Exam 2 Solution

1. Since the original `y` is in the order

```
 2  4  2  3  1  6  7  5 10  1  7
```

and `sort(y)` will sort the data as

```
 1  1  2  2  3  4  5  6  7  7 10
```

we find that

(a) The `y[8]` is 5.
(b) The `sort(y)[8]` is 6.
(c) The `quantile(y, 0.5)` is same as (sample) median, which is 4.

2. To find $\text{se}(\overline{Y})$, we first need to find $\text{Var}(\overline{Y}) = \text{Var}(Y)/n$, where

$$E(Y) = \int y f_Y(y)\, dy = \int_2^\infty y 24 y^{-4}\, dy = \int_2^\infty 24 y^{-3}\, dy = 3$$

and

$$E(Y^2) = \int y^2 f_Y(y)\, dy = \int_2^\infty y^2 24 y^{-4}\, dy = \int_2^\infty 24 y^{-2}\, dy = 12$$

so that

$$\text{Var}(Y) = E(Y^2) - (E(Y))^2 = 12 - 3^2 = 12 - 9 = 3$$

and

$$\text{se}(\overline{Y}) = \sqrt{\text{Var}(\overline{Y})} = \sqrt{\text{Var}(Y)/n} = \sqrt{3/n}$$

3. If $Y_1, \ldots, Y_n$ are iid Poisson(9), then $E(Y) = 9$ and $\text{Var}(Y) = 9$, so that $E(\overline{Y}) = 9$ and $\text{Var}(\overline{Y}) = \text{Var}(Y)/n = 9/36$, which follows that $\text{se}(\overline{Y}) = \sqrt{\text{Var}(\overline{Y})} = \sqrt{9/36} = 1/2$. Since

$$Z = \frac{\overline{Y} - E(\overline{Y})}{\text{se}(\overline{Y})} \sim N(0,1)$$

for large n, we have that

$$P(\overline{Y} \le 8.5) = P\left(\frac{\overline{Y} - E(\overline{Y})}{\text{se}(\overline{Y})} \le \frac{8.5 - E(\overline{Y})}{\text{se}(\overline{Y})}\right) \approx P\left(Z \le \frac{8.5 - E(\overline{Y})}{\text{se}(\overline{Y})}\right) = P\left(Z \le \frac{8.5 - 9}{1/2}\right) = P(Z \le -1)$$

Hence, we can compute this probability in R by

```
pnorm(-1)
```

4. Because both sample sizes n_1 and n_2 are large, $E(\overline{Y}_1) = \theta_1$ and $E(\overline{Y}_2) = \theta_2$, and the null hypothesis consists of $H_0 : \theta_1 = \theta_2 \iff H_0 : \theta_1 - \theta_2 = 0$, this will be a two-sample Z-test with the statistic $U = \overline{Y}_1 - \overline{Y}_2$, so that the test statistic will be based on

$$Z = \frac{U - E(U)}{\mathrm{se}(U)}$$

where

$$E(U) = E(\overline{Y}_1 - \overline{Y}_2) = E(\overline{Y}_1) - E(\overline{Y}_2) = \theta_1 - \theta_2$$

and since $\mathrm{Var}(\overline{Y}_1) = \theta_1^2/n_1$ and $\mathrm{Var}(\overline{Y}_2) = \theta_2^2/n_2$,

$$\mathrm{se}(U) = \sqrt{\mathrm{Var}(U)} = \sqrt{\mathrm{Var}(\overline{Y}_1 - \overline{Y}_2)} = \sqrt{\mathrm{Var}(\overline{Y}_1) + \mathrm{Var}(\overline{Y}_2)} = \sqrt{\frac{\theta_1^2}{n_1} + \frac{\theta_2^2}{n_2}}$$

Now, under $H_0 : \theta_1 = \theta_2 \iff H_0 : \theta_1 - \theta_2 = 0$, we get

$$Z = \frac{U - E(U)}{\mathrm{se}(U)} = \frac{(\overline{Y}_1 - \overline{Y}_2) - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1^2}{n_1} + \frac{\theta_2^2}{n_2}}} = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{\theta_1^2}{n_1} + \frac{\theta_2^2}{n_2}}}$$

and for large n_1 and n_2, we have that $\overline{Y}_1 \approx \theta_1$ and $\overline{Y}_2 \approx \theta_2$, from which we conclude that

$$Z = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{\overline{Y}_1^2}{n_1} + \frac{\overline{Y}_2^2}{n_2}}}$$

is the test statistic, with the rejection region $RR = \{Z > z_\alpha\}$.

5. Since $n < 30$, this will be a T-test (one-sample), so that the test statistic

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$$

can be coded in R as

```
(mean(exam2$Y, na.rm=T)-5)/(sd(exam2$Y, na.rm=T)/sqrt(length(na.omit(exam2$Y))))
```

and the critical value $t_{\alpha/2,n-1}$ from $RR = \{|T| > t_{\alpha/2,n-1}\}$ can be coded in R as

```
qt(0.005, df=length(na.omit(exam2$Y))-1, lower.tail=F)
```

There are other acceptable answers.