STA 114: Statistics

Practice Problems 3

1. For each question below <u>record</u> which stated option is correct and <u>give</u> a very brief explanation why.

   (a) Consider the model $X_1, \cdots, X_n \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, $\sigma \in (0, \infty)$. The p-value for testing $H_0 : \mu \leq 0$ vs $H_1 : \mu > 0$ is always exactly the half of the p-value for testing $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$.

   A. True B. False

   (b) If the p-value of observed data equals 0.03 based on a collection of level-α_0 tests $\{\delta_{\alpha_0} : 0 < \alpha_0 < 1\}$, then $\delta_{0.01}$ should reject H_0.

   A. True B. False

   (c) For a vector of counts $\underline{N} = (N_1, N_2, N_3, N_4)$ modeled as $\underline{N} \sim \mathsf{Multinomial}(n, p)$, $p = (p_1, p_2, p_3, p_4) \in \Delta_4$, the Pearson's chi-squared test statistic $Q(x)$ for testing $H_0 : p_1 = p_2 = p_3$ is approximately a χ^2_r random variable under H_0 for

   A. $r = 1$ B. $r = 2$ C. $r = 3$

   (d) In a three-way classification, 200 university students are simultaneously classified according to their college year (4 categories), major type (5 categories) and presidential candidate choice (3 categories). To test H_0: "college year is independent of major type and presidential candidate choice", one could use a Pearson's chi-squared test statistic $Q(x)$, whose distribution under H_0 is χ^2_r with

   A. $r = 59$ B. $r = 42$ C. $r = 24$ D. $r = 21$

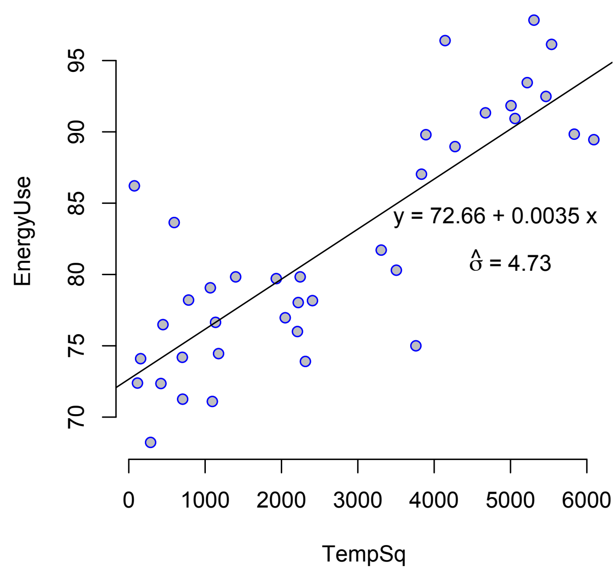
2. Consider data consisting of $n = 15$ observations $X_1, X_2, \cdots, X_n$. Below are three possible statistical models for this data and a pair of hypotheses for each model. For each setting <u>record</u> whether or not to reject H_0 according to the size-5% ML test if we observe $\bar{X} = 3.33$, $\sum_{i=1}^{n}(X_i - \bar{X})^2 = 55.73$.

   (a) *Model.* $X_i \overset{\text{IID}}{\sim} \mathsf{Poisson}(\mu)$ with $\mu > 0$ unknown.
   Hypotheses. $H_0 : \mu \leq 2.5$ against $H_1 : \mu > 2.5$.

   (b) *Model.* $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ with $-\infty < \mu < \infty$ unknown and $\sigma^2 = 2.5$.
   Hypotheses. $H_0 : \mu \leq 2.5$ against $H_1 : \mu > 2.5$.

(c) *Model.* $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ with both $-\infty < \mu < \infty$ and $\sigma^2 > 0$ unknown.
Hypotheses. $H_0 : \mu \leq 2.5$ against $H_1 : \mu > 2.5$.

3. The figure below shows data on the square of mean temperature (in square degrees F) and energy consumption (in KWH) for a certain building in University of Minnesota campus, for 38 months in 1988-1992.



Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, $\epsilon_i \overset{\text{IID}}{\sim} \mathsf{Normal}(0, \sigma^2)$ where (X_i, Y_i) denotes the square mean temperature and energy consumption for month i with $i = 1, \cdots, n = 38$. The least-squares line is shown in the plot along with the usual estimate $\hat{\sigma}$ of σ. For the observed data, the *standardized residuals* $e_i = (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)/\hat{\sigma}$, $i = 1, \cdots, n$, (arranged from smallest to largest) equal

```
-2.105 -1.565 -1.245 -1.111 -0.993 -0.871 -0.848 -0.833 -0.821 -0.737
-0.685 -0.634 -0.415 -0.342 -0.302 -0.294 -0.245 -0.242 -0.176  0.113
 0.146  0.259  0.269  0.281  0.325  0.373  0.434  0.458  0.478  0.515
 0.556  0.605  0.734  1.113  1.519  1.592  1.729  2.965
```

Perform a level-5% Pearson-χ^2 godness-of-fit test of how well the simple linear model fits the observed data.