1. (a) $\overline{X} = 5.5$, $\overline{Y} = 50$

$\sum_{i=1}^{7} X_i^2 = 218.75$, $\sum_{i=1}^{7} Y_i^2 = 18154$, $\sum_{i=1}^{7} X_i Y_i = 1991.5$

$S_{xx} = 218.75 - (7)(5.5)^2 = 7$, $S_{yy} = 18154 - (7)(50)^2 = 654$

$S_{xy} = 1991.5 - (7)(5.5)(50) = 66.5$

$$b = \frac{S_{xy}}{S_{xx}} = \frac{66.5}{7} = 9.5 \quad , \quad a = \overline{Y} - b\overline{X} = 50 - (9.5)(5.5) = -2.25$$

Fitted regression line : $\widehat{Y} = -2.25 + 9.5X$

(b) $SS_T = S_{yy} = 654$

$SS_R = b^2 S_{xx} = (9.5)^2(7) = 631.75$

$SS_E = 654 - 631.75 = 22.25$

ANOVA table :

Source	**SS**	**d.f.**	**MS**	**F-ratio**
Regression	631.75	1	631.75	141.97
Error	22.25	5	4.45	
Total	654			

(c) $R^2 = \dfrac{SS_R}{SS_T} = \dfrac{631.75}{654} = 96.60\%$

(d) Point prediction of Y_0 at $X_0 = 5$: $\widehat{Y}_0 = -2.25 + 9.5 \times 5 = 45.25$

A 95% prediction interval for Y_0 at $X_0 = 5$ is given by

$$\widehat{Y}_0 \pm t_{5,0.025}\sqrt{MS_E\left(1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{S_{xx}}\right)} = 45.25 \pm (2.571)\sqrt{(4.45)\left(1 + \frac{1}{7} + \frac{(5 - 5.5)^2}{7}\right)}$$
$$= 45.25 \pm 5.8879$$
$$= [39.3621, 51.1379]$$

(e) Yes. The prediction based on a single observation is less accurate than the prediction based on all the data, especially when the regression effect is so significant.

(f) We use the linear regression model $Y = \alpha + \beta X + \varepsilon$ to model the relationship between the variables X and Y only in the range of X from 4.0g to 7.0g. We wouldn't extrapolate this relationship to X = 0g which is out of the observed range. Therefore even when α is not equal to zero, the model is still reasonable and not "ridiculous" as stated by the experimenter.

2. (a) For circuit (i),

$$\Pr(\text{circuit functions properly}) = \Pr(\text{both amplifiers function}) = 0.9^2 = 0.81$$

For circuit (ii),

$$\begin{aligned}\Pr(\text{circuit functions properly}) &= 1 - \Pr(\text{circuit failed})\\ &= 1 - \Pr(\text{both parallel components failed})\\ &= 1 - (1-0.81)^2 = 0.9639\end{aligned}$$

For circuit (iii),

$$\begin{aligned}\Pr(\text{circuit functions properly}) &= \Pr(\text{both components in series function})\\ &= \left(1-(1-0.9)^2\right)^2 = 0.9801\end{aligned}$$

Circuit (iii) is the most reliable circuit.

(b) $X \sim b(4, 0.9)$, $E(X) = 4 \times 0.9 = 3.6$, $Var(X) = 4 \times 0.9 \times 0.1 = 0.36$

(c) $\Pr(\text{two broken}) = \Pr(\text{two functions}) = \Pr(X = 2) = \binom{4}{2}(0.9)^2(0.1)^2 = 0.0486$

The circuit can function with only 2 functioning amplifiers in one of the following cases : 1,2 functions and 3,4 failed; 1,4 functions and 2,3 failed; 2,3 functions and 1,4 failed; 3,4 functions and 1,2 failed. Hence

$$\Pr(\text{two amplifiers broken} \cap \text{circuit functions}) = 4 \times 0.9^2 \times 0.1^2 = 0.0324$$

$$\Pr(\text{circuit functions} \mid \text{two amplifiers broken}) = \frac{0.0324}{0.0486} = \frac{2}{3} = 0.6667$$

3. (a) "Team": Nominal scale
"Points": Ratio scale

(b) Sorted points: 33, 40, 45, 45, 45, 46, 49, 49, 51, 56, 57, 59, 60, 61, 66, 68

$Q_L = X_{(4.25)} = X_{(4)} + 0.25\left(X_{(5)} - X_{(4)}\right) = 45 + 0.25(45-45) = 45$

80th percentile = $X_{(13.6)} = X_{(13)} + 0.6\left(X_{(14)} - X_{(13)}\right) = 60 + 0.6(61-60) = 60.6$

(c) Test $H_0 : \mu_B = \mu_D$ vs $H_1 : \mu_B \neq \mu_D$.

For team B, $m = 8$, $\bar{X} = 48.75$, $S_x^2 = 67.0714$

For team D, $n = 8$, $\bar{Y} = 55$, $S_y^2 = 109.7143$

$$S_{pool}^2 = \frac{(8-1)(67.0714) + (8-1)(109.7143)}{8+8-2} = 88.3929$$

$$T = \frac{\bar{X} - \bar{Y}}{S_{pool}\sqrt{1/m + 1/n}} = \frac{48.75 - 55}{\sqrt{(88.3929)(1/8 + 1/8)}} = -1.3295$$

$|T| = 1.3295 < 2.145 = t_{14,0.025}$

Do not reject H_0 at $\alpha = 0.05$, i.e. the guesses on the two teams are not significantly different.

(d) Assumptions : (i) Normal populations. (ii) Independent samples. (iii) Equal variances.

4. (a) $T \sim \Gamma(3, 0.002) \Rightarrow 0.004T \sim \Gamma(3, 0.5) \equiv \chi_6^2$

$\Pr(T \geq 2660) = \Pr(0.004T \geq 10.64) = \Pr(\chi_6^2 \geq 10.64) = 1 - 0.9 = 0.1$

(b) $H_0 : \mu \leq 1500$ vs $H_1 : \mu > 1500$.

μ is the population mean lifetime (in days) of the products manufactured after the quality control requirements were tightened.

(c) $\overline{T} = 1681$, $S = 759$

Since sample size $n = 100$ is large, we use the Z-test by normal approximation.

$$Z = \frac{\overline{T} - 1500}{S/\sqrt{n}} = \frac{1681 - 1500}{759/\sqrt{100}} = 2.3847 > 1.645 = Z_{0.05}$$

Reject H_0 at $\alpha = 0.05$, i.e. the data provide strong evidence to support the manufacturer's claim.

5. (a) $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$; $j = 1, 2, \ldots, 26$; $i = 1, 2, 3, 4$

$$\sum_{i=1}^{4} \alpha_i = 0 \quad , \quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

where Y_{ij} is jth rating on the ith commercial, μ is the overall population mean of rating, α_i is the ith treatment effect of the ith commercial on the rating, ε_{ij} is the (ij)th random error.

Assumptions : (i) Normal populations. (ii) Independent samples. (iii) Equal variances.

(b) (i) 4 – 1 = 3

(ii) 104 – 4 = 100

(iii) 104 – 1 = 103

(iv) $6378.159 \times 3 = 19134.477$

(v) $93.7 \times 100 = 9370$

(vi) $19134.477 + 9370 = 28504.477$

(vii) $68.07 \times 93.7 = 6378.159$

(c) $p - value = 0.000 < 0.01$

Reject H_0 at 1% significance level.

(d) Advantage : Combining the results from the two samples t-tests will largely increase the type I error probability in making the overall conclusion. By a single test we can easily control the type I error probability to a desired level.

Disadvantage : If the single test rejected the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, we can only conclude that not all the treatment effects are the same, but have no idea on which treatment effects are different and which treatment effects are the same. Hence it is less informative that the test results by pairwise comparison based on the two samples t-tests.

(e) 90% C.I. for $\alpha_1 - \alpha_2$: $\left(\overline{Y}_1 - \overline{Y}_2\right) \pm t_{100,0.05}\sqrt{MS_E\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

$$= (50.640 - 36.815) \pm (1.645)\sqrt{(93.7)\left(\frac{1}{26}+\frac{1}{26}\right)}$$

$$= 13.825 \pm 4.416 = [9.409\,, 18.241] \qquad (\alpha_1 > \alpha_2)$$

90% C.I. for $\alpha_1 - \alpha_3$: $(50.640 - 56.917) \pm 4.416$

$$= -6.277 \pm 4.416 = [-10.693\,, -1.861] \qquad (\alpha_1 < \alpha_3)$$

90% C.I. for $\alpha_1 - \alpha_4$: $(50.640 - 74.560) \pm 4.416$

$$= -23.92 \pm 4.416 = [-28.336\,, -19.504] \qquad (\alpha_1 < \alpha_4)$$

90% C.I. for $\alpha_2 - \alpha_3$: $(36.815 - 56.917) \pm 4.416$

$$= -20.102 \pm 4.416 = [-24.518\,, -15.686] \qquad (\alpha_2 < \alpha_3)$$

90% C.I. for $\alpha_2 - \alpha_4$: $(36.815 - 74.560) \pm 4.416$

$$= -37.745 \pm 4.416 = [-42.161\,, -33.329] \qquad (\alpha_2 < \alpha_4)$$

90% C.I. for $\alpha_3 - \alpha_4$: $(56.917 - 74.560) \pm 4.416$

$$= -17.643 \pm 4.416 = [-22.059\,, -13.227] \qquad (\alpha_3 < \alpha_4)$$

Hence we may conclude (with less than 90% confidence) that $\alpha_2 < \alpha_1 < \alpha_3 < \alpha_4$, i.e. the commercial having a jingle got the highest rating, then followed by the commercial featuring a celebrity, the commercial featuring a product package shot, and the commercial featuring a cast got the lowest rating.

6. (a) $H_0 : p_T \geq p_U$ vs $H_0 : p_T < p_U$.

(b) Test statistic : $Z = \dfrac{\widehat{p}_T - \widehat{p}_U}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_T}+\frac{1}{n_U}\right)}}$

Reject H_0 if $Z < -Z_{0.03} = -1.88$.

From the data, $\widehat{p}_T = \dfrac{84}{100} = 0.84$, $\widehat{p}_U = \dfrac{132}{150} = 0.88$, $\widehat{p} = \dfrac{84+132}{100+150} = 0.864$

$$Z_{obs} = \frac{0.84 - 0.88}{\sqrt{(0.864)(0.136)\left(\frac{1}{100}+\frac{1}{150}\right)}} = -0.9039 > -1.88 \ .$$

Do not reject H_0 at $\alpha = 0.03$.

$p - value = \Phi(-0.9039) = 1 - 0.8170 = 0.183$

(b) (**Alternative calculation**)

Test statistic : $$Z = \frac{\hat{p}_T - \hat{p}_U}{\sqrt{\frac{\hat{p}_T(1-\hat{p}_T)}{n_T} + \frac{\hat{p}_U(1-\hat{p}_U)}{n_U}}}$$

Reject H_0 if $Z < -Z_{0.03} = -1.88$.

From the data, $\hat{p}_T = \frac{84}{100} = 0.84$, $\hat{p}_U = \frac{132}{150} = 0.88$, $\hat{p} = \frac{84+132}{100+150} = 0.864$

$$Z_{obs} = \frac{0.84 - 0.88}{\sqrt{\frac{(0.84)(0.16)}{100} + \frac{(0.88)(0.12)}{150}}} = -0.8839 > -1.88 \ .$$

Do not reject H_0 at $\alpha = 0.03$.

$p - value = \Phi(-0.8839) = 1 - 0.8117 = 0.1883$

(c) A 95% confidence interval for $p_T - p_U$ is given by

$$(\hat{p}_T - \hat{p}_U) \pm Z_{0.025}\sqrt{\frac{\hat{p}_T(1-\hat{p}_T)}{n_T} + \frac{\hat{p}_U(1-\hat{p}_U)}{n_U}}$$

$$= (0.84 - 0.88) \pm (1.96)\sqrt{\frac{(0.84)(0.16)}{100} + \frac{(0.88)(0.12)}{250}}$$

$$= -0.04 \pm 0.0887 = [-12.87\% , 4.87\%]$$

(d) From the results in (b), the data did not provide strong evidence that $p_T < p_U$. Moreover, since the confidence interval in (c) contains zero, p_T and p_U are not significantly different. Hence we don't have evidence that the chemical treatment has effect on the rate of seed germination.