Exam 3
Psych 3101, Fall 14

Vocabulary

1. Main effect: The effect of one factor by itself, averaging over all levels of all other factors

2. Grand mean: The mean of all data across all groups or conditions

3. Scatterplot: A plot showing the relationship between two variables, with each point indicating one subject and having coordinates equal to that subject's scores on the two variables

4. Intercept: A coefficient in a regression equation that determines the prediction when all predictors are equal to zero

5. Independence: A relationship between two variables, such that the value of one gives no information about the value of the other

Conceptual questions

1. For each of the following studies, write which would be the appropriate analysis: regression, one-way ANOVA, repeated-measures ANOVA, or factorial ANOVA.

(a) Monkeys are trained to identify objects by name. Each subject is then tested with 30 objects: 10 identical to what s/he was trained on, 10 differing in color from a training item, and 10 differing in size from a training item. The number correct is recorded for each subject for each type of item (same, different color, or different size). The question we want to ask is whether naming performance is different for the three types of items.
Repeated-measures ANOVA

(b) Students are grouped according to their majors. Average hours of sleep per night is measured for each subject. The question we want to ask is whether amount of sleep depends on major.
One-way ANOVA

(c) Subjects are asked how many courses they're enrolled in this semester and how many siblings they have. Then each subject is asked to describe his or her plans for Thanksgiving break, and the subject's rate of speech is measured in words per minute. The question we want to ask is whether speech rate depends on number of courses and/or number of siblings.
Regression

2. When X is used to predict Y (i.e., we fit a regression Y ~ X), X explains 25% of the variance of Y. What's the correlation between X and Y?

The proportion of explained variance equals r^2, so the correlation equals ±.5

A study of alcohol use measures mean drinks per week as a function of sex (male or female), and job status (employed, unemployed, or self-employed). A factorial ANOVA is run on the data.

3. List all the components that the total sum of squares is broken into.

$SS_{\text{total}} = SS_{\text{sex}} + SS_{\text{job}} + SS_{\text{sex:job}} + SS_{\text{residual}}$

4. How many groups of subjects are there in the study?

2 values for sex times 3 values for job: 6 groups
(employed men, unemployed men, self-employed men, employed women, unemployed women, self-employed women)

5. The results show an interaction between sex and job status. Explain what this means or give an example of what the interaction might look like.

The difference in alcohol use between men and women depends on job status. For example, maybe employed women drink more than employed men, but unemployed women drink less than unemployed men.

Math questions

1. Calculate the correlation between X = [12, 19, 15, 14] and Y = [54, 59, 65, 62]. To save you time, the means and standard deviations are $M_X = 15$, $s_X = 2.94$, $M_Y = 60$, $s_Y = 4.69$.

$$z_X = \frac{X - M_X}{s_X} = \frac{[12,19,15,14]-15}{2.94} \approx [-1.02, 1.36, 0, -.34]$$

$$z_Y = \frac{X - M_Y}{s_Y} = \frac{[54,59,65,62]-60}{4.69} \approx [-1.28, -.21, 1.07, .43]$$

$$r = \frac{\sum z_X z_Y}{n-1} = \frac{(-1.02)(-1.28)+1.36(-.21)+0-.34*.43}{3} \approx \underline{.29}$$

2. Imagine a new subject has a score of X = 16. Based on your answer to Question 1, what would you predict that person's z-score for Y to be?

$$z_X = \frac{16-15}{2.94} \approx .34$$

$$z_Y = r \cdot z_X \approx .29 \cdot .34 \approx \underline{.097}$$

3. A new plant food is being tested on three kinds of trees. Elms, oaks, and aspens are each given either 0, 5, or 10 kg in the spring, and the growth of each tree is measured through the rest of the year. Below are mean growths in inches for six of the nine groups. Fill in the rest so that there's no interaction.

	Dosage		
Tree	0	5 kg	10 kg
Elm	3	6	12
Oak	8	11	17
Aspen	12	15	21

For every species, the effect of the food is the same: 5 kg increases growth by 3 inches, and 10 kg increases growth by another 6 inches. Another explanation: The differences among species are the same regardless of the food provided; aspens always grow 4 inches more than oaks, and oaks always grow 5 inches more than elms.

4. The table below shows the results of a regression on four subjects. The Y column shows the actual outcome for each subject, and the $\hat{Y}$ column shows the predicted outcome when Y is regressed on X_1 and X_2. What is R^2 for this regression? To save you time, the total sum of squares for Y is $SS_Y = 1449$.

X_1	X_2	Y	$\hat{Y}$	$(Y - \hat{Y})^2$
12	4	70	60	100
15	6	46	74	784
11	9	79	78	1
17	7	99	82	289

$$SS_{\text{residual}} = \sum(Y - \hat{Y})^2 = 100 + 784 + 1 + 289 = 1174$$

$$SS_{\text{regression}} = SS_{\text{total}} - SS_{\text{residual}} = 1449 - 1174 = 275$$

$$R^2 = \frac{SS_{\text{regression}}}{SS_{\text{total}}} = \frac{275}{1449} \approx \underline{.19}$$

5. The table below shows the data from a spatial detection experiment, in which each subject's average response time is measured in three conditions (congruent cue, incongruent cue, and no cue). Does response time reliably differ as a function of condition?

The total sum of squares is $SS_{\text{total}} = 570$.
The critical value is $F_{\text{crit}} = 5.14$.
The degrees of freedom are $df_{\text{treatment}} = 2$, $df_{\text{subject}} = 3$, and $df_{\text{residual}} = 6$.

	Cue			
Subject	Congruent	Incongruent	None	M_{sub}
1	335	345	343	341
2	338	343	345	342
3	334	344	336	338
4	325	328	328	327
$M_{\text{condition}}$	333	340	338	337

$$SS_{\text{treatment}} = \sum_{\text{cond}} n\left(M_{\text{cond}} - \overline{M}\right)^2 = 4(333-337)^2 + 4(340-337)^2 + 4(338-337)^2 = 64 + 36 + 4 = 104$$

$$SS_{\text{subject}} = \sum_{\text{sub}} k\left(M_{\text{sub}} - \overline{M}\right)^2 = 3(341-337)^2 + 3(342-337)^2 + 3(338-337)^2 + 3(327-337)^2 = 48 + 75 + 3 + 300 = 426$$

$$SS_{\text{residual}} = SS_{\text{total}} - SS_{\text{treatment}} - SS_{\text{subject}} = 570 - 104 - 426 = 40$$

$$MS_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}} = \frac{104}{2} = 52$$

$$MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}} = \frac{40}{6} \approx 6.67$$

$$F = \frac{MS_{\text{treatment}}}{MS_{\text{residual}}} \approx \frac{52}{6.67} \approx 7.80$$

$F > F_{\text{crit}}$, so we reject the null hypothesis and conclude <u>response time reliably differs among conditions</u>.

R questions

1. Fill in the missing values in the following output from the `anova()` function. (It's easiest to do them in order.)

```
              Df   Sum Sq  Mean Sq  F value   Pr(>F)
X              E    685.5        D   1.8993  0.11248
Y              3        C   334.17        B  0.01279
X:Y           12    873.1    72.75   0.8064  0.64363
Residuals    180  16240.4        A
```

```
A: 16240.4/180 = 90.22
B: 334.17/90.22 = 3.70
C: C/3 = 334.17 → C = 3*334.17 = 1002.51
D: D/90.22 = 1.8993 → D = 1.8993*90.22 = 171.3548
E: 685.5/E = 171.3548 → E = 685.5/171.3548 = 4
```

2. Based on the previous question, what is the output of the following command?

```
> pf(B,3,180,lower.tail=FALSE)
```

This command calculates the p-value for an F equal to `B` (3.70) on 3 & 180 degrees of freedom. That p-value is shown in the ANOVA table above: 0.01279.

3. Fill in the missing t value below.

```
> summary(lm(Y~X1+X2))

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)    0.5535      2.2716    0.244   0.80800
X1             5.3471      1.6878  _______   0.00205
X2             2.6434      1.5432    1.712   0.09006
```

$t = b/\sigma_b = 5.3471/1.6878 \approx 3.168$

4. What conclusion does the number .09006 in the previous question lead to (assuming $\alpha = 5\%$)?

`X2` gives no information about `Y` once we already know `X1`.

5. There's a 17-ounce squirrel in my backyard with a 6-inch tail who looks to be about 2 years old. How far do you think can he jump?

```
> lm(jumpInFeet ~ ageInYears + ounces + tailInInches)

Coefficients:
(Intercept)   ageInYears       ounces       tailInInches
    -2.6948      -2.9584       2.0192       0.3822
```

-2.6948 - 2.9584*2 + 2.0192*17 + 0.3822*6 = 28.008 feet