

STAT310 - HWK Solution 13

1. (Bayesian Review)

(a) $E(X) = E(E(X|\theta))$

$$E(X|\theta) = \int_{-\infty}^{\infty} \frac{x\theta}{2} e^{-\theta|x|} dx = \frac{\theta}{2} \int_0^{\infty} x e^{-\theta x} dx + \frac{\theta}{2} \int_{-\infty}^0 x e^{\theta x} dx$$

$$\text{Let } x = -y, \text{ then } E(X|\theta) = \frac{\theta}{2} \int_0^{\infty} x e^{-\theta x} dx + \frac{\theta}{2} \int_{\infty}^0 y e^{-\theta y} dy$$

$$= \frac{\theta}{2} \int_0^{\infty} x e^{-\theta x} dx - \frac{\theta}{2} \int_0^{\infty} y e^{-\theta y} dy$$

$$= 0$$

Thus $E(X) = E(0) = 0$

(b) $\pi(\theta|s) \propto \pi(\theta) f(\theta|s) = e^{-\theta} \left(\frac{\theta}{2}\right)^n e^{-\theta \sum_{i=1}^n |x_i|} \propto \theta^n e^{-\theta(1 + \sum_{i=1}^n |x_i|)}$
 Thus, $\theta|s \sim \text{Gamma}(n + 1, 1 + \sum_{i=1}^n |x_i|)$

(c) Posterior density: $\theta|s \sim \text{Gamma}(6, 7.46)$

Mean: $\frac{\alpha}{\lambda} = \frac{6}{7.46} = 0.8043$

Mode: The value of θ that occurs the most frequently. In this problem, it is the maximum likelihood estimator of θ , mode = $\frac{\alpha-1}{\lambda} = \frac{5}{7.46} = 0.6702$

2. (Likelihood ratio test review)

$H_0 : \theta_1 = \theta_2 = \theta_3 = \theta$ vs. $H_A : \text{not all } \theta_i|s \text{ are equal}$

Under H_0 : $L(\theta|s) = \prod_{i=1}^3 \theta(1-\theta)^{x_i} = \theta^3(1-\theta)^{\sum_{i=1}^3 x_i} \Rightarrow \hat{\theta} = \frac{3}{3 + \sum_{i=1}^3 x_i} = 0.3$

Then, $l(\theta|s)_0 = \log L(\theta|s) = -6.1086$

Under H_A : $L(\theta|s) = \prod_{i=1}^3 \theta_i(1-\theta_i)^{x_i} \Rightarrow \hat{\theta}_i = \frac{1}{x_i+1}$, where $\hat{\theta}_1 = 0.5, \hat{\theta}_2 = 1, \hat{\theta}_3 = \frac{1}{7}$

Then, $l(\theta|s)_f = -4.2571$

The test statistics is $-2\log(\Lambda) = -2(l_0 - l_f) = -2(-6.1086 + 4.2571) = 3.7031$

p -value = $P(X^2 > 3.7031) = 0.157$, where $X^2 \sim \chi^2(2)$.

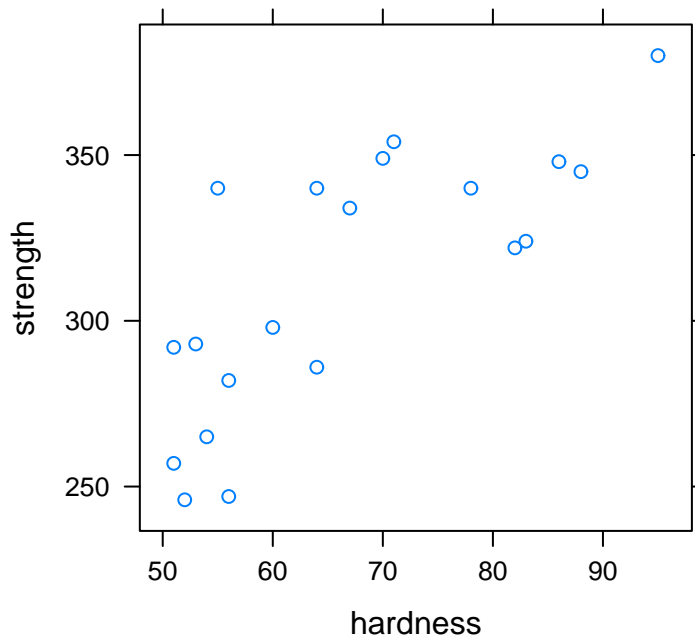
Thus, we have no evidence against the null hypothesis.

3. (Simple linear regression)

```
(a) > hardness = c(53, 70, 78, 55, 64, 71, 82, 67, 56, 86, 60, 51, 95, 88, 51, 54, 52, 64, 83, 56)
> library(lattice)
> xyplot(strength~hardness, main='Plot of strength vs. hardness')
```

The plot tells us that there is a positive relationship between strength and hardness, and this relationship seems to be linear. It makes sense to fit a linear regression model.

Plot of strength vs. hardness



```
(b) > model=lm(strength~hardness)
> summary(model)
```

Call: lm(formula = strength ~ hardness)

Residuals:

Min	1Q	Median	3Q	Max
-41.940	-20.376	-2.395	15.704	53.204

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	168.854	28.981	5.826	1.61e-05 ***
hardness	2.144	0.425	5.046	8.41e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.08 on 18 degrees of freedom Multiple R-squared: 0.5858, Adjusted R-squared: 0.5628 F-statistic: 25.46 on 1 and 18 DF, p-value: 8.405e-05

Estimate of the coefficients: $\hat{\beta}_1 = 168.854, \hat{\beta}_2 = 2.144$

Estimate of the standard deviation of the coefficients: $se(\hat{\beta}_1) = 28.981, se(\hat{\beta}_2) = 0.425$

(c) The means, standard deviations and correlation coefficient are showed below

```
> mean(strength)
[1] 312.1
```

```

> mean(hardness)
[1] 66.8
> sd(strength)
[1] 39.45004
> sd(hardness)
[1] 14.08097
> cor(strength, hardness)
[1] 0.765406

```

Verify the equations: $\hat{\beta}_2 = \rho \frac{s_y}{s_x} = 0.765406 * \frac{39.45004}{14.08097} = 2.144$, where ρ is the correlation coefficient between hardness and strength.

- (d) A 95% confidence interval is $\hat{\beta}_2 \pm t_{0.025,18} * se(\hat{\beta}_2) = 2.144 \pm 2.101 * 0.425 = [1.252, 3.037]$
- (e) Verify that the numerical estimates of the standard errors of the regression coefficients from summary() match those found by plugging into equations from Theorem 10.3.6.
- (1) $B_1 \sim N(\beta_1, \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i - \bar{x})^2}))$, where σ^2 is estimated by 26.08 from the R output. Then, $sd(B_1) = \sqrt{26.08^2 * (\frac{1}{20} + \frac{66.8^2}{\sum(x_i - 66.8)^2})} = 28.981 = se(\hat{\beta}_1)$ from the R output.
- (2) $B_2 \sim N(\beta_2, \sigma^2/\sum(x_i - \bar{x})^2)$. Then, $sd(B_2) = \sqrt{26.08^2/\sum(x_i - 66.8)^2} = 0.425 = se(\hat{\beta}_2)$ from the R output.
- (f) $H_0 : \beta_2 = 0$
 Test statistic of the t test is $\frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = 2.144/0.425 = 5.046$, then $p - value = 2 * P(T > 5.046) = 8.41e - 05$. Thus, we reject the null hypothesis. The slope is significantly different from zero.
- (g)

```

> ratio = var(residuals(model))/var(strength)
> ratio
[1] 0.4141538
> cor(strength, hardness)^2
[1] 0.5858462

```

Thus, ratio = 1 - variance of correlation coefficient