Example memory-imagery experiment: data for reference

How does mental imagery affect memory?
Two groups of subjects (say G1, G2) were asked to memorize pairs of words and then recall the pair given the first word as a cue after a 24 hours gap. G1 was instructed to actively use imagery to link the two words and given examples of such linking. G2 was just instructed to remember the pairs in whatever way they could. The number of pairs recalled was the response variable. The independent variable is the group (or treatment) and levels are G1-imagery, G2-control.
The results for 10 subjects with 5 subjects randomly assigned to each of G1, G2 are given below for 20 word pairs.

G1	G2
8	1
8	2
9	5
11	6
14	6

G1 mean=10; G2 mean=4; Grand mean=7.

Exercise 1: Memory, imagery experiment I

- If we work out the details of exercise 1: $SS_{\text{betweengroups}} = 90.0$, $SS_{\text{withingroups}} = 48.0$, $SS_{\text{total}} = 138.0 = 90.0 + 48.0$. Also, verify that: $SS_{\text{total}} = SS_{\text{betweengroups}} + SS_{\text{withingroups}}$.
- $df_{\text{total}} = (\ell \times N) - 1 = 2 \times 5 - 1 = 9$, $df_{\text{betweengroups}} = \ell - 1 = 2 - 1 = 1$, $df_{\text{withingroups}} = \ell(N - 1) = 2 \times 4 = 8$. Verify that $df_{\text{total}} = df_{\text{betweengroups}} + df_{\text{withingroups}}$.
- $s^2_{\text{betweengroups}} = \frac{SS_{\text{betweengroups}}}{df_{\text{betweengroups}}} = 90.0/1 = 90.0$
- $s^2_{\text{withingroups}} = \frac{SS_{\text{withingroups}}}{df_{\text{withingroups}}} = 48.0/8 = 6.0$

Exercise 1: Memory, imagery experiment II

- $F = \frac{s^2_{betweengroups}}{s^2_{withingroups}} = 90.0/6.0 = 15.0$. The corresponding $F_{critical}$ values with $\nu_1 = 1$, $\nu_2 = 6$ are: at $\alpha = 0.05$, $F_{critical} = 5.987$; at $\alpha = 0.01$, $F_{critical} = 13.745$. So, H_0 is rejected at both α levels.

Detour: how are ANoVAs reported

Source	**df**	**SS**	**MS**	**F**
Between	1	90.0	90.0	15.0**
Within	8	48.0	6.0	
Total	9	138.0		

** $p < 0.01$

$R^2_{Y \cdot X}$ and F

- $R^2_{Y \cdot X} = \frac{SS_{\text{betweengroups}}}{SS_{\text{total}}} = 90.0/138.0 \approx 0.6522.$
- Calculate F:

$$
\begin{aligned}
F &= \frac{R^2_{Y \cdot X}}{1 - R^2_{Y \cdot X}} \times \frac{df_{\text{withingroups}}}{df_{\text{betweengroups}}} \\
&= \frac{90/138}{48/138} \times \frac{8}{1} \\
&= 15
\end{aligned}
$$

- We get exactly the same F value of 15.

Connection between ANoVA and regression I

- ANoVA creates ℓ groups based on the levels of a categorial factor.
- To see the connection between ANoVA and regression we need to convert the categorial factor into a numeric variable X that can be used to predict the value of the dependent variable Y via a regression equation.
- We use the following trick: Identify a group by its group mean μ_l for the lth group. So, instead of identifying a group by the value of the categorial variable we choose to identify it by the corresponding group mean μ_l.

Connection between ANoVA and regression II

- Simultaneously, the best predictor for any member of a group is the corresponding group mean μ_I. So, one way to look at the X-value in a group is to use the corresponding μ_I for each X-value. Note that in the ANoVA view there is no X-value - so this an invented X-value but it is the most logical choice.
- So X has been transformed into a numeric independent variable and we can now construct a regression equation using the data.
  $\hat{Y} = a_r + bX$
- We have: $a_r = \mu_Y - b\,\mu_X$ and $b = \frac{SCP_{X \cdot Y}}{SS_{\text{X}}}$.
- Our choice μ_I for the group categorial variable and for the predictor values in a group means that $\mu_Y = \mu_X = \mu_{\text{grand}}$. Also, $SCP_{X \cdot Y} = SS_{\text{X}}$, so $b = 1$ and $a_r = 0$. (Proved more formally later)

Connection between ANoVA and regression III

- So, $\hat{Y} = X = \mu_I$. So, ANoVA and regression are essentially the same thing.

One way to see this is to look at the memory-imagery data as data for regression - (superimposed with the factor levels)

	I =control					I =imagery				
X	4	4	4	4	4	10	10	10	10	10
Y	1	2	5	6	6	8	8	9	11	14

We see exactly why $\mu_X = \mu_Y = \mu_{\text{grand}}$.

- The previous also means:

$$SS_{\text{total-regr}} = SS_{total-anova}$$
$$SS_{\text{regression}} = SS_{\text{betweengroups}} \quad \text{and}$$
$$SS_{\text{residual}} = SS_{\text{withingroups}}$$

Connection between ANoVA and regression IV

- The degrees of freedom also have similar equivalences:

$$\begin{aligned} df_{\text{total-regr}} &= df_{\text{total-anova}} \\ df_{\text{regression}} &= df_{\text{betweengroups}} \\ df_{\text{residual}} &= df_{\text{withingroups}} \end{aligned}$$

- The quality of the regression and the intensity of the dependence also align.

$$r^2_{\text{X}\cdot\text{Y}} = R^2_{\text{Y}\cdot\text{X}}$$

$$\frac{SS_{\text{regression}}}{SS_{\text{total-regr}}} = \frac{SS_{\text{betweengroups}}}{SS_{\text{total-anova}}}$$

Comparison chart: Regression, ANoVA

Regression	**ANoVA** (l-group label, j-subject in group l)
Y	Y_{lj}
$\hat{Y}$	μ_l
X	μ_l
μ_Y	μ_{grand}
μ_X	μ_{grand}
$SS_{\text{regression}}$	$SS_{\text{betweengroups}}$
SS_{residual}	$SS_{\text{withingroups}}$
$R^2_{\text{Y}\cdot\text{ X}}$ or $r^2_{\text{Y}\cdot\text{ X}}$	$R^2_{\text{Y}\cdot\text{ X}}$
N - total subjects	$\ell \times N$, N=no. of subjects in a group

The regression-ANoVA connection is more general I

- To see the connection between regression and ANoVA we identified each group by the corresponding ANoVA group mean μ_I and for each X-value in the group we again used the same mean μ_I.
- However, the choice of the group mean is the most intuitively natural one it is not necessary for the correspondence between regression and ANoVA. We can use any value in place of μ_I to identify the group and for the X-values within a group provided the values for each group are distinct.
- using an arbitrary value instead of μ_I changes the values of the regression coefficients a_r and b which were respectively 0 and 1 when we used μ_I. But the principle and the other correspondences remain unchanged.

The regression-ANoVA connection is more general II

- The reason for having two different ways of doing the same thing is historical. The psychological literature uses ANoVA while the clinical literature uses regression. ANoVA makes it easier to handle the analysis when more factors or independent variables are involved.

Exercise 8

Repeat the memory-imagery analysis by using +1 for the imagery group and -1 as the value for the control group. Repeat again with values 1 and 0. Observe how the values of a_r and b change based on how the groups are represented numerically. Also, verify via calculation that the correspondences remain unchanged.

Some proofs-1

For ANoVA we only have Y-values for each of ℓ groups written Y_{lj} for the value of subject j in group l. Each group has N subjects.

μ_l is the mean of the lth group $\mu_l = \frac{\sum_{j=1}^{N} Y_{lj}}{N}$, and $\mu_{\text{grand}} = \frac{\sum_{l=1}^{\ell} \mu_l}{\ell}$ is the grand mean.

$SS_{\text{betweengroups}} = \sum_{l=1}^{\ell} \sum_{j=1}^{N} (\mu_l - \mu_{\text{grand}})^2 = N \sum_{l=1}^{\ell} (\mu_l - \mu_{\text{grand}})^2$,

$SS_{\text{withingroups}} = \sum_{l=1}^{\ell} \sum_{j=1}^{N} (Y_{lj} - \mu_l)^2$,

$SS_{\text{total}} = \sum_{l=1}^{\ell} \sum_{j=1}^{N} (Y_{lj} - \mu_{\text{grand}})^2$.

For regression we have the pairs:
$\{(X_{11}, Y_{11}), \ldots, (X_{1N}, Y_{1N}), \ldots, (X_{\ell 1}, Y_{\ell 1}), \ldots, (X_{\ell N}, Y_{\ell N})\}$. The sample size is $N_r = \ell \times N$ (we use N_r for regression instead of N to avoid confusion). For each $l = 1..\ell$ and each $j = 1..N$, $X_{lj} = \mu_l$ so $\mu_X = \mu_Y = \mu_{\text{grand}}$ (see the ANoVA formulae).

Let the regression equation be: $\hat{Y} = a_r + bX$. We know:
$a_r = \mu_Y - b\mu_X$ and $b = \frac{SCP_{X.Y}}{SS_{\text{X}}}$. So we calculate $SCP_{X.Y}$ and SS_{X}.

Some proofs-2

$$
\begin{aligned}
SCP_{X \cdot Y} &= \sum_{l=1}^{\ell} \sum_{j=1}^{N} (X_{lj} - \mu_X)(Y_{lj} - \mu_Y) \\
&= \sum_{l=1}^{\ell} \sum_{j=1}^{N} (X_{lj} - \mu_{\text{grand}})(Y_{lj} - \mu_{\text{grand}}), \quad \mu_X = \mu_Y = \mu_{\text{grand}} \\
&= \sum_{l=1}^{\ell} \sum_{j=1}^{N} (\mu_l - \mu_{\text{grand}})(Y_{lj} - \mu_{\text{grand}}), \quad X_{lj} = \mu_l \\
&= \sum_{l=1}^{\ell} (\mu_l - \mu_{\text{grand}}) \sum_{j=1}^{N} (Y_{lj} - \mu_{\text{grand}}) \\
&= \sum_{l=1}^{\ell} (\mu_l - \mu_{\text{grand}}) [\sum_{j=1}^{N} Y_{lj} - \mu_{\text{grand}} \sum_{j=1}^{N} 1] \\
&= \sum_{l=1}^{\ell} (\mu_l - \mu_{\text{grand}}) [N\mu_l - N\mu_{\text{grand}}] \\
&= N \sum_{l=1}^{\ell} (\mu_l - \mu_{\text{grand}})^2
\end{aligned}
$$

Some proofs-3

Similarly,

$$
\begin{aligned}
SS_{\mathrm{X}} &= \sum_{l=1}^{\ell}\sum_{j=1}^{N}(X_{lj} - \mu_{\text{grand}})^2 \\
&= \sum_{l=1}^{\ell}\sum_{j=1}^{N}(\mu_l - \mu_{\text{grand}})^2, \quad X_{lj} = \mu_l \\
&= \sum_{l=1}^{\ell}(\mu_l - \mu_{\text{grand}})^2 \sum_{j=1}^{N} 1 \\
&= N\sum_{l=1}^{\ell}(\mu_l - \mu_{\text{grand}})^2
\end{aligned}
$$

So, $SCP_{X \cdot Y} = SS_{\mathrm{X}}$ which implies $b = 1$ and $a_r = 0$. This means $\hat{Y} = X = \mu_l$.

Some proofs-4

We show $SS_{\text{regression}} = SS_{\text{betweengroups}}$ and $SS_{\text{residual}} = SS_{\text{withingroups}}$.

$$
\begin{aligned}
SS_{\text{regression}} &= \sum_{i=1}^{N_r=\ell \times N} (\hat{Y}_i - \mu_Y)^2, \quad \text{see regression lecture} \\
&= \sum_{l=1}^{\ell} \sum_{j=1}^{N} (\hat{Y}_{lj} - \mu_{\text{grand}})^2, \quad \mu_Y = \mu_{\text{grand}} \\
&= \sum_{l=1}^{\ell} \sum_{j=1}^{N} (\mu_l - \mu_{\text{grand}})^2, \quad \hat{Y}_{lj} = \mu_l \\
&= SS_{\text{betweengroups}} \\
SS_{\text{residual}} &= \sum_{i=1}^{N_r=\ell \times N} (Y_i - \hat{Y}_i)^2 \quad \text{see regression lecture} \\
&= \sum_{l=1}^{\ell} \sum_{j=1}^{N} (Y_{lj} - \hat{Y}_{lj})^2 \\
&= \sum_{l=1}^{\ell} \sum_{j=1}^{N} (Y_{lj} - \mu_l)^2, \quad \hat{Y}_{lj} = \mu_l
\end{aligned}
$$

Exercise

Exercise 9

For the Romeo-Juliet experiment repeat the calculation by using the group mean to represent the groups and the individual X-values in each group as was done for the memory-imagery experiment and use regression analysis to confirm that you get the same results as the ANoVA.