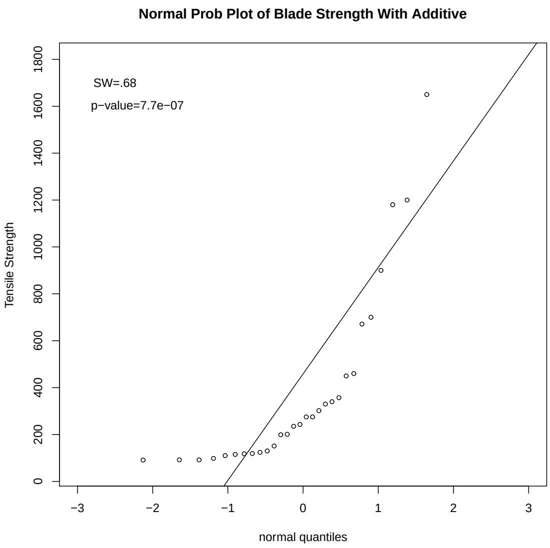
Stat 641
Solutions for Homework 7

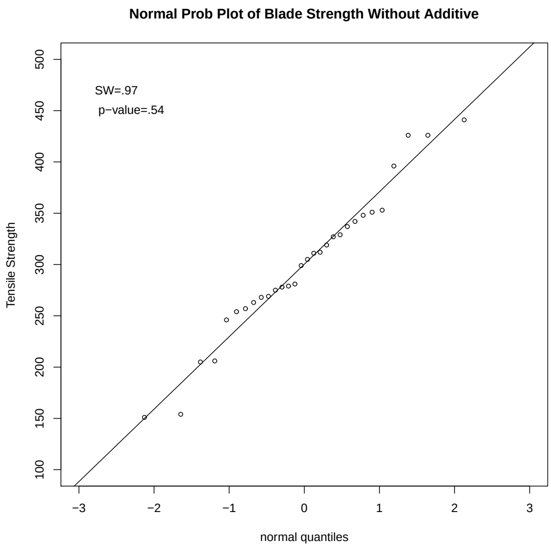
Problem I. (16 points)

Need to determine if the two data sets come from a normal distribution and if not can we transform data to normality:

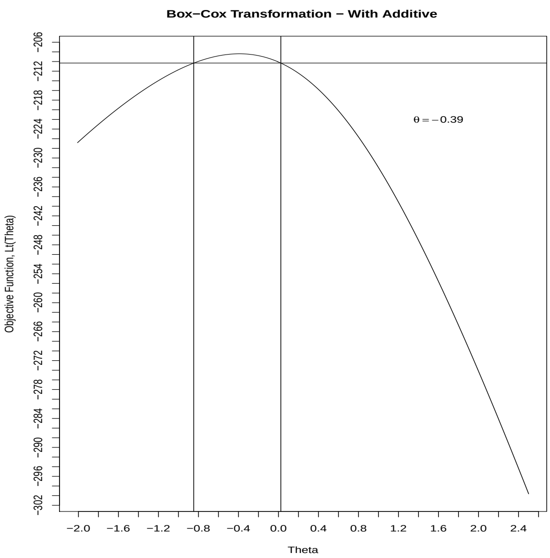
i. Blades With Additive: non-normal from plot, SW=.679 and p-value=0.0000008 from Shapiro-Wilk.

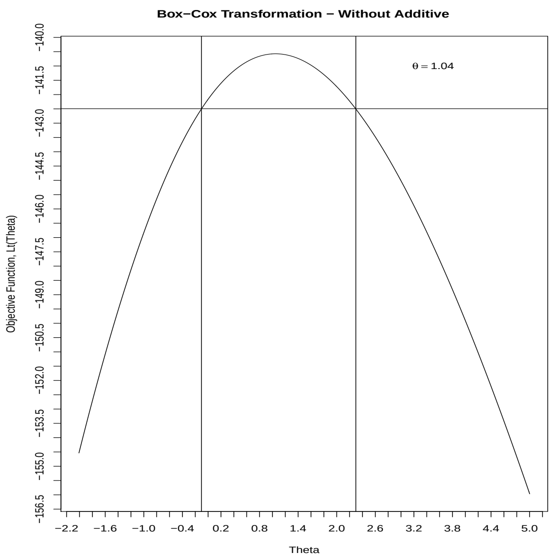
ii. Blades Without Additive: normal from plot, SW=.970 and p-value=0.539 from Shapiro-Wilk.





iii. Box-Cox transformation for With Additive and Without Additive Data:



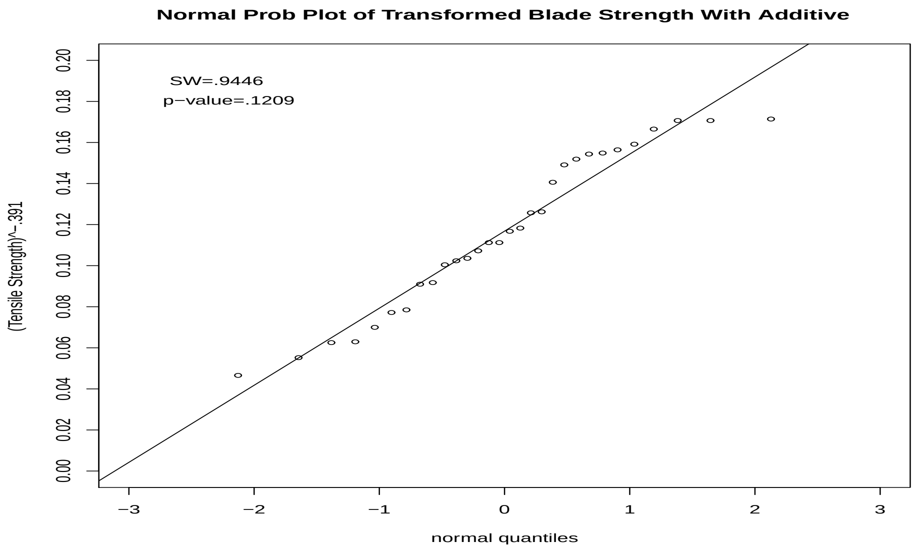


1. For a population with a normal distribution, 95/95 lower tolerance bound and 95/95 upper tolerance bound are respectively,

$$L_{0.95,0.95} = \bar{X} - K_{0.95,0.95}S, \qquad L_{0.95,0.95} = \bar{X} + K_{0.95,0.95}S,$$

where for n=30, $K_{0.95,0.95} = 2.220$ and $S = \sqrt{\frac{1}{n-1}\sum_i (X_i - \bar{X})^2}$.

i. For the With Additive data (Y_1), using the Box-Cox transformation we have that the Shapiro-Wilk's test of $X = g(Y_1) = Y_1^{-.391}$ has a p-value = .121 along with the following normal reference plot indicates that the transformed blade strenght has approximately a normal distribution.



This is a decreasing function, therefore, a 95/95 lower tolerance bound for the distribution of Y_1 is the inverse transformation of the 95/95 upper tolerance bound for the distribution of X:

$$\bar{X} - K_{0.95,0.95}S = 0.1168 + (2.220)(0.0382) = 0.202.$$

Therefore, 95/95 lower tolerance interval on the strength of blades With Additive, Y_1, is

$$(g^{-1}(0.2016) = (0.2016)^{-1/0.39}, \infty) = (60.72, \infty).$$

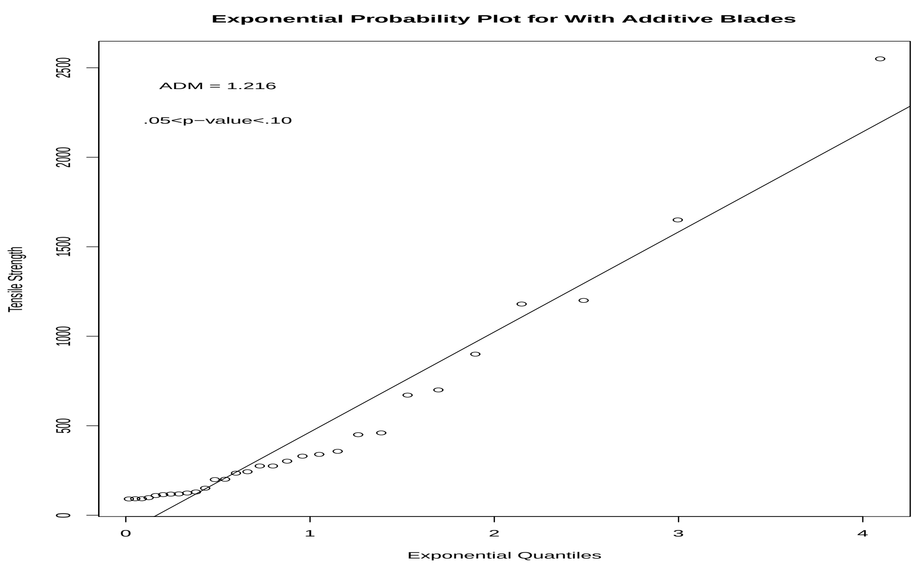
Note that you can use distribution free technique but the sample size of our data is small, n=30, and that the largest achievable value for γ is 0.45.

ii. Without Additive data (Y_2) is normally distributed. Thus, 95/95 lower tolerance bound is

$$\bar{Y_2} - K_{0.95,0.95}S = 300.2667 - (2.220)(71.3031) = 141.97.$$

Therefore, 95/95 lower tolerance interval on the strength of blades Without Additive is $(141.97, \infty)$.

iii. Another approach for obtaining the lower tolerance interval for blades With Additive is to fit an exponential distribution to the data.



From the above exponential reference distribution plot and with $.05 < p - value < .10$ from the Anderson-Darling statistic, the exponential distribution provides an adequate fit to the With Additive data. Therefore, the lower tolerance interval is

$$WP, \gamma = -\hat{\beta}\left[\frac{2n}{\chi^2_{1-\gamma,2n}}\right] log(P) = -458.2\left[\frac{60}{79.082}\right] log(.95) = 17.83$$

There is a noticeable difference between the lower tolerance bound based on the transformed normal based bound and the exponential distribution based bound. The exponential distribution does not appear to fit the data very well so I would be inclined to use the normal based bound.

2. i. Because the With Additive data is not normally distributed and transformations are not in general appropriate for generating C.I. it is necessary to use bootstrap methods:

```
 x =  c(151, 450, 124, 235, 357, 110, 302, 671, 118, 115,
275, 275, 2550, 243, 201, 199, 130, 119, 92, 91, 92, 98,
1650, 1200, 1180, 900, 700, 460, 340, 330 )
n =  length(x)
thestD  =  mean(x)
a=(1-.95)/2
B  =  9999
thestB  =  numeric(B)
thestB  =  rep(0,B)
RS  =  numeric(B)
RS  =  rep(0,B)
for (i in 1:B) {
samp = sample(x,replace=TRUE)
thestB  =  mean(samp)
RS[i] = thestB - thestD
}
RS  =  sort(RS)
LRS  =  RS[(B+1)*a]
URS  =  RS[(B+1)*(1-a)]
thL  =  thestD - URS
thU  =  thestD - LRS
thL
thU
```

the 95% CI for the mean (μ_1) strength of blades With Additive is (thL, thU) = (243.93, 631.73).

ii. Under the normality of Without Additive data, the 95% CI for the mean (μ_2) strength of blades Without Additive is

$$\bar{Y_2} \pm t_{29,0.025}\frac{S_{y_2}}{\sqrt{n}} = 300.2667 \pm (2.045)\frac{71.3031}{\sqrt{30}} = (273.64, 326.89).$$

3. i. Using our R-code or Table VII.3 in Handout 11, we have $r = 10$ so the 95% CI on the median strength of blades With Additive is

$$(Y_{(r)}, Y_{(n-r+1)}) = (Y_{(10)}, Y_{(21)}) = (130,\ 357).$$

ii. For Without Additive, the data is from normal distribution and so the mean is equal to the median. Thus, the answer should be the same as in part 2, (273.64, 326.89).

However, for comparing distribution-free to parametric, we would have 95% CI on the median strength of blades With Additive is

$$(Y_{(r)}, Y_{(n-r+1)}) = (Y_{(10)}, Y_{(21)}) = (269,\ 329).$$

4. i. 95/95 lower tolerance interval:
We are 95% confident that 95% of blades With Additive would have strength readings in the interval $(60.09, \infty)$. And we are 95% confident that 95% of the blades Without Additive would have strength readings in the interval $(141.97, \infty)$.

ii. 95% CI on the average strength (μ) or 95% CI on the median strength ($Q(0.5)$):
Note that *before* this CI was calculated, one could make the statement that the probability is 0.95 that the random interval $\left[\bar{y_2} \pm t_{29,0.025}\frac{s_{y_2}}{\sqrt{n}}\right]$ will include μ_2, which is fixed but unknown. However, *after* the limits of the interval are calculated, the interval is no longer random because its limits are now fixed numbers. μ_2 either does or dose not lie in the calculated interval. Thus, $P(273.64 \leq \mu_2 \leq 326.89)$ is either **0** or **1**, but **not 0.95**. Therefore it is **incorrect** to say that the probability is 0.95 that the true μ_2 is in $(273.64, 326.89)$.
In an infinitely long series of trials in which samples of size n are drawn from the same population and 95% CI's for μ (or $Q(0.5)$) are calculated using the same method, the proportion of intervals that actually include μ will be 95%. However, for any particular CI, it is unknown whether or not that CI includes μ since μ is unknown.

Problem II. (10 points) Failure Stress of impregnated carbon fibers:

1. First analysis without specifying the distribution of the stress to failure values.

   i. Using SAS procedure LIFETEST or R-Code:

```
SAS Code:
option ls=75 ps=55 nocenter nodate;
title 'Strength of Carbon Fibers';
data cords;
input S C @@;
label S  = 'Strength of Fibers' C ='Censoring (0=Yes)';
cards;
2.526 1 2.546 1 2.628 1 2.669 1 2.869 1 2.710 1 2.731 1 2.751 1 2.771 1
2.772 1 2.782 1 2.789 1 2.793 1 2.834 1 2.844 1 2.854 1 2.875 1 2.876 1
2.895 1 2.916 1 2.919 1 2.957 1
2.977 1 2.988 1 3 0 3 0 3 0 3 0
run;
proc lifetest data=cords outsurv=outcen;
time S*C(0);
run;
proc print data=outcen;
run;


R Code:
library(MASS)
library(survival)
stress = c(2.526 , 2.546 , 2.628 , 2.669 , 2.869 , 2.710 , 2.731 , 2.751 , 2.771 ,
2.772 , 2.782 , 2.789 , 2.793 , 2.834 , 2.844 , 2.854 , 2.875 , 2.876 ,
2.895 , 2.916 , 2.919 , 2.957 , 2.977 , 2.988 , 3 , 3 , 3 , 3 )
cens = c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0)
Surv(stress, cens)
fibers.surv <- survfit(Surv(stress, cens) ~ 1,conf.type="log-log")
summary(fibers.surv)
print(fibers.surv,print.rmean=TRUE)
postscript("u:/meth1/homework/solutions/assign7F12PII_surv.ps",height=8,horizontal=F)
plot(fibers.surv,conf.int=F,log=TRUE,
main="Kaplan-Meier Estimator of Survival Function",xlab="Strength of Fibers",
ylab="Survival Function")
graphics.off()
#output:
# records      n.max     n.start     events     *rmean *se(rmean)     median   0.95LCL    0.95UCL
# 28.0000    28.0000     28.0000    24.0000     2.8311     0.0249     2.8490    2.7720     2.8950
```

   The estimate of the average stress to failure for the carbon fibers from the SAS procedure LIFETEST is 2.82943 with a standard error of 0.02504. We could obtain a 95% CI for μ using the expression

$$\hat{\mu} \pm z_{\alpha/2}\widehat{SE}(\hat{\mu}) = 2.82943 \pm 1.96(.02504) = (2.78, 2.88)$$

   The estimate of the average stress to failure for the carbon fibers from the R code is 2.8311 with a standard error of 0.0249. We could obtain a 95% CI for μ using the expression

$$\hat{\mu} \pm z_{\alpha/2}\widehat{SE}(\hat{\mu}) = 2.8311 \pm 1.96(.0249) = (2.78, 2.88)$$

   The problem with the two CI is that they are asymptotic results and with $n = 28$ an asymptotic result is questionable.

   ii. From the same output we have that an estimate of the median is $\hat{\tilde{\mu}} = 2.849$ with 95% C.I. (2.772, 2.895).

   iii. The studentized bootstrap procedure could be implemented using the following R Code:

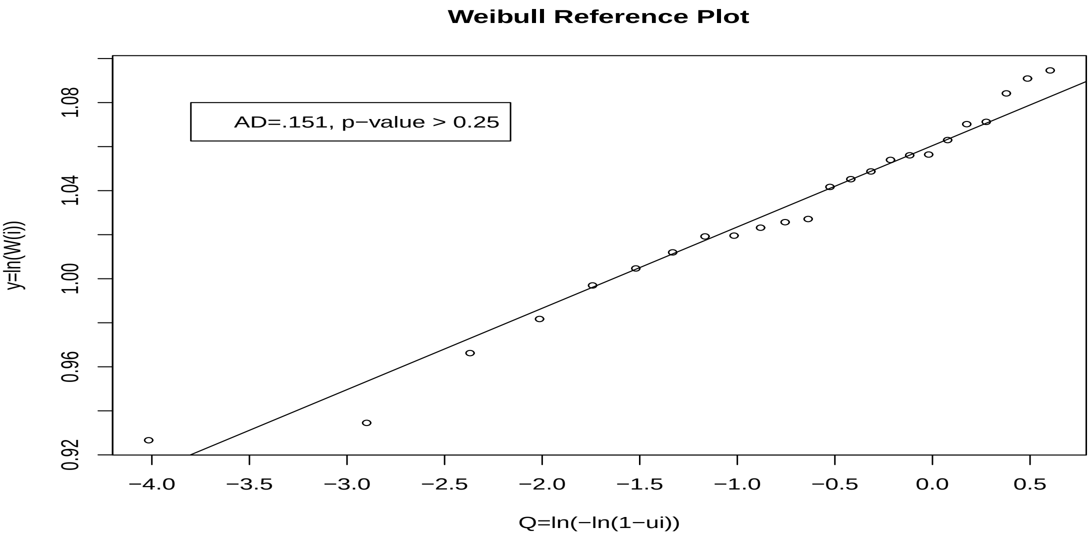
```
 w = c(2.526,  2.546,  2.628,  2.669,  2.869,  2.710,  2.731,  2.751,  2.771,
      2.772,  2.782,  2.789,  2.793, 2.834, 2.844, 2.854,  2.875,  2.876,
      2.895,  2.916,  2.919,  2.957,  2.977,  2.988,  3,  3,  3,  3)
n =  length(w)
thestD  =  mean(w)
a=.025
```

```
B  =  9999
thestB  =  numeric(B)
thestB  =  rep(0,B)
RS  =  numeric(B)
RS  =  rep(0,B)
for (i in 1:B) {
samp = sample(w,replace=TRUE)
thestB  =  mean(samp)
RS[i] = (thestB - thestD)
}
RS  =  sort(RS)
LRS  =  RS[(B+1)*a]
URS  =  RS[(B+1)*(1-a)]
thL  =  thestD-URS
thU  =  thestD-LRS
```

The above yields (thL, thU) = (2.784, 2.880) as a 95% for the mean which is nearly identical to the asymptotic result.

2. Evaluate the fit of a Weibull model to the data. The probability plot and GOF test will need to be modified to take into account the fact that 4 of the 28 data values are Type I censored.

- For the probability plot, make the transformation $Y_i = log(W_i)$ and then plot $\left(Q(u_i), Y_{(i)}\right)$ for the 24 uncensored values using $u_i = \frac{i-.5}{28}$ $i = 1, \ldots, 24$ and the quantile function for the standard extreme value distribution: $Q(u_i) = log(-log(1 - u_i))$. Note that 28 was used in the denominator of u_i. Based on the plot the Weibull model appears to be appropriate.



Because we do not have tables for obtaining the p-values for the GOF tests, we will apply the Anderson-Darling test to just the 24 uncensored data values.

The following R code will produce the value of the adjusted Anderson-Darling statistic and the Weibull Reference distribution plot:

```
library(MASS)
xall = c(2.526, 2.546,  2.628,  2.669,  2.869,  2.710,  2.731,
2.751,  2.771, 2.772,  2.782,  2.789,  2.793, 2.834, 2.844, 2.854,
2.875,  2.876,  2.895,  2.916,  2.919,  2.957, 2.977, 2.988,  3,  3,  3,  3)

x = c(2.526, 2.546,  2.628,  2.669,  2.869,  2.710,  2.731,  2.751,  2.771,
2.772,  2.782,  2.789,  2.793, 2.834, 2.844, 2.854,  2.875,  2.876,
2.895,  2.916,  2.919,  2.957, 2.977, 2.988)
#obtain estimates of Weibull parameters using just the 24 data values with MLE procedure:

fiber.surv = fitdistr(x ,"weibull")
fiber.surv
Shape = fiber.surv$estimate[1]
Scale = fiber.surv$estimate[2]
#calculate AD using just the 24 uncensored values:

n = length(x)
y = -log(x)
ys = sort(y)
i =  seq(1:n)
# Anderson-Darling: For Weibull Model
a = -log(Scale)
b = 1/Shape
#a     #mle of location in Extreme Value Dist. Form
#b     #mle of scale in Extreme Value Dist. Form
z   = exp(-exp(-(ys-a)/b))
zs   = sort(z)
A1i = (2*i-1)*(log(zs)-log(1-zs))
A2i = log(1-zs)
s1  = (-1/n)*sum(A1i)
s2  = -2*sum(A2i)
AD  = -n+s1+s2
ADM  = (1+.2/sqrt(n))*AD
ADM
shapiro.test(x)
#Modified Weibull Reference Distribution Plot for 24 Uncensored Values:

# 0.1449
n = length(xall)
m = length(x)
weibs = sort(log(x))
i = 1:m
ui = (i-.5)/n
QW = log(-log(1-ui))
postscript("u:/meth1/homework/solutions/hw7_2012probII(2).ps",height=6,horizontal=F)
plot(QW,weibs,abline(lm(weib~QW)),
       main="Weibull Reference Plot",cex=.75,lab=c(7,11,7),
       xlab="Q=ln(-ln(1-ui))",
       ylab="y=ln(W(i))")
legend(-3.8,1.08,"AD=.151, p-value > 0.25")
graphics.off()
```

- Applying the Anderson-Darling statistic to the 24 uncensored values we obtain: AD=.151 and from Table 5 on page 34 of Handout 9, $p - value > 0.25$. Thus, we have an excellent fit of the Weibull model to the 24 uncensored data values. This conclusion is consistent with the reference distribution plot.

i. **Exact C.I.** To obtain a confidence interval explicitly for the mean from a Weibull distribution is very involved due to the complex relationship between the mean μ and the parameters β and γ:

$$\mu = \beta^{1/\gamma}\Gamma\left(1+\frac{1}{\gamma}\right) = \alpha\Gamma\left(1+\frac{1}{\gamma}\right) \text{ and variance } \sigma^2 = \alpha^2\Gamma\left(1+\frac{2}{\gamma}\right) - \left(\alpha\Gamma\left(1+\frac{1}{\gamma}\right)\right)^2.$$

ii. **Asymptotic C.I.** Applying an asymptotic C.I. to the 24 uncensored values we have $\bar{W} \pm 1.96S/\sqrt{n} = 2.803 \pm 1.96(.124)/\sqrt{24} = (2.753,\ 2.853)$. This approach would generally not be valid because $n = 24$ is too small to overcome the skewness of the Weibull distribution in using the asymptotic result. Also, we have not used any of the information from the 4 censored values.

iii. **Parametric Bootstrap C.I.** Using the estimates of α and γ from the following R code which takes into account that we have censored values:

```
library(survival)
w = c(2.526,  2.546,  2.628,  2.669,  2.869,  2.710,  2.731,  2.751,  2.771,
      2.772,  2.782,  2.789,  2.793, 2.834, 2.844, 2.854,  2.875,  2.876,
      2.895,  2.916,  2.919,  2.957,  2.977,  2.988,  3,  3,  3,  3)
wc = c(rep(1,24),0,0,0,0)
stress <- survreg(Surv(w,wc) ~ 1, dist='weibull')
summary(stress)
stress$coefficients
stress$scale
#output from R:
#Intercept = 1.065932
#Scale= 0.04393189
```

we obtain the following estimates of γ and α:

$$\hat{\gamma} = 1/Scale = 1/.04393189 = 22.76251; \quad \text{and} \quad \hat{\alpha} = e^{\text{Intercept}} = e^{1.065932} = 2.903544$$

Next do a parametric bootstrap procedure to obtain the 95% confidence interval on the mean where all the values greater than 3 in the 28 randomly generated values are set equal to 3 in order to duplicate the censoring in the original data set:

```
wuc = c(2.526,  2.546,  2.628,  2.669,  2.869,  2.710,  2.731,  2.751,  2.771,
      2.772,  2.782,  2.789,  2.793, 2.834, 2.844, 2.854,  2.875,  2.876,
      2.895,  2.916,  2.919,  2.957,  2.977,  2.988)
n = length(wuc)
a = 2.903544
g = 22.76251
thest = mean(wuc)
# V is the estimated variance of based on Weibull Dist.
V  =  a^2*(gamma(1+2/g) - (gamma(1+1/g))^2)
R  =  9999
Z  =  numeric(R)
Z  =  rep(0,times =R)
W  =  numeric(R)
W  =  rep(0,times =R)
for (i in 1:R){
SAMP = rweibull(28,g,a)
for (j in 1:28){
if(SAMP[j] > 3) SAMP[j] = 3
if(SAMP[j] <= 3) SAMP[j] = SAMP[j]
}
W[i]  =  mean(SAMP)
}
# Z is the pivot for the C.I.
Z  = sqrt(n)*(W-thest)/sqrt(V)
Z  =  sort(Z)
L  =  Z[250]
U  =  Z[9750]
thL  =  thest-U*sqrt(V)/sqrt(n)
thU  =  thest-L*sqrt(V)/sqrt(n)
CI = c(thL, thU)
CI
#output from the above
# 2.714897 2.823974
```

Our 95% C.I. for the mean would be (2.715, 2.824). This C.I. has endpoints within 1.4% of the asymptotic C.I.: (2.753, 2.853). Why?

If we apply a Shapiro-Wilk test to the 24 uncensored values, we obtain a p-value of .3158 so there is an excellent fit of the normal distribution to the data. Thus, the normal based procedure would be appropriate.

Problem III. (9 points) Let X be the number of breaks on a given bar. Then, the average number of breaks per bar is

$$\bar{X} = \frac{(0)(121) + (1)(110) + (2)(38) + (3)(7) + (4)(3) + (5)(1)}{280} = .8$$

Next evaluate if the distribution of X is Poisson(λ). Under the Poisson model, $p_i = e^{-\lambda}\lambda^i/i!$, for $i = 0, 1, 2, 3, 4$ and $p_5 = P[X \geq 5]$. Since λ is unknown use $\hat{\lambda} = \bar{X} = .8$. An initial calculation of $E_i = 280p_i$ shows that E_5 is less than 1. Therefore, combine the last two cells and then compute the following using the R-function $pi = dpois(i, .8)$ for $i = 0, 1, 2, 3$ and $p4 = P[X \geq 4] = 1 - P[X \leq 3] = 1 - ppois(3, .8)$, finally compute $E_i = 280 * p_i$.

```
     i     pi          Ei         Oi    (Oi-Ei)^2/Ei
[1,] 0 0.449328964 125.81211     121       0.184
[2,] 1 0.359463171 100.64969     110       0.869
[3,] 2 0.143785269  40.25988      38       0.127
[4,] 3 0.038342738  10.73597       7       1.300
[5,] 4 0.009079858   2.54236       4       0.836
```

The test statistic is

$$Q^* = \sum_{i=1}^{4} \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i} = .184 + .869 + .127 + 1.300 + .836 = 3.316$$

and Q^* has approximately a chi-squared distribution with df=$5 - 1 - 1 = 3$.
The p-value= $Pr(\chi^2_3 \geq 3.316) = 1 - pchisq(3.316, 3) = 0.345$ and so we conclude that there is an excellent fit of the Poisson model to the data. Notice that the expected counts for the five cells under the Poisson model are very close to the observed counts and all but one expected count are greater than 5 with the remaining value greater than one.

i. The Wald approximate $100(1-\alpha)\%$ C.I. for λ using the Pivot, $Z = \frac{\sqrt{n}(\bar{Y}-\lambda)}{\hat{\lambda}}$, with $\hat{\lambda} = \bar{Y}$ is

   $\bar{Y} \pm 2.576\frac{\sqrt{\hat{\lambda}}}{\sqrt{n}} = .8 \pm 2.576\frac{\sqrt{.8}}{\sqrt{280}} = (.662, .938)$

ii. Using the Wilson approach an approximate $100(1-\alpha)\%$ for λ is obtained as follows:

   $P[\frac{\sqrt{n}(\bar{Y}-\lambda)}{\sqrt{\lambda}} \leq Z_{\alpha/2}] \approx 1 - \alpha$ Next, solve the following inequality for λ:

   $\sqrt{n}|\bar{Y} - \lambda| \leq \sqrt{\lambda}Z_{\alpha/2} \Rightarrow (\bar{Y}^2 - 2\lambda\bar{Y} + \lambda^2) - \frac{1}{n}\lambda Z^2_{\alpha/2} = 0 \Rightarrow$

   $\lambda = \left(2\bar{Y} + \frac{1}{n}Z^2_{\alpha/2} \pm \sqrt{(2\bar{Y} + \frac{1}{n}Z^2_{\alpha/2})^2 - 4\bar{Y}^2}\right)/2$

   Approximate, $100(1-\alpha)\%$ C.I. for λ is $\bar{Y} + \frac{1}{2n}Z^2_{\alpha/2} \pm Z_{\alpha/2}\sqrt{(\frac{1}{n}\bar{Y} + \frac{1}{4n^2}Z^2_{\alpha/2}} =$

   $.8 + \frac{1}{560}(2.576)^2 \pm 2.576\sqrt{(\frac{.8}{280} + \frac{1}{4(280)^2}(2.576)^2} = .81185 \pm .13820 = (.674, .950)$

With n=280 very large, there is very little differences in the Wald and Wilson C.I.'s. For small n, there would be a much more noticeable difference.

Problem IV. (15 points)

A. Let Y be the lifetime of epoxy strands and p be the probability that an epoxy strand will survive for 300 hours. Therefore, the parameter to be estimated is $p = P(Y \geq 300)$.

Because $\min\{n\widehat{p}, n(1-\widehat{p})\} \geq 5$ and $n > 40$, we can use the Agresti-Coull CI for p:

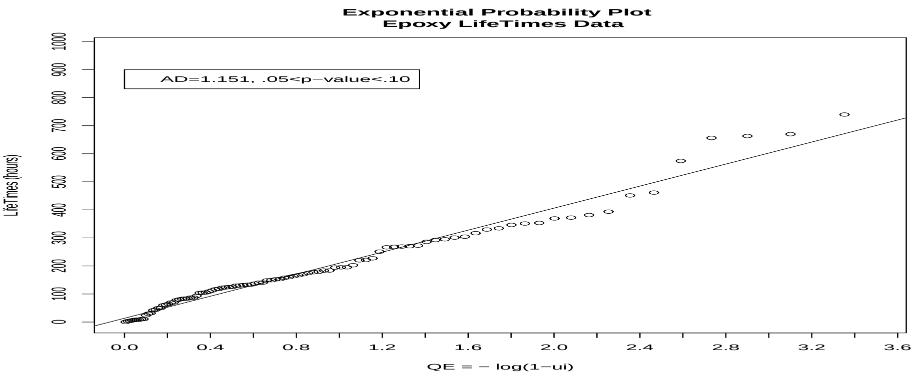
$\widehat{p} = 22/100 = 0.22$ and so $\tilde{X} = 22 + (2.58^2)/2 = 25.3282$, and $\tilde{n} = 100 + 2.58^2 = 106.6564$.

Thus, $\tilde{p} = \tilde{X}/\tilde{n} = 0.2375$ and the 99% Agresti-Coull CI on p is

$$0.2375 \pm (2.58)\sqrt{\frac{(0.2375)(0.7625)}{106.6564}} = (0.1312, 0.3438).$$

B. From the following quantile plot, the exponential model appears to fit with AD=1.151

with $.05 < p - value < .10$ indicates a moderately good fit of exponential model with $\hat{\beta} = \bar{T} = 209.1838$.



i. Using the results for an exponential distribution, a 95/99 lower tolerance bound is

$$L_{0.95,0.99} == -\hat{\beta}\left[\frac{2n}{\chi^2_{1-.99}}\right] log(.95) = -(209.1838)\left[\frac{200}{249.4451}\right](\log 0.95) = 8.603.$$

Therefore, a 95/99 lower tolerance interval on the average lifetime of the epoxy strands is $(8.603, \ \infty)$.

ii. A distribution-free 95/99 lower tolerance interval for n=100 has m=1 from Table on page 51 in Handout 11.

Thus, the distribution-free lower tolerance interval would be $(Y_{(1)}, \ \infty) \ = \ (.18, \ \infty)$. With n=100, the distribution-free interval does not very informative.

C. Using the methodology for an exponential distribution, a 95% PI for Y_{101} is given by

$$(\bar{Y}F_{0.025,2,200}, \ \bar{Y}F_{0.975,2,200}) \ = \ ((209.1838)(0.0253), \ (209.1838)(3.758)) \ = \ (5.3, \ 786.1).$$

Problem V. (20 points) Strength of Braided Cord:

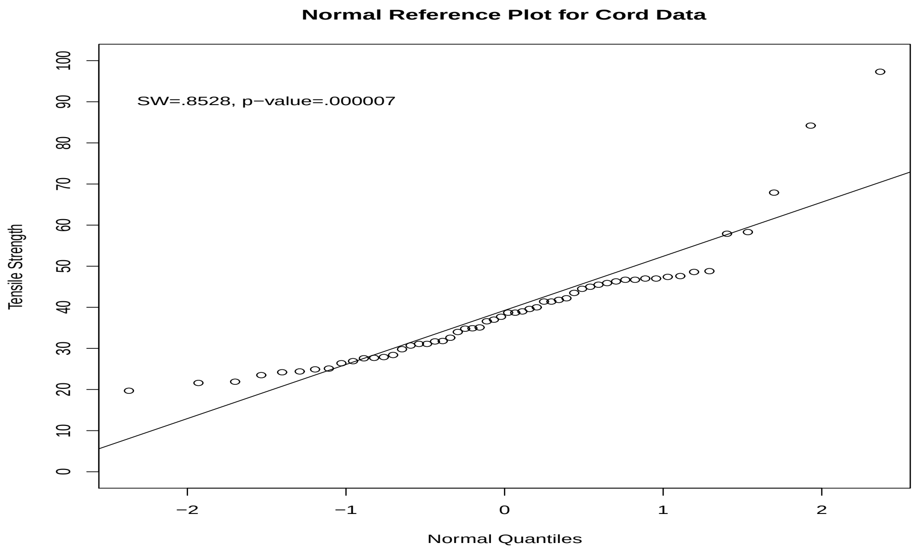
(A) The parameter of interest is $p = P[S < 50]$. From the data, $\hat{p} = \frac{51}{56} = .91$, $n \cdot min(\hat{p}, 1-\hat{p}) = 5.1 > 5$, and $n = 56 > 40 \quad \Rightarrow$ Use Agresti-Coull C.I.

$\tilde{X} = 51 + (1.96^2)/2 = 25.3282$, and $\tilde{n} = 56 + 1.96^2 = 59.8416$.

Thus, $\tilde{p} = \tilde{X}/\tilde{n} = 0.8843$ and the 95% Agresti-Coull CI on p is

$$0.8843 \pm (1.96)\sqrt{\frac{(0.8843)(1-.8843)}{59.8416}} \;=\; (0.803,\ 0.965).$$

(B) The normal probability plot and value of Shapiro-Wilk's test indicates that a normal distribution would not be a good model for data.



Therefore, a nonparametric C.I. will be constructed.

Using Table VII.3 in Handout 11 or the R code on page 31 of Handout 11, we obtain k=21, therefore a 95% C.I. on Median is

$[Y_{(21)},\ Y_{(56-21+1)}] = [Y_{(21)},\ Y_{(36)}] = [32.6,\ 41.8]$ with a coverage of 95.6%.

(C) From the Table in Handout 11 for Nonparametric Tolerance Interval, we obtain for n=56 that m=2 which implies that

the $(P, \gamma) = (.9, .95)$ tolerance interval is $[Y_{(1)}, Y_{(56)}] = [19.7, 97.3]$

(D) From the Table in Handout 11 for Nonparametric Tolerance Interval for $P = .90$ the table yields for $n = 40$ that $\gamma = .92$ and for $n = 50$ that $\gamma = .97$. Therefore, use $n \geq 45$.

Problem VI. (3 points each) Multiple Choice:

1. **- A** See pages 23-25 in Handout 11
2. **- C** The transformation to normality yields an excellent fit therefore just invert the endpoints to obtain the requested Tolerance Interval.
3. **- C** The bootstrap is often used when there is not an appropriate parametric procedure.
4. **- C** $n = \hat{\sigma}^2 \ Z^2_{\alpha/2}/\Delta^2 = ((30)(2.576)/10)^2 = 59.7$. Therefore, take n=60.
5. **- C** Depends on what the engineer was attempting to accomplish.
6. **- E** See page 58 in Handout 11
7. **- B** If the data are highly correlated, then S^2/n underestimates $Var(\bar{X})$. This results in a C.I. for μ which is too narrow and hence has a smaller coverage probability than the nominal level of confidence.
8. **- C** See page 52 in Handout 11.
9. **- B** If the distribution is highly right skewed, $n = 10$ observations would be insufficient to obtain a valid C.I. for σ. A second but less preferable choice is **(E).**
10. **- A** $n = p(1-p)Z^2_{\alpha/2}/\Delta^2 \leq (.5)(1-.5)(1.96)^2/(.03)^2 = 1067.1$**. Therefore, take n=1068.**