# MATH 308 Assignment 18: Exercises 8.5

Nakul Joshi

April 22, 2014

## 4

p=$P(X >= 20)$ = `1-ppois(19,lambda=15)` $\approx$ 12.5%. This implies that more than 10% of the months will have a birth rate as extreme as the one observed, which means that we cannot reject the null ($\lambda = 15$) even at the 5% significance level.

## 6

### a)

Running the test gives a $p$-value of $10^{-4}$. This lets us reject the null hypothesis (that the difference in means is zero) at the 1% significance level.

### b)

The random assignment of seedlings to plots tends to remove confounding effects. Thus, the result hints at a causal relationship between presence of competition and diameter change.

## 11

Running the chi-squared test gives a $p$-value of 0.012, allowing us to reject the null at the 5% significance level.

## 14

A Type I error would be incorrectly concluding that the drug is effective when in fact it is not. This would lead to to patients wasting money on a useless drug.

A Type II error would be failing to conclude that the drug is effective, when in fact it is. This would lead to a waste of the research money put into the drug, since the pharmaceutical company would then not be able to sell it.

## 16

Under the null, $\overline{X} \sim \mathcal{N}(25, 16/30)$. Thus $Z = \frac{\overline{X}-25}{4/\sqrt{30}}$, where $Z$ is a standard normal random variable. Thus, the null is rejected if $Z \geq z_{1-\alpha} \implies \frac{\overline{X}-25}{4/\sqrt{30}} \geq z_{0.95} \implies \overline{X} \geq 25 + \frac{4z_{0.95}}{\sqrt{30}} \equiv C$.

But $\overline{X} \sim \mathcal{N}\left(27, 16/\sqrt{30}\right) \implies Z = \frac{\overline{X}-27}{4/\sqrt{30}}$. So,

$$\begin{aligned}
1 - \beta &= P(\overline{X} \geq C | \mu = 27) \\
&= P\left(\frac{4Z}{\sqrt{30}} + 27 \geq C\right) \\
&= P\left(Z \geq \frac{\sqrt{30}(C-27)}{4}\right) \equiv P(Z \geq C_1) \\
&= 1 - P(Z < C_1) \approx 86.3\%
\end{aligned}$$

## 17

The null is rejected if $\overline{X} \geq \mu_0 + \frac{z_{1-\alpha}\sigma}{\sqrt{n}} \equiv C$, where $\mu_0 = 1, \alpha = 0.01$ and $\sigma = 0.3$. So,

$$\begin{aligned}
1 - \beta &= P(\overline{X} \geq C | \mu = \mu_1) \\
&= P\left(\frac{Z\sigma}{\sqrt{n}} + \mu_1 \geq C\right) \\
&= P\left(Z \geq \frac{\sqrt{n}(C - \mu_1)}{\sigma}\right) \equiv P(Z \geq C_1) \\
&= 1 - P(Z < C_1) \\
&\implies \beta = P(Z < C_1) \\
\implies z_\beta &= C_1 = \frac{\sqrt{n}(C - \mu_1)}{\sigma} \\
\implies \frac{z_\beta \sigma}{\sqrt{n}} + \mu_1 &= C = \frac{z_{1-\alpha}\sigma}{\sqrt{n}} + \mu_0 \\
\implies n &= \left(\frac{\sigma(z_\beta - z_{1-\alpha})}{\mu_0 - \mu_1}\right)^2 = 8
\end{aligned}$$

## 18

$$n = \left(\frac{\sigma(z_\beta - z_{1-\alpha})}{\mu_0 - \mu_1}\right)^2 = 1$$

## 25

$f(x) = \lambda e^{-\lambda x} \implies F(x) = \int_0^x \lambda e^{-\lambda t}\,\mathrm{d}t = 1 - e^{-\lambda t}$

### a)

$$\begin{aligned}
F_{X,\min}(x) &= \prod_{i=1}^{15} P(X_i \leq x) \\
&= (1 - F(x))^{15} \\
&= e^{-15\lambda x} \\
\implies \alpha &= P(X_{\min} \geq 1 | \lambda = 1/5) \\
&= 1 - F_{X,\min}(1)|_{\lambda=\frac{1}{5}} \\
&= 1 - e^{-3} \approx 95\%
\end{aligned}$$

### b)

$$\begin{aligned}
1 - \beta &= P(X_{\min} \geq 1 | \lambda = 1/25) \\
&= 1 - F_{X,\min}(1)|_{\lambda=\frac{1}{25}} \\
&= 1 - e^{-3/5} \approx 45\%
\end{aligned}$$

## 36

The most powerful test simply always rejects the null. Thus, the critical region is $X_1, X_2 \in [0, n]$.

## 37

Under the null, the likelihood is maximised at $\mu = \mu_0$ and $\sigma^2 = \Sigma(X_i - \mu_0)^2/n \equiv \hat{\sigma}^2$. Thus, the likelihood under the null is

$$L_{H,0} = \left(\frac{1}{\hat{\sigma}\sqrt{2\pi}}\right)^n \exp\left(\frac{-\Sigma(X_i - \mu_0)^2}{2\hat{\sigma}^2}\right) = \frac{e^{-n/2}}{(\hat{\sigma}\sqrt{2\pi})^n}$$

To maximise the likelihood, we simply use $\mu = \overline{X}$ and $\sigma = \sqrt{\Sigma(X_i - \overline{X})^2/n}$, giving

$$L = \left(\frac{1}{S\sqrt{2\pi}}\right)^n \exp\left(\frac{-\Sigma(X_i - \overline{X})^2}{2S^2}\right) = \frac{e^{-n/2}}{(S\sqrt{2\pi})^n}$$

The likelihood ratio is therefore $(S/\hat{\sigma})^n$. So, we reject if

$$\begin{aligned}
\left(\frac{\Sigma(X_i - \overline{X})^2}{\Sigma(X_i - \mu_0)^2}\right)^{n/2} \leq c &\implies \frac{\Sigma(X_i - \overline{X})^2}{\Sigma(X_i - \mu_0)^2} \leq c_1 \\
\implies c_1 \leq \frac{\Sigma(X_i - \mu_0)^2}{\Sigma(X_i - \overline{X})^2} &= \frac{\Sigma(X_i - \overline{X} + \overline{X} - \mu_0)^2}{\Sigma(X_i - \overline{X})^2} \\
&\leq 1 + \frac{n(\overline{X} - \mu_0)^2 - 2(\overline{X} - \mu_0)\cancel{\Sigma(X_i - \overline{X})}}{\Sigma(X_i - \overline{X})^2} \\
&\leq 1 + \frac{n}{\Sigma(X_i - \overline{X})^2}(\overline{X} - \mu_0)^2 \\
\implies c_2 \leq \sqrt{\frac{n}{\Sigma(X_i - \overline{X})^2}}(\overline{X} - \mu_0) &= \frac{\overline{X} - \mu_0}{S/\sqrt{n}}
\end{aligned}$$

which is a $t$-statistic. Thus, the test is a one-sided $t$-test. $\square$

# MATH 308 Assignment 19:<br>Exercises 9.7

Nakul Joshi

April 22, 2014

## 7

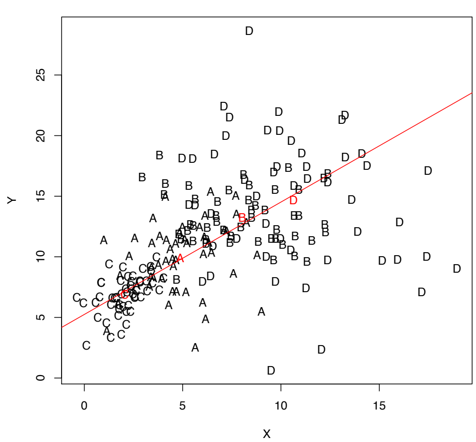
### a)

$\rho \approx 0.50$

### b)

| | A | B | C | D |
|---|---|---|---|---|
| $\overline{X}$ | 4.88 | 8.03 | 2.06 | 10.64 |
| $\overline{Y}$ | 9.91 | 13.24 | 6.89 | 14.68 |

### c)




$\rho = 0.99$, which is much higher than the result from (a).

## 9

$$
\begin{aligned}
g(a,b) &= \Sigma(y_i - \hat{y}_i)^2 \\
\implies \frac{\partial g}{\partial a} &= \Sigma\left(2(y_i - \hat{y}_i)\frac{\partial(y_i - (a + bx_i))}{\partial a}\right) \\
\implies 0 &= -2\Sigma(y_i - \hat{y}_i) \\
\implies \Sigma(y_i - \hat{y}_i) &= 0 \quad \square
\end{aligned}
$$

## 10

### a)

$$
\begin{aligned}
r &= \frac{ss_{xy}}{\sqrt{ss_x ss_y}}, b = \frac{ss_{xy}}{ss_x} \\
\Sigma(\hat{y}_i - \overline{y})^2 &= \Sigma(a + bx_i - (a + b\overline{x}))^2 \\
&= b^2\Sigma(x_i - \overline{x})^2 = b^2 ss_x = \frac{ss_{xy}^2}{ss_x} \\
&= r^2 ss_y \\
\implies s_{\hat{y}} &= \sqrt{\frac{\Sigma(\hat{y}_i - \overline{y})^2}{n-1}} = r\sqrt{\frac{ss_y}{n-1}} \\
&= r\sqrt{\frac{\Sigma(y_i - \overline{y})^2}{n-1}} = rs_y \quad \square
\end{aligned}
$$

## b)

$$\overline{e} = \Sigma(y_i - \hat{y}_i)/n = 0$$
$$\implies s_e = \sqrt{\frac{\Sigma(y_i - \hat{y}_i - 0)^2}{n-1}}$$
$$\implies (n-1)s_e^2$$
$$= \Sigma(y_i - \hat{y}_i)^2$$
$$= \Sigma(y_i - \overline{y} + \overline{y} - \hat{y}_i)^2$$
$$= \Sigma(y_i - \overline{y})^2 + \Sigma(\overline{y} - \hat{y}_i)^2$$
$$- 2\Sigma(y_i - \overline{y})(\hat{y}_i - \overline{y})$$
$$= ss_y + r^2 ss_y - 2\Sigma(y_i\hat{y}_i - \cancel{y_i\overline{y}} - \hat{y}_i\overline{y} + \cancel{\overline{y}^2})$$
$$= ss_y(1 + r^2) - 2\Sigma(\hat{y}_i\cancel{(y_i - \overline{y})})$$
$$\implies s_e = \sqrt{\frac{(1+r^2)ss_y}{n-1}} = \sqrt{1+r^2}s_y \quad \square$$

# 11

## a)

$$b = r\frac{s_w}{s_h} \approx 1.6a = \overline{w} - b\overline{h} \approx 20.1$$
$$\implies \hat{w} = a + bh \approx 20.1 + 1.6h$$

## b)

$\hat{w}(5 \text{ ft}) = \hat{w}(60 \text{ in}) = 116.5$ pounds

## c)

$r^2 \approx 0.56$, which means that the model explains 56% of the variability of the weights of the individuals.

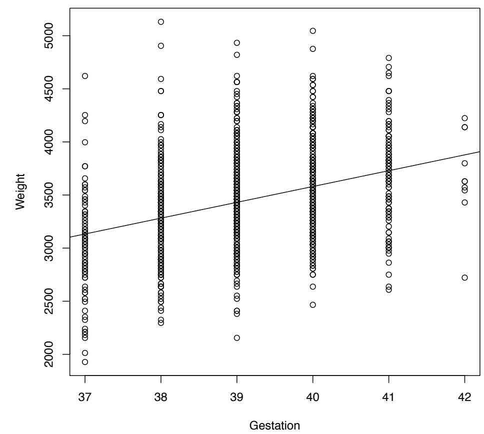
# 17

## a)

Correlation is $\approx 0.35$

## b)

$y = -2380 + 149x$

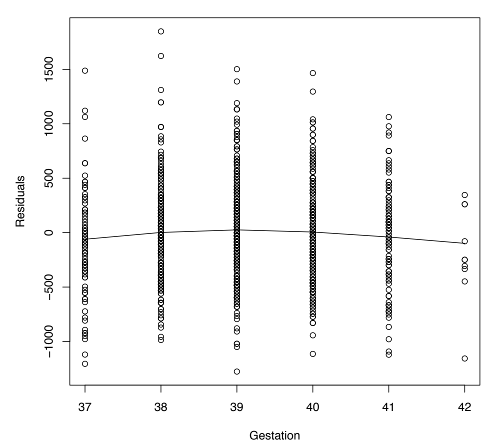



## c)

The slope is interpreted as 'a week increase in gestation period is correlated with a 149g increase in weight'.

The $r^2 \approx 0.12$ is interpreted as 'the linear model explains 12% of the variation in weights'.

## d)




The residual plot shows a curve peaking at week 39. The linear model is therefore inappropriate.

## 18

### a)

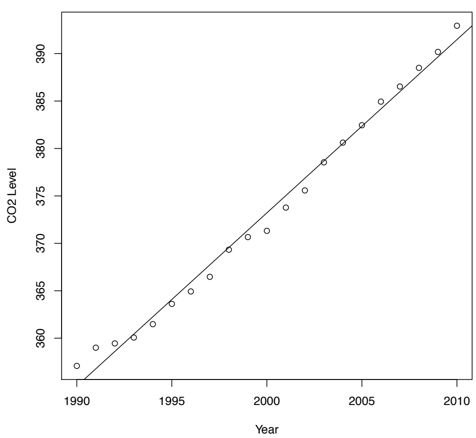
457.4 g

### b)

$1-\alpha = 0.95 \implies 1-\alpha/2 = 0.975$
$\implies t \equiv t_{n-2,1-\alpha/2} = t_{1007,0.975} \approx 1.96$
$\therefore$ 95% CI: $\hat{\beta} \pm \frac{tS}{\sqrt{ss_x}} = 149 \pm 24.8$
$= (124.2, 173.8)$

## 21

### a)

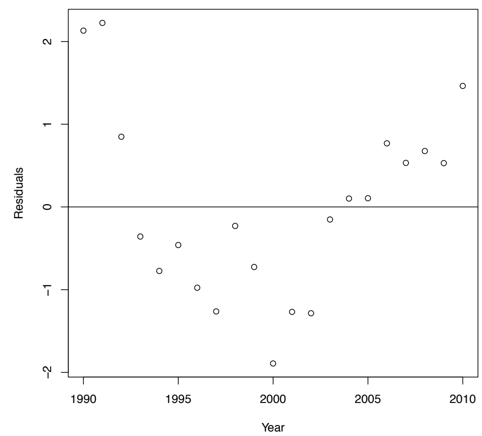



The data is strongly linear, with a correlation of over 0.99.

### b)

$$y = -3279.593 + 1.826x$$

### c)




At first, it appears that there is a pattern to the residuals, decreasing until year 2000 and then increasing again. However, looking at the scale of the $y$-axis we see that the largest residual is only 2, but the $CO_2$ Levels are of order 350.

Thus, the apparent pattern can be put down to random error and ignored, making a linear model appropriate.

# MATH 308 Assignment 20: Power Function

Nakul Joshi

April 22, 2014

## 1

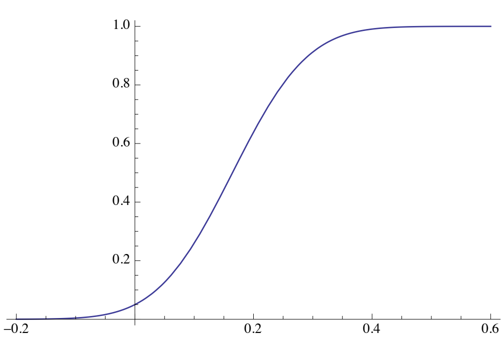
The null is rejected if $\overline{X} > \mu_0 + \frac{z_{1-\alpha}}{\sqrt{n}} \equiv C$. But, $\overline{X} \sim \mathcal{N}\left(\mu, 1^2/n\right) \implies \overline{X} = \mu + \frac{Z}{\sqrt{n}}$

$$\begin{aligned}\therefore \beta(\mu) &= P(\overline{X} > C)\\ &= 1 - P(\overline{X} \leq C)\\ &= 1 - P\left(\mu + \frac{Z}{\sqrt{n}} \leq C\right)\\ &= 1 - P(Z \leq (C-\mu)\sqrt{n})\\ &= 1 - \Phi((C-\mu)\sqrt{n})\\ &= 1 - \Phi(z_{1-\alpha} + (\mu_0 - \mu)\sqrt{n})\end{aligned}$$

## 2

Setting $\alpha = 0.05, \mu_0 = 0$ and $n = 100$, we get

$$\beta(\mu) = 1 - \Phi(z_{0.95} - 10\mu)$$




## 3

$$\begin{aligned}0.8 &= 1 - \Phi(z_{0.95} + (0-1)\sqrt{n})\\ \implies z_{0.2} &= z_{0.95} - \sqrt{n}\\ \implies n &= (z_{0.95} - z_{0.2})^2 \approx 7\end{aligned}$$