To Do

Read Sections 4.6 - 4.7.

Do End-of-Chapter Problems 1-17 in preparation for Tutorial Test 2.

Today's Lecture

(1) Confidence Interval for Gaussian mean μ when standard deviation σ is unknown

(2) Handspan Example

(3) Sample Size Calculation – Gaussian Data

Gaussian data with unknown mean μ and unknown standard deviation σ

Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from a G(μ,σ) distribution where E(Y_i) = μ is unknown and sd(Y_i) = σ is also unknown.

A point estimator for μ is $\tilde{\mu} = \bar{Y}$ (the maximum likelihood estimator).

Point Estimator for σ^2

A point estimator for σ^2 is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

(not the maximum likelihood estimate).

We prefer S^2 because $E(S^2) = \sigma^2$.

See Course Notes page 132.

Theorem

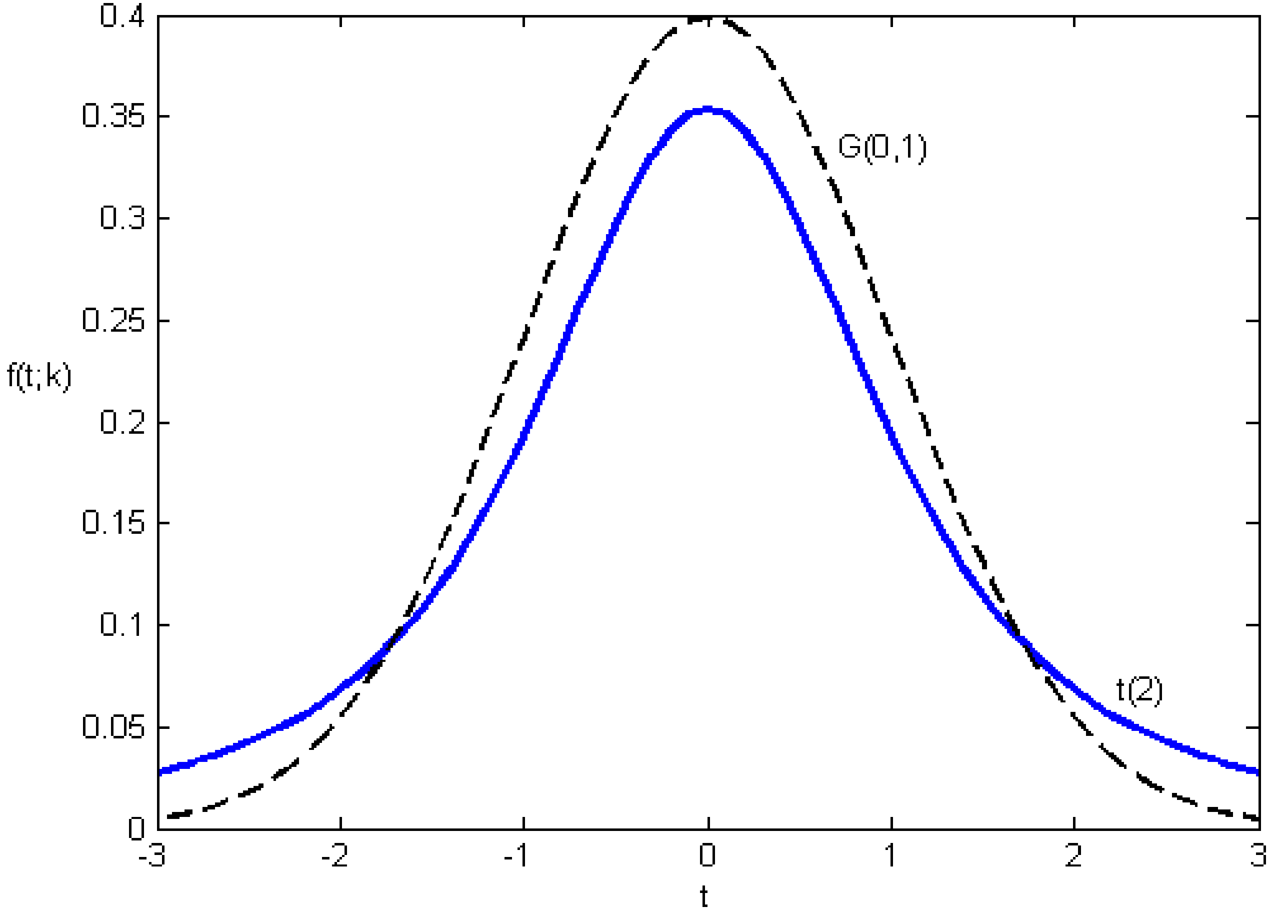
Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from a G(μ,σ) distribution.

Then

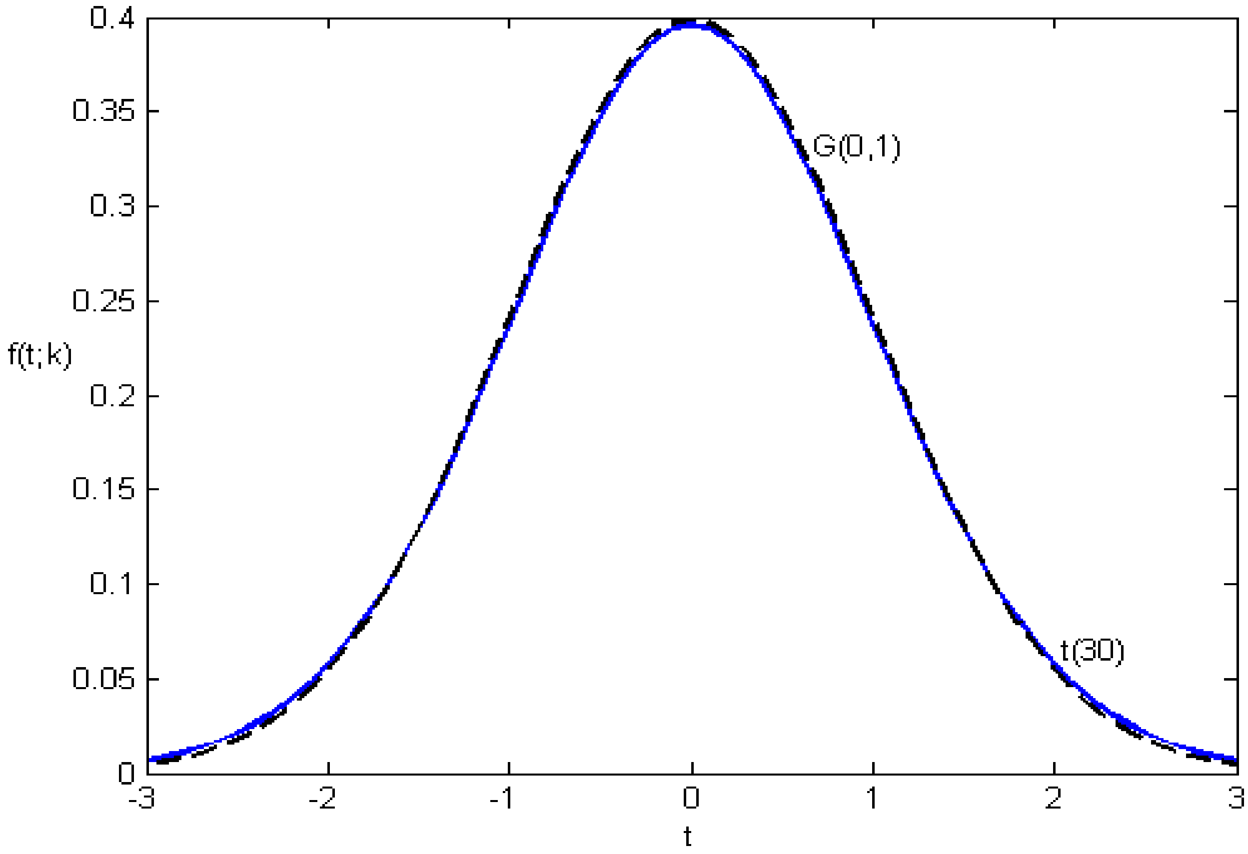
$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

(t distribution with degrees of freedom (parameter) equal to n-1).

t(2) and G(0,1)



t(30) (blue) and G(0,1) (black)



Pivotal Quantity

The random variable $$\frac{\bar{Y}-\mu}{S/\sqrt{n}}$$

is a function of the data *$Y_1, Y_2, \ldots, Y_n$* and the unknown parameter *μ*.

Since $$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

the distribution of $\frac{\bar{Y}-\mu}{S/\sqrt{n}}$ **is completely known.**

We use this pivotal quantity to construct confidence intervals for *μ*.

100p% Confidence Interval for μ, when σ is unknown

Since the t distribution is symmetric about zero, we find the value a from t tables such that

$P(-a \le T \le a) = p$ or equivalently $P(T \le a) = (1+p)/2$

where $T \sim t(n-1)$.

Since

$$\frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

therefore

$$P\left(-a \le \frac{\bar{Y}-\mu}{S/\sqrt{n}} \le a\right) = p$$

100*p*% Confidence Interval for *μ*, when *σ* is unknown

Since

$$P\left(-a \le \frac{\bar{Y}-\mu}{S/\sqrt{n}} \le a\right) = p$$

is equivalent to

$$P\left(\bar{Y} - a\frac{S}{\sqrt{n}} \le \mu \le \bar{Y} + a\frac{S}{\sqrt{n}}\right) = p$$

therefore

$$\left[\bar{y} - a\frac{s}{\sqrt{n}}, \bar{y} + a\frac{s}{\sqrt{n}}\right]$$

is a 100*p*% confidence interval for *μ*

100*p*% Confidence Interval for μ

When σ is known a 100*p*% confidence interval is

$$\bar{y} \pm a \frac{\sigma}{\sqrt{n}}$$

where P(*Z* ≤ a) = (1+*p*)/2 and Z ~ G(0,1).

When σ is unknown a 100p% confidence interval is

$$\bar{y} \pm a \frac{s}{\sqrt{n}}$$

where P(*T* ≤ a) = (1+*p*)/2 and T ~ *t*(n-1).

Useful Result

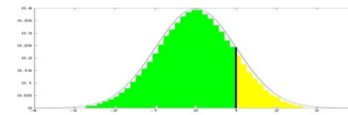
If $p = 0.9$ then $(1+ p)/2 = 0.95$ so use the column labelled 0.95 for a 90% confidence interval.

If $p = 0.95$ then $(1+ p)/2 = 0.975$ so use the column labelled 0.975 for a 95% confidence interval.

If $p = 0.99$ then $(1+ p)/2 = 0.995$ so use the column labelled 0.995 for a 99% confidence interval.

t Tables

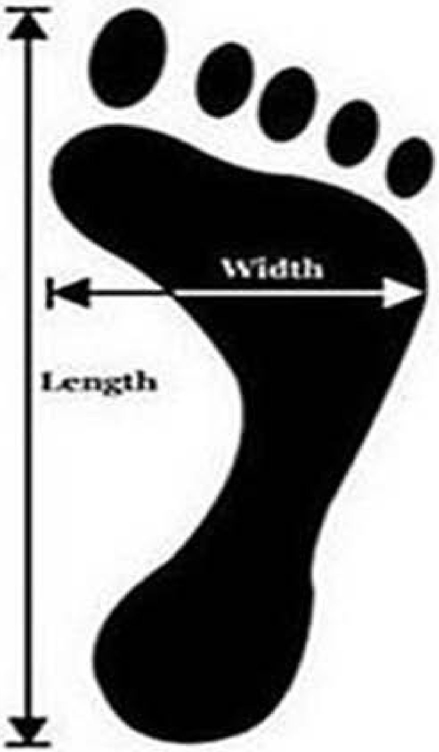
90% 95% 99%



Student t Quantiles

df \ p	0.6	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.3249	1.3764	3.0777	6.3138	12.7062	31.8205	63.6567	318.3088	636.6192
2	0.2887	1.0607	1.8856	2.9200	4.3027	6.9646	9.9248	22.3271	31.5991
3	0.2767	0.9785	1.6377	2.3534	3.1824	4.5407	5.8409	10.2145	12.9240
4	0.2707	0.9410	1.5332	2.1318	2.7764	3.7469	4.6041	7.1732	8.6103
5	0.2672	0.9195	1.4759	2.0150	2.5706	3.3649	4.0321	5.8934	6.8688
6	0.2648	0.9057	1.4398	1.9432	2.4469	3.1427	3.7074	5.2076	5.9588
7	0.2632	0.8960	1.4149	1.8946	2.3646	2.9980	3.4995	4.7853	5.4079
8	0.2619	0.8889	1.3968	1.8595	2.3060	2.8965	3.3554	4.5008	5.0413
9	0.2610	0.8834	1.3830	1.8331	2.2622	2.8214	3.2498	4.2968	4.7809
10	0.2602	0.8791	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437	4.5869
29	0.2557	0.8542	1.3114	1.6991	2.0452	2.4620	2.7564	3.3962	3.6594
30	0.2556	0.8538	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852	3.6460
40	0.2550	0.8507	1.3031	1.6839	2.0211	2.4233	2.7045	3.3069	3.5510
70	0.2543	0.8468	1.2938	1.6669	1.9944	2.3808	2.6479	3.2108	3.4350
80	0.2542	0.8461	1.2922	1.6641	1.9901	2.3739	2.6387	3.1953	3.4163
90	0.2541	0.8456	1.2910	1.6620	1.9867	2.3685	2.6316	3.1833	3.4019
100	0.2540	0.8452	1.2901	1.6602	1.9840	2.3642	2.6259	3.1737	3.3905
>100	0.2535	0.8423	1.2832	1.6479	1.9647	2.3338	2.5857	3.1066	3.3101

Hand and Foot Measurements



Foot measurements WITHOUT shoe!

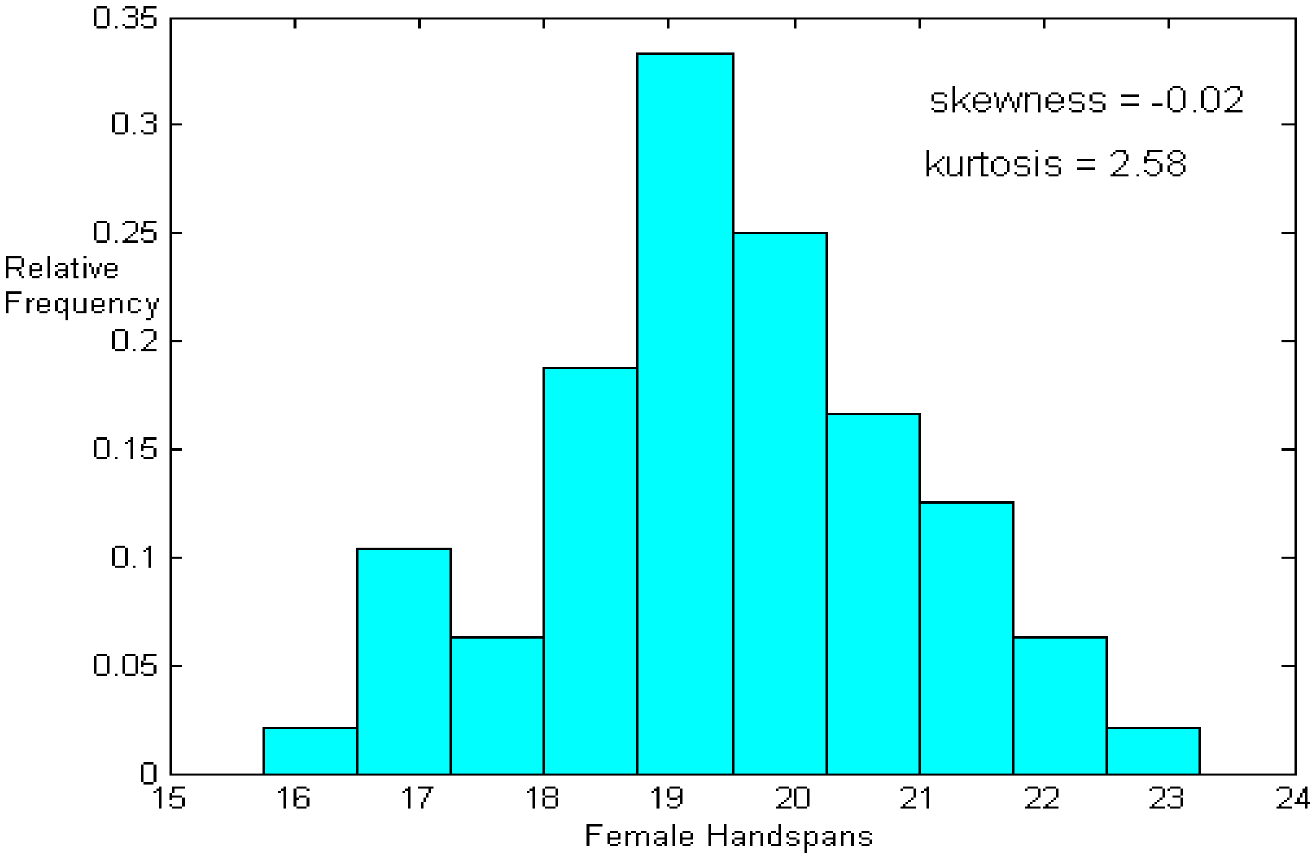
Exercise

Construct a PPDAC for this experiment.

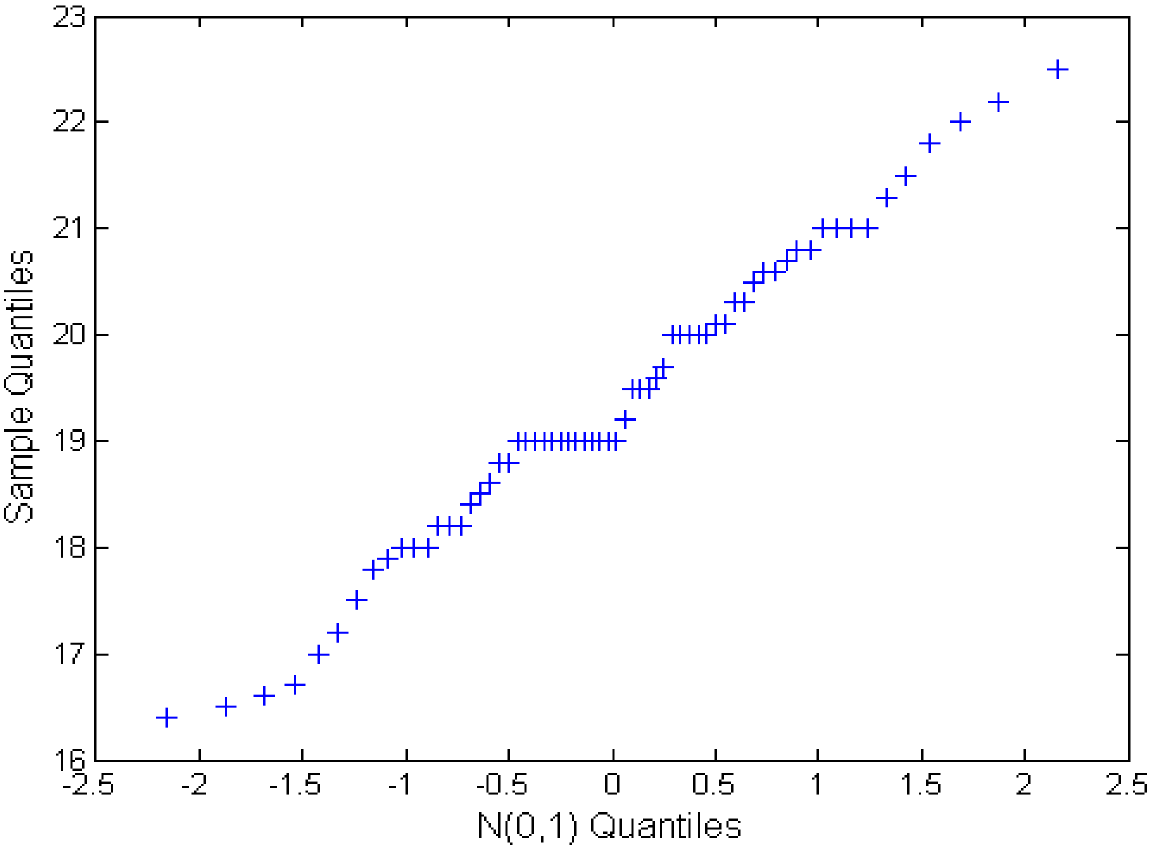
Numerical Summaries for Female Handspans

sample mean = 19.37 cm
sample standard deviation = 1.43 cm
sample median = 19.0 cm
sample skewness = -0.022
sample kurtosis = 2.58

Relative Frequency Histogram for Female Handspans



Qqplot for Female Handspans



Model - Females

It seems reasonable to assume the model

$$Y_i \backsim G(\mu,\sigma), \quad i=1,2,\ldots,64$$

where Y_i = handspan of i'th female.

We wish to estimate the parameter μ, the mean hand span for females registered in STAT 231 in Fall 2016.

Confidence Interval for Mean Handspan for Females

A point estimate for *μ* is $\hat{\mu} = \bar{y} = 19.3656$ **.**

An interval estimate is given by a 95% confidence interval.

Using R, we obtain:
P(*T* ≤ 1.9983) = (1+0.95)/2 = 0.975 for *T* ~ *t*(63).

A 95% confidence interval for *μ* is

$$\bar{y} \pm 1.9983 s / \sqrt{64} = 19.3656 \pm 1.9983(1.3655) / \sqrt{64}$$

$$= 19.3656 \pm 0.3655$$

or [19.00,19.73].

Numerical Summaries for Male Handspans

sample mean = 21.50 cm

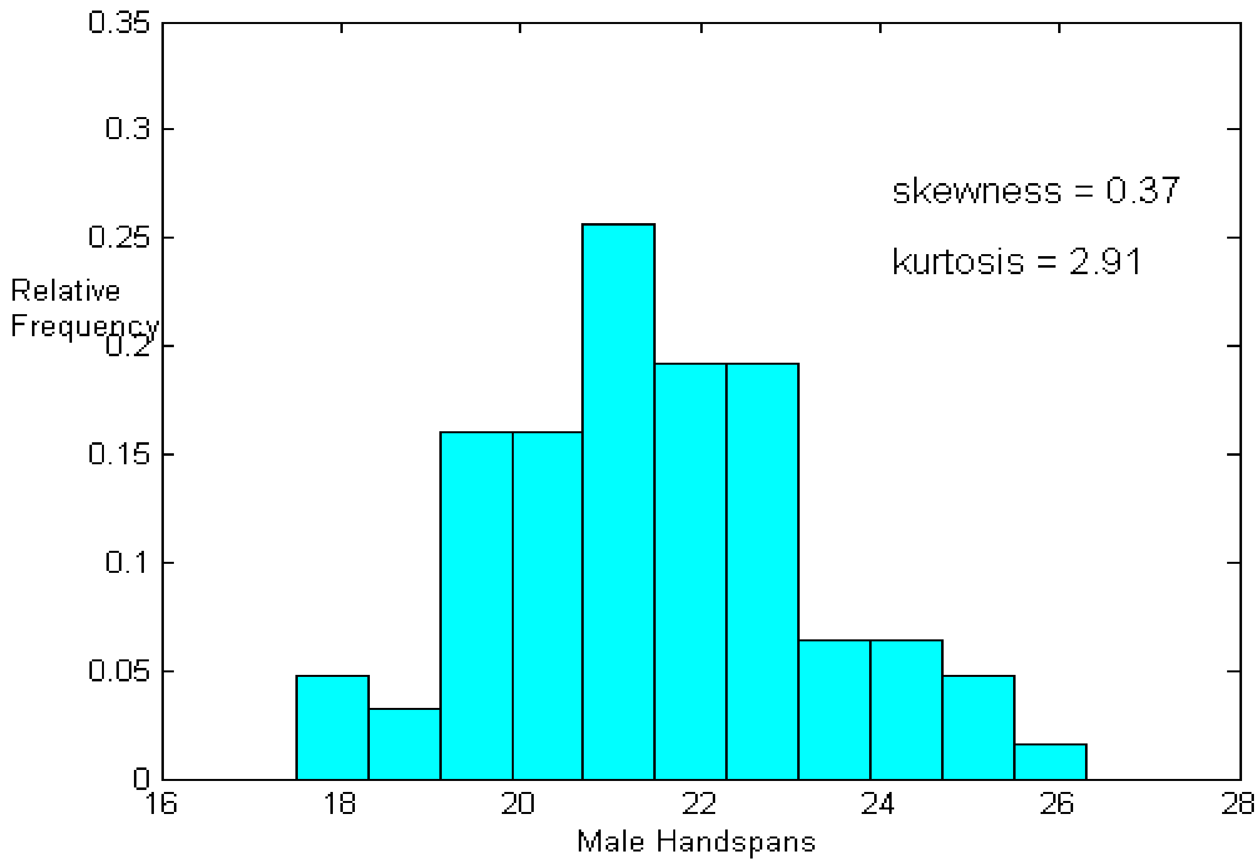
sample standard deviation = 1.85 cm

sample median = 21.2 cm

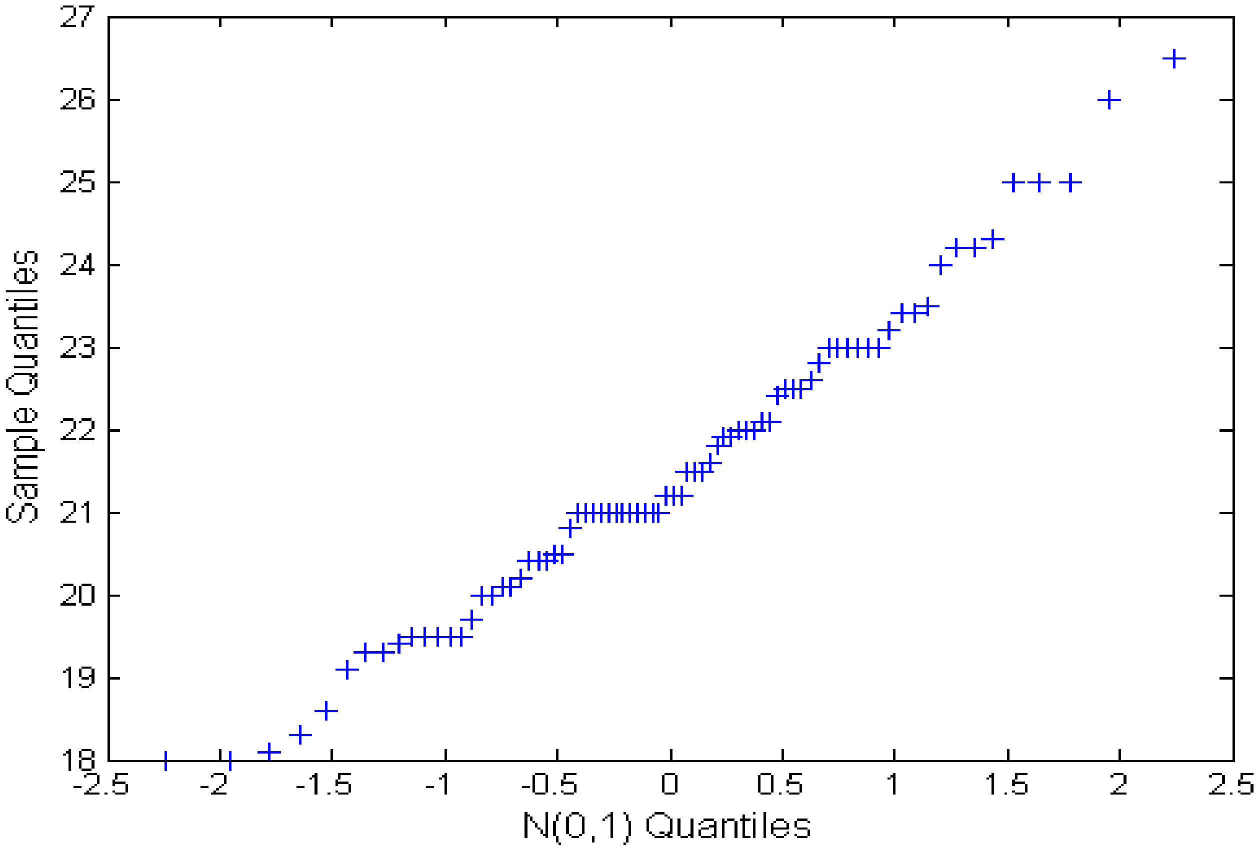
sample skewness = 0.37

sample kurtosis = 2.91

Relative Frequency Histogram for Male Handspans



Qqplot for Male Handspans



Model - Males

It seems reasonable to assume the model

$$X_i \backsim G(\mu_m, \sigma), \quad i=1,2,\ldots,78$$

where X_i = handspan of i'th male.

We wish to estimate the parameter μ_m, the mean hand span for males registered in STAT 231 in Fall 2016.

Confidence Interval for Mean Handspan for Males

A point estimate for μ_m, is $\hat{\mu}_m = \bar{x} = 21.5026$ **.**

An interval estimate is given by a 95% confidence interval.

Using R, we obtain:

P($T \leq 1.9913$) = (1+0.95)/2 = 0.975   for $T \sim t(77)$.

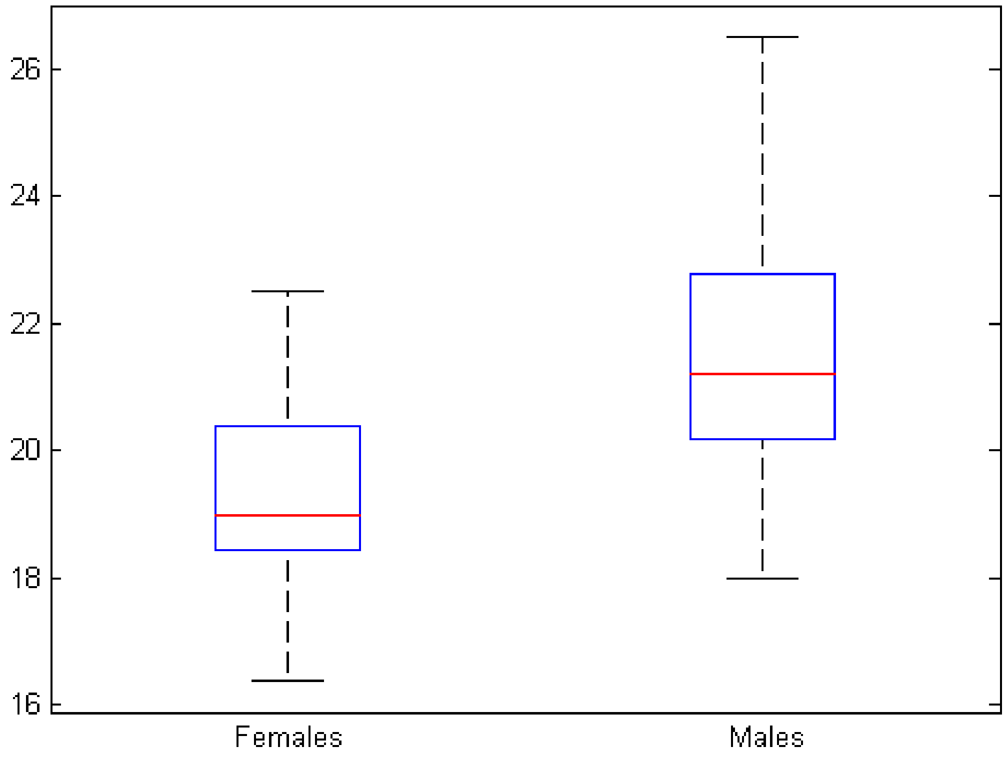
A 95% confidence interval for μ_m is

$$\bar{x} \pm 1.9913 s / \sqrt{78} = 21.5026 \pm 1.9913(1.8523) / \sqrt{78}$$

$$= 21.5026 \pm 0.4176$$

or [21.08,21.92].

Boxplots for Handspans



CBC Documentary

Right hands, wrong piano: a game changer for small-handed pianists

http://www.cbc.ca/radio/docproject/right-hands-wrong-piano-a-game-changer-for-small-handed-pianists-1.3819321

Sample Size Calculation: Gaussian Data

If we know the approximate value of *σ* (possibly from previous experiments), we can determine the sample size *n* needed for a future experiment to ensure a 95% confidence interval has a given width.

When σ is known a 95% confidence for μ is given by

$$\bar{y} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

with width

$$2(1.96)\frac{\sigma}{\sqrt{n}}$$

Sample Size Calculation: Gaussian Data

If we want the confidence interval

$$\bar{y} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

to be of the form $\bar{y}$ ± d then we should choose *n* such that

$$1.96 \frac{\sigma}{\sqrt{n}} \approx d \quad \text{or} \quad n \approx \left(\frac{1.96\sigma}{d} \right)^2$$

In practice, since we usually don't know *σ*, we would choose *n* larger than $(1.96\sigma/d)^2$.