Sergey Ivanov

8.66

$X_1, X_2$ - independent random variables

$p(X_i|\theta) = p\theta$

$p(\theta) = 1$ - uniform on $[0,1]$

a) $p(\theta|X_1,X_2) = \dfrac{p(X_1,X_2|\theta)p(\theta)}{\int_0^1 p(X_1,X_2|\theta)p(\theta)d\theta} =$

$= \dfrac{\theta^2 \cdot 1}{\int_0^1 \theta^2 d\theta} = 3\theta^2$

b)

$p(X_3) = p(X_3|\theta) \cdot p(\theta|X) = \theta \cdot 3\theta^2 = 3\theta^3$

9.24

a) Likelihood $L(p_0 = \frac{1}{2}) = \binom{n}{x} p_0^x \cdot (1-p_0)^{n-x} = \binom{n}{x} .5^n$

$L(p = \hat{p}) = \binom{n}{x} \hat{p}^x (1-\hat{p})^{n-x} \to \max \Rightarrow \hat{p} = \frac{x}{n}$

$\Lambda = \dfrac{\binom{n}{x} .5^n}{\binom{n}{x}\left(\frac{x}{n}\right)^x \left(1-\frac{x}{n}\right)^{n-x}} = \dfrac{n^n .5^n}{n^x \left(\frac{x}{n}\right)^x n^{n-x} \left(1-\frac{x}{n}\right)^{n-x}} =$

$= \dfrac{\left(\frac{n}{2}\right)^n}{x^x (n-x)^{n-x}}$

b)

Let $g(x) = \dfrac{1}{x^x (n-x)^{n-x}}$. Notice $g(x) = g(n-x) \Rightarrow$

$\Rightarrow y = x - \frac{n}{2} \to g\left(\frac{n}{2}+y\right) = g\left(\frac{n}{2}-y\right)$.

Let $h(y) = \log\left(g\left(\frac{n}{2}+y\right)\right)$. Notice if $h(y)$ is decreasing, then likelihood is also decreasing; and, if $y$ is increasing, then $x$ is increasing. So if we show that for $y \geq 0$

$h(y)$ is non-increasing, then due to symmetry of $g$, as $|y|$ gets big, likelihood ratio $\Lambda$ gets small, that means rejecting $H_0$.

$$\log(g(\tfrac{n}{2}+y)) = -(\tfrac{n}{2}+y)\log(\tfrac{n}{2}+y) - (\tfrac{n}{2}-y)\log(\tfrac{n}{2}-y) = h(y)$$

$$h'(y) = -\log(\tfrac{n}{2}+y) - (\tfrac{n}{2}+y) + \log(\tfrac{n}{2}-y) + (\tfrac{n}{2}+y) =$$

$$= \log(\tfrac{n}{2}-y) - \log(\tfrac{n}{2}+y) \leq 0$$ for $y \geq 0$, since log is decreasing

c)

$$\alpha = Pr(|X-\tfrac{n}{2}|>k \mid H_0) = Pr(X<-k+\tfrac{n}{2} \mid H_0) + Pr(X>k+\tfrac{n}{2} \mid H_0),$$

where $X \sim Bi(n, \tfrac{1}{2})$

d) $n=10$
$k=2$ $\Rightarrow$ $\alpha = Pr(X<-2+5 \mid H_0) + Pr(X>2+5) =$

$= Pr(X<3) + Pr(X>7) =$

$= Pr(X=\{0,1,2,8,9,10\}) = 0.065$

e) $X \sim Bi(100, 0.5)$

$\alpha = Pr(X<40 \mid H_0) + Pr(X>60 \mid H_0)$

$E(X) = np = 50$ $Var(X) = np(1-p) = 25$

$Pr(X<40 \mid H_0) = P(\frac{X-50}{5} < \frac{-10}{5} = -2) \approx \Phi(-2) = 0.0228$

Because symmetric around $E(X)$ $\Rightarrow$ $Pr(X>60 \mid H_0) = 0.0228$

So $\alpha = 0.0455$

## 9.36

$X_i$ – number of suicides per month of $i$

$H_0: p_i = \frac{d_i}{365} \; \forall i$ , $d_i$ – number of days in month $i$

$H_A: p_i \neq \frac{d_i}{365}$

$p_i \cdot T$ – expected number of suicides in month $i$
where $T = 23480$, total number

$O_i = n\hat{p}_i$ $E_i = n p_i(\hat{\theta})$
observed expected

then $X^2$ statistic is $\Sigma \frac{(O-E)^2}{E}$. Its null distribution is chi-square with 11 df. The p-value $\approx 0$.

$X^2$:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | ← month |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| −8.11 | −0.08 | −1.26 | 14 | 5.3 | 1.36 | −5.76 | 0.446 | −0.00 | 0.717 | 1.20 | −9.17 | ← |


⇒ for winter season, # of suicides is less than expected. for April, May, June, the number of suicides is higher than expected.

## 14.7

Let $Y = X\beta + e$, where $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_n \end{pmatrix}$

$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$, $e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$

a)

$$Cov(e) = \sigma^2 \begin{pmatrix} \delta_1^2 & & 0 \\ & \delta_2^2 \ddots & \\ 0 & & \delta_n^2 \end{pmatrix} = \sigma^2 W$$

$$W^{-\frac{1}{2}} = \begin{pmatrix} \frac{1}{\delta_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\delta_n} \end{pmatrix}$$

$$W^{-\frac{1}{2}}Y = W^{-\frac{1}{2}}X\beta + W^{-\frac{1}{2}}e \Leftrightarrow Y' = X'\beta + \delta$$

$$Cov(\delta) = Cov(W^{-\frac{1}{2}}e) = Cov(W^{-\frac{1}{2}}Cov(e)W^{-\frac{1}{2}} = W^{-\frac{1}{2}}\sigma^2 W W^{-\frac{1}{2}} = \sigma^2 I_n$$

So, we have $Y' = X'\beta + \delta$ satisfying the standard statistical model

b)

$$\beta' = (X'^T X')^{-1} X'^T Y' = ((W^{-\frac{1}{2}}X)^T(W^{-\frac{1}{2}}X))^{-1}(W^{-\frac{1}{2}}X)^T W^{-\frac{1}{2}}Y = (X^T W^{-1} X)^{-1}(X)^T W^{-1} Y$$

c)

$$\underset{\beta}{\arg\min} \|Y' - X'\beta\|^2 = \underset{\beta}{\arg\min} \|W^{-\frac{1}{2}}(Y - X\beta)\|^2 = \sum_{i=1}^{n} \delta_i^{-2}(y_i - x_i\beta_1 - \beta_0)^2 \to \min$$

d) The covariance of $\beta$ is given by:

$$Cov(\beta) = \sigma^2 (X'^T X')^{-1} = \sigma^2 (X^T W^{-1} X)^{-1}$$

9.47 Sergey Ivanov

$X \sim Pois(\lambda)$

$Y = f(X) = \sqrt{X}$

$\sigma_Y^2 = \sigma^2(\mu) \cdot (f'(\mu))^2 = \mu \cdot \left(\frac{1}{2} \frac{1}{\sqrt{\mu}}\right)^2 = \frac{1}{4}\mu \frac{1}{\mu} = \frac{1}{4} = const$

c) $Y$ is variance-stabilizing

9.64

a,b) Look at the plots

Male temperature looks like as a straight line => it has normal distribution

Female temperature is smaller than expected for a normal distribution and in the right tail they are bigger, indicating that the tails of temp. decrease less quickly than the tails of temp of normal distribution.

Male heart rate deviates a little bit from normal distribution at the tails; while female rate has convex shape in the left tail, and concave shape in the right tail.

c)

$H_0: \mu = \mu_0 = 98.6$

$H_A: \mu \neq \mu_0$

$\Lambda = \frac{f(x_1 \dots x_n \mid \mu_0, \hat{\sigma})}{f(x_1, \dots x_n \mid \bar{X}, \hat{\sigma})}$, where $\bar{X} = \mu_x$ and $\hat{\sigma} = \sqrt{\frac{1}{n}\sum(X_i - \bar{X})^2}$ using MLE.

then $\Lambda = \exp\left(-\frac{1}{2\hat{\sigma}^2}\left(\sum_{i=1}^{n}(X_i - \mu_0)^2 - \sum_{i=1}^{n}(X_i - \bar{X})^2\right)\right) =$

$\sum^{n}(X_i - \mu_0)^2 = \sum^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2$

$= \exp\left(-\frac{1}{2\hat{\sigma}^2} n(\bar{X} - \mu_0)^2\right)$

$$\Rightarrow \quad -2\log\Lambda = \left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\hat{\sigma}}\right)^2 \sim \chi^2_1$$

Also, $\dfrac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t_{n-1}$, where $S = \sqrt{\frac{n}{n-1}}\,\hat{\sigma}$

$$-2\log\Lambda = \left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{\hat{\sigma}}\right)^2 = \frac{n}{n-1}\left(\frac{\sqrt{n}(\bar{X}-\mu_0)}{S}\right)^2 \overset{d}{=} \frac{n}{n-1}t^2_{n-1}$$

If we want significance level of $\alpha$, then we set $\sqrt{\frac{n}{n-1}}\,c = t_{n-1}\left(\frac{\alpha}{2}\right)$

$$P\left(\frac{n}{n-1}t^2_{n-1} \geq c\right) = P\left(|t_{n-1}| \geq \sqrt{\frac{n}{n-1}}\,c\right)$$

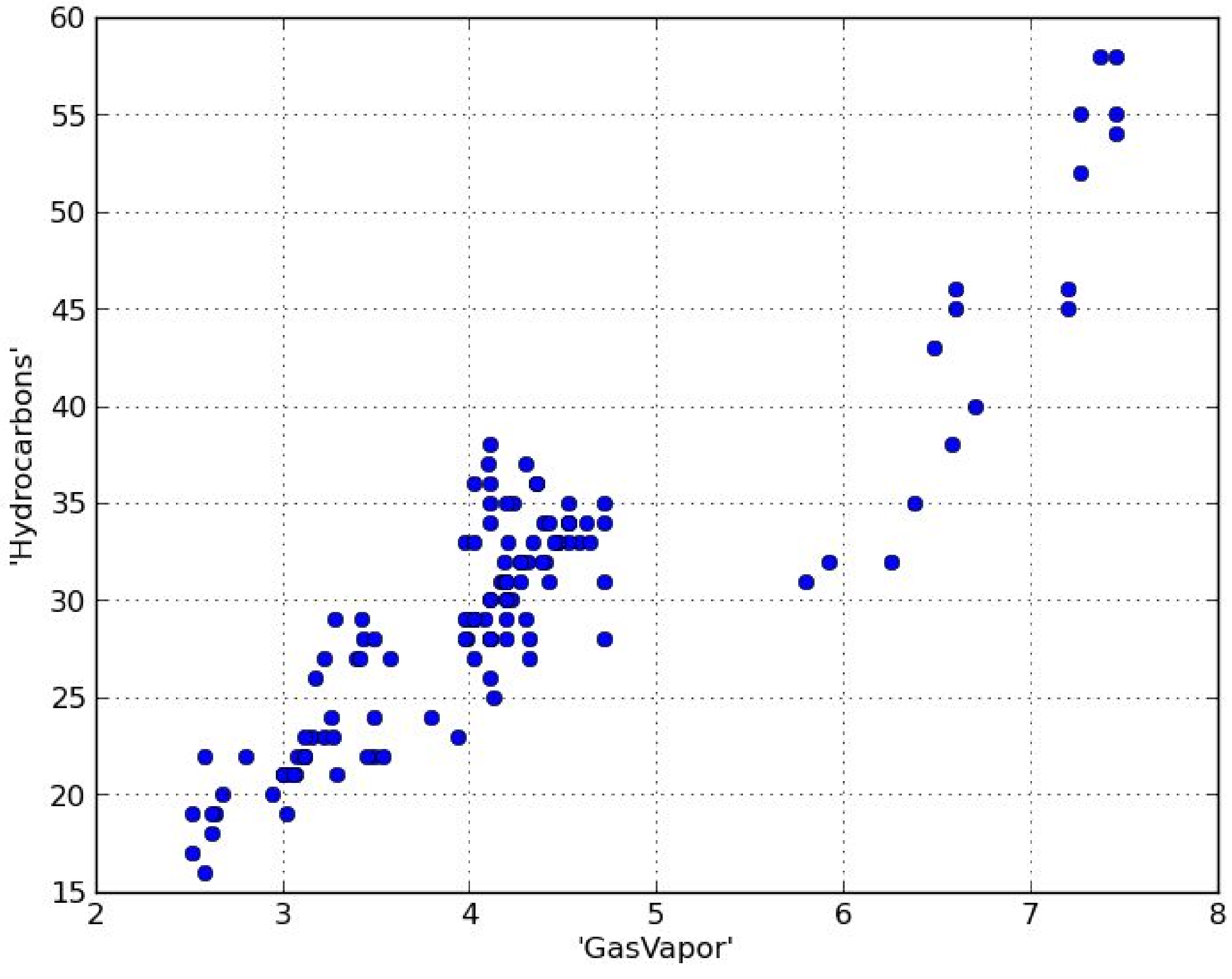
Then, rejection region: $\left|\dfrac{\sqrt{n}(\bar{X}-\mu_0)}{S}\right| > t_{n-1}\left(\frac{\alpha}{2}\right)$ exact test

for asymptotic test we have:

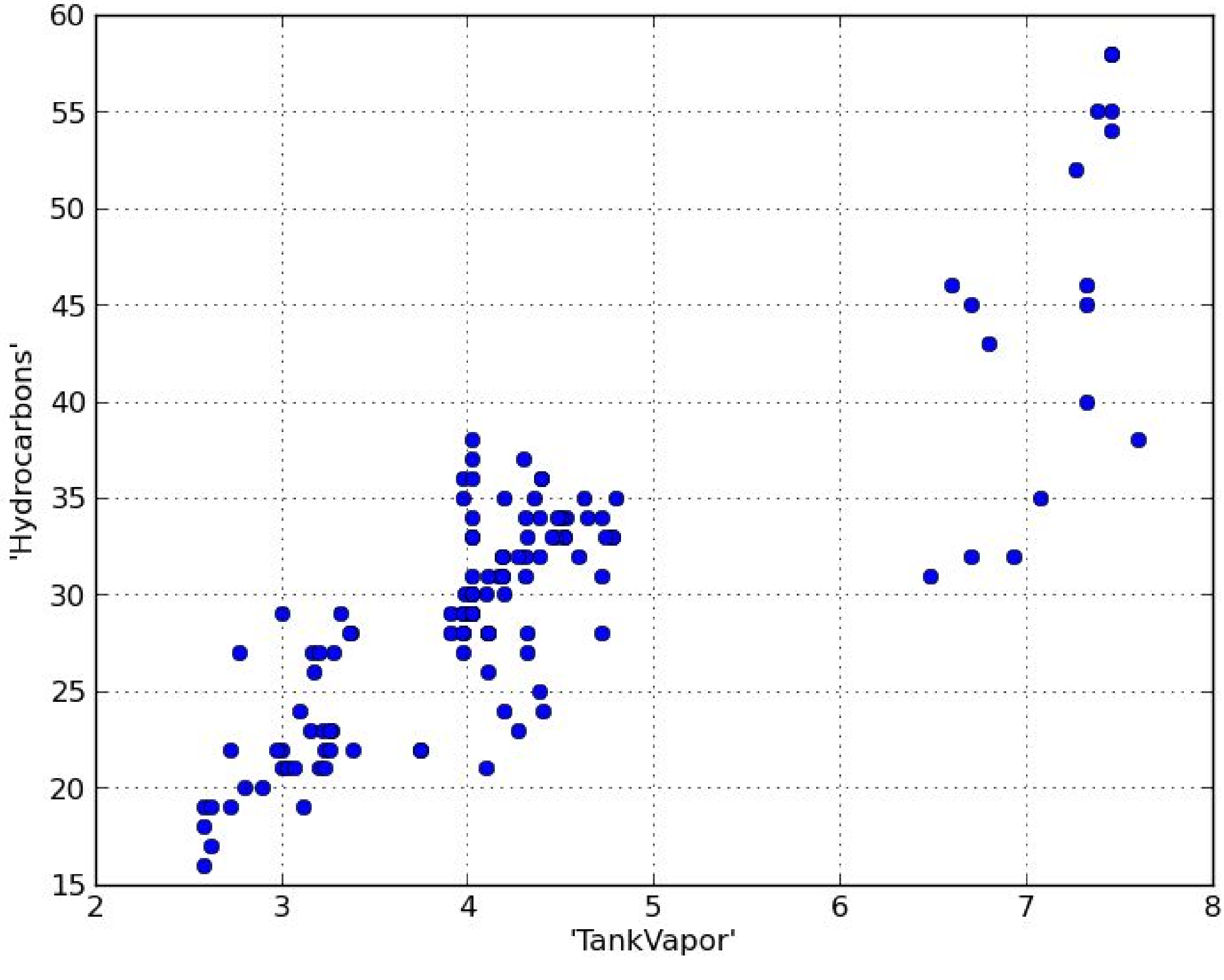
$$\left\{\left|\frac{\sqrt{n}(\bar{X}-\mu_0)}{\hat{\sigma}}\right| > z\left(\frac{\alpha}{2}\right)\right\} = \left\{\left|\frac{\sqrt{n}(\bar{X}-\mu_0)}{S}\right| > \sqrt{\frac{n}{n-1}}\,z\left(\frac{\alpha}{2}\right)\right\}$$

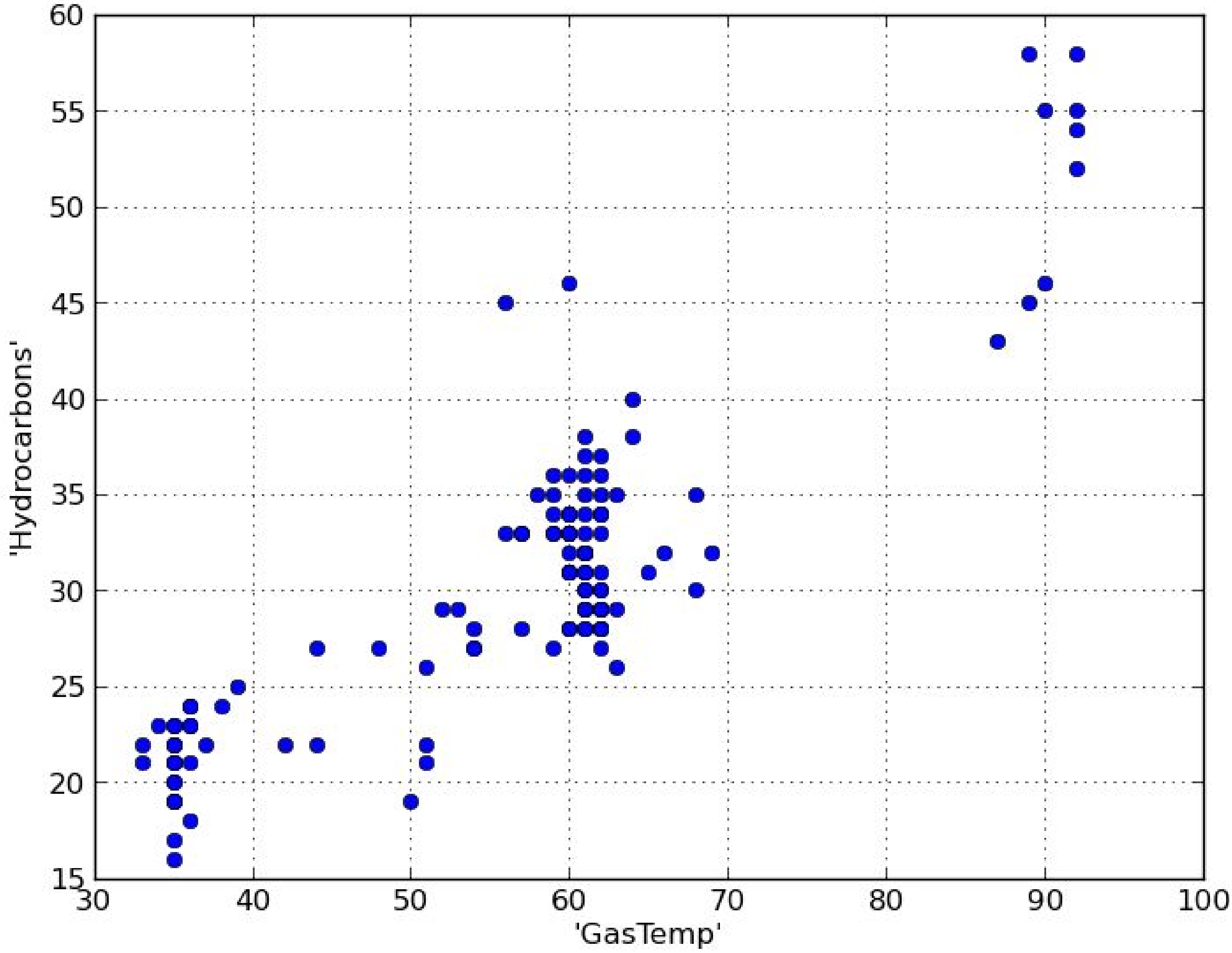
If $\alpha = 0{,}05$, then $n=65$, and cutoff points are 2.0003 and 1.975 for exact and asymptotic results.

For male and female results are 5.715 and 2.254 $\Rightarrow$ we reject $H_0$ for both cases.

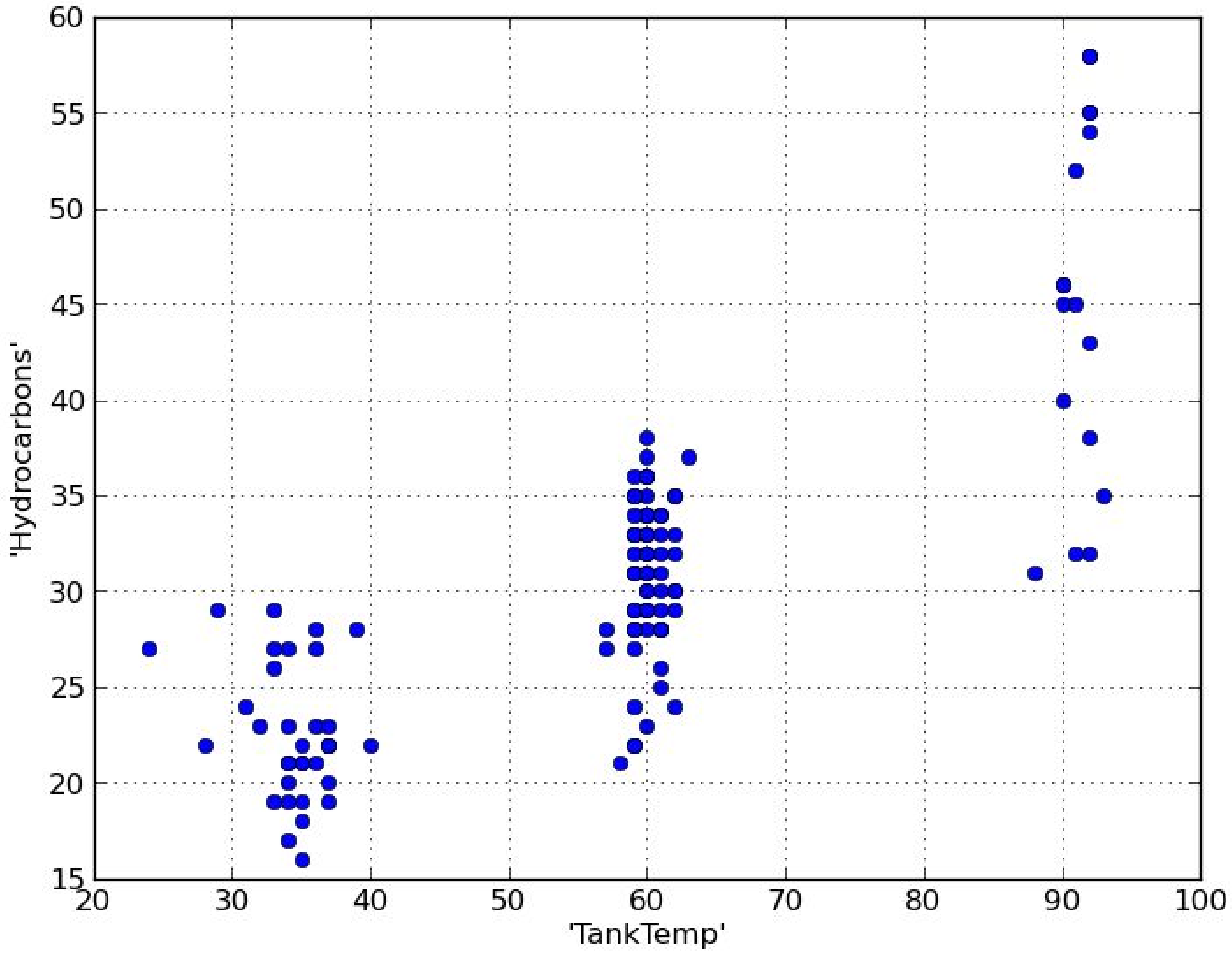


60
55
50
45
40
35
30
25
20
15
2
3
4
5
6
7
8
'Hydrocarbons'
'GasVapor'

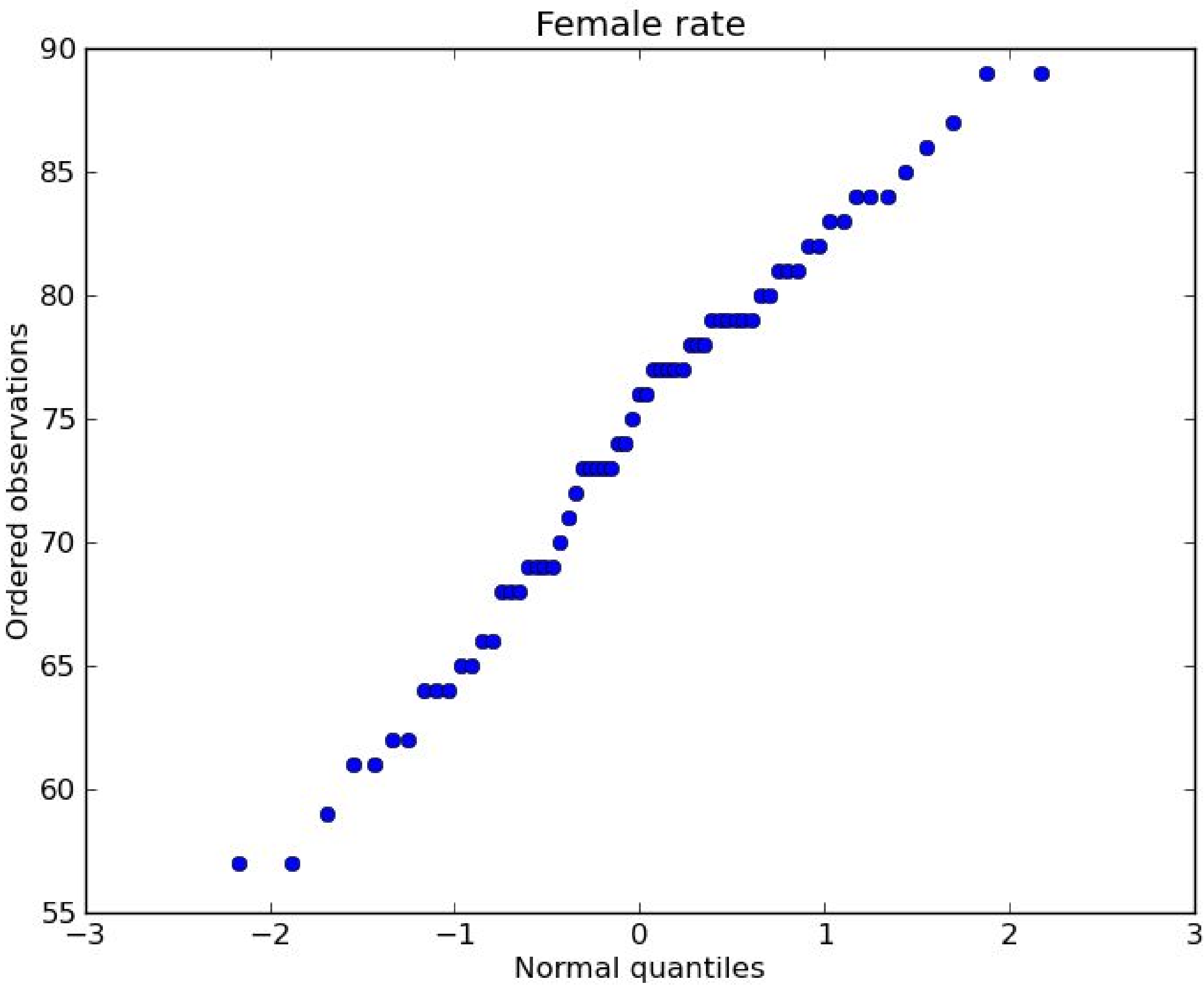

'Hydrocarbons'
'TankVapor'
60
55
50
45
40
35
30
25
20
15
2
3
4
5
6
7
8

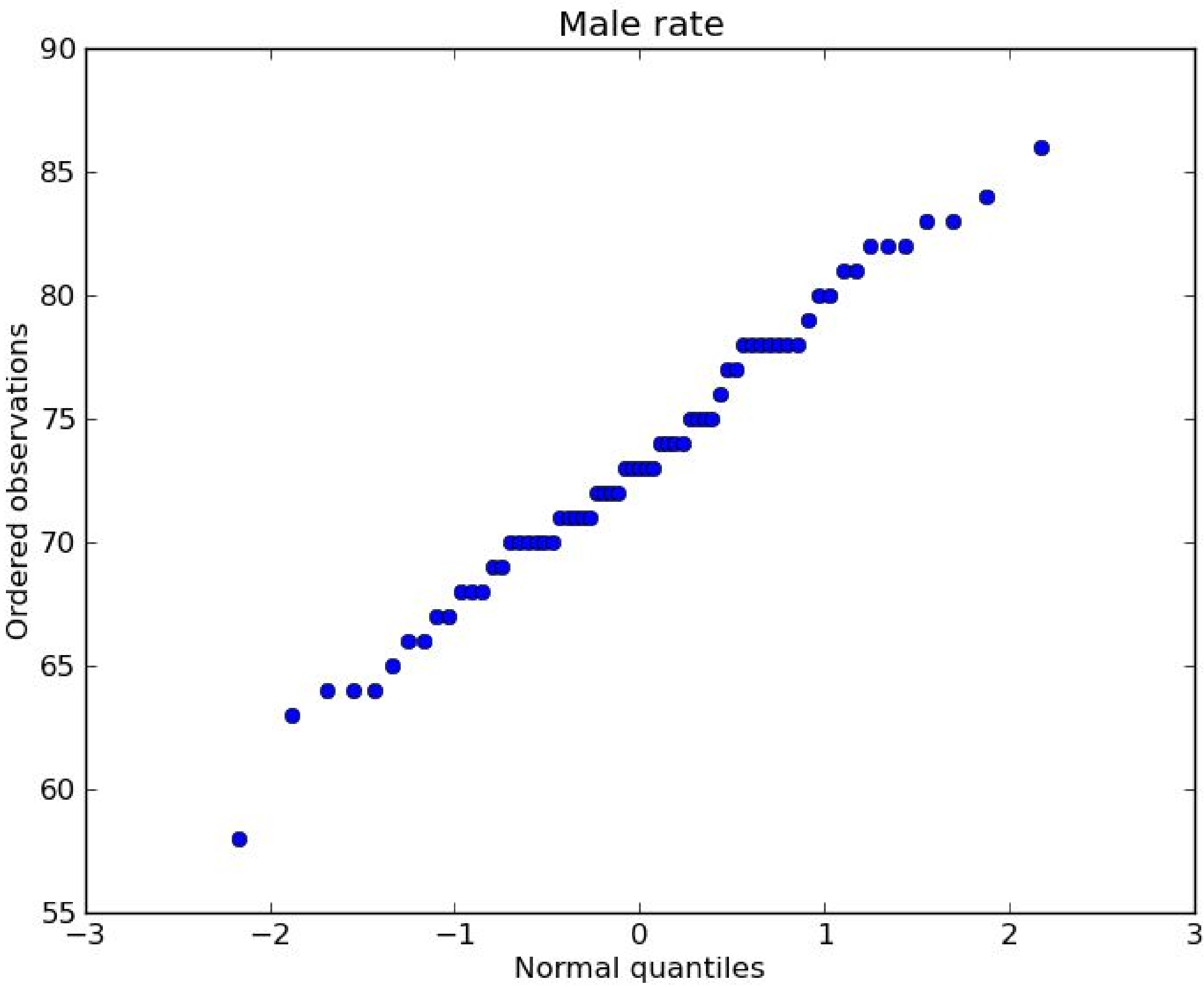


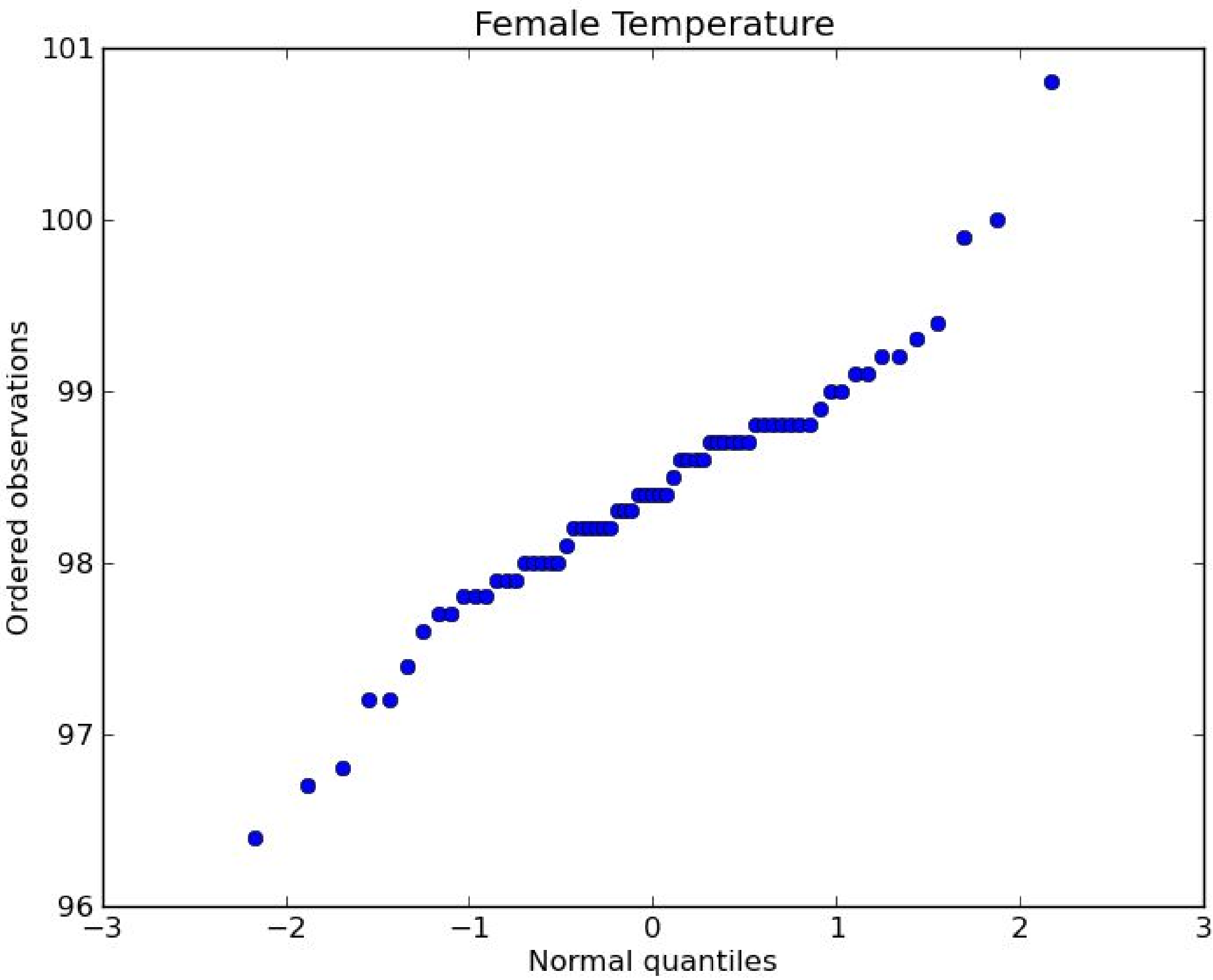
60
55
50
45
40
35
30
25
20
15
'Hydrocarbons'
30
40
50
60
70
80
90
100
'GasTemp'

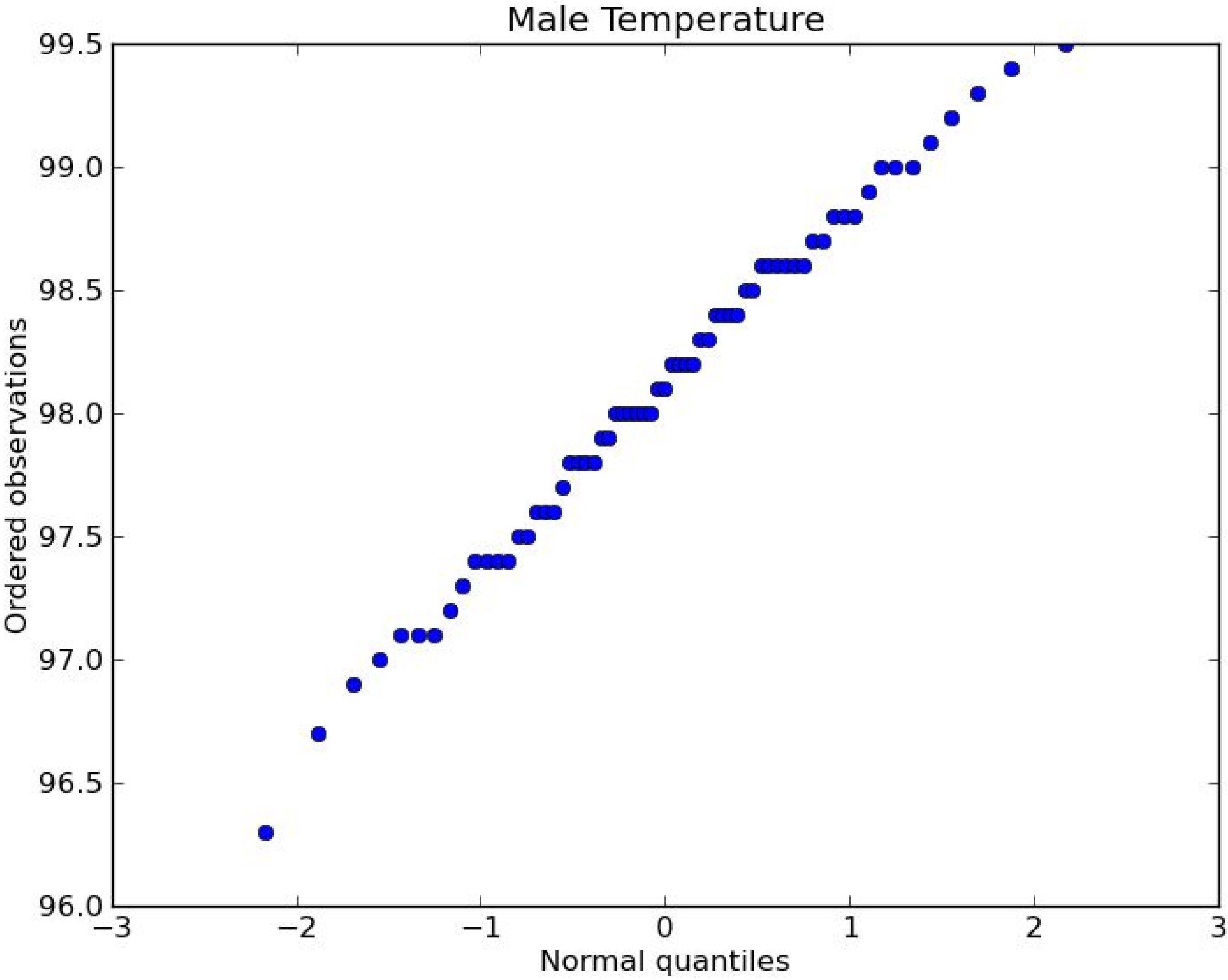


'Hydrocarbons'
'TankTemp'
60
55
50
45
40
35
30
25
20
15
20
30
40
50
60
70
80
90
100


Female rate
Ordered observations
Normal quantiles
90
85
80
75
70
65
60
55
−3
−2
−1
0
1
2
3


Male rate
Ordered observations
Normal quantiles
90
85
80
75
70
65
60
55
−3
−2
−1
0
1
2
3


Female Temperature
Ordered observations
Normal quantiles
101
100
99
98
97
96
−3
−2
−1
0
1
2
3


Male Temperature
99.5
99.0
98.5
98.0
97.5
97.0
96.5
96.0
Ordered observations
−3
−2
−1
0
1
2
3
Normal quantiles