# 

# **ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ**

## ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

## ΠΜΣ Εφαρμοσμένη Οικονομική & Ανάλυση Δεδομένων

FIRST ASSIGMENT IN R 2022-2023

Εργασία στα πλαίσια του μαθήματος : Ειδικά Θέματα Οικονομικής των Επιχειρήσεων.

Του

Μουστάκα Ανδρέα

Πάτρα, Οκτώβριος 2022

Πίνακας Περιεχομένων

[Ερώτημα 1a 3](#_Toc117360324)

[Ερώτημα 1b 4](#_Toc117360325)

[Ερώτημα 1c 4](#_Toc117360326)

[Ερώτημα 1d 5](#_Toc117360327)

[Ερώτημα 2 6](#_Toc117360328)

[Ερώτημα 3a 7](#_Toc117360329)

[Ερώτημα 3b 8](#_Toc117360330)

[Ερώτημα 3c 9](#_Toc117360331)

[Ερώτημα 4 10](#_Toc117360332)

[Ερώτημα 5 12](#_Toc117360333)

# Ερώτημα 1a

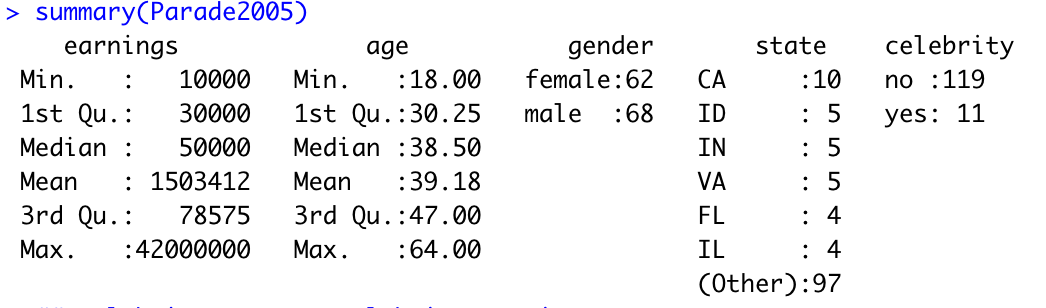
1.The “Parade2005” dataset https://cran.r-project.org/web/packages/AER/AER.pdf is the Sunday newspaper magazine supplementing the Sunday or weekend edition of some 500 daily newspapers in the USA. An important yearly feature is an article providing information on some 120-150 “randomly” selected US citizens, indicating their profession, hometown and state, and their early earnings. The Parade2005 data contain the 2005 version, amended by a year indicating celebrity status. For Parade2005,

a. Determine the mean earnings in California and explain the result.

Πίνακες συνοπτικών περιγραφικών στατιστικών στοιχείων, για την California και ολόκληρου του δείγματος αντίστοιχα:

Table

Description automatically generated with low confidence



Μέσος όρος των αποδοχών για την California = 6241430

Μέσος όρος δείγματος = 1503412

Ο μέσος όρος των αποδοχών είναι αρκετά μεγαλύτερος από αυτόν στο σύνολο. Επίσης παρατηρούμε ότι στο σύνολο των 11 celebrity, οι 4 ανήκουν στην πολιτεία της California.

## Ερώτημα 1b

b. Determine the number of individuals residing in Idaho. What does this say about the dataset?

Table

Description automatically generated

Άτομα που διαμένουν στο Idaho = 5

Συνολικό δείγμα = 130

## Ερώτημα 1c

c. Determine the mean and the median earnings of celebrities. Comment.

Table

Description automatically generated

Μέσος όρος αποδοχών celebrity : 17107273

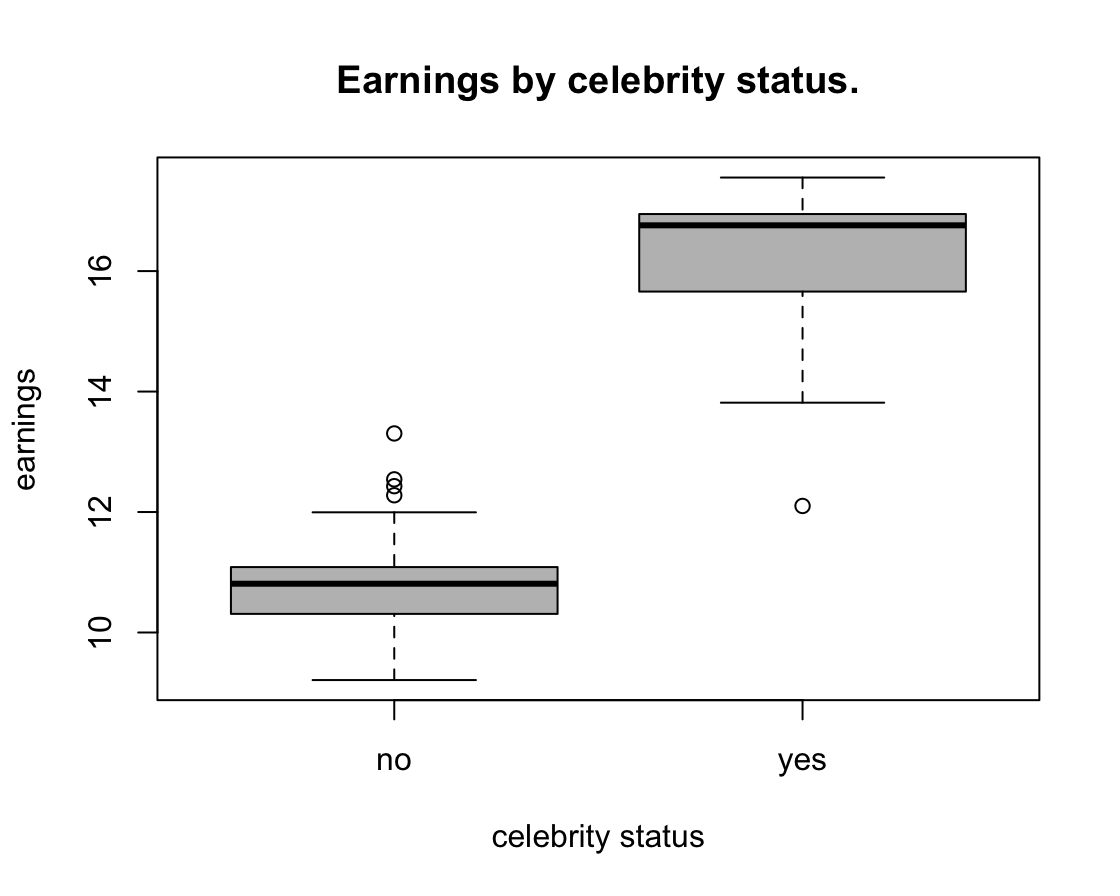
Διάμεσος αποδοχών celebrity : 19000000

Όπως παρατηρούμε, η διάμεσος είναι υψηλότερη από τον μέσο όρο, άρα ακραίες τιμές θεωρούνται οι πιο χαμηλές αποδοχές που «τραβάνε κάτω» τον μέσο όρο.

Επίσης, οι τιμές των αποδοχών αποτελούν τις υψηλότερες από όλο το δείγμα.

## Ερώτημα 1d

d. Obtain boxplots of log(earnings) stratified by celebrity. Comment.



Στο παραπάνω γράφημα βλέπουμε τον τρόπο κατανομής των αποδοχών των

celebrities και μη. Όσον αφορά τα άτομα που δεν είναι celebrities, η κατανομή τους φαίνεται να είναι περίπου συμμετρική, οι απολήξεις έχουν περίπου ίδιο μήκος και η διάμεσος είναι κοντά στο κέντρο του πλαισίου. Για τους celebrities, η κατανομή είναι αρνητικά ασύμμετρη καθώς η κάτω απόληξη είναι μεγαλύτερη από την πάνω και η διάμεσος είναι πάνω από το κέντρο του πλαισίου.

# Ερώτημα 2

2. For the Parade2005 data of the preceding exercise, obtain a kernel density estimate of the earnings for the full data set. It will be necessary to transform the data to logarithms (why?). Comment on the result. Be sure to try out some arguments to density() particular the plug-in bandwidth bw.

Για το Ερώτημα 2, κάνουμε μετασχηματισμό των δεδομένων

στους λογαρίθμους τους, για να έχουμε όσο το δυνατόν πιο «κανονικά» δεδομένα.

Το παρακάτω διάγραμμα αποτελεί την απεικόνιση της μαθηματικής διαδικασίας

εκτίμησης πυκνότητας Kernel μιας τυχαίας μεταβλητής (earnings), καθώς δείχνει πως

συγκλίνουν οι εκτιμήσεις της πυκνότητας. Η συγκεκριμένη καμπύλη πυκνότητας

είναι θετικά ασύμμετρη, ο μέσος όρος είναι υψηλότερος από την διάμεσο.

Chart, histogram

Description automatically generated

# Ερώτημα 3a

3. Consider the CPS1988 data, taken from Bierens and Ginther (2001).

a. Obtain scatterplots of the logarithm of the real wage (wage) versus experience and versus education. Please comment.

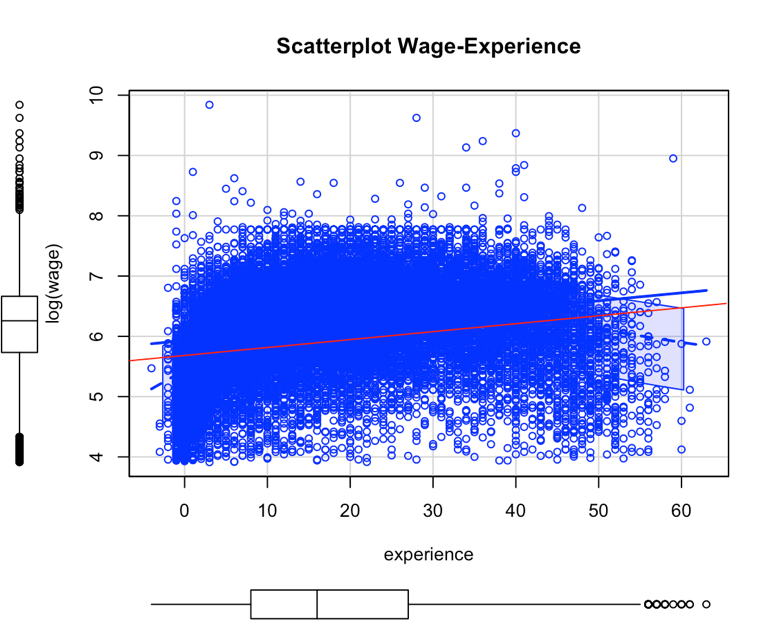
Chart

Description automatically generated

Scatterplot Wage-Education:

Το παραπάνω διάγραμμα εκφράζει την σχέση μεταξύ του μισθού και της εκπαίδευσης.

Η κλίση της ευθείας που διαπερνά τις τιμές είναι θετική, δηλαδή υπάρχει θετική σχέση μεταξύ των δύο μεταβλητών, όση περισσότερη εκπαίδευση έχει κάποιος, τόσο υψηλότερους μισθούς θα λαμβάνει.



Scatterplot Wage-Experience:

Στο συγκεκριμένο διάγραμμα απεικονίζεται η σχέση μεταξύ του μισθού και των ετών εμπειρίας, σύμφωνα με την κλίση της καμπύλης, η σχέση των δύο μεταβλητών είναι θετική.

Συγκρίνοντας το συγκεκριμένο διάγραμμα (Μισθός-Εμπειρία) με το προηγούμενο

(Μισθός-Εκπαίδευση), παρατηρούμε ότι η εκπαίδευση έχει μεγαλύτερη επιρροή στην διαμόρφωση (βελτίωση) του μισθού καθώς η καμπύλη της έχει μεγαλύτερη κλίση σε σχέση με αυτή της εμπειρίας.

## Ερώτημα 3b

b. In fact, education corresponds to years of schooling and therefore takes οα οηlΥ a limited number of values. Transform education into a factor and obtain parallel boxplots of wage stratified by the levels of education.

Στον κάτωθι πίνακα προβάλλεται η μεταβολή του μισθού στις διαφορετικές τιμές των επιπέδων εκπαίδευσης η οποία μετράτε σε χρόνια παρακολούθησης/συμμετοχής στην πρωτοβάθμια/δευτεροβάθμια/τριτοβάθμια εκπαίδευσης και ούτω καθεξής. Βλέπουμε ότι κατά μέσο όρο, οι αυξήσεις στα έτη εκπαίδευσης έχουν ως αποτέλεσμα την αύξηση του μισθού. Όσο περισσότερα έτη εκπαίδευσης έχει κάποιος εργαζόμενος τόσο πιο υψηλούς μισθούς θα λαμβάνει.

Chart, box and whisker chart

Description automatically generated

## Ερώτημα 3c

c. The data set contains four additional factors, ethnicity, smsa, region, and parttime. Obtain suitable graphical displays of log( wage) versus each of these factors. Comment.

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

Στα τέσσερα παραπάνω γραφήματα εκφράζονται τέσσερις επιπλέον παράγοντες που πιθανώς να μεταβάλλουν τον μισθό ενός εργαζομένου. Οι κατανομές τους είναι συμμετρικές καθώς οι απολήξεις τους έχουν περίπου ίδιο μήκος και η διάμεσος είναι κοντά στο κέντρο του πλαισίου. Σύμφωνα με τα γραφήματα, οι παράγοντες ethnicity , smsa και region δεν φαίνεται να έχουν μεγάλες διαφορές σε όρους μισθών. Αντίθετα, ο παράγοντας part time παρουσιάζει σημαντικές αποκλίσεις στους μισθούς, δηλαδή, κατά μέσο όρο κάποιος που δουλεύει part time αναμένεται να έχει σημαντικά χαμηλότερο μισθό από κάποιον που δε δουλεύει part time.

# Ερώτημα 4

4. In this exercise, please calculate simple statistics of wage in the CPS1985 data-frame. First, use the summary function to see the summary statistics of wage; then, calculate mean, median, variance, and standard deviation of the wage. Graphical illustration of statistics would be also helpful. Since wage is numerical, histograms (density plot) and box-plots would be helpful. In this exercise, create a histogram for a wage parameter in the CPS1985 data-frame. As for categorical variables, it is not possible to calculate mean and variance. Instead, we can produce frequency tables, bar-plots, and pie charts. In this exercise, use a table function to produce a frequency table for sector and occupation variables. Along with these, produce bar- plots and pie charts for these variables

Πίνακας περιγραφικών στατιστικών στοιχείων του Wage:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Min | Max | Median | Mean | Variance | St. D | 1stQu | 3rdQu |
| 1 | 44.5 | 7.78 | 9.02 | 26.41 | 5.13 | 5.25 | 11.25 |

Density plot – Box plot:

Το διάγραμμα πυκνότητας αναπαριστά την κατανομή του ωρομισθίου, η οποία είναι θετικά ασύμμετρη με μέσο όρο μεγαλύτερο από τη διάμεσο.

Όσον αφορά το Θηκόγραμμα, μπορούμε να δούμε τις ακραίες τιμές και όπως και στο διάγραμμα πυκνότητας ,παρατηρούμε ότι η κατανομή είναι θετικά ασύμμετρη .

Chart, line chart

Description automatically generatedDiagram, box and whisker chart

Description automatically generated

Περιγραφικά στατιστικά για Sector και Occupation.

Πίνακας συχνοτήτων Sector:

|  |  |  |
| --- | --- | --- |
| Manufacturing | Construction | Other |
| 99 | 24 | 411 |

Chart, waterfall chart

Description automatically generatedChart, pie chart

Description automatically generated

Πίνακας συχνοτήτων Occupation:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Worker | Technical | Services | Office | Sales | Management |
| 156 | 105 | 83 | 97 | 38 | 55 |

Chart, bar chart

Description automatically generatedChart, pie chart

Description automatically generated

# Ερώτημα 5

5.Using the CPS1985 dataset and for categorical variables, it is possible to calculate mean and variance. Moreover, please produce frequency tables, bar-plots, and pie charts. In this exercise, use a table function to produce a frequency table for sector and occupation variables. Along with these, produce bar-plots and pie charts for these variables. It would also be useful to see the relationship between two categorical variables and consider, also, the gender and occupation variables. Use the xtabs function to show the contingency table, then use the plot function to plot a bar-plot showing gender proportion for each occupation type. In advance, calculate the correlation between wage and education to investigate how wage changes by education, then calculate the mean wage for each gender. Finally, compare the wage between males and females using box-plots and qq plots.

Πίνακας συχνοτήτων Occupation by Gender:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gender | Worker | Technical | Services | Office | Sales | Management |
| Male | 126 | 53 | 34 | 21 | 21 | 34 |
| Female | 30 | 52 | 49 | 76 | 17 | 21 |

Chart, bar chart

Description automatically generated

Υπολογισμός συσχέτισης μεταξύ ωρομισθίου και εκπαίδευσης:

Αποτέλεσμα με την βοήθεια της R: r=0.3819221.

Άρα το 38.19% της μεταβλητότητας του ωρομισθίου ερμηνεύεται από την μεταβλητότητα της εκπαίδευσης.

Μέσο ωρομίσθιο για κάθε φύλο:

|  |  |
| --- | --- |
| Male | Female |
| 9.99 | 7.87 |

Chart, box and whisker chart

Description automatically generated

Chart, line chart

Description automatically generated

Όπως φαίνεται και από τα παραπάνω γραφήματα, οι τιμές των ωρομισθίων μεταξύ ανδρών και γυναικών κατανέμονται διαφορετικά. Παρατηρείται ότι οι άνδρες εργάζονται με μεγαλύτερο ωρομίσθιο από τις γυναίκες καθώς ο μέσος όρος και η διάμεσος των ανδρών είναι υψηλότερος.