

**Handout: Problem Set #2 Solutions**

***PREDICT 401: Introduction to Statistical Analysis***

These problem sets are meant to allow you to practice and check the accuracy of your work. Please do not review the solutions until you have finalized your work. Although these problem sets are not submitted and graded, treat them as if they were. It is to your great benefit to work on and even struggle with the problem sets. Looking at the solutions before finalizing your work will, quite simply, make for a less meaningful learning experience.

**1. Shelter Demographics**

A survey of the homeless shelters in your city asked people where they had stayed the night before. In your shelter, the results were as follows:

Men (150 people) Women (70 people)

Street 40% Street 28.57%

Friend 15.33% Friends 18.57%

Family 14.0% Family 17.14%

Own Apartment 12.0% Own Apartment 14.3%

Other Institution 18.67% Other Institution 21.43%

Assume that each person's decision about where to spend the night is independent of all the others. Based on this information, answer the following questions.

1. What is the probability that a person in your shelter is a man?

**P(Man) = 150/220 = 0.682**

1. What is the probability that a person in your shelter is a man who spent last night on the street?

**We can restate this question in probability terms as:  
“What is the probability that a person 1) is a man and 2) spent last night on the street?”**

**P(Man and Street) = P(Man) \* P(Street given that the person is a Man) = P(Man) \* P(Street | Man)**

**So P(Man and Street) = 0.682\*0.40 = 0.273**

**Alternatively, we could calculate the number of men who spent the night on the street (.4\*150) and divided that by the total number of people (220):**

**P(Man and Street) = (0.4\*150)/220 = 0.273**

1. What is the probability that a person in your shelter spent last night on the street?

**Answering this question requires that we use both the “and” and “or” rules. We want P(Street). You could have spent last night on the street if you spent last night on the street and are a man OR if you spent last night on the street and are a woman.**

**P(Street) = P(Street and Man OR Street and Woman)**

**For OR, recall:**

**P(A or B) = P(A) + P(B) – P(A and B)**

**Since a person cannot be both a man and woman, the possibilities here are mutually exclusive. Thus, P(A and B) = 0.**

**Above, we found that P(Street and Man) = 0.273. Also, P(Woman) = 1-P(Man) = 0.318.**

**[Alternatively, P(Woman) = 70/220 = 0.318]**

**P(Street and Woman) = P(Woman)\*P(Street | Woman) = 0.318\*0.2857 = 0.091**

**Now, we combine the probabilities for the two groups who spent the night on the street:**

**P(Street) = P(Street and Man) + P(Street and Woman) = 0.273+0.091 = 0.364**

**Alternatively, you could have calculated the total number of people spending the night on the street: P(Man and Street) = 0.4\*150 = 60; P(Woman and Street) = (0.286\*0.70) = 20. Then add these and divide by the total number of people: (60+20)/220 =80/220 = 0.364**

1. What is the probability that a person in your shelter is either a man or someone who spent last night on the street?

**This is an “either or both” probability. We want P(Man or Street).**

**P(A or B) = P(A) + P(B) – P(A and B)**

**P(Man or Street) = P(Man) + P(Street) – P(Man and Street)**

**Here A and B aren’t mutually exclusive (someone on the street can be a man). This means that P(Man and Street) exists. We already calculated these values, so just plug them in:**

**P(Man or Street) = 0.682+0.364-0.273 = 0.773**

1. What is the probability that a woman in your shelter either spent last night in her own apartment or with her family (not in her own apartment)?

**Note that this is only about women here. So, of the women, we want P(Apt or Family).**

**P(Apt or Family) = P(Apt) + P(Family) – P(Apt and family).**

**But “apt” and “family” are mutually exclusive (P(Apt and family) = 0)**

**So, of women:**

**P(Apt or Family) = 0.143+0.171 = 0.314**

1. Three people come into your shelter. What is the probability that none of them spent last night on the street?

**What is probability of P(not street and not street and not street) = P(A and B and C)? These are independent events, so P(not street and not street and not street) = P(not street) \* P(not street) \* P(not street). We previously found that P(street) = 0.364. Thus, P(not street) = 1-0.364 = 0.636. So, P(not street)\* P(not street)\* P(not street)= 0.636\*0.636\*0.636 = 0.26**

**2. Accounting for Absences**

You are staffing a day program at a Community Mental Health Center. Suppose that prior experience shows that on any given day the probability of at least one client being absent from the program (which takes place on both weekdays and weekends) is 0.10. Assume that the probability of at least one client being absent is the same for every day regardless of what happens on the other days. Based on this information, answer the following questions.

1. What is the probability that no one is absent during the first four days of a week?

**On each day, the probability that at least one person is absent is 0.10. On any given day, either someone is absent or no one is absent. There are no other possibilities.**

**Here, “A” denotes absent, “” denotes no one absent (the complementary event).**

**P(A) = 0.1; P() = 0.9**

**P( and  and  and ) = 0.9\*0.9\*0.9\*0.9 = 0.656**

1. What is the probability that Thursday is the first day that someone is absent? (Assume that the week begins on a Monday.)

**P(Mon and Tues and Weds and AThurs) = 0.9\*0.9\*0.9\*0.1 =0.073**

1. What is the probability that, during one week, someone is absent on Tuesday and someone is absent on Thursday but there are no absences the other days?

**P(monday and Atuesday and wednesday and Athursday and friday and saturday and sunday)**

**= 0.9\*0.1\*0.9\*0.1\*0.9\*0.9\*0.9 =0.006**

1. What is the probability that in a week (7 days), there is only one day that someone is absent?

**There are seven different ways we can get this outcome:**

**Amonday and tuesday and wednesday and thursday and friday and saturday and sunday = 0.96 \* 0.1**

**monday and Atuesday and wednesday and thursday and friday and saturday and sunday = 0.96 \* 0.1**

**monday and tuesday and Awednesday and thursday and friday and saturday and sunday = 0.96 \* 0.1**

**monday and tuesday and wednesday and Athursday and friday and saturday and sunday = 0.96 \* 0.1**

**monday and tuesday and wednesday and thursday and Afriday and saturday and sunday = 0.96 \* 0.1**

**monday and tuesday and wednesday and thursday and friday and Asaturday and sunday = 0.96 \* 0.1**

**monday and tuesday and wednesday and thursday and friday and saturday and Asunday = 0.96 \* 0.1**

**There are seven mutually exclusive outcomes that meet the criteria set forth above. The probability of each is 0.053**

**Since the probability of A or B for mutually exclusive events is P(A or B) = P(A) + P(B), then we can add the probabilities for each:**

**P(one absence in 7 days) = 7 (0.053) = 0.371**

**NOTE: Later in the course we will learn a more concise way to arrive at this answer via the BINOMIAL PROBABILITY DISTRIBUTION: P(one absence in 7 days) = 7C1 \* 0.96 \* 0.1 = 7 (0.053) = 0.371**

1. Consider the first two days. What is the probability of absences on both days? What is the probability that someone is absent on one or both days?

**The probability that someone is absent on both days = P(A and A) = .1\*.1 = .01**

**The event that someone is absent on one or both days can occur in three ways (outcomes):**

**A and  = .1\*.9 = .09**

** and A = .9\*.1 = .09**

**A and A = .1\*.1 = .01**

**These 3 outcomes are mutually exclusive, so we use the “OR” rule for mutually exclusive outcomes:**

**P(A or B or C) = P(A) + P(B) + P(C) = 0.09+0.09+0.01 = 0.19. We’re lucky here that the outcomes are in fact mutually exclusive. With 3 or more outcomes that are NOT mutually exclusive, we have to work quite a bit harder with a formula that gets more elaborate and unwieldy as the number of outcomes grows.**

**Alternatively, we can use the Complement Rule: P(absence on one or both days) = 1-P(no absence on both days) = 1 – 0.92 = 1 – 0.81 = 0.19. That’s pretty slick!**

**3. STDs among high-risk pregnant women**

Suppose a survey of the incidence of sexually transmitted diseases (STDs) among pregnant women in a high risk clinical setting found that disease A (STDA) occurred in 10% of the sample and that disease B (STDB) occurred among 5% of women in the sample. In addition, suppose that among those women with STDA, 30% had STDB.

1. Are A and B mutually exclusive? Why or why not?

**Since some people have both, they are not mutually exclusive.**

1. Are A and B independent? Why or why not?

**No, they aren’t. If they were independent, then P(B) would equal P(B | A) would equal P(B | ).**

**Here, though, P(B) = 0.05 and P(B | A) = 0.30. So, they are not independent. In other words, the probability of one outcome is influenced by the other outcome.**

Suppose we choose a woman randomly from the survey participants.

1. What is the probability she has both diseases?

**P(A and B) = P(A) \* P(B | A) since they are not independent.**

**Here, P(A and B) = 0.1\*0.3=0.03.**

**Note that without a value for P(A | B) in the problem set-up, we cannot use the alternative of**

**P(A and B) = P(B) \* P(A | B)**

1. What is the probability that she has neither disease?

**P(neither) = P( and ) = 1 – P(A or B). (\*\*Note that this is NOT 1 – P(A and B)\*\*)**

**P(A or B) = P(A) + P(B) – P(A and B) since they aren’t mutually exclusive.**

**P(A or B) = 0.1+0.05-0.03 = 0.12**

**P(neither) = 1-0.12 = 0.88**

**4. Employee Data**

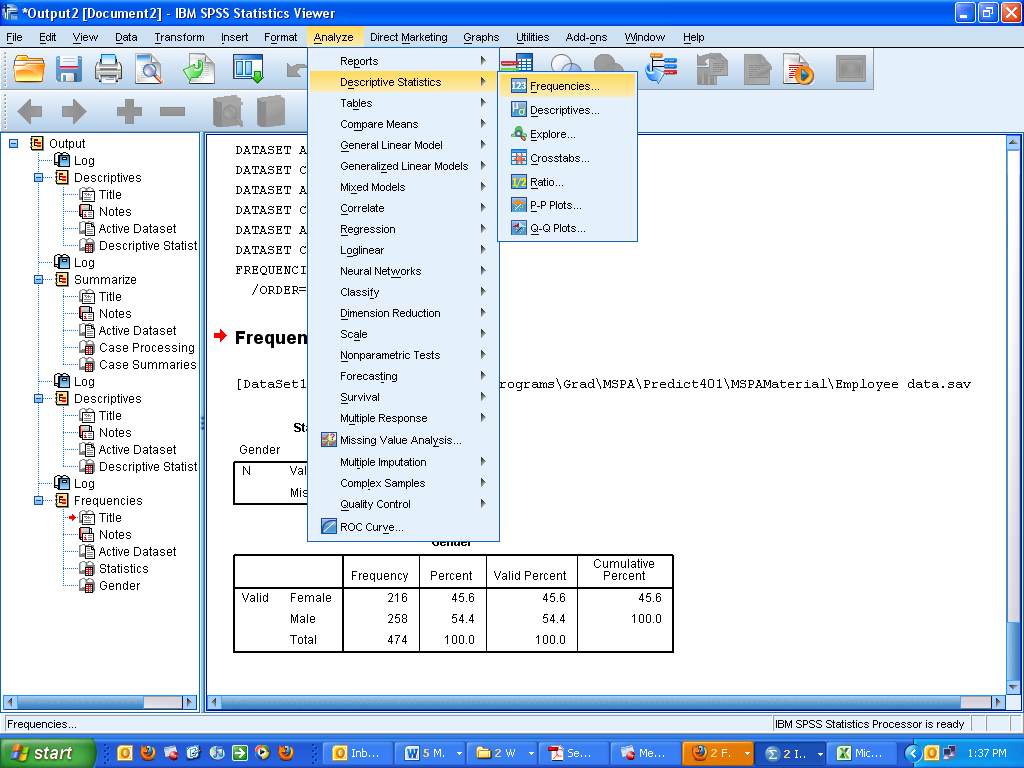
For this problem, use the SPSS database called “Employee data.sav.” Assume that this is a set of data on all employees in a particular company.

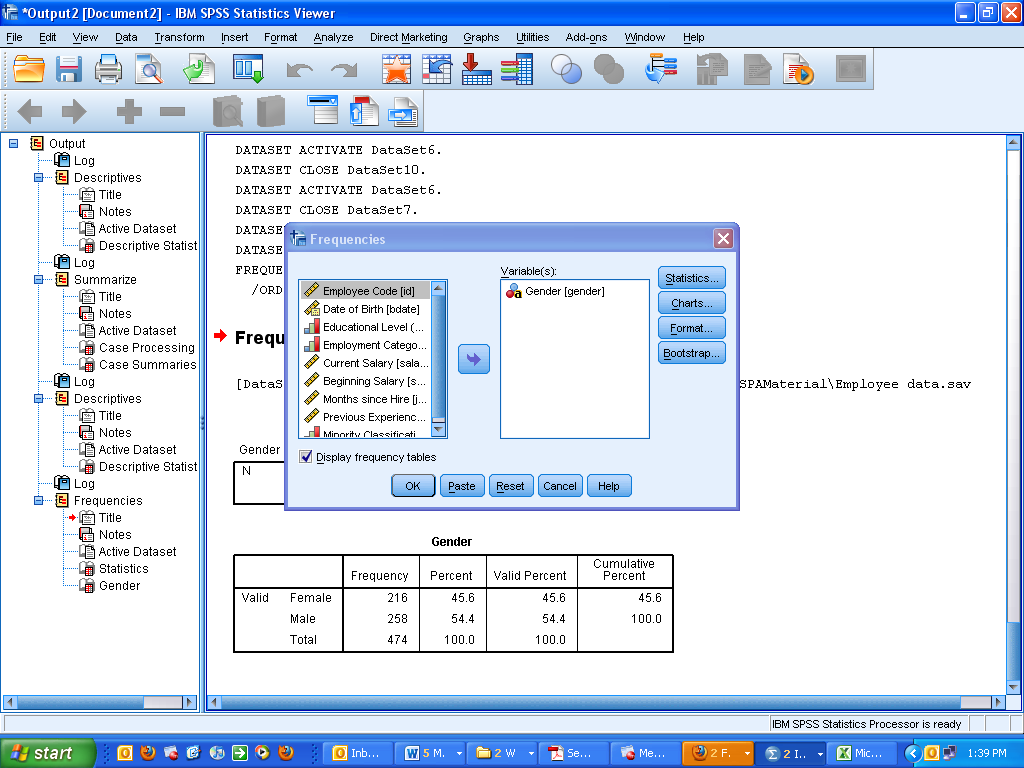
1. How many total observations comprise this data set?

**Simply scrolling to the bottom of the data view shows 474 total observations.**

1. If a person were randomly selected from this sample, what are the chances of selecting a male?

**Using the Descriptive Statistics\Frequencies tool with “gender” as the relevant variable, we can see that 54.4% of the observations are male. Thus, the chances are 54.4% that a person selected at random is male.**

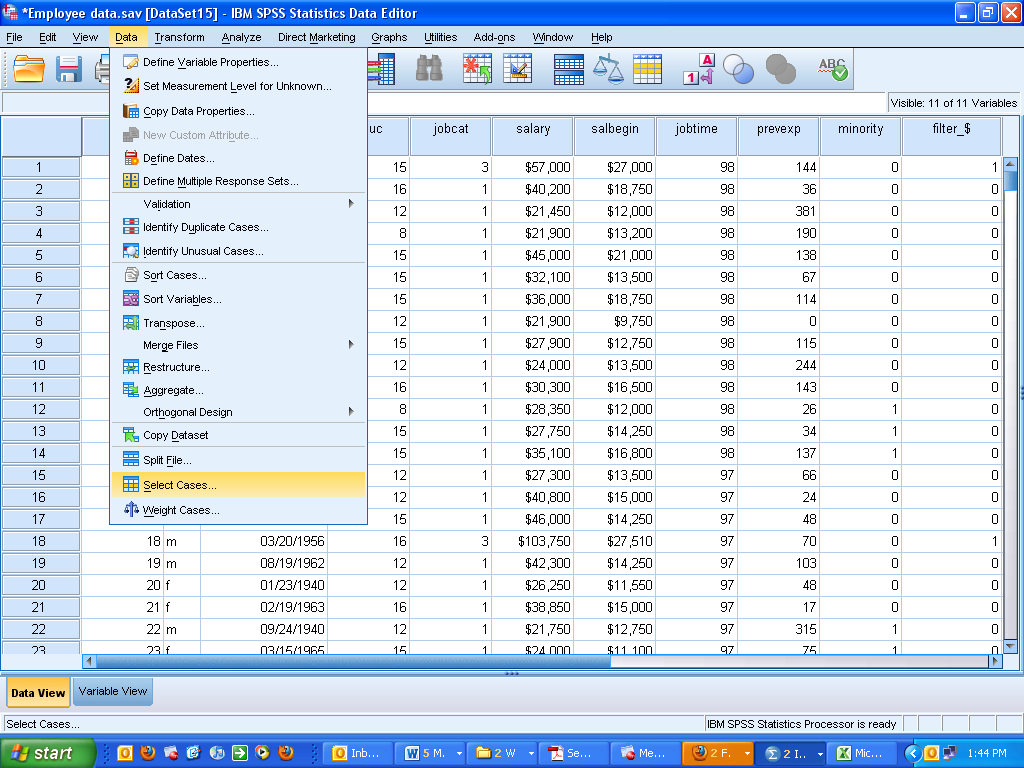
****

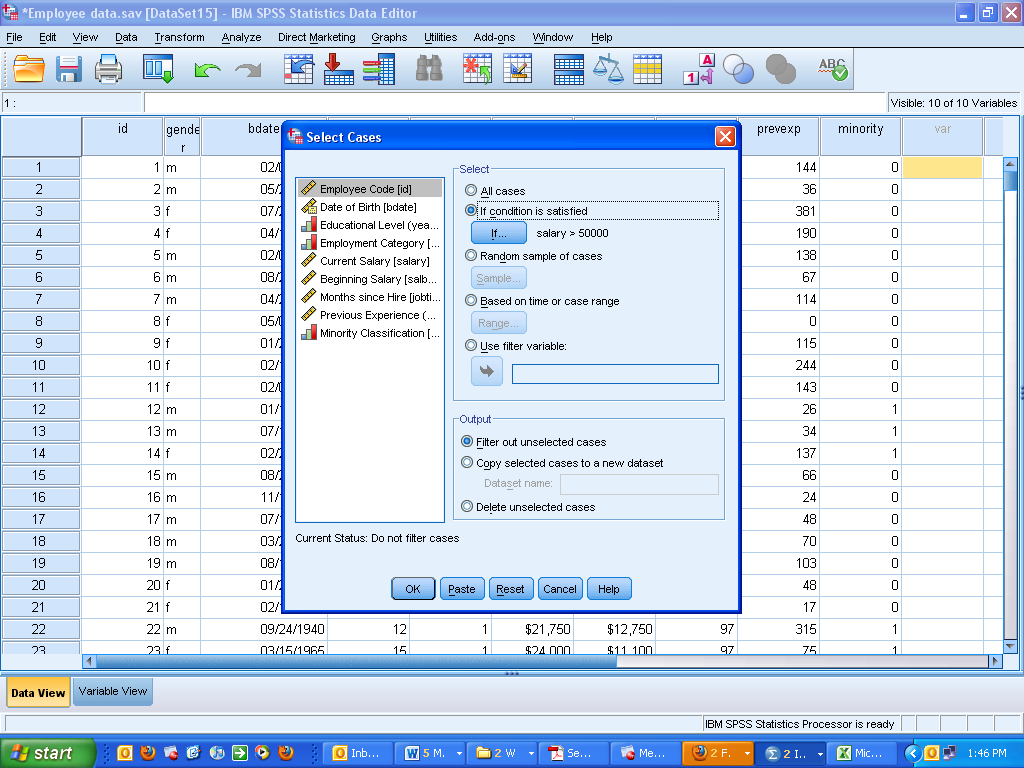
****

| **Gender** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Female | 216 | 45.6 | 45.6 | 45.6 |
| Male | 258 | **54.4** | 54.4 | 100.0 |
| Total | 474 | 100.0 | 100.0 |  |

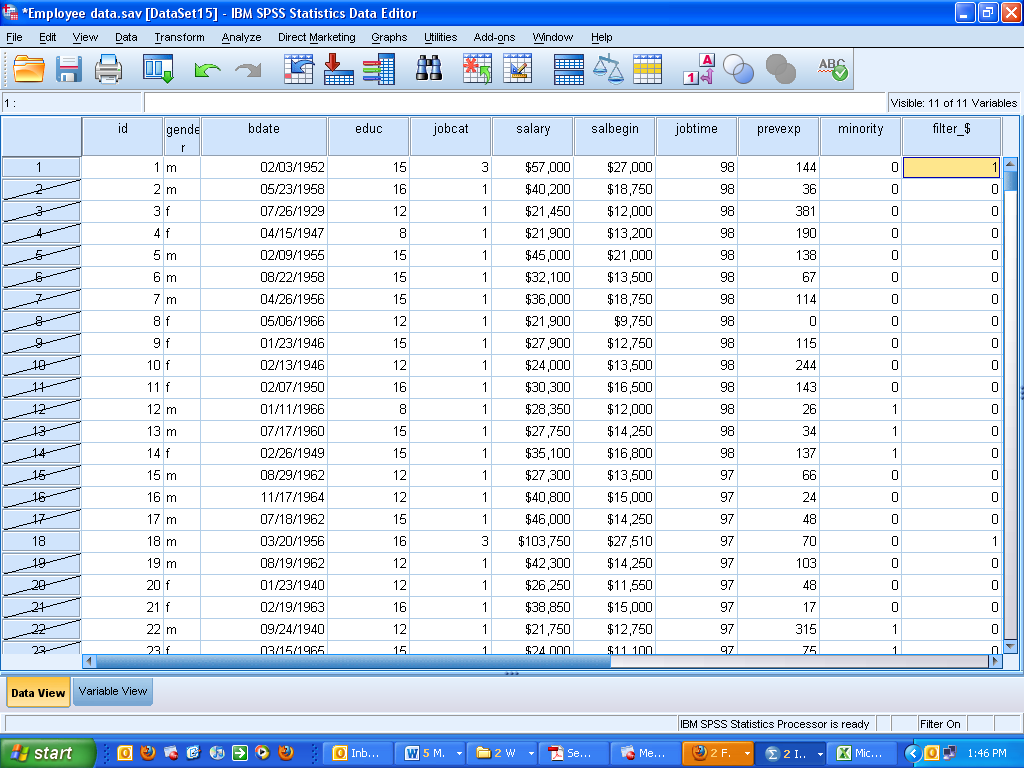
1. Of those who make more than $50,000, what are the chances that someone chosen at random would be male?

**First, we must filter out all cases that fall at $50,000 or below by using the Data\Select Cases function and setting salary > $50,000. Be sure to only “filter” the unselected cases, under ‘Output’. Don’t delete them unless you never want them back!**





**Note what happens in the resultant data set below. Observations with salaries at $50,000 or less are shown with a line through them at the far left. Plus, there’s a new variable on the far right—a “filter” variable showing a 1 for all selected cases and a 0 for all unselected cases.**



**Now, we can run the same analysis as in part (b). We’re just running it on a subset of observations. The SPSS output is:**

| **Gender** | | | | | |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Female | 6 | 8.5 | 8.5 | 8.5 |
| Male | 65 | 91.5 | 91.5 | 100.0 |
| Total | 71 | 100.0 | 100.0 |  |

**Thus, we see that a person chosen at random from a sample of those making more than $50,000 has a 91.5% chance of being male.**