**Exam 3 Practice Questions**
**Psych 3101-100, Fall 9**

## Vocabulary

Rather than choosing some practice terms at random, I suggest you go through all the terms in the vocabulary lists. The real exam will ask for definitions of about 5 terms taken from these lists.

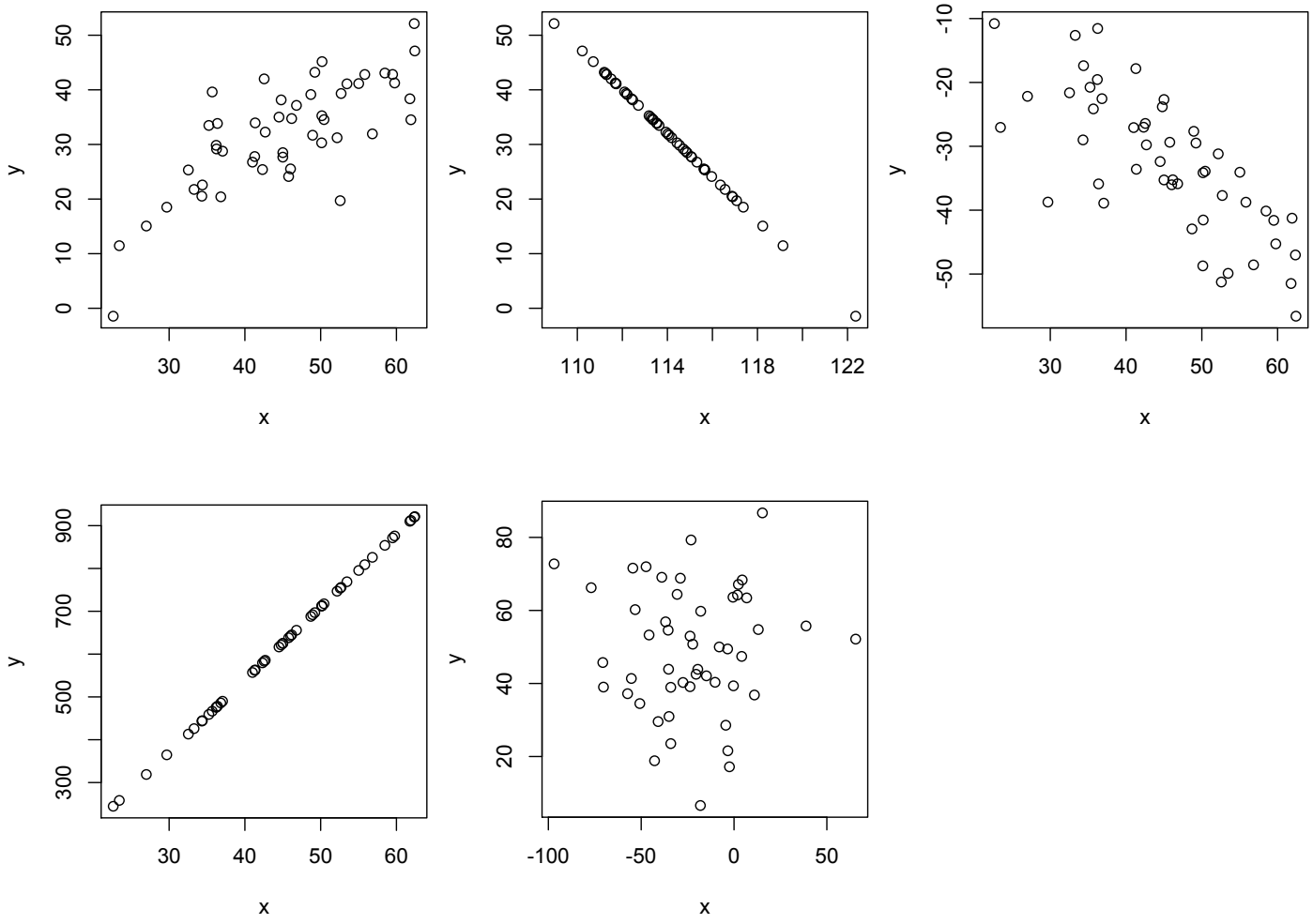
## Conceptual questions

1. Research has shown that people in Western (individualist) societies tend to think more analytically compared to people in Eastern (collectivist) societies, who tend to think more holistically. Because this is a cultural difference, you suspect the effect is stronger for adults than for children. To test this prediction, you recruit children and adults from both this country and China. Some subjects are tested on an analytic reasoning task, and others are tested on a holistic reasoning task. List the factors for this experiment. Then list all the possible interactions you could test.

2. What is the expected value of $MS_{\text{treatment}}$, according to the null hypothesis that the group means are all equal? (You can answer in words or in symbols.)

3. Repeated-measures ANOVA differs from regular ANOVA in that we remove the variability due to _____ before calculating _____.

4. The five scatterplots below show correlations of -1. -.5, 0, .5, and 1, in a scrambled order. Write the correlation under each plot.




5. Draw a picture of an F distribution being used for a hypothesis test (such as regression or ANOVA). Include the following: F statistic obtained from the data, p-value (as a shaded region), 0 (location on the horizontal axis).

**Math questions**

You don't need to show your work, but I will give partial credit for partial answers.

1. Assume you have data from three groups and want to know if the group means are reliably different from each other. The scores from the three groups are [4,8,2,5,6], [3,9,5,7], and [4,5,2,5]. Calculate the total sum of squares, the treatment sum of squares, and the residual sum of squares (in whatever way you want).

2. Use your answers from Question 1 to calculate $F$. The degrees of freedom are $df_{treatment} = 2$ and $df_{residual} = 10$. (Even if you were to have the wrong $SS$ answers for Question 1, you would get full credit for Question 2 if you did all the steps correctly going from $SS$s to $F$.)

3. Six subjects are measured on two variables, and the results are converted to z-scores. The z-scores for the two variables are [-.68, .85, 1.36, .34, -1.18, -.68] and [.18, .52, 1.29, .31, -.76, -1.53]. Calculate the correlation.

4. Researchers studying childhood language acquisition measure several variables for a sample of children, including age, number of siblings, and years of education for both the mother and father. They use these variables as predictors in a regression to try to explain the children's vocabulary size (in number of words known). The regression coefficient for age (in months) is 25. My son is currently 8½. How many new words would we predict him to learn between now and age 9 (i.e., in 6 months)?

5. A new educational program designed to teach kindergarteners the alphabet is tested in a controlled experiment. Half of the children in the study receive the new curriculum and half do not. This is done in both urban and rural schools. Therefore there are two factors with two levels each (experimental vs. control and urban vs. rural). At the conclusion of the study, the children are tested for how many letters they can correctly write. The urban-experimental children score [17, 13, 21, 9, 20, 16]. The urban-control children score [9, 15, 16, 10, 14, 20]. The rural-experimental children score [19, 16, 20, 9, 11, 15]. What would the mean for the rural-control children need to be for there to be no interaction in the sample?

**R questions**

1. What is the result of the following command?

```
> pf(0,1,2)
```

2. Write a number that could <u>not</u> be the result of the following command.

```
> cor(x,y)
```

3. Find $MS_{regression}$ using the output of the regression analysis below. Remember that the residual standard error is the square root of $MS_{residual}$. (Hint: Use the formula for $F$.)

```
> summary(lm(y ~ x1+x2))

Call:
lm(formula = y ~ x1 + x2)

Residuals:
     Min       1Q   Median       3Q      Max
-2.00427 -0.59131  0.07437  0.53894  1.69545

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1310     0.1227  -1.068  0.29080
x1            0.4747     0.1295   3.664  0.00063 ***
x2            0.3350     0.1108   3.023  0.00405 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8461 on 47 degrees of freedom
Multiple R-squared: 0.3703,Adjusted R-squared: 0.3435
F-statistic: 13.82 on 2 and 47 DF,  p-value: 1.905e-05
```

4. Use the regression output below to find the predicted depression score for a 47-year-old mother of three.

```
Call:
lm(formula = depression ~ age.years + children)

Coefficients:
(Intercept)    age.years     children
     0.4395       0.3240      -0.1570
```

**Answers**

Conceptual

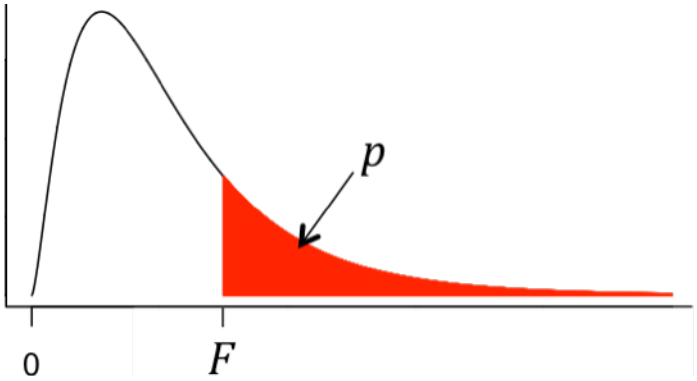
1. The factors are Country (with levels US and China), Age (with levels Child and Adult), and Task (with levels Analytic and Holistic). The possible interactions are Country:Age, Country:Task, Age:Task, and Country:Age:Task.

2. The expected value of $MS_{treatment}$ is the population variance, $\sigma^2$.

3. Repeated-measures ANOVA differs from regular ANOVA in that we remove the variability due to individual differences before calculating the residual variability or $SS_{residual}$.

4. The order of correlations is .75, -1, -.75, 1, 0.

5.




Math

1.

$$\overline{M} = \frac{4+8+2+5+6+3+9+5+7+4+5+2+5}{13} = 5$$

$$M_1 = \frac{4+8+2+5+6}{5} = 5$$

$$M_2 = \frac{3+9+5+7}{4} = 6$$

$$M_3 = \frac{4+5+2+5}{4} = 4$$

$$SS_{total} = (4-5)^2 + (8-5)^2 + (2-5)^2 + (5-5)^2 + (6-5)^2$$
$$+(3-5)^2 + (9-5)^2 + (5-5)^2 + (7-5)^2$$
$$+(4-5)^2 + (5-5)^2 + (2-5)^2 + (5-5)^2$$
$$= 1+9+9+0+1+4+16+0+4+1+0+9+0$$
$$= 54$$

$$SS_{\text{residual}} = (4-5)^2 + (8-5)^2 + (2-5)^2 + (5-5)^2 + (6-5)^2$$
$$+(3-6)^2 + (9-6)^2 + (5-6)^2 + (7-6)^2$$
$$+(4-4)^2 + (5-4)^2 + (2-4)^2 + (5-4)^2$$
$$= 1+9+9+0+1+9+9+1+1+0+1+4+1$$
$$= 46$$
$$SS_{\text{treatment}} = 5(5-5)^2 + 4(6-5)^2 + 4(4-5)^2$$
$$= 0+4+4$$
$$= 8$$

You can also calculate $SS_{\text{residual}} = SS_{\text{total}} - SS_{\text{treatment}} = 54-8 = 46$ or $SS_{\text{treatment}} = SS_{\text{total}} - SS_{\text{residual}} = 54-46 = 8$

2.

$$MS_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}} = \frac{8}{2} = 4 \qquad MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{46}{10} = 4.6 \qquad F = \frac{MS_{\text{treatment}}}{MS_{\text{residual}}} = \frac{4}{4.6} = .87$$

3.

$$r = \frac{-.68\cdot.18 + .85\cdot.52 + 1.36\cdot1.29 + .34\cdot.31 + (-1.18)(-.76) + (-.68)(-1.53)}{6-1}$$
$$= \frac{-.1224 + .442 + 1.7544 + .1054 + .8968 + 1.0404}{5}$$
$$= \frac{4.1166}{5}$$
$$= .82332$$

4. The regression coefficient for age tells how much the prediction increases for every additional month of age. Therefore the prediction increases by $6\cdot 25 = 150$ words.

5.

$$M_{\text{urban, experimental}} = \frac{17+13+21+9+20+16}{6} = 16$$
$$M_{\text{urban, control}} = \frac{9+15+16+10+14+20}{6} = 14$$
$$M_{\text{rural, experimental}} = \frac{19+16+20+9+11+15}{6} = 15$$

There is no interaction if the effect of the program is the same for children in both locations. That is,

$$M_{\text{urban, experimental}} - M_{\text{urban, control}} = M_{\text{rural, experimental}} - M_{\text{rural, control}}$$
$$16-14 = 15 - M_{\text{rural, control}}$$
$$M_{\text{rural, control}} = 13$$

R

1. The command asks for the probability that $F$ is less than 0. $F$ can never be negative, so the probability is 0.

2. 27.6 (or any other number bigger than 1 or less than -1)

3.

$$MS_{\text{residual}} = .8461^2 = .7159$$

$$F = \frac{MS_{\text{regression}}}{MS_{\text{residual}}}$$

$$MS_{\text{regression}} = F \cdot MS_{\text{residual}} = 13.82 \cdot .7159 = 9.89$$

4. $.4395 + .3240 \cdot 47 - .1570 \cdot 3 = 15.1965$