with a stick, called *club*, towards a flag on another short grass, called *green*.

The aim of the game is to put the ball in a 10-centimetre hole, under the flag, with a minimum number of strokes.

I. Part 1 :

The first stroke, hit from the start, is called a *drive*, allows strong men to show their strength and club the ball as far as possible.

Let X be the distance, in metres, of an average player's drive. We assume that X is a normal random variable $\mathcal{N}(M, \sigma=10)$. A n=50 sized sample gives a mean $\overline{x} = 197$ (metres).

Calculate a confidence interval of M, with a level $1-\alpha = 0{,}9$.

II. Part 2 :

Roger is a retired employee of the Statistics Institute who has just discovered golf. His concern for perfection makes him note everythings down.

In order to improve his putting (i.e. when he plays the ball from the green with a special club called putter), he counted the number of times he needed three putts to put the ball in the hole.

He proudly announces to his friends "I need 3 putts in 5% of the cases".

During the season he completed the whole course 60 times, which means 1080 holes and he needed three putts 61 times.

1°) Do you agree with him, with an α-risk = 0.05 ?

2°) He says to his wife Monique: "As for you, you did the whole course only ten times and you needed 14 times 3 putts".

Monique concludes with fatalism that she is worse than Roger at putting. Not at all says the statistician.

Carry out a comparison test of the two proportions to know who is right.

III. Part 3 :

The sport commission of the club has to choose between Robert and Roger to complete the senior team. It has both players' scores of the previous competitions. X = Roger's score Y = Robert's score.

We suppose that X and Y are normal, $\mathcal{N}(M_1, \sigma_1)$ for X and $\mathcal{N}(M_2, \sigma_2)$ for Y .

X (Roger)	102	91	85	86	88	94	84	90	90
Y (Robert)	95	87	92	92	89				

1°) Give estimations for M_1, M_2, σ_1 and σ_2 .

2°) Test the equality of the variances.

3°) If the variances are the same, give an estimation of the common value.

4°) Can we accept the hypothesis $M_1 = M_2$ with a risk $\alpha = 0.05$?

IV. Part 4 :

Roger is now a member of the team, and his coach wants him to improve at putting. He places the ball three metres from the hole and performs a number of series of 10 consecutive putts. Let X be the number of times he succeeds (puts the ball in the hole) in one putt.

After 100 series, he gets the following data.

X =	0	1	2	3	4	5	6	7	8	9	10
Number of series	3	15	22	23	17	12	5	2	1	0	0

1°) Give an estimation of the mean.

2°) Test the hypothesis : " X is a Poisson distribution with parameter $\lambda = 3$", with $\alpha = 0.05$.

V. Part 5 :

The coach does not want his players to miss the putts at less than one metre. He asks his players to place 10 balls one metre from the hole and to succeed with the ten balls on the first putt.

Let p the probability for Roger to succeed with one ball, and X the number of successes for Roger.

1°) What is the distribution of X ? Give its expectation and its variance.

If a player fails, he starts again from scratch, and begins a new series of 10 balls.

2°) Let Y be the number of series Roger needs to fulfil the conditions. What is the distribution of Y ? Give its expectation and its variance.

It is easy to ask the difficult question.

Wystan Hugh AUDEN (United States 1907 – 1973)

SQ 28 Final June 2007 .CORRECT VERSION.

-PART 1-

$X \sim \mathcal{N}(M = ?, \sigma^2 = 100)$ sample $n = 50$ $\bar{x} = 197$

Confidence variable : $Y = \dfrac{\bar{X} - M}{\frac{\sigma}{\sqrt{50}}} \sim \mathcal{N}(0,1)$ (σ^2 known) table $\mathcal{N}(0,1)$ $0{,}95 \longrightarrow 1{,}645$; $0{,}05 \longrightarrow -1{,}645$

$\Rightarrow I_M = \left]\bar{X} - 1{,}645 \times \frac{10}{\sqrt{50}}, \bar{X} - 1{,}645 \frac{10}{\sqrt{50}}\right[$ and observation $\underline{I_M =]194{,}67 ; 199{,}33[}$

PART 2

1°) $p = p(\text{needs 3 putts})$ estimator $F = \bar{X}$ observation $f = \frac{61}{1080} \simeq 0{,}0565$

one trial X_k $\begin{cases} p & \text{success} \\ & \text{failure} \end{cases}$ 1080 trials $\sum_{k=1}^{1080} X_k \sim \mathcal{B}(n = 1080, p) \simeq \mathcal{N}(1080p ; \sigma^2 = 1080p(1-p))$

↳ Central limit theorem.

and $F \simeq \mathcal{N}\left(p ; \sigma^2 = \frac{p(1-p)}{1080}\right)$

$H_0 : p = 0{,}05$ against $H_1 : p \neq 0{,}05$ and test variable $Y = \dfrac{F - 0{,}05}{\sqrt{\frac{0{,}05 \times 0{,}95}{1080}}} \sim \mathcal{N}(0,1)$

Table $\mathcal{N}(0,1)$ $0{,}975 \longrightarrow 1{,}96$

$+0{,}025 \longrightarrow -1{,}96 \Rightarrow D_0 = \left]0{,}05 - 1{,}96\sqrt{\frac{0{,}05 \times 0{,}95}{1080}} ; 0{,}05 + 1{,}96\sqrt{\ }\right[= \underline{]0{,}037 ; 0{,}063[= D_0}$

observation $0{,}0565 \in D_0 \Rightarrow \underline{H_0 : p = 0{,}05}$

2°) Comparison test : $p_R = p(\text{success for Roger})$ $p_M = p(\text{success for Monique})$

Test $H_0 : p_R = p_M = p$ against $H_1 : p_R < p_M$ observation $f_r = 0{,}0565$; $f_M = \frac{14}{180} = 0{,}078$

~~Test~~ Estimator of p : $T = \dfrac{n_R F_R + n_M F_M}{n_R + n_M} \longrightarrow$ observation $\underline{f = 0{,}0595}$

Test variable : $Y = \dfrac{F_R - F_M - \cancelto{0}{(p_R - p_M)}}{\sqrt{f(1-f)\left(\frac{1}{1080} + \frac{1}{180}\right)}} \sim \mathcal{N}(0,1) \Rightarrow 0{,}95 \longrightarrow 1{,}645$

$D_0 =]-1{,}645\sqrt{\ }, 0[=]-0{,}0313 ; 0]$ $f_R - f_M = -0{,}05 \notin D_0 \Rightarrow H_1$: she is right

M is worse than R

PART 3

1°) observations → estimations M_1 : $\bar{x}_1 = 90$ σ_1 : $s_1 = 5{,}50$ $n_1 = 9$ (Roger)

M_2 : $\bar{x}_2 = 91$ σ_2 : $s_2 = 3{,}08$ $n_2 = 5$ (Robert)

2°) Variances : $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma^2$ against $H_1 : \sigma_1^2 > \sigma_2^2$

Test variable $Y = \dfrac{S_1^2}{S_2^2} \sim F(8,4)$ Table F : $0{,}95 \longrightarrow 6{,}04 \rightarrow \underline{D_0 = [1 ; 6{,}04[}$

observation : $y = \dfrac{s_1^2}{s_2^2} = 3{,}2 \in D_0 \rightarrow$ We accept $\underline{H_0 : \text{same variance}}$

3°) Common value : estimated by $S^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \longrightarrow$ observation $\underline{s^2 = 23{,}3 = 4{,}83^2}$

4°) Means : $H_0 : M_1 = M_2$ against $H_1 : M_1 \neq M_2$

Test variable $Z = \dfrac{\bar{X}_1 - \bar{X}_2 - \cancelto{0}{(M_1 - M_2)}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{n_1 + n_2 - 2} = T_{12}$ (variances unknown)

Table T_{12} : $0{,}975 \longrightarrow 2{,}179$

$0{,}025 \longrightarrow -2{,}179 \Rightarrow D_0 = \left]-2{,}179 \times 4{,}83\sqrt{\frac{1}{9} + \frac{1}{5}}, +2{,}179 \times 4{,}83\sqrt{\ }\right[$

$D_0 =]-5{,}87 ; 5{,}87[$ and $\bar{x}_1 - \bar{x}_2 = -1$ $\in D_0 \Rightarrow \underline{H_0 : \text{same mean}}$

Part 4 :

X = number of success / 10 trials

1°) Estimation of the mean $\overline{x} = \dfrac{0\times3 + 1\times15 + \ldots + 10\times0}{100} = 3{,}08$

2°) Test $H_0 : X \sim \mathcal{P}(\lambda=3)$ againsts $H_1 : X \not\sim \mathcal{P}(\lambda=3)$ which means $X \not\sim \mathcal{P}$ or $\lambda \neq 3$

k	n_k	p_k	$100p_k$	$\frac{(n_k - np_k)^2}{np_k}$
0	3	0,0498	4,98	
1	15] 18	0,1494	14,94] 19,9	0,19
2	22	0,224	22,4	0,007
3	23	0,224	22,4	0,16
4	17	0,168	16,8	0,002
5	12	0,1	10	0,365
6	5	0,05	5	
7	2] 8	0,0216	2,16] 8,26	0,018
≥ 8	1	0,011	1,1	
				0,59

Test variable $D^2 = \sum \dfrac{(N_k - np_k)^2}{np_k} \sim \chi^2_{6-1-0} = \chi^2_5$

Table χ^2_5 : $0{,}95 \longrightarrow 11{,}07$

$D_0 = [0 ; 11{,}07[$ and $d^2 = 0{,}59 \in D_0$

then we accept $\underline{H_0 : X \sim \mathcal{P}(\lambda=3)}$

Remarque: the result d^2 is so small that we may be doubtful as far as the fairness of the data are concerned. The outcome is too beautiful to be true!

Part 5

1°) p success / $1-p$ failure on one putt. ten times $\longrightarrow X \sim \mathcal{B}(n=10, p)$ $\quad E(X) = 10p$

$Var(X) = 10p(1-p)$

2°) $p(\text{failure on 10 putts}) = 1 - p(X=10) = 1 - p^{10} \Rightarrow p(\text{success 10 times}) = p^{10}$

p^{10} succesful series / $1-p^{10}$ un ——— - the experiment is performed until it succeeds $\Rightarrow Y \sim \mathcal{G}(p^{10})$

$\forall k \in \mathbb{N}^* \quad p(Y=k) = p^{10}(1-p^{10})^{k-1}$

$E(Y) = \dfrac{1}{p^{10}} \qquad Var(Y) = \dfrac{1-p^{10}}{p^{20}}$