Επανάληψη ΒΙΟΣΤΑΤΙΣΤΙΚΗ 18/06/2015

Θέμα 1ο: **α.** Το ποσοστό μιας συγκεκριμένης ασθένειας σε ένα πληθυσμό είναι 8%. Το 80% από εκείνους που έχουν την ασθένεια εμφανίζουν ένα συγκεκριμένο εργαστηριακό εύρημα (θετικό τεστ), ενώ μόνο το 10% από τους μη-ασθενείς παρουσιάζουν το ίδιο εύρημα (ψευδή θετικά).

1. Να βρεθεί η πιθανότητα ένα τυχαίο άτομο του πληθυσμού να εμφανίσει αυτό το εργαστηριακό εύρημα.
2. Ποια είναι η πιθανότητα ένα άτομο που δεν εμφανίζει το εργαστηριακό εύρημα να μην έχει πράγματι την ασθένεια;

β. Το βάρος ενός χαπιού, για τη ρύθμιση της οφθαλμικής πίεσης, είναι τυχαία μεταβλητή με μέση τιμή 8gr και διασπορά 0.25gr^2. Παίρνουμε δείγμα 30 χαπιών.

1. Πόσο κατά μέσο όρο περιμένουμε να είναι το συνολικό τους βάρος; (να δικαιολογήσετε)
2. Να υπολογιστεί προσεγγιστικά η πιθανότητα το συνολικό τους βάρος να είναι από 230 έως 243.8 gr.

ΘΕΜΑ 1, Λύση: **α.** Εάν συμβολίσουμε με

$\mathbf{A}$: το ενδεχόμενο επιλογής ασθενούς ατόμου

$\mathbf{Y}$: το ενδεχόμενο επιλογής μη-ασθενούς ατόμου, δηλαδή $\mathbf{Y}'=\mathbf{A}$

$\mathbf{T}^+$: το ενδεχόμενο το τεστ να είναι θετικό

$\mathbf{T}^-$: το ενδεχόμενο το τεστ να είναι αρνητικό, δηλαδή $\mathbf{T}^{-'}=\mathbf{T}^+$

$\mathbf{T}^+ \mid \mathbf{A}$: το ενδεχόμενο το τεστ να είναι θετικό, όταν όντος το άτομο αυτό έχει τη νόσο

$\mathbf{T}^+ \mid \mathbf{Y}$: το ενδεχόμενο το τεστ να είναι θετικό, όταν το άτομο αυτό δεν πάσχει από τη νόσο

Από τα δεδομένα του προβλήματος γνωρίζουμε ότι

$$P(\mathbf{A}) = 0.08 \Rightarrow P(\mathbf{Y}) = 0.92 \quad P(\mathbf{T}^+ \mid \mathbf{A}) = 0.8 \Rightarrow P(\mathbf{T}^- \mid \mathbf{A}) = 0.2 \quad P(\mathbf{T}^+ \mid \mathbf{Y}) = 0.1 \Rightarrow P(\mathbf{T}^- \mid \mathbf{Y}) = 0.9$$

1. Επειδή τα ενδεχόμενα **A, Y** αποτελούν μια διαμέριση του Ω, εφαρμόζοντας το θεώρημα ολικής πιθανότητας παίρνουμε ότι:

$$P(\mathbf{T}^+) = P(\mathbf{T}^+ \mid \mathbf{A})P(\mathbf{A}) + P(\mathbf{T}^+ \mid \mathbf{Y})P(\mathbf{Y}) = 0.8 \times 0.08 + 0.1 \times 0.92 = 0.064 + 0.092 = \underline{\underline{0.156}}$$

2. Χρησιμοποιώντας τον τύπο του Bayes παίρνουμε ότι:

$$P(\mathbf{Y} \mid \mathbf{T}^-) = \frac{P(\mathbf{T}^- \mid \mathbf{Y})P(\mathbf{Y})}{P(\mathbf{T}^-)} = \frac{0.9 \times 0.92}{1 - P(\mathbf{T}^+)} = \frac{0.828}{0.844} = \underline{\underline{0.9810}}$$

β. Συμβολίζουμε με την τ.μ. X_i: βάρος (σε gr) του i χαπιού $i = 1, \ldots, 30$, για τη ρύθμιση της οφθαλμικής πίεσης. Τότε, σύμφωνα με τα δεδομένα του προβλήματος, $\mu := E(X_i) = 8$, και $\sigma^2 := Var(X_i) = 0.25 = 0.5^2$. Η τ.μ. $Y = \sum_{i=1}^{30} X_i$ είναι το συνολικό βάρος των 30 χαπιών.

1. Για αυτήν ισχύει ότι: $\mu_Y := E(Y) = E\left(\sum_{i=1}^{30} X_i\right) = \sum_{i=1}^{30} E(X_i) = 30 \times \mu = \underline{\underline{240 \text{ gr}}}$
2. Επίσης, $\sigma_Y^2 := Var\left(\sum_{i=1}^{30} X_i\right) \xlongequal[\text{είναι ανεξ.}]{\text{οι τ.μ. } X_i} \sum_{i=1}^{30} Var(X_i) = 30 \times \sigma^2 = \underline{\underline{7.5}}$

   Επειδή οι $X_1, X_2, \ldots, X_{30}$ είναι $n = 30$ ανεξάρτητες και ισόνομες τ.μ., για την τ.μ. $Y = \sum_{i=1}^{30} X_i$ επειδή $n = 30 > 25$ (οριακά μεγάλο πλήθος), μπορούμε να χρησιμοποιήσουμε το κεντρικό οριακό θεώρημα,

   δηλ. $Y \overset{\text{προσεγ.}}{\sim} N\left(\mu_Y = 240, \sigma_Y^2 = 7.5 = 2.738^2\right) \Rightarrow \dfrac{Y - \mu_Y}{\sigma_Y} \overset{\text{προσεγ.}}{\sim} N(0,1)$

   και άρα για την α.σ.κ. της τ.μ. $Z = \frac{Y-\mu_Y}{\sigma_Y} = \frac{\sum_{i=1}^{30} X_i - 240}{2.738}$ έχουμε ότι $P(\frac{Y-\mu_Y}{\sigma_Y} \le z) \simeq \Phi(z)$.

   Έτσι

$$
\begin{aligned}
P\left(230 \le \sum_{i=1}^{30} X_i \le 243.5\right) &= P\left(\frac{230-240}{2.738} \le \frac{\sum_{i=1}^{30} X_i - 240}{2.738} \le \frac{243.5-240}{2.738}\right) \\
&= P(-3.651 \le Z \le 1.387) \\
&\simeq \Phi(1.387) - (1 - \Phi(3.651)) \approx \underline{\underline{0.919 - 0 (= 0.917)}}
\end{aligned}
$$

Θέμα 2ο:

β. Η διάμετρος του κορμού ενός είδους κωνοφόρων δένδρων, μετρημένη σε ύψος 1.5m από το έδαφος, ακολουθεί κανονική κατανομή με μέσο $\mu = 20$cm και τυπική απόκλιση $\sigma = 6$cm.

1. Να βρεθεί η πιθανότητα η διάμετρος ενός δένδρου να είναι μεγαλύτερη των 17cm αλλά να μην υπερβαίνει τα 29cm.
2. Αν πάρουμε ένα δείγμα από 33 δέντρα αυτού του είδους, ποια είναι κατά προσέγγιση η πιθανότητα, σε ακριβώς 20 από αυτά η διάμετρος να είναι μεγαλύτερη των 20cm (να δικαιολογήσετε);

Λύση

β. Συμβολίζουμε με την τ.μ. X: τη διάμετρο (σε cm) του κορμού αυτού του είδους των κωνοφόρων δένδρων. Τότε, σύμφωνα με τα δεδομένα του προβλήματος, $X \sim N(\mu = 20, \sigma^2 = 6^2)$, οπότε η τυποποιημένη τ.μ. $Z = \frac{X-\mu}{\sigma} = \frac{X-20}{6} \sim N(0,1)$ και άρα την α.σ.κ. τη συμβολίζουμε με $P(Z \le z) = \Phi(z)$.

1. Έτσι:

$$
\begin{aligned}
P(17 < X \le 29) &= P\left(\frac{17-20}{6} < \frac{X-20}{6} \le \frac{29-20}{6}\right) = P(-0.5 < Z \le 1.5) \\
&= P(Z \le 1.5) - P(Z \le -0.5) = \Phi(1.5) - \Phi(-0.5) \xlongequal{\text{N(0,1) συμμ. ως προς 0}} \Phi(1.5) - (1 - \Phi(0.5)) \\
&= 0.933 - 1 + 0.691 = 0.624
\end{aligned}
$$

Θέμα 1ο: Δεκαέξι νεογέννητα γουρουνάκια (με περίπου το ίδιο βάρος) μοιράστηκαν τυχαία σε δύο διαφορετικές ομάδες, για την καθεμία εκ των οποίων επιλέχθηκε διαφορετική μέθοδος διατροφής. Μετά από κάποιο χρονικό διάστημα, τα βάρη των ζώων (σε κιλά) μετρήθηκαν και δίδονται στον παρακάτω πίνακα:

Ομάδα 1	60.8	67.0	65.0	68.6	61.7	59.9	69.3	64.6
Ομάδα 2	68.7	67.7	75.0	73.3	71.8	78.2	68.5	71.8

α. Εάν μ και σ^2 είναι αντίστοιχα το μέσο και η διασπορά του βάρους των ζώων (ανεξάρτητα ομάδας), να δοθεί εκτίμηση για αυτές τις παραμέτρους: $\widehat{\mu}=$ _______ $\widehat{\sigma^2}=$ _______

β. Ποια είναι η διάμεσος $\delta=$ _______ και το τρίτο τεταρτημόριο $Q_3=$ _______ για αυτά τα δεδομένα;

γ. Δώστε το ιστόγραμμα συχνοτήτων για το βάρος των ζώων, σχεδιάζοντας σε αυτό και τη συνάρτηση πυκνότητας της κατανομής που πιστεύετε ότι αυτός ακολουθεί. Τι παρατηρείτε; ______________________

δ. Εάν μ_1 και μ_2 είναι τα μέσα του βάρους των ζώων στις αντίστοιχες ομάδες, να κατασκευάσετε διάστημα εμπιστοσύνης (δ.ε.) συντελεστού εμπιστοσύνης (σ.ε.) 90% για το $\mu_1-\mu_2$: (_______, _______). Υποθέσατε ότι οι διασπορές του βάρους στις δύο ομάδες των ζώων είναι ίσες;_______

ε. Να γίνει ο έλεγχος της υπόθεση $H_o:\mu_1=\mu_2$ έναντι της $H_1:\mu_1\neq\mu_2$ και να δοθεί το αντίστοιχο p-value_______. Να συμπεράνετε εάν γίνεται δεκτή η H_o σε επίπεδο σημαντικότητας (ε.σ.) 1%: _______, και σε ε.σ. 10%: _______

στ. Με βάση το δ.ε. του ερωτήματος δ. ποιο είναι το συμπέρασμά σας για τον έλεγχο της $H_o:\mu_1\overset{(\leq)}{=}\mu_2$ έναντι της $H_1:\mu_1>\mu_2$; _________________________ και σε ποιο ε.σ.; _______

ζ. Για να απαντήσετε στα προηγούμενα τρία ερωτήματα είναι απαραίτητο οι δύο ομάδες των δεδομένων να προέρχονται από κανονική κατανομή; ___________ και γιατί _________________________________.

Θέμα 2ο: Ένας ερευνητής, για να μελετήσει την αφθονία και των δύο φύλων ενός είδος πτηνού στις τέσσερεις εποχές του έτους, επέλεξε τυχαία 249, 212, 111 και 81 πτηνά αυτού του είδους την άνοιξη, το καλοκαίρι το φθινόπωρο και τον χειμώνα αντιστοίχως. Τα αποτελέσματα που πήρε δίνονται στον παρακάτω πίνακα

	Εποχή του έτους			
Φύλο	Άνοιξη	Καλοκαίρι	Φθινόπωρο	Χειμώνα
Αρσενικά	163	135	71	43
Θηλυκά	86	77	40	38
	249	212	111	81

α. Εάν συμβολίσουμε με p_Φ το ποσοστό των αρσενικών από το δείγμα των πτηνών του φθινοπώρου και με p_X το ποσοστό των αρσενικών από το δείγμα των πτηνών του χειμώνα, να δοθεί εκτίμηση για αυτές τις παραμέτρους: $\widehat{p_\Phi}=$ _______ $\widehat{p_X}=$ _______

β. Δώστε τα τομεογράμματα για τις τέσσερεις διαφορετικές εποχές του έτους, εμφανίζοντας σε αυτά τα αντίστοιχα ποσοστά για τα δύο διαφορετικά φύλα των πτηνών.

γ. Εάν συμβολίσουμε p_A το συνολικό ποσοστό των αρσενικών πτηνών, μπορούμε να δεχτούμε την υπόθεση $H_o:p_A=0.5$ έναντι της $H_1:p_A\neq 0.5$ σε επίπεδο σημαντικότητας (ε.σ.) 5%; _______ Γιατί;

____

δ. Να γίνει X^2 έλεγχος για να διαπιστώσετε εάν οι τέσσερεις εποχές του έτους είναι ομοιογενείς ως προς το φύλο των πτηνών αυτών. Να δοθεί το αντίστοιχο p-value_______. Ποια η τιμή του κριτηρίου _______ και η κατανομή του _______; Ποια είναι η απόφασή σας σε ε.σ. 10%;__________________

____

1 : Weight | 60,8

	Weight	Group	var
1	60,80	1	
2	67,00	1	
3	65,00	1	
4	68,60	1	
5	61,70	1	
6	59,90	1	
7	69,30	1	
8	64,60	1	
9	68,70	2	
10	67,70	2	
11	75,00	2	
12	73,30	2	
13	71,80	2	
14	78,20	2	
15	68,50	2	
16	71,80	2	
17			

	Sex	Season	Bird_Number	var
1	1	1	163	
2	2	1	86	
3	1	2	135	
4	2	2	77	
5	1	3	71	
6	2	3	40	
7	1	4	43	
8	2	4	38	
9				
10				
11				

OMADA B

```
FREQUENCIES VARIABLES=Weight
  /FORMAT=NOTABLE
  /PERCENTILES=75.0
  /STATISTICS=VARIANCE MEAN MEDIAN
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```

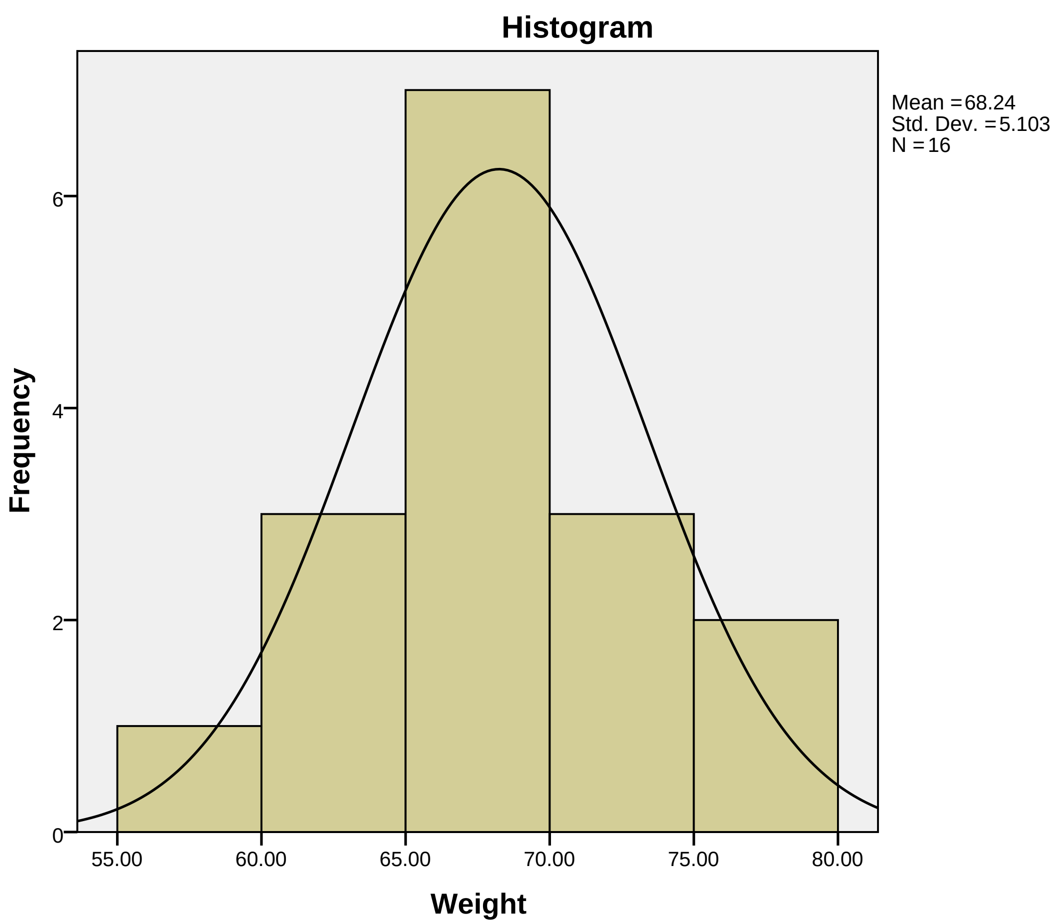
Frequencies

```
[DataSet1] C:\Users\Violeta Piperigkou\Desktop\Biost_Lab_S14\Data_B1.sav
```

Statistics

Weight

N	Valid	16
	Missing	0
Mean		68.2438
Median		68.5500
Variance		26.043
Percentiles	75	71.8000



```
T-TEST GROUPS=Group(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=Weight
  /CRITERIA=CI(.90).
```

T-Test

[DataSet1] C:\Users\Violeta Piperigkou\Desktop\Biost_Lab_S14\Data_B1.sav

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Weight	Group 1	8	64.6125	3.56468	1.26031
	Group 2	8	71.8750	3.59990	1.27276

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
Weight	Equal variances assumed	.021	.887	-4.055	14
	Equal variances not assumed			-4.055	13.999

Independent Samples Test

		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	90% Confidence ... Lower
Weight	Equal variances assumed	.001	-7.26250	1.79117	-10.41730
	Equal variances not assumed	.001	-7.26250	1.79117	-10.41732

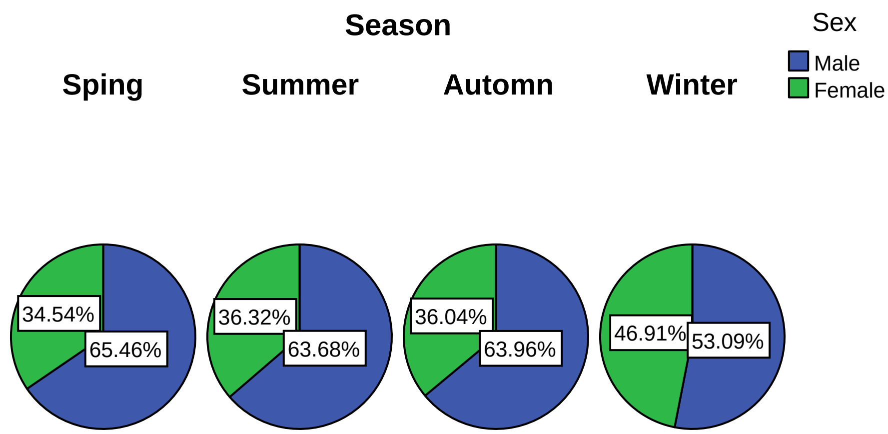
Independent Samples Test

		t-test for Equality of ... 90% Confidence ... Upper
Weight	Equal variances assumed	-4.10770
	Equal variances not assumed	-4.10768

```
GRAPH
  /PIE=PCT BY Sex
  /PANEL COLVAR=Season COLOP=CROSS.
```

Graph

[DataSet2]



Cases weighted by Bird_Number

```
NPAR TESTS
  /BINOMIAL (0.50)=Sex (1)
  /MISSING ANALYSIS.
```

NPar Tests

[DataSet2]

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (2-tailed)
Sex	Group 1	<= 1	412	.63	.50	.000
	Group 2	> 1	241	.37		
	Total		653	1.00		

```
WEIGHT BY Number.
T-TEST
  /TESTVAL=0.5
  /MISSING=ANALYSIS
  /VARIABLES=Male
  /CRITERIA=CI(.95).
```

T-Test

[DataSet2]

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Male	653	.6309	.48292	.01890

One-Sample Test

	Test Value = 0.5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Male	6.928	652	.000	.13093	.0938	.1680

```
 WEIGHT BY Bird_Number.
CROSSTABS
  /TABLES=Sex BY Season
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ
  /CELLS=COUNT EXPECTED COLUMN
  /COUNT ROUND CELL.
```

Crosstabs

[DataSet2]

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Sex * Season	653	100.0%	0	0.0%	653	100.0%

Sex * Season Crosstabulation

			Season				Total
			Sping	Summer	Automn	Winter	
Sex	Male	Count	163	135	71	43	412
		Expected Count	157.1	133.8	70.0	51.1	412.0
		% within Season	65.5%	63.7%	64.0%	53.1%	63.1%
	Female	Count	86	77	40	38	241
		Expected Count	91.9	78.2	41.0	29.9	241.0
		% within Season	34.5%	36.3%	36.0%	46.9%	36.9%
Total		Count	249	212	111	81	653
		Expected Count	249.0	212.0	111.0	81.0	653.0
		% within Season	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.151[a]	3	.246
Likelihood Ratio	4.055	3	.256
Linear-by-Linear Association	2.806	1	.094
N of Valid Cases	653		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 29.89.