TA’S NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Problem Set #7**

1. I have 5 cats. Below I have reported my cats’ ages and my cats’ level of cuteness, each measured as continuous variables. The age measure is expressed in years; the cuteness measure ranges from 0 (Ugly) to 10 (Insufferably Cute). For these data, **(a)** report the correlation between my cats’ ages and their level of cuteness; **(b)** compute the least squares prediction equation for the regression of level of cuteness on age; **(c)** compute the predicted value of level of cuteness for a 10 year old cat; **(d)** report and interpret the value of R2YX; **(e)** interpret the value of the intercept of the prediction equation; and **(f)** describe the strength and direction of the association between my cats’ ages and their levels of cuteness based on the results you have produced. **SHOW YOUR WORK**

CAT AGE (X) CUTENESS (Y)

Fluffy 1 9

Boots 12 3

Tabitha 7 4

Mr. Snookums 2 5

Snowball III 5 6

* 1. First, we must compute the mean and standard deviation of each variable. I should show the work here, but I won’t: For X: Mean=5.40; SD=4.39. For Y: Mean=5.40; SD=2.30.







* 1. The prediction equation is where and .

In this example, and . So the prediction equation is 

* 1. The predicted value of level of cuteness for a 10 year old cat is 
  2. R2YX equals -0.812, or 0.656. This means that 65.6% of the variation in my cats level of cuteness is due to (or explained by) their ages.
  3. The intercept, 7.67, is interpreted as the predicted value of Y when X equals zero. So we expect a cat to have a “level of cuteness” score of 7.67 when it is born.
  4. There is a fairly strong negative association between my cats’ ages and their level of cuteness. We see this from the correlation … -0.81 … and the slope of the regression prediction equation; my cats get about 0.42 points less cute each year.

1. The General Social Survey asks people how much education they have completed (X) and it also includes a 10-item vocabulary test (Y). Here are the means and variances of X and Y, along with the correlation and sample size:

**N** = 21,638 ; **rYX** = 0.5

**Y-bar** = 5.996; **s2Y** = 4.688

**X-bar** = 12.795; **s2X** = 9.246

**(a)** compute the least squares prediction equation; **(b)** compute R2YX; **(c)** test the null hypothesis that 2YX equals zero in the population; **(d)** test the null hypothesis that YX equals zero in the population; and **(e)** test the null hypothesis that YX equals zero in the population. Use =0.05 for all hypothesis tests. *Show your work*.

(a)  and 

so 

(b) R2YX = 0.52 = 0.25

(c) H0: 2YX=0 ; H1: 2YX>0

With dfNUM=1 and dfDENOM=N-1=21,637 and =0.05, the critical value F\*=3.85

To compute the F statistic, we need:

= (4.688)(21,637) = 101,434.256

= (0.25)(101,434.256) = 25,358.564

= 101,434.256-25,358.564 = 76,075.692



Since F exceeds the critical value F\*, we reject the null hypothesis

(d) H0: YX=0 ; H1: YX≠0

With =0.05 for this two-sided test, the critical value Z\*=1.96





Since Z exceeds the critical value Z\*, we reject the null hypothesis

(e) H0: YX=0 ; H1: YX≠0

With =0.05 for this two-sided test, the critical value t\* with df=N-2 is 1.96



Since t exceeds the critical value t\*, we reject the null hypothesis