### Notation and Equations for Exam 3

- **$r$**  
  Sample correlation

- **$m$**  
  The number of predictor variables in a regression

- **$X_i$**  
  A predictor variable in a regression. The subscript $i$ represents any number from 1 through $m$.

- **$Y$**  
  The outcome variable that is being predicted or explained in a regression

- **$\hat{Y}$ (Y-hat)**  
  The estimated outcome value as predicted by the regression equation

- **$b_i$**  
  The regression coefficient for predictor $X_i$ (sometimes written as $b_{\text{predictor name}}$)

- **$b_0$**  
  The intercept in the regression equation

- **$\sigma_{b_i}$**  
  The standard error of a regression coefficient

- **$SS_Y$**  
  The total sum of squares for the outcome in a regression

- **$SS_{\text{regression}}$**  
  The sum of squares explained by the predictors in a regression

- **$R^2$**  
  The proportion of variability explained by a regression

- **$SS_{\text{total}}$**  
  The total variability in the data for an ANOVA

- **$SS_{\text{treatment}}$**  
  Variability explainable by differences among groups (simple ANOVA) or measurements (repeated measures)

- **$SS_{\text{factor}}$**  
  Variability explainable by the main effect of some factor

- **$SS_{A:B}$**  
  Variability explainable by interaction between factors A and B

- **$SS_{\text{residual}}$**  
  The residual sum of squares, representing the variability that can’t be explained in regression or ANOVA

- **$MS_{\text{effect}}$**  
  Mean square for any effect we might want to test; the subscript can be regression, treatment, Factor, A:B, etc.

- **$df_{\text{effect}}$**  
  Degrees of freedom for $SS_{\text{effect}}$ and $MS_{\text{effect}}$, where $effect$ is any effect we might want to test

- **$MS_{\text{residual}}$**  
  The residual mean square; used as an estimate of the population variance, $\sigma^2$ or $\sigma^2_Y$

- **$df_{\text{residual}}$**  
  The degrees of freedom for $SS_{\text{residual}}$ and $MS_{\text{residual}}$

- **$F$**  
  F statistic

- **$k$**  
  The number of levels of a factor (treatment) in an ANOVA; written as $k_{\text{Factor}}$ when there are multiple factors

- **$M_i$**  
  The sample mean of Group $i$ in a simple ANOVA or Measurement $i$ in a repeated-measures ANOVA ($i = 1$ to $k$)

- **$M_s$**  
  The mean of all measurements from Subject $s$ in a repeated-measures ANOVA ($s = 1$ to $n$)

- **$n_i$**  
  The number of data (e.g., subjects) in Group $i$

- **$\overline{M}$**  
  The grand mean, i.e. the mean of all the data in all groups taken together
<table>
<thead>
<tr>
<th>Formula for correlation</th>
<th>$r = \frac{\sum(z_X \cdot z_Y)}{n - 1}$</th>
</tr>
</thead>
</table>
| Interpreting correlation | $r = -1 \rightarrow$ perfect negative relationship  
$r < 0 \rightarrow$ negative relationship  
$r = 0 \rightarrow$ no linear relationship  
$r > 0 \rightarrow$ positive relationship  
$r = 1 \rightarrow$ perfect positive relationship |
| Relationship between correlation and prediction | $z_Y = r \cdot z_X$ |
| Regression equation | $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + ... + b_mX_m = b_0 + \sum b_iX_i$ |
| Total variability in a regression | $SS_Y = \sum(Y - M_Y)^2$ |
| Residual variability in a regression | $SS_{residual} = \sum(Y - \hat{Y})^2$ |
| Variability explained by a regression | $SS_{regression} = SS_Y - SS_{residual}$ |
| Proportion of variability explained by regression | $R^2 = \frac{SS_{regression}}{SS_Y}$ |
| Explained variability with one predictor | $R^2 = r^2$ |
| t statistic for a regression coefficient | $t = \frac{b_i}{\sigma_{b_i}}$ |
| Total sum of squares in an ANOVA | $SS_{total} = \sum(X - \bar{M})^2$ |
| Residual sum of squares in a simple ANOVA | $SS_{residual} = \sum(X_1 - M_i)^2 + \sum(X_2 - M_2)^2 + ... + \sum(X_k - M_k)^2 = \sum(\sum(X_i - M_i)^2)$ |
| Treatment sum of squares for simple ANOVA | $SS_{treatment} = \sum n_\cdot (M_i - \bar{M})^2$ |
### Mean squares from sum of squares

- $MS_{\text{regression}} = \frac{SS_{\text{regression}}}{df_{\text{regression}}}$
- $MS_{\text{treatment}} = \frac{SS_{\text{treatment}}}{df_{\text{treatment}}}$
- $MS_{\text{factor}} = \frac{SS_{\text{factor}}}{df_{\text{factor}}}$
- $MS_{A:B} = \frac{SS_{A:B}}{df_{A:B}}$
- $MS_{\text{residual}} = \frac{SS_{\text{residual}}}{df_{\text{residual}}}$

### F statistic for explained variability

- $F = \frac{MS_{\text{regression}}}{MS_{\text{residual}}}$
- $F = \frac{MS_{\text{treatment}}}{MS_{\text{residual}}}$
- $F_{\text{factor}} = \frac{MS_{\text{factor}}}{MS_{\text{residual}}}$
- $F_{A:B} = \frac{MS_{A:B}}{MS_{\text{residual}}}$

### p-value for explained variability

- $p = p\left(F_{df_{\text{regression}},df_{\text{residual}}} \geq F\right)$
- $p = p\left(F_{df_{\text{treatment}},df_{\text{residual}}} \geq F\right)$
- $p_{\text{factor}} = p\left(F_{df_{\text{factor}},df_{\text{residual}}} \geq F_{\text{factor}}\right)$
- $p_{A:B} = p\left(F_{df_{A:B},df_{\text{residual}}} \geq F_{A:B}\right)$

### Partitioning variability for regression

- $SS_{Y} = SS_{\text{regression}} + SS_{\text{residual}}$

### Partitioning variability for simple ANOVA

- $SS_{\text{total}} = SS_{\text{treatment}} + SS_{\text{residual}}$

### Partitioning variability for repeated measures

- $SS_{\text{total}} = SS_{\text{treatment}} + SS_{\text{subject}} + SS_{\text{residual}}$

### Partitioning variability for factorial ANOVA

- $SS_{\text{total}} = SS_{A} + SS_{B} + SS_{C} + \ldots$ [every main effect]
- $+ SS_{A:B} + SS_{A:C} + SS_{B:C} + \ldots$ [every 2-way interaction]
- $+ SS_{A:B:C} + \ldots$ [every higher-order interaction, up to the total number of factors]
- $+ SS_{\text{residual}}$

### Recognizing an interaction

- $M_{a_{1},b_{1}} - M_{a_{2},b_{1}} \neq M_{a_{1},b_{2}} - M_{a_{2},b_{2}} \implies$ Interaction