Least Squares Fit and ANOVA Table

1. Here is a scatterplot of the sizes (in $100\ \text{ft}^2$) and prices (in $1000) for $n = 18$ apartments in the Village.
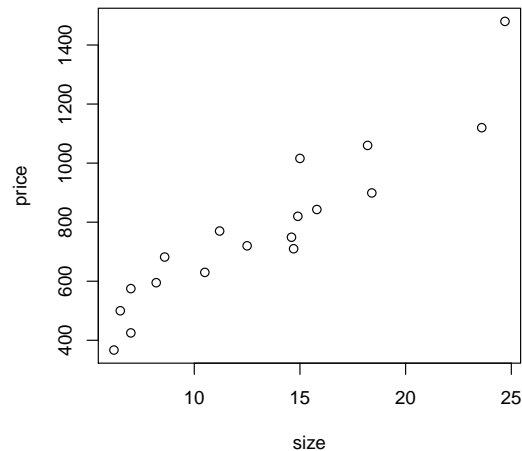
Here is the Minitab output for the least squares regression fit to the housing data. Some of the entries have been redacted (replaced by question marks).

```
The regression equation is
price = 182 + 45.0 size

Predictor          Coef  SE Coef       T       P
Constant         182.27    62.43    2.92   0.010
Size(100sqft)    44.950     4.370   10.29   0.000

S = 101.4    R-Sq = 86.9%    R-Sq(adj) = ????

Analysis of Variance
                 DF          SS            MS          F       P
Regression        1   1087556.8    1087556.84     ??????   ?????
Residual Error   16    164431.4      10276.96
Total            17   1251988.3
```
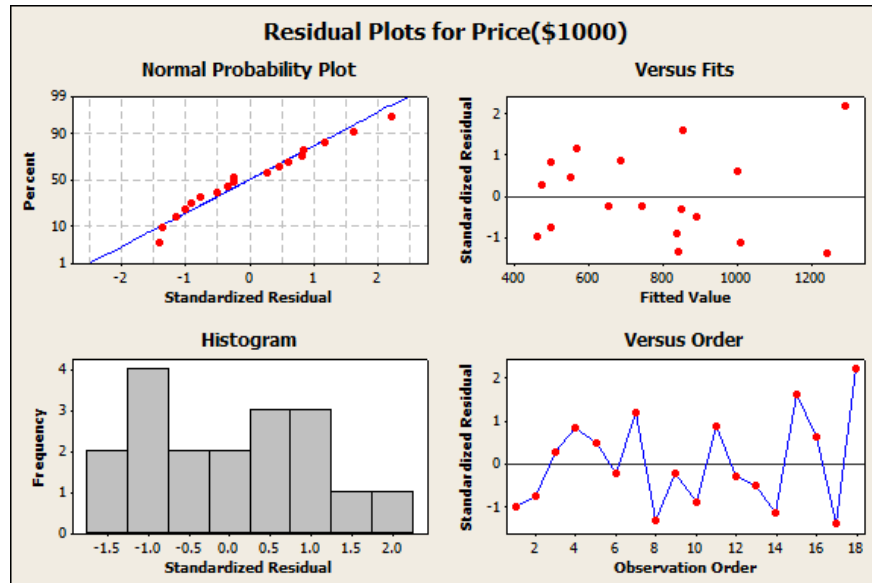
(a) Explain the meaning of the non-redacted values in the first 6 lines of the output (ending before the "Analysis of Variance" table).

(b) Explain the meaning of the values in the Analysis of Variance table.

Model Assumptions

2. Here are plots of the residuals from the least squares fit to the housing data.

Do the plots indicate any potential violations in assumptions? Specifically, answer the following questions.

(a) Do the residual errors look approximately normal?

> **Solution:** The normal probability plot and the histogram show that the residuals are approximately normal.

(b) Does the error variance look constant?

> **Solution:** The plot of residuals versus fitted value and residuals versus order hint that the variance of the residuals might be larger when the fitted value is big, but there is not enough data to say for certain.

(c) Is there any apparent dependence in the residuals?

> **Solution:** There is no clear pattern in the plot of residual versus fit or the plot of residual versus observation order. Thus, there is no apparent dependence in the residuals.

3. What is a "standardized" residual?

> **Solution:** A "standardized" residual is defined by $\hat{\varepsilon}_i/s$. This is "standardized" so that the sample standard deviation of the standardized residuals is approximately equal to 1.

4. Here is a plot of the residuals versus Size (x). Why is this plot nearly identical to the plot of residuals versus fits?

Residuals Versus Size(100sqft)
(response is Price($1000))
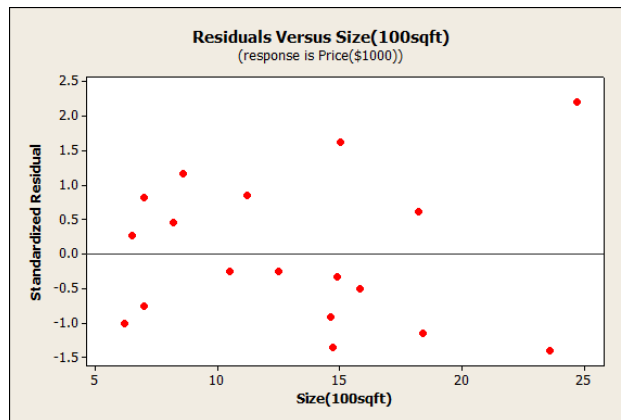
> **Solution:** Both plots have the same Y-axis. The X-axis on the plot of residuals versus size is x_i. The X-axis on the plot of residuals versus fits is $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, which is an affine transformation of x_i. Thus, the only difference in the plots is the values on the X-axis scale.

5. Here are some plots of the residuals from the fit of Price to Food for the Zagat data:
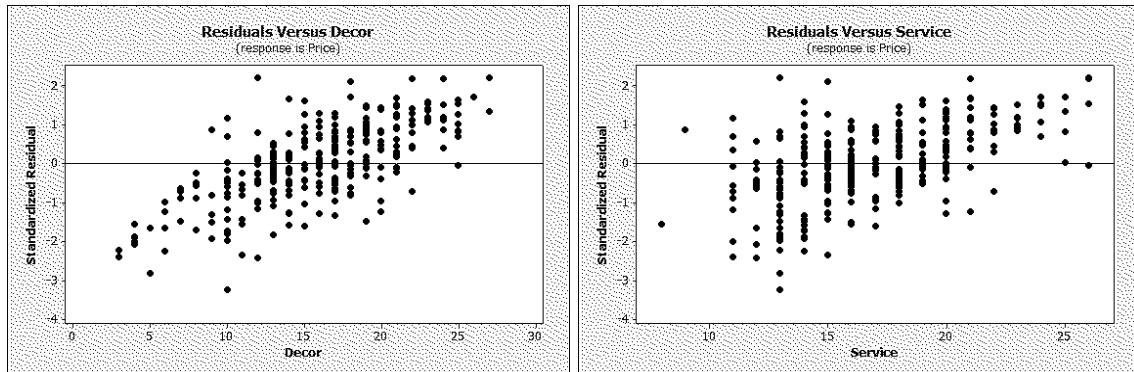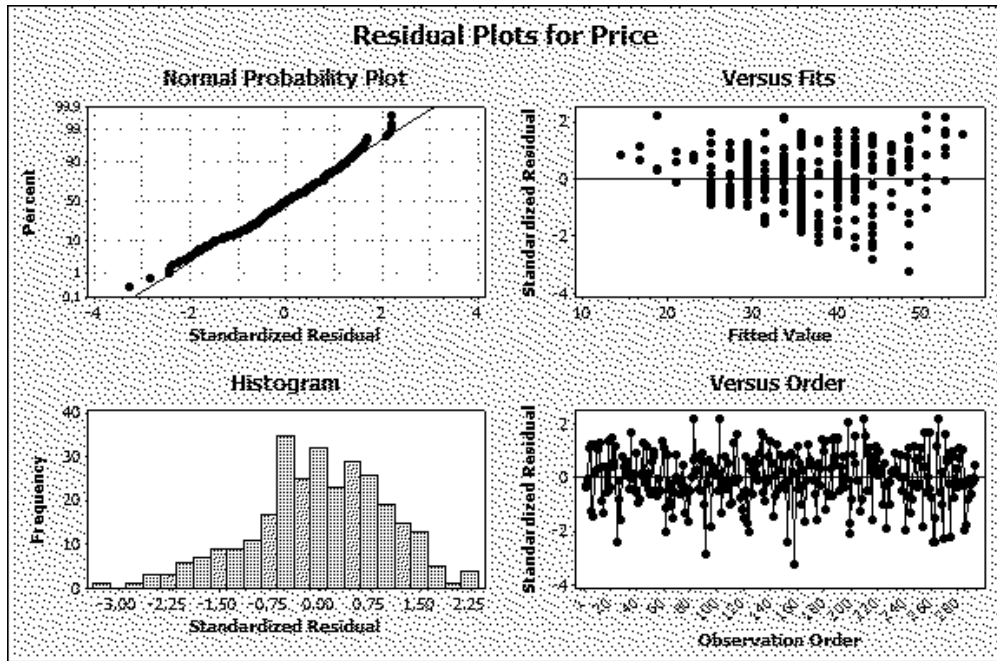
Use the plots to assess whether or not the four regression assumptions hold.

Solution: The normal probability plot and the histogram indicate that the residuals, are approximately normally-distributed. In the Residual verses Fitted Values, it looks like the mean value of the residual is approximately 0. This plot also shows that the error variance tends to increase when the fitted value increases. There is no apparaent pattern in the "Versus Order" plot, but there are clear trends in the "Versus Decor" and the "Versus Service" plots.

In summary, two assumptions are plausible: that the errors are normally distributed, and that the mean value of the error is zero. One assumption is violated, but only mildly so: that the error variance is constant. One assumption is in clear violation: that the errors are independent.