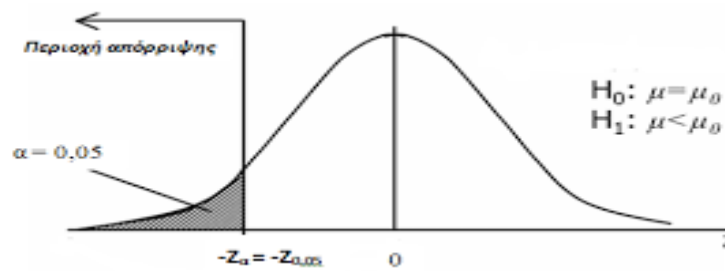




Ακαδημία Εμπορικού Ναυτικού Ασπροπύργου
Σχολή Μηχανικών

Πτυχιακή εργασία
Έλεγχος υποθέσεων



Σπουδαστής: **Κανέλλης Ιωάννης** (ΑΜ 7204)

Επιβλέπων Καθηγητής: **Στέφανος Ι. Καρναβάς**, Μαθηματικός (M.Ed.),
Επίκουρος Καθηγητής

Ακαδημαϊκό έτος: **2020 – 2021**

Ημερομηνία ανάθεσης: 14.06.2019

Ημερομηνία κατάθεσης: 18.02.2021

Ημερομηνία εξέτασης: 08.03.2021

A/A	Ονοματεπώνυμο	Χαρακτηρισμός	Υπογραφή
1	Στέφανος Ι. Καρναβάς Μαθηματικός (M.Ed.) Επίκουρος Καθηγητής	Ιδιαίτερα ικανοποιητική	
2	Αντωνάτος Χαράλαμπος Ναυπηγός Μηχανολόγος Μηχανικός Καθηγητής		
3	Χρηστίδου Αθανασία Ηλεκτρολόγος Μηχανικός Επίκουρη Καθηγήτρια		
Τελικός χαρακτηρισμός			

Περιεχόμενα

Εισαγωγή	3
1. Το εμβόλιο της πολιομυελίτιδας.....	3
2. Η ομάδα ελέγχου.....	4
3. Ο έλεγχος μέσω placebo	4
4. Στατιστικά σημαντικές διαφορές	5
5. Η ανάγκη λήψης ενός πολύ μεγάλου δείγματος	6
6. Αποτελέσματα.....	6
7. Η πολιομυελίτιδα	7
8. Ασπιρίνη και έμφραγμα	7
9. Κάπνισμα και καρκίνος του πνεύμονα	8
10. Ταυτοποίηση και σφάλμα	10
11. William Sealy Gosset (Student).....	11
12. Αστικοί μύθοι.....	12
12.1 Διατήρησή του ανθρακικού σε ένα μπουκάλι σαμπάνιας.....	13
12.2 Εξ όψεως επιλογή πεπονιού	13
12.3 Απώθηση μυγών με σακούλες νερού	14
13. Στατιστικές υποθέσεις και έλεγχος αυτών	14

Εισαγωγή

Η στατιστική είναι χρήσιμη όταν παρουσιάζονται ερωτήματα στα οποία για να δώσουμε απαντήσεις πρέπει να συλλέξουμε και να αναλύσουμε δεδομένα. Τέτοια ερωτήματα μπορεί να είναι αν ένα εμβόλιο είναι ασφαλές, αν μία θεραπεία είναι αποτελεσματική ή αν ένα σύστημα κόλλησης είναι ανθεκτικό. Ένα από τα πιο κοινά προβλήματα είναι ότι η συλλογή των δεδομένων απαιτεί κόπο (και χρήμα) και πρέπει να βρούμε τον ιδανικότερο τρόπο ώστε να μην δαπανήσουμε άσκοπα τους δυσεύρετους πόρους μας. Άλλο πρόβλημα είναι ότι στο τέλος δεν έχουμε ποτέ όλα τα δεδομένα που θα θέλαμε και πρέπει να εκμεταλλευτούμε αυτά που διαθέτουμε στο έπακρο. Στα παραπάνω, θα πρέπει να προσθέσουμε και την παρουσία μεγάλης διακύμανσης, όπου τα δεδομένα δεν ακολουθούν μαθηματικές εκφράσεις αλλά κάτω από ίδιες συνθήκες μπορούμε να πάρουμε διαφορετικά αποτελέσματα.

Έστω ότι το ερώτημα είναι «*Βοηθά η τακτική λήψη μίας δόσης ασπιρίνης στη μείωση του κινδύνου εμφράγματος;*». Μπορούμε να απαντήσουμε επιχειρηματολογώντας πάνω στις επιδράσεις της ασπιρίνης στον οργανισμό αλλά πολλές φορές η πραγματικότητα εκπλήσσει. Ο πιο σίγουρος τρόπος είναι η συλλογή δεδομένων. Η συνήθης πρακτική είναι ο διαχωρισμός ενός συνόλου δεδομένων σε δύο όσο το δυνατόν πιο όμοιες ομάδες. Στη μία χορηγείται ασπιρίνη ενώ στην άλλη όχι και στη συνέχεια συγκρίνονται τα δεδομένα. Γνωρίζουμε ότι όλα τα άτομα που συμμετέχουν στην έρευνα δεν είναι όμοια ούτε έχουν ίδια αντίδραση στην λήψη ασπιρίνης. Πρέπει να γνωρίζουμε πως να ενεργούμε και να βγάζουμε συμπεράσματα δηλώνοντας και το βαθμό αξιοπιστίας που έχουν αυτά. Αυτό είναι στατιστική.

1. Το εμβόλιο της πολιομυελίτιδας

Η επίτευξη πλήρους ανοσίας απέναντι σε ορισμένες μεταδοτικές ασθένειες είναι σίγουρα από τις σημαντικότερες επιτυχίες της ιατρικής στη μάχη για τη βελτίωση των συνθηκών υγιεινής και την αύξηση του προσδόκιμου ζωής. Ωστόσο, κάθε ασθένεια απαιτεί το δικό της εμβόλιο και η ανακάλυψη του δεν είναι πάντα εύκολη υπόθεση. Οι διαδικασίες παρασκευής εμβολίων είναι περίπλοκες και απαιτούν πειράματα σε ζώα ή ακόμα και σε ανθρώπους (σε μικρότερη κλίμακα) για να αποφανθούμε για την αποτελεσματικότητά τους. Πριν εγκριθεί επισήμως η κυκλοφορία ενός εμβολίου πρέπει να επιβεβαιωθεί ότι τα οφέλη είναι τέτοια ώστε να αξίζουν το κόστος και τις παρενέργειες που αναμφίβολα προκύπτουν. Και σε αυτούς τους ελέγχους αποτελεσματικότητας η στατιστική παίζει πολύ μεγάλο ρόλο.

Το 1954 πραγματοποιήθηκε μία έρευνα ευρείας κλίμακας για την αξιολόγηση της αποτελεσματικότητας ενός εμβολίου απέναντι στην πολιομυελίτιδα (το επονομαζόμενο εμβόλιο Salk, που επινοήθηκε από τον επιδημιολόγο Jonas Salk).

Η διαδικασία που ακολουθήθηκε αναλύεται λεπτομερώς στο βιβλίο «Στατιστική, ένας οδηγός στο άγνωστο» στο οποίο 29 περιπτώσεις εφαρμογής της στατιστικής από διαφορετικά πεδία αναλύονται από ειδικούς πάνω στο εκάστοτε πεδίο. Το κεφάλαιο που αντιστοιχεί στις έρευνες του εν λόγω εμβολίου συγγράφηκε από τον καθηγητή Paul Meier, του Πανεπιστημίου του Σικάγο.

Η αποτελεσματικότητα των εμβολίων που έχουν παρασκευαστεί έχει ως αποτέλεσμα τη σχεδόν πλήρη εξάλειψη της πολιομυελίτιδας η οποία μέχρι προσφάτως ήταν μία από τις πιο τρομακτικές ασθένειες. Χτυπούσε παιδιά προκαλώντας τους παραλυσία ή άλλες επιπτώσεις για όλη τους την ζωή. Επιπλέον, ερχόταν κατά απρόβλεπτα επιδημικά κύματα και επηρέαζε περισσότερο κοινωνικές ομάδες με υψηλό επίπεδο ζωής ενώ σπανίως εμφανιζόταν στα πιο φτωχά στρώματα ή χώρες. Φαίνεται πως στους λιγότερο ευνοημένους πληθυσμούς, η ασθένεια εμφανιζόταν νωρίτερα όταν τα μωρά προστατεύονταν ακόμα από τα αντισώματα της

μητέρας τους και έτσι όταν χτυπούσε ο ιός δεν εκδήλωναν την ασθένεια αλλά παρουσίαζαν ανοσία. Μία σύμπτωση που επηρέασε τις προσπάθειες για την εξάλειψη της ασθένειας ήταν το γεγονός ότι ο πρόεδρος Ρούσβελτ είχε εμφανίσει ο ίδιος την ασθένεια και ως εκ τούτου υποστήριξε την έρευνα στο πεδίο αυτό. Στις αρχές της δεκαετίας του 1950 οι υγειονομικές αρχές των ΗΠΑ θεώρησαν ότι το νέο εμβόλιο που είχε παρασκευαστεί από τον Jonas Salk είχε αποδειχθεί ασφαλές και αποτελεσματικό στις μελέτες που είχαν πραγματοποιηθεί σε μικρή κλίμακα, όμως πριν τη γενικευμένη κυκλοφορία του έπρεπε να υπάρχουν αδιάσειστες αποδείξεις τόσο για την αποτελεσματικότητα του όσο και για την απουσία επικίνδυνων παρενεργειών. Έτσι, αποφασίστηκε να πραγματοποιηθεί το μεγαλύτερο πείραμα που είχε γίνει μέχρι τότε στον τομέα της δημόσιας υγείας.

2. Η ομάδα ελέγχου

Έστω ότι δοκιμάζεται ένα νέο φάρμακο για μία ασθένεια και διαπιστώνεται ότι σε όποιον χορηγείται, η ασθένεια θεραπεύεται σε 7 ημέρες. Μπορούμε να πούμε ότι είναι αποτελεσματικό; Ίσως σκέφτεστε ότι από τη στιγμή που όποιος παίρνει το φάρμακο θεραπεύεται, έτσι το φάρμακο πρέπει να είναι αποτελεσματικό, όμως η αλήθεια είναι ότι αυτή η δοκιμή δε δείχνει κάτι τέτοιο. Είναι πιθανόν ότι οι ασθενείς μπορεί να θεραπεύονται και χωρίς να πάρουν τίποτα. Μπορεί ακόμα και να θεραπευόταν συντομότερα αν δεν έπαιρναν το φάρμακο.

Για το λόγο αυτό, όταν δοκιμάζεται η αποτελεσματικότητα ενός καινούριου φαρμάκου ή εμβολίου, επιλέγεται ένα αντιπροσωπευτικό δείγμα από άτομα προς τα οποία απευθύνεται αυτό και χωρίζεται σε δύο ομάδες κατά τυχαίο τρόπο, έτσι ώστε να εξασφαλίσουμε ότι δεν θα υπάρχουν συστηματικές διαφορές ανάμεσα στις δύο ομάδες. Το φάρμακο δίνεται μόνο στη μία ομάδα, ενώ η άλλη ομάδα χρησιμεύει μόνο για τη σύγκριση των επιπτώσεων του φαρμάκου. Η ομάδα στην οποία δεν χορηγείται το φάρμακο, ονομάζεται «ομάδα ελέγχου». Στην περίπτωση της πολιομυελίτιδας η εμφάνιση της ασθένειας παρουσιάζει έντονη διακύμανση.

Για παράδειγμα το 1952, χρονιά με τα περισσότερα κρούσματα από το 1930 μέχρι το 1956, υπήρχαν περίπου 60.000 κρούσματα στις ΗΠΑ ενώ το 1953 ήταν μόνο 35.000, μία μείωση μεγαλύτερη του 40%. Αν το 1953 είχε δοκιμαστεί ένα μη αποτελεσματικό εμβόλιο, θα είχε θεωρηθεί αποτελεσματικό λόγω της σημαντικής μείωσης των κρουσμάτων. Και η χρονιά αυτή δεν ήταν μοναδική, το 1932 ο αριθμός των περιστατικών μειώθηκε παραπάνω από το μισό και το ίδιο συνέβη το 1936, το 1938, το 1942, το 1947 και το 1956. Ούτε θα ήταν καλή ιδέα να εμβολιαστούν μόνο τα παιδιά μίας συγκεκριμένης περιοχής π.χ. της Νέας Υόρκης και να μείνουν ανεμβολίαστα π.χ. του Σικάγο καθώς η εμφάνιση της ασθένειας δεν ήταν ομοιογενής και θα μπορούσε κάλλιστα σε μία πολιτεία να υπάρχει υψηλή εμφάνιση και σε άλλη χαμηλή. Ήταν απαραίτητο, οι δύο ομάδες να επιλεγούν από ολόκληρο το σύνολο των συμμετεχόντων με τα ίδια κριτήρια. Στην μία ομάδα θα χορηγούταν κανονικά το εμβόλιο, ενώ η άλλη θα χρησίμευε ως ομάδα ελέγχου.

3. Ο έλεγχος μέσω placebo

Αν σε κάποια άτομα χορηγηθεί ένα φάρμακο (είτε αυτό είναι ένα χάπι μία φορά την ημέρα, είτε μία ένεση που γίνεται εφάπαξ, όπως στην περίπτωση του εμβολίου) και σε άλλα δεν χορηγηθεί τίποτα, αυτοί που πήραν το φάρμακο, αν είναι πεπεισμένοι ότι αυτό θα έχει θετική επίπτωση θα παρατηρήσουν βελτίωση ακόμα και αν το προϊόν που θα λάβουν δεν έχει καμία θεραπευτική δράση. Είναι το επονομαζόμενο «φαινόμενο placebo». Εκεί στηρίζεται σίγουρα και η επιτυχία πολλών από τα αποκαλούμενα εναλλακτικά φάρμακα καθώς και στο γεγονός ότι

πολλές ασθένειες θεραπεύονται από μόνες τους χωρίς την επίδραση κάποιου φαρμάκου. Στην περίπτωση του εμβολίου της πολιομυελίτιδας, αυτό έχει δύο επακόλουθα: είτε το άτομο θα νοσήσει είτε όχι και δεν τίθεται θέμα βελτίωσης του ασθενή. Επομένως, δεν υπάρχει το ενδεχόμενο υποκειμενικής εκτίμησης του επιπέδου βελτίωσης της ασθένειας. Ωστόσο, δεν ήταν όλα τα κρούσματα εξίσου σοβαρά και δεν άφηναν όλα μόνιμες επιπτώσεις στα παιδιά και:

- αν ένα εμβολιασμένο παιδί παρουσίαζε ελαφρά συμπτώματα της ασθένειας, οι γονείς ή ο γιατρός πίστευαν ότι το παιδί ήταν δύσκολο να νοσήσει (μιας και είχε κάνει το εμβόλιο) ενώ

- αν επρόκειτο για μια ελαφρά περίπτωση που εξελισσόταν καλά, θεωρούσαν ότι είχε γίνει σύγχυση με άλλη ασθένεια και ως εκ τούτου είχαμε ένα μη καταγεγραμμένο κρούσμα της ασθένειας.

Αντιθέτως, όσοι δεν είχαν εμβολιαστεί αισθάνονταν πιο απροστάτευτοι και θεωρούσαν κάθε σύμπτωμα ως ένδειξη της ασθένειας οπότε και θεωρούσαν πως είχαν νοσήσει με αποτέλεσμα τη λανθασμένη εντύπωση ότι υπήρχαν περισσότερα κρούσματα ανάμεσα στους ανεμβολίαστους. Για να αποφευχθεί η περίπτωση το φαινόμενο placebo να λειτουργήσει ευνοϊκά μόνο σε όσους λαμβάνουν τη θεραπεία, είναι προτιμότερο να δεχθούν όλοι μία φαινομενικά όμοια θεραπεία έτσι ώστε να μη γνωρίζουν αν λαμβάνουν το πραγματικό ή το εικονικό φάρμακο, το οποίο έχει ίδια γεύση και όψη και αποκαλείται placebo.

Όμως, δεν είναι μόνον ο συμμετέχων που δε γνωρίζει αν ανήκει στην ομάδα που δέχεται την θεραπεία ή στην ομάδα ελέγχου (στην περίπτωση μας παίζει μεγαλύτερο ρόλο ο γονέας παρά το παιδί), αλλά και ο γιατρός που παρακολουθεί το άτομο. Αυτό δε σημαίνει ότι οι γιατροί δεν είναι έμπιστοι, αλλά υπάρχει η πιθανότητα να επηρεαστούν από τη γνώση τους αυτή και αν ο ασθενής ανήκει στην ομάδα θεραπείας να τον προιδεάσουν για μεγαλύτερη βελτίωση, ενώ αν ανήκει στην ομάδα ελέγχου να τον προιδεάσουν αρνητικά. Για να εξαλειφθεί η πιθανότητα να συμβεί αυτό, οι έρευνες αυτού του τύπου σχεδιάζονται με τρόπο που ούτε ο ασθενής ούτε ο γιατρός που τον παρακολουθεί να γνωρίζουν αν παίρνει το πραγματικό φάρμακο ή το placebo και για αυτό το λόγο η συγκεκριμένη ερευνητική μέθοδος ονομάζεται «διπλά – τυφλή» μέθοδος.

4. Στατιστικά σημαντικές διαφορές

Όταν γίνονται συγκρίσεις, δίνεται μεγάλη σημασία στο αν οι παρατηρούμενες διαφορές είναι στατιστικά σημαντικές ή όχι. Στην πράξη, αυτό στοχεύουν να αποσαφηνίσουν οι στατιστικοί έλεγχοι. Παρόλο που μπορεί να φαίνεται αντιφατικό, το γεγονός ότι μία διαφορά είναι στατιστικά σημαντική δε σημαίνει ότι είναι και σημαντική ως προς το μέγεθος. Μία διαφορά λέγεται ότι είναι στατιστικά σημαντική, όταν θεωρείται ότι δεν μπορεί να αποδοθεί στην τύχη, όταν τα δύο σύνολα που συγκρίνονται δίνουν διαφορετικά αποτελέσματα. Και μπορούμε να είμαστε σίγουροι ότι είναι διαφορετικά, παρόλο που αυτή η διαφορά μπορεί να είναι τόσο μικρή που στην πράξη να φαίνεται αμελητέα. Για παράδειγμα, από μία μελέτη για την οποία χρησιμοποιήθηκε η μέθοδος των πολλαπλών δοκιμών μπορεί να προκύψει ότι η αντοχή που πετυχαίνομε με έναν τύπο κόλλας είναι μεγαλύτερη από αυτήν που πετυχαίνομε με έναν άλλο τύπο, αλλά αυτή η διαφορά μόλις που γίνεται αισθητή.

Αντίστροφα, μπορεί, λόγω του μικρού όγκου των δεδομένων ή της μεγάλης διακύμανσης, να παρατηρηθεί μία μεγάλη διαφορά που όμως μπορεί να αποδοθεί στην τύχη. Με άλλα λόγια, δε μπορούμε να είμαστε σίγουροι ότι το ένα υπό μελέτη αντικείμενο είναι καλύτερο από το άλλο. Ωστόσο, η χρήση της ομάδας ελέγχου παρουσιάζει κάποια προβλήματα, όπως για παράδειγμα ότι απαιτεί περίπλοκη

οργάνωση. Στην περίπτωση του εμβολίου του Salk, έπρεπε να παρασκευαστούν ενέσεις από το πραγματικό φάρμακο πανομοιότυπες με εκείνες που περιείχαν απλό διάλυμα άλατος και έπρεπε να αριθμηθούν έτσι ώστε να μπορούν να ταυτοποιηθούν, χωρίς ούτε το υγειονομικό προσωπικό ούτε ο γιατρός που παρακολουθούσε τον ασθενή να γνωρίζουν αν περιείχαν ή όχι το πραγματικό φάρμακο. Άλλο ζήτημα αφορούσε στον τομέα της ηθικής, το πρόβλημα ήταν ότι σε κάποιους δεν άρεσε το γεγονός να εμβολιαστούν τα παιδιά που συμμετείχαν στην έρευνα με ένα απλό αλατούχο διάλυμα αντί για ένα εμβόλιο το οποίο θεωρούσαν αποτελεσματικό.

Προτάθηκε ως εναλλακτική λύση να εμβολιαστούν τα παιδιά της δεύτερης τάξης και να θεωρηθούν αυτά της πρώτης και της τρίτης ως ομάδες ελέγχου. Αυτό είχε κάποια μειονεκτήματα όπως ότι δεν ακολουθούσε το διπλά τυφλό μοντέλο. Αυτή η μέθοδος, εφαρμόστηκε στις μισές περιοχές ενώ στις άλλες μισές χρησιμοποιήθηκε η λογική της ομάδας ελέγχου με placebo.

5. Η ανάγκη λήψης ενός πολύ μεγάλου δείγματος

Το ποσοστό εμφάνισης της ασθένειας ήταν μόνο 50 κρούσματα ανά 100.000 άτομα και η αναμενόμενη αποτελεσματικότητα του εμβολίου ήταν να μειώσει αυτό τον αριθμό στο μισό. Από αυτό, προκύπτει ότι δε μπορούν να χρησιμοποιηθούν μικρές ομάδες. Αν για παράδειγμα, εμβολιάσουμε 1.000 παιδιά και χρησιμοποιήσουμε άλλα 1.000 ως ομάδα ελέγχου, το πιο πιθανό είναι να μην υπάρχει κανένα κρούσμα ανάμεσα στις δύο ομάδες και το πείραμα να είναι εντελώς άχρηστο.

Αν χρησιμοποιηθούν 10.000 σε κάθε ομάδα, μπορεί να έχουμε 5 κρούσματα στην ομάδα ελέγχου και μόνο 2 στην εμβολιασμένη ομάδα. Η διαφορά εδώ είναι τόσο μικρή που θα μπορούσε άνετα να αποδοθεί στην τύχη (δε θα είναι δυνατόν να απορρίψουμε την μηδενική υπόθεση ότι το ποσοστό εμφάνισης είναι το ίδιο).

Έτσι, ήταν απαραίτητο σε κάθε ομάδα να υπάρχουν εκατοντάδες χιλιάδες άτομα ώστε τα αποτελέσματα να μπορούν να θεωρηθούν αξιόπιστα. Ήταν απαραίτητο να πραγματοποιήσουμε ένα πείραμα ευρείας κλίμακας.

6. Αποτελέσματα

Η έρευνα έδειξε ότι, το εμβόλιο ήταν αποτελεσματικό πέραν από κάθε αμφιβολία. Στην ομάδα των εμβολιασμένων, το ποσοστό εμφάνισης ήταν μικρότερο από το μισό του αντίστοιχου ποσοστού στην ομάδα που είχε λάβει placebo. Η p value της μελέτης ήταν της τάξης του 10^{-9} , δηλαδή στην περίπτωση που δεν υπήρχαν στατιστικά σημαντικές διαφορές ανάμεσα στις δύο ομάδες, η πιθανότητα μία τόσο μεγάλη διαφορά να είναι αποτέλεσμα τύχης ήταν μία στο δισεκατομμύριο.

		Κρούσματα πολιομυελίτιδας	
	Πληθυσμός της μελέτης	Πλήθος	Ποσοστό (ανά 100.000)
Έλεγχος μέσω πλασέμπο			
Εμβολιασμένοι	200.745	57	28
Πλασέμπο	201.229	142	71
Μέθοδος χωρισμού ανά τάξεις			
Εμβολιασμένοι	221.998	56	25
Ομάδες ελέγχου	725.173	391	54

Τα αποτελέσματα, συμφωνούσαν και στην περίπτωση που χρησιμοποιήθηκαν οι μαθητές της πρώτης και της τρίτης τάξης ως ομάδα ελέγχου και όλοι ικανοποιήθηκαν από τον τρόπο που είχε πραγματοποιηθεί το πείραμα, επιλέγοντας

την καλύτερη δυνατή μέθοδο. Ωστόσο, παρόλο που η διαφορά ήταν μεγάλη, αν δεν είχε υπάρξει ομάδα ελέγχου μέσω placebo, θα υπήρχαν πάντα οι αμφιβολίες για άλλες πιθανές ερμηνείες των αποτελεσμάτων.

7. Η πολιομυελίτιδα

Το εμβόλιο Salk, αν και ήταν μία σημαντική πρόοδος στη μάχη απέναντι στην ασθένεια, δεν ήταν απόλυτα ικανοποιητικό και μέσα σε λίγα χρόνια αντικαταστάθηκε με άλλο πιο αποτελεσματικό εμβόλιο, το οποίο πριν κυκλοφορήσει μαζικά υποβλήθηκε σε στατιστικές δοκιμές οι οποίες σχεδιάστηκαν και πραγματοποιήθηκαν με τον πλέον κατάλληλο τρόπο. Σήμερα, η πολιομυελίτιδα είναι μία ασθένεια που οδεύει προς την οριστική εξάλειψη. Μένουν μόνο τέσσερις χώρες στον κόσμο στις οποίες η ασθένεια εξακολουθεί να είναι ενδημική, Νιγηρία, Ινδία, Πακιστάν, και Αφγανιστάν. Ο Παγκόσμιος Οργανισμός Υγείας (ΠΟΥ), η UNICEF και άλλοι διεθνείς οργανισμοί έχουν ανακοινώσει ότι γίνονται προσπάθειες σε αυτές τις χώρες και εκτιμούν ότι σύντομα δεν θα υπάρχουν νέα κρούσματα της ασθένειας. Θα πρέπει να περιμένουμε ακόμα μερικά έτη μέχρι η πολιομυελίτιδα να κηρυχθεί επισήμως εξαφανισμένη από προσώπου Γης.

8. Ασπιρίνη και έμφραγμα

Το 1983 ξεκίνησε στις ΗΠΑ μία μελέτη ευρείας κλίμακας, για την αξιολόγηση της επίδρασης της ασπιρίνης στις ασθένειες της καρδιάς. Έρευνες μικρότερης κλίμακας, είχαν δείξει ότι ένα άτομο το οποίο έχει υποστεί καρδιακό επεισόδιο μπορεί να μειώσει τον κίνδυνο δεύτερου εμφράγματος παίρνοντας ασπιρίνη, αλλά δεν υπήρχε απόδειξη ότι η ευεργετική της επίδραση μπορούσε να γενικευθεί στο σύνολο του αντρικού πληθυσμού. Στην έρευνα, κλήθηκαν να συμμετάσχουν 261.248 άρρενες ιατροί ηλικίας άνω των 40 ετών, των οποίων τα στοιχεία βρέθηκαν από τα αρχεία της Αμερικάνικης Ιατρικής Ένωσης.

Από αυτούς, δέχθηκαν να συμμετάσχουν 59.285, αλλά έπρεπε να αποκλεισθούν όσοι είχαν βεβαρημένο ιατρικό ιστορικό, όσοι έπαιρναν ήδη ασπιρίνη ή όσοι παρουσίαζαν επιπλοκές κατά τη λήψη της. Τελικά, έμειναν 22.071 ιατροί οι οποίοι μπορούσαν να θεωρηθούν υγιείς ενήλικες χαμηλού κινδύνου, οι οποίοι θα έπαιρναν κάθε δεύτερη μέρα μία δόση 325 mg ασπιρίνης (ή placebo). Παράλληλα με την έρευνα για την επίδραση της ασπιρίνης εγκρίθηκε μία έρευνα για την επίδραση της β-καροτίνης (ένα συστατικό που ο οργανισμός μετατρέπει σε βιταμίνη Α) στην πρόληψη συγκεκριμένων μορφών καρκίνου. Έτσι, οι συμμετέχοντες χωρίστηκαν κατά τυχαίο τρόπο σε τέσσερις ομάδες ανάλογα με το αν έπαιρναν:

- πραγματική ασπιρίνη και β-καροτίνη,
- placebo ασπιρίνης και πραγματική β-καροτίνη,
- πραγματική ασπιρίνη και placebo β-καροτίνη ή
- placebo τόσο ασπιρίνης όσο και β-καροτίνης.

Παρά τα αυστηρά ιατρικά κριτήρια επιλογής για τη συμμετοχή στην έρευνα, οι συμμετέχοντες επιλέχθηκαν κατά τύχη όσον αφορά την ηλικία, το ιατρικό ιστορικό, το χαρακτήρα, το αν ήταν καπνιστές ή όχι κτλ. Για αυτό δόθηκε μεγάλη σημασία στην τυχαία κατανομή των ατόμων μέσα στις 4 ομάδες καθώς μόνο με αυτό τον τρόπο θα εξασφαλιζόταν ότι και οι 4 ομάδες αποτελούνταν από άτομα με ίδια χαρακτηριστικά. Θα μπορούσε να υποστηριχτεί ότι, άτομα που είναι στα πρόθυρα του εμφράγματος μπορεί κατά τύχη να συγκεντρωθούν σε μία από τις 4 ομάδες, όμως η θεωρία των πιθανοτήτων μας διαβεβαιώνει ότι αν η κατανομή γίνει εντελώς στην τύχη, η πιθανότητα να συμβεί κάτι τέτοιο σε τόσο μεγάλες ομάδες είναι αμελητέα.

Έτσι, καθώς οι 4 ομάδες είχαν την ίδια σύνθεση και υπόκεινταν στις ίδιες συνθήκες, αν παρατηρούνταν σημαντικές διαφορές ανάμεσα στα αποτελέσματα, αυτές θα οφείλονταν στη διαφορετική επίδραση κάθε σκευάσματος και όχι στην τύχη.

Αυτή είναι η λογική των πειραμάτων που αποσκοπούν στη σύγκριση των φαρμακευτικών σκευασμάτων με τη χρήση ανεξάρτητων τυχαίων δειγμάτων. Η έρευνα, στηρίχτηκε στο «διπλό – τυφλό» μοντέλο, δηλαδή ούτε τα υποκείμενα, ούτε οι επιστήμονες που τους παρακολουθούσαν γνώριζαν σε ποια από τις 4 ομάδες ανήκαν. Μία επιτροπή επιτήρησης, είχε επιφορτιστεί με την ανάλυση των αποτελεσμάτων κάθε 6 μήνες και παρόλο που είχε προβλεφθεί ότι η έρευνα θα κρατούσε 7 χρόνια, μετά από 5 χρόνια τα αποτελέσματα από τη λήψη της ασπιρίνης ήταν τόσο καλά που το πείραμα διακόπηκε για να κοινοποιηθούν τα αποτελέσματα όσο το δυνατόν πιο γρήγορα τόσο στους συμμετέχοντες όσο και στην ιατρική κοινότητα γενικότερα.

	Ομάδα ασπιρίνης (n=11.037)	Ομάδα placebo (n=11.034)
Εμφράγματα μυοκαρδίου		
Θάνασμα	5	18
Μη Θάνασμα	99	171
Σύνολο	104	189

Η ομάδα της ασπιρίνης συμπεριλάμβανε όσους έπαιρναν ασπιρίνη και β-καροτίνη και όσους έπαιρναν ασπιρίνη και placebo β-καροτίνης και η ομάδα του placebo τις άλλες δύο. Ένας στατιστικός έλεγχος σύγκρισης των ποσοστών, δείχνει ότι αν η ασπιρίνη δεν είχε καμία επίδραση (αν η πιθανότητα εμφράγματος ήταν η ίδια και στις δύο ομάδες), μία διαφορά ίση με την παρατηρούμενη ή μεγαλύτερη θα είχε πιθανότητα εμφάνισης 2 στις 100.000. Επομένως, είναι λογικό να θεωρήσουμε ότι η ασπιρίνη μειώνει την πιθανότητα εμφάνισης εμφράγματος.

Η είδηση, έγινε πρωτοσέλιδο στους New York Times και είχε μεγάλη απήχηση στα ΜΜΕ. Αναφορικά με την β-καροτίνη, η έρευνα συνεχίστηκε για το προβλεπόμενο διάστημα και δεν υπάρχουν αναφορές για τα τελικά αποτελέσματα. Μάλλον τα αποτελέσματα δεν ήταν καλά, γιατί, ανατρέχοντας σε ότι ξέρουμε σήμερα για τις επιδράσεις της β-καροτίνης, φαίνεται ότι όχι μόνο δε μειώνει την πιθανότητα εμφάνισης καρκίνου αλλά και την αυξάνει, στην περίπτωση των καπνιστών.

Όμως η ασπιρίνη δεν είναι πανάκεια. Πιστεύεται, ότι αναστέλλει τη διαδικασία της συσσώρευσης των αιμοπεταλίων και έτσι αποτρέπεται η δημιουργία θρόμβων, όμως αυτός ο μηχανισμός αποτελεί επίσης έναν εν δυνάμει κίνδυνο. Πράγματι, η έρευνα έδειξε ότι υπήρχε μία μικρή, στατιστικά μη σημαντική αύξηση της θνησιμότητας από εμβολή σε όσους πήραν ασπιρίνη. Για αυτό, δε συνιστάται να λαμβάνει ανεξέλεγκτα και δίνεται μεγάλη έμφαση στο να ακολουθείται η συμβουλή ιατρού, ο οποίος πρέπει να ζυγίσει τα υπέρ και τα κατά στην κάθε περίπτωση.

9. Κάπνισμα και καρκίνος του πνεύμονα

Σήμερα, είναι αποδεδειγμένο και πέραν πάσης αμφιβολίας ότι «το κάπνισμα βλάπτει σοβαρά την υγεία». Όμως, δεν ήταν πάντα έτσι, σήμερα γνωρίζουμε ποιά είναι τα συστατικά του καπνού του τσιγάρου που μπορούν να προκαλέσουν καρκίνο και γνωρίζουμε πως επιδρούν για να μετατρέψουν τα υγιή κύτταρα σε καρκινικά, κάτι που έχει αποδειχθεί πειραματικά σε ζώα. Ωστόσο, όπως και σε τόσες άλλες περιπτώσεις, τον κώδωνα του κινδύνου ότι κάτι δεν πάει καλά, έκρουσαν τα στατιστικά δεδομένα τα οποία οδήγησαν σε πιο εμπεριστατωμένες έρευνες. Τα διαθέσιμα στοιχεία κατά την δεκαετία του 1950, έδειχναν μεγαλύτερη τάση εμφάνισης καρκίνου του πνεύμονα

ανάμεσα σε καπνιστές από ότι σε μη καπνιστές, όμως έπρεπε να διεξαχθούν πιο λεπτομερείς μελέτες για να επιβεβαιωθούν οι υποψίες. Για να επιβεβαιωθεί αυτή η σχέση του καρκίνου του πνεύμονα ή και άλλων ασθενειών, με τη συνήθεια του καπνίσματος, διεξήχθησαν 7 εκτεταμένες έρευνες στον αγγλοσαξονικό κόσμο (μία στο Ηνωμένο Βασίλειο, μία στον Καναδά και 5 στις ΗΠΑ), όλες έρευνες ευρείας κλίμακας όπου ο αριθμός των συμμετεχόντων κυμάνθηκε ανάμεσα σε 34.000 και 448.000. Η μέθοδος που ακολουθήθηκε, ήταν κατά βάση ίδια σε όλες.

Στα άτομα που επιλέχθηκαν για να συμμετέχουν, εστάλη ένα ερωτηματολόγιο όπου ερωτούνταν σχετικά με τις τρέχουσες και τις παρελθοντικές καταναλωτικές τους συνήθειες, καθώς και για άλλα δημογραφικά στοιχεία και δημιουργήθηκε ένα σύστημα που εγγυόταν ότι όταν ένα άτομο από τους συμμετέχοντες πέθαινε, το γεγονός θα καταγραφόταν μαζί με τη διάγνωση της αιτίας θανάτου. Οι έρευνες αυτές, επέτρεψαν τη συλλογή δεδομένων σχετικά με την επίδραση της ηλικίας που το άτομο ξεκινούσε το κάπνισμα, το είδος και την ποσότητα καπνού καθώς και τι συνέβαινε όταν κάποιος έκοβε το κάπνισμα. Ένα από τα συμπεράσματα ήταν ότι ανάμεσα στους καπνιστές, η εμφάνιση καρκίνου ήταν από 11 έως 20 φορές μεγαλύτερη από ότι ανάμεσα στους μη καπνιστές.

Στη συγκεκριμένη έρευνα, υπήρξαν πολλές αντιρρήσεις. Μία από αυτές, οι οποίες εμφανίζονται συχνά σε τέτοιου είδους έρευνες και εκφραστής της οποίας ήταν και ο Fisher, στηρίζεται στο ότι μπορεί να αποδεικνύεται ότι ανάμεσα στους καπνιστές υπάρχει μεγαλύτερο ποσοστό εμφάνισης μίας ασθένειας, αυτό όμως δε σημαίνει απαραίτητα ότι η αιτία ήταν ο καπνός.

Θα μπορούσε για παράδειγμα να σημαίνει ότι τα άτομα που καπνίζουν είναι πιο ανήσυχα ή πιο νευρικά και αυτή η ιδιαιτερότητα που τα ωθεί στο κάπνισμα να είναι και αυτή που τα καθιστά πιο επιρρεπή σε τέτοιου είδους ασθένειες, καθώς επίσης και ότι υπάρχει κάτι στη γενετική δομή κάποιων ατόμων που τα κάνει επιρρεπή στον εθισμό από το κάπνισμα και παράλληλα πιο επιρρεπή στον καρκίνο του πνεύμονα χωρίς αυτό να σημαίνει ότι η αιτία είναι ο καπνός. Οι αντιρρήσεις αυτές, έχουν βάση γιατί οι έρευνες που διεξήχθησαν δεν ακολουθούσαν το πειραματικό όπως στην περίπτωση του εμβολίου της πολιομυελίτιδας ή της λήψης ασπιρίνης και τις καρδιακές ασθένειες.

Στις περιπτώσεις αυτές, το δείγμα της έρευνας χωριζόταν στην τύχη σε δύο ομάδες, την **ομάδα θεραπείας** και την **ομάδα ελέγχου** και καθώς η μοναδική διαφορά ανάμεσα στις δύο ομάδες ήταν η παρουσία του υπό μελέτη παράγοντα, αν υπήρχαν διαφορετικά αποτελέσματα ανάμεσα τους, αυτά θα μπορούσαν να αποδοθούν στην επίδραση του εν λόγω παράγοντα. Οι έρευνες για τις επιπτώσεις του καπνού δεν αποτελούσαν πειραματικές αλλά προοπτικές μελέτες, όπου υπάρχουν ήδη δύο ομάδες που θέλουμε να μελετήσουμε και απλά παρατηρούμε την εξέλιξη τους.

Εδώ δε μπορούμε να υποχρεώσουμε τους μη καπνιστές να αρχίσουν το κάπνισμα ή τους ήδη καπνιστές να το κόψουν. Το ιδανικό θα ήταν, αν θέλαμε να εφαρμόσουμε τη θεωρία στο έπακρο, να δώσουμε στους μισούς να καπνίσουν κανονικό καπνό και στους άλλους μισούς μία ακίνδυνη ουσία που θα είχε μορφή και γεύση πανομοιότυπη του καπνού.

Οι καπνοβιομηχανίες, θα μπορούσαν να υποστηρίξουν ότι αυτή είναι η μόνη ορθή πρακτική και θα είχαν δίκιο. Ορθή και παράλληλα μη εφικτή. Ωστόσο, με τα διαθέσιμα στοιχεία είναι προφανές, πέραν από κάθε αμφιβολία ότι ο καπνός αποτελεί σημαντικό παράγοντα κινδύνου για εμφάνιση καρκίνου του πνεύμονα, της ουροδόχου κύστης, καρδιακά νοσήματα και άλλες ασθένειες. Η συσχέτιση ανάμεσα στον καρκίνο του πνεύμονα και στην κατανάλωση καπνού, έχει επιβεβαιωθεί μέσω πολυάριθμων ερευνών σε διάφορες χώρες και περιβάλλοντα, κάτι που αποκλείει την

προδιάθεση μίας συγκεκριμένης ομάδας ανθρώπων. Εκτός αυτού, έχει διαπιστωθεί ποιές είναι οι βλαβερές ουσίες του καπνού που μπορεί να προκαλέσουν καρκίνο.

Η γενετική υπόθεση δε μπορεί να εξηγήσει την αύξηση των κρουσμάτων που έχει παρατηρηθεί σε γυναίκες, όταν αυτές υιοθετούν τη συνήθεια του καπνίσματος ή σε μη καπνιστές οι οποίοι εκτίθενται στον καπνό άλλων ατόμων. Προκύπτει ότι, το συμπέρασμα είναι αδιαμφισβήτητο αλλά δεν ήταν πάντα έτσι. Ωστόσο, η στατιστική βρισκόταν πάντα στην πρώτη γραμμή παρέχοντας επιχειρήματα στη σωστή κατεύθυνση.

10. Ταυτοποίηση και σφάλμα

Όταν σχεδιάζονται πειράματα για τη σύγκριση παραγόντων (είτε αυτοί είναι δύο φάρμακα που επιδρούν στην εμφάνιση μίας ασθένειας, είτε δύο καταλύτες που επιδρούν στην απόδοση μίας χημικής αντίδρασης) το κλειδί βρίσκεται στη συλλογή δύο συνόλων δεδομένων στα οποία η μοναδική μεταβλητή κατά την οποία διαφέρουν είναι ο παράγοντας τον οποίο εξετάζουμε. Στο πεδίο της ιατρικής, αυτό που συγκρίνεται (οι θεραπευτικές αγωγές) μπορεί να είναι δύο φάρμακα ή η λήψη ενός φαρμάκου, απέναντι στη μη λήψη του όπως στην περίπτωση του εμβολίου της πολιομυελίτιδας ή της επίδρασης της ασπιρίνης, στα καρδιακά νοσήματα.

Το κλειδί, είναι ο διαχωρισμός των συμμετεχόντων στην έρευνα σε δύο ομάδες, οι οποίες έχουν όσο το δυνατόν πιο όμοια σύνθεση και στις οποίες όλες οι μεταβλητές, εκτός από τον παράγοντα τον οποίο μελετάμε, έχουν ακριβώς την ίδια επίδραση (η τύχη, παρόλο που μπορεί να φαίνεται παράδοξο είναι ένα καλό εργαλείο για να εξασφαλίσουμε την ισορροπία ανάμεσα στις δύο ομάδες).

Για αυτό το λόγο, αν παρατηρούνται στατιστικά σημαντικές διαφορές ανάμεσα στις δύο ομάδες (διαφορές πέραν από αυτές που οφείλονται στην επίδραση της τύχης), αυτές αποδίδονται στον υπό μελέτη παράγοντα. Ωστόσο, αν εκτός από τον υπό μελέτη παράγοντα έχουμε και άλλους παράγοντες που επιδρούν διαφορετικά σε κάθε ομάδα, δε μπορούμε να ξέρουμε αν οι παρατηρούμενες διαφορές αποδίδονται στον υπό μελέτη παράγοντα ή σε κάποιον από τους τρίτους παράγοντες.

Ας δούμε ένα παράδειγμα. Ένα από τα κείμενα που αναφέρονται στον σχεδιασμό πειραμάτων είναι και το βιβλίο *Statistics for Experiments* των Box, Hunter και Hunter στο οποίο εξηγείται πως σχεδιάστηκε ένα πείραμα για τη σύγκριση της φθοράς δύο υλικών που χρησιμοποιούνται στην κατασκευή παπουτσιών για νέους.

Αν διαθέτομε 10 νέους για την υλοποίηση του πειράματος, μία ιδέα θα ήταν να τους κατατάξουμε στην τύχη σε δύο ομάδες των 5 ατόμων, στη μία ομάδα δίνουμε παπούτσια με σόλες από το υλικό Α και στην άλλη παπούτσια με σόλες από το υλικό Β. Μετά από ένα χρονικό διάστημα (έστω 6 μήνες) τους ζητείται να εμφανιστούν με τα παπούτσια, μετράται η φθορά που έχει υποστεί το καθένα και πραγματοποιείται ο αντίστοιχος στατιστικός έλεγχος (στην περίπτωση αυτή καταλληλότερο θα ήταν ο λεγόμενος έλεγχος *t* του Student για ανεξάρτητα δείγματα). Το σωστό θα ήταν ο διαχωρισμός να γίνει στην τύχη.

Δε θα είχε νόημα να πάμε σε ένα σχολείο και να ζητήσουμε από 10 παιδιά να διαγωνιστούν στο τρέξιμο και στους 5 πρώτους να δίνουμε παπούτσια με την σόλα Α και στους υπόλοιπους παπούτσια με την σόλα Β, αφού σίγουρα οι πρώτοι θα τρέχουν περισσότερο, άρα θα κινούνται και περισσότερο και κατ' επέκταση τα παπούτσια τους θα φθείρονται περισσότερο. Ο συγκεκριμένος τρόπος συλλογής δεδομένων, έχει ένα πρόβλημα. Η φθορά της σόλας των παπουτσιών, επηρεάζεται από το υλικό αλλά επηρεάζεται και από το ίδιο το παιδί, ίσως υπάρχουν παιδιά που τρέχουν πολύ ή παίζουν ποδόσφαιρο με τα παπούτσια και άλλα που τρέχουν λίγο.

Επιπλέον μπορεί κάποια παιδιά να μην τα φορούν γιατί δεν τους αρέσουν ή τους χτυπούν τα πόδια. Έτσι, η φθορά της σόλας επηρεάζεται όχι μόνο από το υλικό αλλά και από άλλους παράγοντες και στην περίπτωση που παρατηρήσουμε διαφορές δε μπορούμε να ξέρουμε αν οφείλονται στο υλικό ή σε κάποιον παράγοντα. Μπορεί ακόμα να μην παρατηρήσουμε διαφορές από του υπό μελέτη παράγοντες ενώ στην πραγματικότητα υπάρχουν. Πως μπορούμε να λύσουμε αυτό το πρόβλημα; Κάθε παιδί θα φορέσει στο ένα πόδι τον ένα τύπο σόλας και στο άλλο τον άλλο τύπο. Εφόσον και τα δύο πόδια κινούνται μαζί, οι διαφορές στη φθορά θα οφείλονται αποκλειστικά και μόνο στο υλικό. Στο συγκεκριμένο πειραματικό σχήμα, δε συγκρίνεται ο μέσος όρος του ενός δείγματος με το μέσο όρο από το άλλο, αλλά οι συγκρίσεις γίνονται ξεχωριστά για το κάθε παιδί. Αν γενικά, φθείρεται περισσότερο ο ένας τύπος σόλας από τον άλλο (δεν έχει σημασία αν η φθορά είναι μικρή η μεγάλη) αυτό θα οφείλεται στη διαφορά των υλικών.

11. William Sealy Gosset (Student)

Όλοι όσοι έχουν ασχοληθεί έστω και λίγο με τη μελέτη της στατιστικής, έχουν ακουστά την κατανομή t του Student (σίγουρα πιο δημοφιλής ακόμα και από την κανονική κατανομή) ή τους ελέγχους t του Student για τη σύγκριση των μέσων όρων.

«Student» είναι το ψευδώνυμο με το οποίο υπέγραφε της εργασίες του ο Gosset (1876–1937).



Ένας επιστήμονας με μεγάλη συνεισφορά στη στατιστική ο οποίος πέρασε όλη την επαγγελματική του ζωή στο εργοστάσιο μπίρας της Guinness στο Δουβλίνο. Στις αρχές του 20^{ου} αιώνα, όταν ο Gosset αποκτούσε πτυχία χημείας και μαθηματικών από το πανεπιστήμιο της Οξφόρδης, η Guinness είχε ήδη περάσει στα χέρια ενός νεαρού κληρονόμου που αποφάσισε να κάνει ένα βήμα πέρα από τις παραδοσιακές μεθόδους και να προσλάβει επιστήμονες για να εισάγουν εξελιγμένες τεχνικές, ένας από τους οποίους ήταν και ο Student.

Ο Student, κατάλαβε γρήγορα τη σημασία της χρήσης των στατιστικών μεθόδων στην έρευνα για τη βελτίωση των μεθόδων επεξεργασίας της μπίρας. Ήταν απαραίτητο να μελετηθεί η δράση κάποιων πρώτων υλών που είχαν πολύ μεταβλητά χαρακτηριστικά και ήταν και ήταν πολύ ευαίσθητες στις συνθήκες του περιβάλλοντος. Ήταν απαραίτητο να γίνουν δοκιμές, όμως αυτό σπάνια γινόταν, έτσι τα διαθέσιμα στοιχεία παρέμεναν λιγοστά.

Μέχρι τότε, θεωρούνταν ότι τα δείγματα ήταν αρκετά μεγάλα, ώστε να υπολογιστούν με ακρίβεια οι παράμετροι του πληθυσμού που χρησιμοποιούνταν για τον υπολογισμό πιθανοτήτων. Ωστόσο, αν τα δείγματα ήταν μικρά, οι εκτιμήσεις είναι ανακριβείς. Ο Gosset μελέτησε πολύ, ώστε να βρει ποια ήταν η λύση στο πρόβλημα αυτό και δημοσίευσε τα συμπεράσματα του σε ένα άρθρο με το ψευδώνυμο «Student» γιατί ήταν κανόνας της εταιρείας οι εργαζόμενοι της να μη δημοσιεύουν άρθρα με τα αποτελέσματα των ερευνών τους. Υπάρχουν πολλές εκδοχές σχετικά με το πως και το γιατί χρησιμοποίησε το εν λόγω ψευδώνυμο.

Μία φήμη, λέει ότι στο εργοστάσιο μαθεύτηκε το χόμπι του Gosset όταν αυτός είχε ήδη πεθάνει, ενώ κάποιες άλλες λένε ότι όχι μόνο το γνώριζαν από πριν, αλλά ότι ήταν και ο ίδιος ο διευθυντής της εταιρείας που του πρότεινε να χρησιμοποίησει το Student ως ψευδώνυμο. Φαίνεται πως η πρόθεση της εταιρείας δεν ήταν τόσο να κρατήσει μυστικές τις θεωρίες που αναπτύσσονταν, αλλά να μη μαθευτεί από τις άλλες ζυθοποιίες ότι η Guinness χρησιμοποιούσε στατιστικές τεχνικές για τη βελτίωση των προϊόντων.

Ο στατιστικός έλεγχος που χρησιμοποιείται για τη σύγκριση μέσων όρων όταν τα δεδομένα έχουν ληφθεί με τον παραπάνω τρόπο, ονομάζεται **έλεγχος t** (t-test) του Student για ζευγαρωτές παρατηρήσεις. Σίγουρα, θα ήταν κακή ιδέα να τοποθετούμε πάντα τη σόλα Α στο δεξί πόδι και τη σόλα Β στο αριστερό, γιατί ίσως φθείρεται περισσότερο το ένα από τα δύο παπούτσια. Αυτό μπορεί να εξισορροπηθεί τυχαιοποιώντας τη θέση (για κάθε παιδί ρίχνεται νόμισμα και ανάλογα το αποτέλεσμα επιλέγεται το πόδι που θα τοποθετηθεί το κάθε παπούτσι). Με αυτό τον τρόπο σε περίπτωση που υπάρχει πράγματι επίδραση του ποδιού, αυτή διαχέεται ανάμεσα στα διάφορα αποτελέσματα και εξουδετερώνεται.

Η τυχαιοποίηση δεν κοστίζει τίποτα και προστατεύει απέναντι στην επίδραση πιθανών παραγόντων γνωστών και μη. Για παράδειγμα, μία παρόμοια περίπτωση είναι οι μελέτες που διεξάγονται για τη μέτρηση της ανθεκτικότητας του υλικού επίστρωσης που προστατεύει τα κρύσταλλα των γυαλιών όρασης. Αν σε μία ομάδα ατόμων δοθούν γυαλιά που έχουν το ένα είδος επίστρωσης και σε μία άλλη δοθούν γυαλιά με το άλλο είδος και μετά από ένα συγκεκριμένο διάστημα τους ζητηθεί να επιστρέψουν τα γυαλιά για να μετρηθεί η φθορά, είναι σαφές ότι η φθορά δε θα οφείλεται μόνο στο υλικό αλλά και στον τρόπο που ο καθένας μεταχειρίζεται τα γυαλιά του ή στο περιβάλλον εργασίας του κάθε ατόμου.

Ως εκ τούτου, όπως και στην περίπτωση των παπουτσιών, το καλύτερο θα ήταν να δώσουμε σε κάθε άτομο γυαλιά όπου το κάθε κρύσταλλο θα έχει μία επίστρωση διαφορετικού τύπου (σίγουρα αυτό δε μπορεί να γίνει αν η μία είναι κίτρινη και η άλλη μπλε).

Ομως, θα πρέπει να τυχαιοποιήσουμε τη θέση στην οποία θα μπει κάθε κρύσταλλο, ή μπορούμε πάντα να τοποθετούμε το Α στο δεξί μάτι και το Β στο αριστερό; Πάντα είναι καλύτερο να τυχαιοποιούμε. *Οι ειδικοί λένε ότι όταν καθαρίζουμε τα γυαλιά μας, τείνουμε να ξεκινάμε πάντα από το ίδιο κρύσταλλο.* Και το κρύσταλλο που καθαρίζεται πρώτο, καθαρίζεται πάντα καλύτερα. Το συμπέρασμα είναι ότι πάντα είναι καλύτερο να τυχαιοποιούμε.

12. Αστικοί μύθοι

Υπάρχουν κάποιοι αστικοί μύθοι (ή μήπως δεν πρόκειται για μύθους;) των οποίων η αλήθεια μπορεί να επαληθευτεί χρησιμοποιώντας τη στατιστική. Ας δούμε κάποια παραδείγματα.

12.1 Διατήρησή του ανθρακικού σε ένα μπουκάλι σαμπάνιας

Κάποιοι πιστεύουν ότι τοποθετώντας ένα κουταλάκι στο στόμιο ενός μπουκαλιού σαμπάνιας αποτρέπουμε τη διαφυγή του ανθρακικού (ή τουλάχιστον δε διαφεύγει τόσο όσο αν το στόμιο ήταν ελεύθερο) και με αυτό τον τρόπο διατηρείται καλύτερα για την επόμενη μέρα. Ένας τρόπος να λύσουμε τις απορίες, είναι μέσω της δοκιμής (ή αλλιώς του πειράματος). Το πρόβλημα αυτό, θυμίζει το πρόβλημα με τη δοκιμή του τσαγιού και μπορεί να υλοποιηθεί με τη βοήθεια ενός ατόμου το οποίο θα δοκιμάσει δείγματα σαμπάνιας από δύο διαφορετικά μπουκάλια.

Γνωρίζουμε εξ' αρχής ότι το πείραμα δε μπορεί να γίνει με ένα ποτήρι από το κάθε μπουκάλι. Πρέπει κατ' ελάχιστον να έχουμε 3 ποτήρια από το ένα και 3 ποτήρια από το άλλο, τα οποία θα είναι σε όλα όμοια συμπεριλαμβανομένου και του τρόπου συντήρησης της σαμπάνιας, με μόνη εξαίρεση ότι στο ένα θα έχει τοποθετηθεί ένα κουτάλι και στο άλλο όχι. Αν ο δοκιμαστής αναγνωρίσει τα 3 ποτήρια που προέρχονται από το μπουκάλι με το κουτάλι και αν έχουμε εκτελέσει σωστά το πείραμα, η πιθανότητα να το έχει βρει στην τύχη είναι ακριβώς 5% (3 αντικείμενα μπορούν να επιλεγθούν ανάμεσα σε 6, με 20 διαφορετικούς τρόπους από τους οποίους μόνο ένας είναι σωστός). Αν θέλουμε να μειώσουμε την πιθανότητα να επιλέξει στην τύχη, πρέπει να του δώσουμε να δοκιμάσει περισσότερα ποτήρια αλλά πρέπει να έχουμε υπόψη μας ότι κάποια στιγμή θα αρχίσει να χάνει τις γευστητικές του ικανότητες (μαζί με όλες τις υπόλοιπες).

Μία ιδέα, είναι να χρησιμοποιήσουμε περισσότερα άτομα αλλά χρειάζεται προσοχή διότι κάτι το οποίο είναι απίθανο στη μία επανάληψη, δεν είναι και τόσο απίθανο αν το πείραμα επαναληφθεί περισσότερες φορές. Αν η πιθανότητα να μαντέψει ένα άτομο στην τύχη είναι 5%, στην περίπτωση που επαναλάβουν το πείραμα 5 άτομα, η πιθανότητα να μαντέψει ένας από αυτούς κάτι στην τύχη είναι 40%, κάτι που δεν μας εξυπηρετεί και τόσο.

Σίγουρα για να κάνουμε τα πράγματα πιο απλά θα μπορούσαμε να χρησιμοποιήσουμε ένα μηχάνημα που μετρά την περιεκτικότητα ανθρακικού στα μπουκάλια και το πρόβλημα λύθηκε. Είναι αλήθεια ότι, αν είχαμε ένα τέτοιο μηχάνημα θα ήταν μία καλή επιλογή παρόλο που το να εντοπίσει ένα μηχάνημα τέτοιες διαφορές, δε μας λέει και πολλά καθώς αυτοί που πίνουν την σαμπάνια είναι άνθρωποι και αν ένας άνθρωπος δε μπορεί να εντοπίσει τη διαφορά, τότε το κουτάλι είναι άχρηστο. Για τον ίδιο λόγο, δεν είναι απαραίτητο η δοκιμή να γίνει από έναν δοκιμαστή με εξαιρετικές ικανότητες.

12.2 Εξ όψεως επιλογή πεπονιού

Η επιλογή ενός ώριμου πεπονιού, μοιάζει ακόμα περισσότερο με την περίπτωση της δοκιμασίας του τσαγιού. Κάποια άτομα, υποστηρίζουν ότι είναι σε θέση να διαλέξουν το καλύτερο πεπόνι απλά αγγίζοντας το ή πιάνοντας με τα χέρια τους. Για να διαπιστώσουμε αν είναι αλήθεια, μπορούμε να οργανώσουμε το εξής πείραμα, ο «ειδικός» επιλέγει ανάμεσα σε 5 πεπόνια αυτό που του φαίνεται καλύτερο.

Στη συνέχεια, του δίνεται να δοκιμάσει ένα κομμάτι από το καθένα (φυσικά, χωρίς να ξέρει ποιό αντιστοιχεί σε που) και καλείται να επιλέξει πάλι το καλύτερο αυτή τη φορά με κριτήριο τη γεύση. Θα συμπέσει η επιλογή του με αυτό που επέλεξε την πρώτη φορά; Το πρόβλημα είναι, ότι η πιθανότητα να το επιλέξει στην τύχη είναι 1/5 (δηλ. 20%) και άρα δε μπορούμε να κρίνουμε. Όμως, αν εκτελέσει το πείραμα 2 φορές και πέσει μέσα και στις 2, η πιθανότητα να έχει μαντέψει στην τύχη θα είναι 4%, ενώ αν το εκτελέσει 3 φορές τότε η αντίστοιχη πιθανότητα θα είναι 8%, κάτι που είναι πολύ δύσκολο αν δεν έχει την εν λόγω ικανότητα.

12.3 Απώθηση μυγών με σακούλες νερού

Ένα αρκετά δημοφιλές αυτοσχέδιο μέσο που φημολογείται ότι διώχνει τις μύγες, είναι οι γεμάτες με νερό πλαστικές σακούλες (λατινική Αμερική, Ταϊλάνδη). Κάποιοι πιστεύουν ότι έχει αποτέλεσμα, ενώ άλλοι όχι. Το περίεργο είναι, ότι αυτοί που πιστεύουν πως λειτουργεί, το πιστεύουν για εντελώς διαφορετικούς λόγους, κάποιοι υποστηρίζουν ότι το φως διαχέεται περνώντας από το νερό και αποπροσανατολίζει τις μύγες λόγω του ότι έχουν σύνθετα μάτια. Άλλοι, λένε ότι οι μύγες αποφεύγουν να πλησιάσουν το νερό γιατί ξέρουν ότι θα βραχούν, ενώ υπάρχουν και εκείνοι που πιστεύουν ακριβώς το αντίθετο, τις τοποθετούν στα μαγαζιά για να προσελκύουν τις μύγες και να μην ενοχλούν τους πελάτες.

Τελικά, λειτουργεί ή δε λειτουργεί; Χωρίς να αναλύσουμε το γιατί, ο μόνος τρόπος για να απαντήσουμε σε αυτή την ερώτηση, είναι να συλλέξουμε δεδομένα μέσω ενός προσεχτικά σχεδιασμένου πειράματος και να δούμε σε τι συμπεράσματα θα φτάσουμε. Η αλήθεια είναι ότι δεν είναι εύκολο. Η γενική ιδέα είναι να μετρήσουμε τις μύγες που υπάρχουν στον ίδιο χώρο, όταν υπάρχουν και όταν δεν υπάρχουν σακούλες.

Κάποιες ημέρες, κρεμάμε τις σακούλες και κάποιες όχι (τις επιλέγουμε στην τύχη) και μετράμε πόσες μύγες υπάρχουν κάθε μέρα. Ωστόσο, δεν είναι εύκολη υπόθεση να μετρήσουμε μύγες, παρόλο που η τεχνολογία μπορεί να βοηθήσει, κάποιες φωτογραφικές μηχανές, μπορούν να προγραμματιστούν ώστε να βγάζουν φωτογραφίες σε τακτά διαστήματα και ίσως με μία καλή μηχανή και με τη χρήση λευκού φόντου να μπορέσουμε να βγάλουμε μερικές φωτογραφίες και να μετρήσουμε πόσες μύγες υπήρχαν σε κάθε στιγμιότυπο. Αν υποθέσουμε ότι αυτή η μέθοδος είναι εφικτή, έχει το μειονέκτημα ότι δε μπορούμε να διαπιστώσουμε αν οι μύγες είναι οι ίδιες ή όχι. Ένας άλλος τρόπος να μετρήσουμε έμμεσα πόσες μύγες υπάρχουν στην περιοχή, είναι να χρησιμοποιήσουμε τις κολλητικές ταινίες που παγιδεύονται οι μύγες. Αν το πείραμα δεν είναι προσεχτικά σχεδιασμένο, τότε δεν θα ξέρουμε αν η χρήση της σακούλας έχει αποτέλεσμα ή όχι.

13. Στατιστικές υποθέσεις και έλεγχος αυτών

Θα ασχοληθούμε με την έννοια της στατιστικής υπόθεσης και με τη μέθοδο του ελέγχου των στατιστικών υποθέσεων, όχι με τη μορφή της ολικής ανάπτυξης του θέματος αυτής της ενότητας της στατιστικής ανάλυσης, με σκοπό την ανάλυση και την επεξήγηση του αντικειμένου στο στατιστικό λογισμό. Θα αναλύσω ένα παράδειγμα (S. Goldberg, *Probability: Introduction circa 1960*) με το οποίο απλά αλλά και λεπτομερώς, εμφανίζεται η μέθοδος ελέγχου των υποθέσεων, χρησιμοποιώντας τη διωνυμική κατανομή.

Ο δήμαρχος κ. Α μίας πόλης Π πρόκειται να βάλει υποψηφιότητα στις προσεχείς δημοτικές εκλογές. Η επιτροπή που έχει αναλάβει την οργάνωση της επανεκλογής, συνέρχεται πολύ πριν από την ημέρα των εκλογών. Στη συζήτηση για το θέμα της προεκλογικής εκστρατείας του κ. Α, φαίνεται να υπερέχει του αντιπάλου του κ. Β όμως η επιτροπή συζητά το κατά πόσο θα μπορούσε να έχει ασφαλέστερη πληροφόρηση ως προς αυτό, ώστε να είναι σε θέση να αποφασίσει αν η εξόρμηση που θα οργανώσει θα είναι έντονη και δαπανηρή ή όχι. Ο αντίπαλος του κ. Α, ο κ. Β θα διαθέσει κάθε δυνατότητα για να κερδίσει, προσπαθώντας συγχρόνως να μειώσει το κύρος και τα άλλα πλεονεκτήματα του κ. Α. Η επιτροπή, αποφασίζει ότι αν είναι υπέρ του κ. Α, ποσοστό μέχρι του 60% του πληθυσμού θα αυξήσει τις δαπάνες για την εξόρμηση, αν όμως το ποσοστό αυτό είναι μεγαλύτερο θα ελαττώσει τις δαπάνες και την έντονη εξόρμηση.

Αν η πραγματική αναλογία p των ψηφοφόρων του κ. Α ήταν γνωστή, τότε δεν θα υπήρχε πρόβλημα για την επιτροπή, διότι θα ήταν σε θέση αν μεν $p \leq 0.60$ να πάρει την πρώτη απόφαση, αν δε $p > 0.60$ να πάρει τη δεύτερη. Άρα, το πρόβλημα εντοπίζεται στο ότι η p είναι άγνωστη και επομένως πρέπει να εξασφαλιστεί και κάποια άλλη πληροφορία ώστε να οδηγηθεί η επιτροπή προς τη μία ή την άλλη απόφαση. Στη στατιστική, ονομάζουμε τις περιπτώσεις $p \leq 0.60$ και $p > 0.60$ «υποθέσεις», η διαδικασία δε, με την οποία γίνεται η εκλογή μίας από τις δύο λέγεται «έλεγχος (test)» της μίας απέναντι από την άλλη. Η υπόθεση που ελέγχεται, ονομάζεται «μηδενική υπόθεση», η δε άλλη ως προς την οποία ελέγχεται η μηδενική ονομάζεται «εναλλακτική ή δεύτερη υπόθεση». Συνηθίζουμε την έκφραση για το παράδειγμα η επιτροπή είναι μεταξύ της αποδοχής ή απόρριψης της H_0 .

Παρακάτω, θα εξετάσουμε ποια από τις δύο υποθέσεις θα παίρναμε ως μηδενική. Έστω ότι $H_0: p \leq 0.60$ και $H_1: p > 0.60$. Η απόφαση να δεχτούμε ή όχι την H_0 θα βασιστεί στην εκλογή ενός δείγματος από πρόσωπα–ψηφοφόρους, έτσι ώστε η διαδικασία της δειγματοληψίας να μπορεί να χαρακτηριστεί ως διαδικασία Bernoulli, στην οποία κάθε δοκιμή (= ερώτημα σε κάθε πρόσωπο για την ψήφο του) καταλήγει σε επιτυχία (απάντηση θα ψηφίσω τον κ. Α) ή αποτυχία (απάντηση δεν θα ψηφίσω τον κ. Α) με πιθανότητα επιτυχίας p , για κάθε δοκιμή.

Το μέγεθος του δείγματος επειδή είναι χωρίς επανατοποθέτηση, πρέπει να είναι πολύ μικρό σε σύγκριση με τον όλο πληθυσμό–αριθμό ψηφοφόρων. Έστω ότι το μέγεθος του δείγματος είναι $n=20$ δοκιμές του Bernoulli, η μεταβλητή X είναι διωνυμικά κατανομημένη με παραμέτρους $n=20$ και p (άγνωστο). Οι δύο υποθέσεις H_0, H_1 αναφέρονται και σε μία παράμετρο μίας πιθανοτικής κατανομής.

Οι υποθέσεις αυτού του είδους, ονομάζονται (στατιστικές υποθέσεις). Η απόφαση της επιτροπής να δεχτεί ή όχι την H_0 , θα βασιστεί στο δείγμα $n=20$ και κυρίως την τιμή της X . Δηλαδή, η επιτροπή θα δεχτεί την H_0 αν η p είναι μικρή, όταν η τιμή της X είναι μεγάλη. Αυτό όμως δεν είναι συγκεκριμένο και χρειάζεται να εκφραστεί ένας κανόνας, που σε κάθε περίπτωση να υπαγορεύει την απόφαση της επιτροπής, ανάλογα με το αποτέλεσμα του δείγματος. Παίρνουμε λοιπόν ως **κανόνα απόφασης** «Απορρίπτεται η H_0 αν και μόνο αν τουλάχιστον 13 από τα $n=20$ πρόσωπα δηλώνουν ότι είναι υπέρ του κ. Α». Ο κανόνας αυτός, συμπληρώνεται αν δίνονται και οι τιμές της X , που οδηγούν στην απόρριψη της H_0 . Το σύνολο των τιμών αυτών της X , ονομάζεται **κριτικό σύνολο** (k). Αν το αποτέλεσμα ανήκει στο κριτικό σύνολο, η H_0 απορρίπτεται ή αλλιώς η H_0 είναι δεκτή, που σημαίνει αντίστοιχα με την απόφαση ότι η H_0 θα είναι αληθινή ή ψεύτικη. Δηλαδή σύμφωνα με τον κανόνα:

α΄. Η H_0 είναι πράγματι αληθινή και η τιμή της X δεν ανήκει στο κριτικό σύνολο, δηλαδή η H_0 δεν είναι αποδεκτή.

β΄. Η H_0 είναι πράγματι αληθινή και η τιμή της X ανήκει στο k , δηλαδή η H_0 απορρίπτεται.

γ΄. Η H_0 είναι πραγματικά ψεύτικη και η τιμή της X δεν ανήκει στο k , δηλαδή η H_0 είναι αποδεκτή.

δ΄. Η H_0 είναι πραγματικά ψεύτικη και η τιμή της X ανήκει στο k , δηλαδή η H_0 απορρίπτεται.

Βλέπουμε ότι κατά τις **α΄** και **δ΄** δυνατότητες αναπτύσσεται σωστή δραστηριότητα, επειδή η επιτροπή θα αποδεχθεί μεν την H_0 όταν είναι αληθινή, θα την απορρίψει δε όταν αυτή είναι ψεύτικη. Όμως, η ενέργεια στις δυνατότητες **β΄** και **γ΄** είναι λάθος. Και η περίπτωση **β΄** λέμε αποτελεί **σφάλμα (πλάνη) 1^{ου} είδους** ή **τύπο σφάλματος I**, η περίπτωση δε **γ΄** αποτελεί **σφάλμα (πλάνη) 2^{ου} είδους** ή **τύπο σφάλματος II**. Άρα, αν η επιτροπή πέσει σε σφάλμα 1^{ου} είδους (**β΄**), τότε θα

αποφασίσει για σχετικά μικρές και λιγότερο δαπανηρές προεκλογικές ενέργειες αν και ο κ. Α έχει το πολύ 60% των ψηφοφόρων με το μέρος του. Το σφάλμα αυτό ενώ θα έχει οικονομικό όφελος, μπορεί να οδηγήσει στην ήττα του κ. Α.

Αν η επιτροπή κάνει σφάλμα 2^{ου} είδους, τότε πρέπει να αποφασίσει έντονη δραστηριότητα παρόλο που ο κ. Α έχει με το μέρος του περισσότερους από 60% των ψηφοφόρων, πράγμα που οδηγεί σε άσκοπη οικονομική δαπάνη που θα μπορούσε να αποφύγει, γνωρίζοντας την τόσο ισχυρή στήριξη των ψηφοφόρων. Αλλά οι συνέπειες από σφάλμα 1^{ου} είδους θα είναι σοβαρότερες από εκείνες ενός σφάλματος 2^{ου} είδους.

Για το λόγο αυτό ορίζουμε την $\rho \leq 0.60$ σαν H_0 και την $\rho > 0.60$ σαν εναλλακτική (H_1) της H_0 . Άρα, σαν H_0 θα παίρνουμε την υπόθεση εκείνη, που αν απορριφτεί ενώ είναι αληθινή (σφάλμα 1^{ου} είδους) συνεπάγεται κατά πολύ σοβαρότερες συνέπειες από εκείνες του να γινόταν δεκτή ενώ ήταν ψεύτικη (π.χ. κατά τον έλεγχο ενός νέου φαρμάκου υπάρχουν 2 υποθέσεις «Το φάρμακο είναι τοξικό» και «Το φάρμακο δεν είναι τοξικό». Αυτή που πρέπει να πάρουμε σαν H_0 είναι η 1^η γιατί αν είναι αληθινή και την απορρίψουμε, οι συνέπειες ασφαλώς θα είναι σοβαρότερες από ότι αν ενώ είναι ψέματα τη δεχτούμε, διότι στην 1^η περίπτωση θα έχουμε πολλούς θανάτους αρρώστων, ενώ στη 2^η περίπτωση οικονομική ζημιά ή και απομάκρυνση του φαρμάκου. Στην περίπτωση που και τα 2 είδη είναι της ίδιας σπουδαιότητας, μπορούμε να εκλέξουμε ως μηδενική (H_0) οποιαδήποτε από τις δύο.

Για να επανέλθουμε στο παράδειγμα, η εξέταση του παραπάνω κανόνα απόφασης (δηλαδή για ≥ 20 και με την H_0 να απορρίπτεται αν και μόνο αν η τιμή της $X \geq 13$) μας οδηγεί στον ορισμό μίας συνάρτησης π που η τιμή της για κάθε δυνατή τιμή της παραμέτρου ρ , δίνει την πιθανότητα να απορρίπτεται η H_0 . Δηλαδή είναι $\pi(\rho) = P(X \geq 13) = \sum b(k/20, \rho)$. Τη συνάρτηση π ονομάζουμε **δυναμοσυνάρτηση** του κανόνα απόφασης. Από τον πίνακα διωνυμικής κατανομής παίρνομε το μέρος που χρειαζόμαστε για να ελέγξουμε τις τιμές της $\pi(\rho)$: –

ρ	0	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1
$\pi(\rho)$	0	0+	0+	0.001	0.021	0.132	0.416	0.772	0.968	1-	1

Όπου (0+) αριθμός θετικός και μικρότερος του 0.0005 και (1-) αριθμός μεταξύ του 0.9995 και του 1. Η δυναμοσυνάρτηση αυτή, έχει διάγραμμα στο οποίο σημειώνουμε και δύο ευθύγραμμα τμήματα ΟΑ, ΒΓ που είναι το διάγραμμα της δυναμοσυνάρτησης ενός ιδανικού κανόνα απόφασης, δηλαδή ενός κανόνα που οι πιθανότητες σφαλμάτων 1^{ου}, 2^{ου} είδους είναι μηδέν. Σε τέτοια διαγράμματα βλέπουμε ότι $\pi(\rho)=0$ όταν η H_0 είναι Α και $\pi(\rho)=1$, όταν η H_0 είναι γ. Οπότε, $\rho \leq 0.6$, η διαφορά ύψους του πραγματικού διαγράμματος είναι η πιθανότητα σφάλματος 1^{ου} είδους και άρα όταν $0 < \rho \leq 0.6$, η πιθανότητα σφάλματος 1^{ου} είδους αυξάνει από 0 μέχρι 0.416 αλλά και $\rho > 0.6$, η διαφορά ύψους των δύο διαγραμμάτων θα δίνει την πιθανότητα σφάλματος 2^{ου} είδους και άρα όταν $1 > \rho > 0.6$, η πιθανότητα για σφάλμα 2^{ου} είδους αυξάνει από 0 μέχρι της τιμής $1 - 0.416 = 0.584$.

Από τα παραπάνω βγάζομε το συμπέρασμα ότι «Η επιτροπή δεν πρέπει να είναι ικανοποιημένη από τον κανόνα αυτό, της απόφασης, γιατί και αν ακόμη ο κ. Α έχει με το μέρος του το 50% των ψηφοφόρων, η πιθανότητα να υπάρξουν 13 άτομα υπέρ αυτού από τα 20 του δείγματος είναι 0.132, πράγμα που θα οδηγήσει την επιτροπή σε άτονη εξόρμηση, ενώ είναι φανερό ότι χρειάζεται έντονος αγώνας και το ίδιο θα συμβαίνει για οποιοδήποτε ποσοστό μικρότερο μέχρι πολύ κοντά στο 60%».

Ακόμα και στην περίπτωση που ο κ. Α υποστηρίζεται από το 70% των ψηφοφόρων, πάλι η πιθανότητα θα είναι $1 - 0.772 = 0.228$ ότι στο δείγμα $n=20$ θα είναι λιγότεροι από 13 ψηφοφόροι με το μέρος του κ. Α. Άρα, το πρόβλημα της επιτροπής

εντοπίζεται στο να βρεθεί κανόνας απόφασης τέτοιος, που και τα δύο προηγούμενα σφάλματα να γίνουν μικρότερα. Διατηρώντας το μέγεθος του δείγματος το ίδιο $n=20$, οι κανόνες απόφασης που θα μπορούσε να έχει δεχθεί η επιτροπή του δείγματος ίδιο $n=20$, οι κανόνες απόφασης που θα μπορούσε να δεχθεί η επιτροπή είναι « H_0 απορρίπτεται αν και μόνο αν ο X (ο αριθμός των ψηφοφόρων υπέρ του κ. Α) είναι τουλάχιστον κάποιος ορισμένος αριθμός, έστω σ ». Κάθε εκλογή άρα του σ καθορίζει και έναν κανόνα απόφασης και έχουμε ήδη εξετάσει αυτόν με $\sigma=13$.

Η δυναμοσυνάρτηση $\pi(\rho)=P(X \geq \sigma)=\sum_{k=\sigma}^n b(k/20, \rho)$ είναι ο γενικός τύπος για όλες τις δυναμοσυναρτήσεις ($V\sigma = \leq X$) κάθε κανόνα απόφασης. Χρησιμοποιώντας και πάλι τον πίνακα της διωνυμικής κατανομής:

ρ		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\pi(\rho)$	$\sigma=15$	0	0+	0+	0+	0.002	0.021	0.126	0.416	0.804	0.989	1
	$\sigma=16$	0	0+	0+	0+	0+	0.006	0.051	0.238	0.630	0.857	1
	$\sigma=17$	0	0+	0+	0+	0+	0.001	0.016	0.107	0.411	0.867	1

Από τον πίνακα βλέπουμε ότι όταν το σ αυξάνει, η πιθανότητα σφάλματος 1^{ου} είδους ελαττώνεται για όλα τα $\rho \leq 0.6$, δηλαδή τα διαγράμματα κατεβαίνουν και πλησιάζουν το ιδανικό διάγραμμα όπου η πιθανότητα του σφάλματος είναι μηδέν. Συγχρόνως όμως και η πιθανότητα σφάλματος 2^{ου} είδους ανεβαίνει άρα τα διαγράμματα απομακρύνονται από το ιδανικό για όλα τα $\rho > 0.6$.

Έτσι, η επιτροπή παίρνει απάντηση αρνητική για την ταυτόχρονη ελάττωση και των δύο σφαλμάτων, με δείγμα μεγέθους $n=20$. Άρα, στην περίπτωση αυτή, η προσοχή μας συγκεντρώνεται στα σφάλματα του 1^{ου} είδους, που είναι πιο σοβαρά από αυτά του 2^{ου} είδους. Η επιτροπή, εκλέγει έναν αριθμό που συνήθως παριστάνεται με το (α) και που αυτός κάνει την πιθανότητα ενός σφάλματος, μέγιστη, που η επιτροπή θα ανεχθεί. Στις συνηθισμένες εφαρμογές, ο (α) είναι ένας από τους αριθμούς 0.01, 0.05, 0.10. Μετά τον προσδιορισμό του α , καταλήγουμε στον αντίστοιχο κανόνα απόφασης ο οποίος όχι μόνο θα υπακούσει στο ότι η μέγιστη πιθανότητα σφάλματος 1^{ου} είδους θα είναι $\leq \alpha$, αλλά και θα οδηγήσει και στις κατώτατες δυνατές πιθανότητες σφαλμάτων 2^{ου} είδους.

Έστω ότι η επιτροπή εκλέγει $\alpha=0.02$. Ζητάμε από το διάγραμμα, την τιμή του σ που το ύψος της δυναμοκαμπύλης για $\rho=0.60$ (μέγιστη πιθανότητα σφάλματος 1^{ου} είδους), δεν υπερβαίνει την $\alpha=0.02$ αλλά είναι πολύ κοντά στην $\alpha=0.02$. Από το σχήμα ή από τον πίνακα βρίσκουμε ότι αυτή είναι η $\sigma=17$, που οδηγεί και στον αντίστοιχο κανόνα απόφασης. Δηλαδή, η επιτροπή καθορίζει την τιμή του X από το δείγμα $n=20$ και απορρίπτει την H_0 αν και μόνο αν $X \geq 17$. Έτσι, υπάρχει η βεβαιότητα στην επιτροπή ότι το πολύ 2% είναι ενδεχόμενο να γίνει σφάλμα 1^{ου} είδους (όχι έντονος αγώνας) παρόλο ότι υπάρχει μεγάλη πιθανότητα η επιτροπή να κάνει σφάλμα 2^{ου} είδους (οικονομική επιβάρυνση, όχι απαραίτητη). Αν τώρα πάρουμε την περίπτωση, που η επιτροπή δε θέλει να έχει άνοιγμα σφάλματος 2^{ου} είδους 2%, από τον πίνακα έχουμε $P(X \geq 17)=0.107$ όταν $\rho=0.70$. Δηλαδή, κατά 90% περίπου θα είναι αποδεκτή η H_0 (άσκοπη οικονομική δαπάνη) ακόμα και αν ο κ. Α έχει με το μέρος του το 70% των ψηφοφόρων.

Τι μπορεί λοιπόν να γίνει ώστε να διατηρηθεί μεν η μέγιστη πιθανότητα σφάλματος 1^{ου} είδους για $\alpha=0.02$ και συγχρόνως να ελαττωθεί ο κίνδυνος σφαλμάτων του 2^{ου} είδους; Επειδή από τα προηγούμενα συμπεραίνουμε ότι το δείγμα $n=20$ είναι μικρό για το σκοπό μας, παίρνουμε δείγματα μεγέθους $n=50$, $n=100$ και $n=300$ οπότε η πορεία απόφασης μας διατυπώνεται γενικά ως εξής «Παίρνουμε δείγμα μεγέθους n από τον πληθυσμό μας (ψηφοφόροι). Αν X η τυχαία μεταβλητή (η τιμή της είναι ο

αριθμός των υπέρ του κ. Α ψηφοφόρων) με H_0 ($\rho \leq 0.60$) να απορρίπτεται αν και μόνο αν $X > \sigma$ (έστω $\alpha = 0.02$), οι πιθανότητες σφαλμάτων 2^{ου} είδους να είναι όσο γίνεται πιο μικρές». Όπως και πριν, για να ορίσουμε το σ , παίρνουμε πρώτα $\rho = 0.60$, δηλαδή την τιμή που η πιθανότητα σφάλματος 1^{ου} είδους γίνεται μέγιστη. Οποιοδήποτε σ για τον οποίο $P(X \geq \sigma) = \sum b(k/\eta, 0.60) \leq \alpha$, (1) ορίζει κανόνα απόφασης, που η μέγιστη πιθανότητα σφάλματος 1^{ου} είδους είναι το πολύ α . Για να ελαττώσουμε επίσης όσο μπορούμε περισσότερο την πιθανότητα να κάνουμε σφάλμα 2^{ου} είδους, εκλέγουμε τον σ τέτοιον που να είναι ο μικρότερος αριθμός–στοιχείο στο σύνολο.

Το σύνολο αυτό που περιέχει τις τιμές της X , για τις οποίες απορρίπτεται η H_0 , ονομάζεται **κριτικό σύνολο** των τιμών της X για τον αντίστοιχο κανόνα απόφασης. Χρησιμοποιώντας τους πίνακες της διωνυμικής κατανομής με το $\alpha = 0.02$ βρίσκουμε ότι είναι:

$\sigma = 17$ για $n = 20$,
 $\sigma = 38$ για $n = 50$,
 $\sigma = 71$ για $n = 100$ και
 $\sigma = 198$ για $n = 300$.

Άρα, έχουμε 4 κανόνες απόφασης με $\alpha = 0.02$ που οι τιμές της δυναμοσυνάρτησης αυτών δίνονται από τον παρακάτω πίνακα.

ρ		0,5	0,6	0,7	0,8	0,9	1
$\Pi(\rho)$	η						
	20	17	0.001	0.016	0.107	0.411	0.867
	50	38	0+	0.013	0.223	0.814	0.999
	100	71	0+	0.015	0.462	0.989	1
	300	19	0+	0.019	0.941	1	1

Όπως βλέπουμε, όταν αυξάνει το δειγματικό μέγεθος, ελαττώνεται ο κίνδυνος του σφάλματος 2^{ου} είδους. Δηλαδή, όταν αυξάνεται το n , $\rho > 0.60$ οι «καμπύλες» ανεβαίνουν και τείνουν προς το ιδανικό διάγραμμα που η πιθανότητα για σφάλμα 2^{ου} είδους είναι μηδέν. Στον τελευταίο πίνακα, βλέπουμε ότι για $n = 300$ η πιθανότητα να βρεθούν 198 ψηφοφόροι υπέρ του κ. Α όταν $\rho = 0.70$ είναι 0.941.

Έτσι, η πιθανότητα σφάλματος 2^{ου} είδους ελαττώνεται κατά 0.059. Έτσι, βλέπουμε με ποιο τρόπο η επιτροπή, μπορεί να διατηρήσει μεν την μέγιστη ανεκτή, για αυτήν, πιθανότητα σφάλματος 2^{ου} είδους με $\alpha = 0.02$ και συγχρόνως να μπορεί να ελέγχει τους κινδύνους από σφάλματα 2^{ου} είδους παίρνοντας από όλο τον πληθυσμό (όλοι οι ψηφοφόροι) ένα αρκετά μεγάλο δείγμα (πρόσωπα).

Για να τελειώνουμε δε την διερεύνηση του προβλήματος και για να μπορούμε να χρησιμοποιήσουμε τους πίνακες της διωνυμικής κατανομής, ξαναερχόμαστε στον απλό κανόνα απόφασης με $n = 20$ και $\sigma = 17$. Επειδή ο σ έχει εκλεγεί πάρα πολύ μικρός αριθμός, στο σύνολο που ορίζεται από την παραπάνω περιγραφή (2) γνωρίζουμε:

$$P(X \geq \chi) = \sum b(k/20, 0.60) \leq 0.02, \chi \geq 17, \quad (3)$$

$$\text{και } P(X \geq \chi) = \sum b(k/20, 0.60) > 0.12, \chi < 17. \quad (4)$$

Άρα, μπορούμε να συμπεράνουμε ότι από πρακτικής πλευράς δεν είναι απαραίτητο να εργαστούμε με αυτόν ακριβώς τον τρόπο, στην περίπτωση που μοναδικός σκοπός είναι να αποφασίσουμε αν η μαρτυρία που προκύπτει από το πείραμα μας οδηγεί στην αποδοχή ή στην απόρριψη της H_0 (μηδενικής υπόθεσης). Και αυτό, παρά το ότι ο τρόπος αυτός βοηθά πάρα πολύ να καταλάβουμε τη μέθοδο ελέγχου των υποθέσεων, του καθορισμού των κανόνων, των αποφάσεων και δυναμοσυναρτήσεων. Το αποτέλεσμα του πειράματος όταν $n = 20$, κατά την πραγματοποίηση του συμβάντος $X = \chi$, όπου $0 \leq \chi \leq 20$, με $\chi = 0$ όταν κανένας ψηφοφόρος δεν είναι υπέρ του κ. Α και $\chi = 20$ όταν και οι 20 ψηφοφόροι είναι υπέρ

του κ. Α. Όσο μεγαλύτερη είναι η τιμή του X , τόσο δυσμενέστερο θα είναι το αποτέλεσμα του πειράματος μας για την H_0 ότι $\rho \leq 0.60$. Ο αριθμός $P(X \geq \chi)$ όταν υπολογίζεται για την οριακή τιμή $\rho = 0.60$, είναι η πιθανότητα να πάρουμε τιμή του X που να είναι το λιγότερο τόσο δυσμενής για την H_0 , όσο και αυτή που εμφανίστηκε στην πραγματικότητα και λέγεται **στατιστική σημαντικότητα** ή **περιγραφική στάθμη σημαντικότητας** του συμβάντος $X = \chi$. Από τους τύπους (3), (4) συμπεραίνουμε ότι, αν η στάθμη σπουδαιότητας του $X = \chi$ είναι κατώτερη ή ίση της τιμής $\alpha = 0.02$, τότε η H_0 απορρίπτεται επειδή θα πρέπει ο $\chi \geq 17$. Αν η περιγραφική στάθμη σπουδαιότητας του γεγονότος $X = \chi$ είναι μεγαλύτερη του $\alpha = 0.02$ τότε η μηδενική υπόθεση H_0 είναι αποδεκτή επειδή ο $\chi < 17$.

Η τιμή της X που οδηγεί στην απόρριψη της H_0 λέγεται **σημαντική** στην στάθμη $\alpha = 0.02$ ή στην 2% στάθμη σημαντικότητας. Η τιμή της X , που οδηγεί στην αποδοχή της H_0 δεν είναι σημαντική στην στάθμη $\alpha = 0.02$. Άρα, **έλεγχος της σημαντικότητας** της τιμής που παρατηρήσαμε για την X στη στάθμη α (δηλαδή, υπολογισμός της $P(X \geq \chi)$ για $\rho = 0.60$ και σύγκριση της με την α) είναι ο τρόπος εκείνος που καθορίζει τη δραστηριότητα που θα πρέπει να αναλάβουμε χωρίς να βρούμε πρώτα τον κανόνα απόφασης και τη δυναμοσυνάρτηση του. Για αυτό και ο έλεγχος στατιστικής υπόθεσης ονομάζεται και **έλεγχος σημαντικότητας**.

Σαν παράδειγμα, ας υποθέσουμε ότι η επιτροπή αποφάσισε να πάρει δείγμα μεγέθους $n = 20$ ψηφοφόρων και με τιμή $\alpha = 0.02$ σαν μέγιστη ανεκτή από αυτή πιθανότητα για σφάλμα 1^{ου} είδους. Μετά από την έρευνα, προκύπτει το συμβάν $X = 16$ (16 ψηφοφόροι υπέρ του κ. Α). Από τους πίνακες της διωνυμικής κατανομής όταν $\rho = 0.60$ βρίσκουμε $P(X \geq 16) = 0.051$. Επειδή είναι $0.051 > 0.02$ το συμβάν $X = 16$ δεν είναι σημαντικό στη 2% στάθμη σημαντικότητας (ή σπουδαιότητας) και άρα η επιτροπή αποδέχεται τη μηδενική υπόθεση H_0 . Αξίζει να παρατηρήσουμε ότι αν η επιτροπή είχε πάρει $\alpha = 0.06$ τότε η ίδια τιμή της X ($= 16$) θα ήταν σημαντική στην 6% στάθμη σημαντικότητας (επειδή $0.051 < 0.06$) και τότε η επιτροπή θα απέρριπτε την H_0 . Αυξάνοντας η επιτροπή το α , αυξάνει και τις δυνατότητες της να απορρίψει την μηδενική υπόθεση. Αυτό επίσης αυξάνει και τις δυνατότητες για σφάλμα 1^{ου} είδους.