STATISTICS 211
Tutorial 6
Solutions

1. Beth wants to see if a die is weighted. She records the number of times that a number comes up out of 60 throws. She records the following data:

Value	frequency
1	12
2	8
3	15
4	9
5	10
6	6

Is the die weighted? Carry out the appropriate test procedure.

Ho: the die is fair
Ha: the die is not fair

Expected frequency = (1/6)60=10

$$\chi^2 = sum\ of\ \frac{(obs\,freq - \exp\ freq)^2}{\exp freq}$$

$$\chi^2 = \frac{(12-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(15-10)^2}{10} + \frac{(9-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(6-10)^2}{10} = 5$$

df=6-1=5
between 30% and 50% chance of observing these difference or differences more extreme. We can conclude that the die is fair because the differences between what we observe and what we expect is not significant.

2. An official of a plastics industry claimed that the industry employed 30% white women, 5% minority women, 50% white men, and 15% minority men. To test the claim, an affirmative action committee randomly sampled 150 employees and obtained the following information:

Category	observed
White females	40
Minority females	15
White males	80
Minority males	15

Test the official's claim.
Ho: 30% of white women, 5% minority women, 50% white male, 155 minority men
Ha: Ho is false
Total#= 150 expected frequencies: .3(15)=45, .05(150)= 7.5, .5(150) = 75, .15(150)=22.5

$$\chi^2 = \frac{(40-45)^2}{45} + \frac{(15-7.5)^2}{7.5} + \frac{(80-75)^2}{75} + \frac{(15-22.5)^2}{22.5} = 10.89$$

df=3 less than 5% or between 1% and 5%
There is less than a 5% chance of these difference behind due to random sampling. Few can conclude that the data does not support the plastic industry's claim.

3. A shipment of assorted nuts is labeled as having 45% walnuts, 20% hazelnuts, 20% almonds, and 15% pistachios. By randomly picking several scoops of nuts from this shipment, an inspector find the following counts.

	Walnuts	Hazelnuts	Almonds	Pistachios	Total
Counts	92	69	32	42	235

Could these findings be a strong basis for an accusation of mislabeling?

Ho: the bag contains 45% walnuts, 20% hazelnuts, 20% almonds and 15% pistachios
Ha: Ho is false
Expected frequencies: 235(.45)=105.75, 235(.2) = 47, 235(.15) =35.25

$$\chi^2 = \frac{(92-105.75)^2}{105.75} + \frac{(69-47)^2}{47} + \frac{(32-47)^2}{47} + \frac{(42-35.25)^2}{35.25} = 18.165$$

df=3 <1%

There is less than 1% chance of these differences being due to random sampling or chance error. We conclude that the company is guilty of false labeling based on the sample data.

4. A personnel administrator provided the following data as an example of hiring to fill 12 positions from among 40 male and 40 female applicants.

Applicant	Selected	Not Selected	Total
Male	7	33	40
Female	5	35	40

Does this sample indicate a selection bias in favour of males?

Ho: there is not selection bias
Ha: there is a selection bias.

Applicant	Selected	Not Selected	Total
Male	7 (6)	33 (34)	40
Female	5 (6)	35 (34)	40
	12	68	80

Expected freqy 40(12)/80= 6 68(4)/80 = 32

$$\chi^2 = \frac{(7-6)^2}{6} + \frac{(33-34)^2}{34} + \frac{(5-6)^2}{6} + \frac{(35-34)^2}{34} = .392$$

df=1
between 50% and 70% chance
There is between 50% and 70% chance of these difference being due to random sampling or chance error. We conclude that there is not selection bias.

5. Applicants for public assistance are allowed an appeals process when they feel unfairly treated. At such a hearing, the applicant may choose self-representation or representation by an attorney. The appeal may result in an increase, decrease, or no change of the aid recommendation. Court records of 320 appeals cases provided the following data.

	Amount of Aid		
Type of Representation	Increased	Unchanged	Decreased
Self	59	108	17
Attorney	70	63	3

Are the patterns of the appeals decision significantly different between the two types of representation?

Ho: there is no relationship between type of representation and aid recommendation
Ha: there is a relationship between type of representation and aid recommendation.

Amount of Aid

Type of Representation	Increased	Unchanged	Decreased	Total
Self	59 (74.2)	108 (98.3)	17 (11.5)	184
Attorney	70 (54.8)	63 (72.7)	3 (8.5)	136
	129	171	20	320

expected frequencies: 184(129)/320=72.6   184(171)/320=98.3…..

$$\chi^2 = \frac{(59-74.2)^2}{74.2} + \frac{(108-98.3)^2}{98.3} + \frac{(17-11.5)^2}{11.5} + \frac{(70-54.8)^2}{54.8} + \frac{(63-72.7)^2}{72.7} + \frac{(3-8.5)^2}{8.5}$$

= 15.7
df = 2   <1%
There is less than 1% chance of the differences being due to chance error. We conclude that there is a relationship between type of representation and aid recommendation.

6. avg yr of car = 5.3 SD = 1.3
   avg price of car =88 (in hundreds) SD = 29.7 (in hundreds)
   r= -.85
   a. Find the slope of the SD line
   We want to predict price (y) based on age of car (x)
   Slope SD line = -SDy/SDx = -29.7/1.3= -22.85

   b. If the age of the car is 7, find the price of the car that would fall on the SD line.
   1st find the # of SD 7 is away from its average

   $$z=\frac{(7-5.3)}{1.3}=1.31$$

Since this observation falls 1.31 SDs above the avg of x, then y must fall 1.31 SDs below its avg when the relationship is negative.
88 – 1.31(29.7) =49.09 (in hundred) or $4909

7. In a study of the stability of IQ scores, a large group of individuals is tested once at age 18 and again at age 35. The following results are obtained
   Age 18: average score = 100, SD=15
   Age 35: average score =100, SD=15,   r=0.8

   (a) Find the regression equation.
   It makes sense that we'd want to predict the IQ at 35 based on the IQ at 18.

   Slope = $\frac{r \times SDy}{SDx} = \frac{(.8)\times 15}{15} =$ **.8**
   Next find the y intercept when x=0 (IQ at 18)
   Such a person is 100 below the average IQ for 18 year olds.
   Her IQ would be 100(.8) = 80 below average for 35 year olds (since the relationship is positive).
   Avg y – amount below average = (100)- 80 = 20 (this is the y-intercept when x = 0)
   Equation of line: y=.8(x) + 20

   (b) Estimate the average score at age 35 for all the individuals who scored 115 at age 18.
   y=.8(115) + 20= 112

8. A computer program prints out the following data set shown below.

X	y	**x std. Units**	**y std. units**	**product of x and y std. units**
1	6	-1.5	1	(-1.5)(1) = -1.5
2	5	-1	.5	(-1)(.5) = -.5
3	7	-.5	1.5	(-.5)(1.5) =-.75
4	3	0	-.5	0 (-.5) = 0
5	4	.5	0	(.5) 0 = 0
6	1	1	-1.5	(1)(-1.5) = -1.5
7	2	1.5	-1	(1.5)(-1) = -1.5

avg x = 4 **SDx = 2**
avg y = 4 **SDy = 2**

(a) Compute r and interpret it.

Find out how many standard units each x observation and each y observation are away from its respective average.

e.g. x = 1 **y = 6**

$$\frac{(1-4)}{2} = -1.5 \qquad \frac{(6-4)}{2} = 1$$

r = avg of product of x and y standard units
r = [-1.5+(-.5)+(-.75)+ 0 +0+(-1.5) +(-1.5)] / 7 = - .8214

Fairly strong negative linear association (r is close to -1) between x and y

rough estimate of r

$$r \sim \frac{\operatorname{cov}(x,y)}{(SD\ of\ x)\times(SD\ of\ y)} = \frac{avg\ xy - (avg\ x)(avg\ y)}{(SD\ of\ x)\times(SD\ of\ y)} = \frac{12.71-(4)(4)}{(2)\times(2)} = .8225$$

(b) Find the regression equation

Slope = $\frac{r \times SDy}{SDx} = \frac{(-.8214)\times 2}{2}$ **= -.8214**

Next find the y intercept when x=0
X = 0 is 4 below the average
The y intercept should be (-.8214)(4) = -3.29 -> above the average of y since the relationship is negative.
Avg y + amount above average = 4+ 3.29 = 7.29 (this is the y-intercept when x = 0)
Equation of line: y=-.8214(x) + 7.29

(c) Estimate the y value when x = 3.5.

y=-.8214(3.5) + 7.29 = 4.4151

(d) Find the r.m.s error.

r.m.s = $\sqrt{1-r^2} \times SDy = \sqrt{1-(-.8214)^2} \times 2$ =1.1407

(e) Find the error for x = 3

y=-.8214(3) + 7.29 = 4.8258
error = y- expected y = 7 – 4.8258= 2.1742

9. In a study of heights, a large group of fathers and sons heights were recorded.

Father: average height = 68 inches, SD=2.7 inches
Son: average height =69 inches, SD=2.7 inches, r=0.5

(a) Find the r.m.s error of the regression line for predicting son's height from father's height.

r.m.s = $\sqrt{1-r^2} \times SDy$ = $\sqrt{1-(-.5)^2} \times 2.7$ =2.3383

(b) If a father is 72 inches tall, predict his son's height

$\frac{(72-68)}{2.7}$ **=1.48 The father is 1.48 SD above average (x). The son should be r x 1.48 SDy taller than avg (relationship is positive)**
(.5) (1.48)(2.7) = 1.998
69+1.998 = 70.998 or 71 inches

You could also find the regression equation.

Slope = $\frac{r \times SDy}{SDx} = \frac{(.5)\times 2.7}{2.7}$ **= .5**
Next find the y intercept when x=0
X = 0 is 68 below the average
(.5)(68) = 34 below the average of y since the relationship is positive.
Avg y – amount below average = 69- 34 = 35 (this is the y-intercept when x = 0)
Equation of line: y=-.5(x) + 35
Y = -.5(72) + 35 = 71

(c) This prediction is likely to be off by _____ inches or so. If more information is needed, say what it is and why.

r.m.s = 2.3383
more information is not needed.

(d) About what percentage of fathers have sons over 73 inches?

Z = $\frac{(73-69)}{2.7}$ **=1.48 using area under normal curve ~6.9%**

(e) Of the fathers that are 72 inches tall, what percentage of their sons have heights above 73 inches?

Expected son's height when father is 72 inches tall = 71 inches
SE for expected son's height = r.m.s = 2.3383
Z = $\frac{(73-71)}{2.3383}$ **=.86 ~ 19.5%**

(f) Interpret r in the context of the question.

There is not a very strong positive linear relationship between a father's height and his son's height.

Review chapters 28, 8,9,10, 11, and 12 in the text and do as many questions as possible from these chapters.